SAND2017-13274C

A Locally Adapted Reduced Basis Method for Solving Risk-Averse PDE-Constrained Optimization Problems *

Zilong Zou†
Department of Civil and Environmental Engineering, Duke University, Durham, NC, 27705, USA

Drew Kouri‡
Optimization and Uncertainty Quantification, Sandia National Laboratories, Albuquerque, NM, 87185, USA

Wilkins Aquino §
Department of Civil and Environmental Engineering, Duke University, Durham, NC, 27705, USA

The numerical solution of risk-averse PDE-constrained optimization problems requires substantial computational effort resulting from the discretization of the underlying PDE in both the physical and stochastic dimensions. To practically solve problems with high-dimensional uncertainties, one must intelligently manage the individual discretization fidelities throughout the optimization iteration. In this work, we combine an inexact trust-region algorithm with the recently developed local reduced basis approximation to efficiently solve risk-averse optimization problems with PDE constraints. The main contribution of this work is a numerical framework for systematically constructing surrogate models for the trust-region subproblem and the objective function using local reduced basis approximations. We demonstrate the effectiveness of our approach through a numerical example.

I. Introduction

Many science and engineering applications can be formulated as PDE-constrained optimization problems that, more often then not, are riddled with uncertainty. For such scenarios, it is critical that we determine optimal solutions that, in some sense, mitigate the underlying uncertainties. In this paper, we model these problems as risk-averse PDE-constrained optimization problems. Unfortunately, after discretization, these problems become large-scale nonlinear (and often nonconvex) stochastic programming problems. The numerical solution of these fully discretized problems faces tremendous computational challenges due to the shear size of the discretized PDE. Therefore, managing the discretization fidelity in the physical and stochastic dimensions throughout the optimization procedure is essential to tractably solve these problems.

In this work, we focus on PDE-constrained optimization problems in which the coefficients of the governing PDE are uncertain and model risk aversion using the optimized certainty equivalent risk measures [3]. One important optimized certainty equivalent risk measure is the conditional value at risk (CVaR) [12]. CVaR is a coherent risk measure [2] and is commonly used in financial applications to determine risk-averse investment strategies [9]. The difficulty with minimizing CVaR and coherent risk measures in general is that they typically are nonsmooth. Therefore, efficient derivative-based optimization algorithms are often not applicable. Furthermore, accurately computing the risk measure value using, e.g., Monte Carlo methods requires a large number of samples and hence a large number of PDE solves. The recent work [11] has shown the feasibility of minimizing CVaR in the context of PDE-constrained optimization problems. However, the computational cost to numerically obtain a solution is still overwhelming.

Our goal is to devise a computational framework to efficiently solve risk-averse PDE-constrained optimization problems using cheap surrogate models for the PDE solutions. Few methods exist for constructing efficient surrogate models in the context of PDE-constrained optimization under uncertainty. For example, the authors in [10] introduce

*This research was supported by the DARPA EQUiPS program, grant SNL 014150709, "Risk-Averse Optimization of Large-Scale Multiphysics Systems."

†Ph.D. Candidate, Department of Civil and Environmental Engineering, Duke University, Durham, NC, 27705.

‡Principal Member of Technical Staff, Optimization and Uncertainty Quantification, MS-1320, Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185-1320. E-mail: dpkouri@sandia.gov. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

§Professor, Department of Civil and Environmental Engineering, Duke University, Durham, NC, 27705.

an adaptive sparse-grid approach whereas the author in [15] introduces a global reduced basis method. However, to obtain convergence of the optimization algorithm when using surrogate models, the errors associated with the surrogate approximation need to be properly managed. To this end, both [10] and [15] employ an inexact trust-region (TR) framework to manage the error in the surrogate approximations throughout the optimization process. Additionally, under standard assumptions, the authors in [10] show that the inexact TR algorithm is guaranteed to converge from any initial guess, provided that errors in the approximation of the objective function and its gradient are adequately bounded.

In this work, we combine the inexact TR framework of [10] with the adaptive local reduced basis (RB) method proposed in [16]. In [16], we demonstrated that the local RB method produces efficient and accurate approximations of PDE solutions as well as approximations of risk measures of quantities of interest. We use the local RB method to systematically construct the surrogate models for the TR subproblem and objective function evaluations. To this end, we derive error bounds for the objective and gradient approximation using *a posteriori* error indicators and subsequently introduce adaptive sampling schemes for the sequential construction of the surrogate models. Throughout the TR approach, we maintain two separate surrogate models: one to approximate the objective function gradient and one to approximate the objective function value. These two surrogate models employ different sets of basis functions and different error indicators to efficiently achieve the required error bounds of the TR method. We demonstrate the performance of our approach through a 1D advection-diffusion numerical example, which shows that our method can solve risk-averse optimization problems with PDE constraints using far fewer high-fidelity PDE solves than those required with Monte Carlo methods.

II. Problem Formulation

Let $(\Omega, \mathcal{F}, P)$ denote a probability space. Here, Ω is the set of outcomes, $\mathcal{F} \subseteq 2^\Omega$ is a σ-algebra of events and $P : \mathcal{F} \to [0,1]$ is a probability measure. We denote the uncertain inputs of our PDE by the random vector $\xi : \Omega \to \Xi$ where $\Xi := \xi(\Omega) \subseteq \mathbb{R}^m$ with $m \in \mathbb{N}$. Throughout, we abuse notation and let ξ denote the random inputs as well as a realization of the random inputs. We further denote the expectation of a random variable $X : \Omega \to \mathbb{R}$ by

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, \mathrm{d}P(\omega)$$

and we denote the space of random variables $X : \Omega \to \mathbb{R}$ with p-finite moments for $p \in [1, \infty)$, i.e., $\mathbb{E}[|X|^p] < \infty$, by $L^p(\Omega, \mathcal{F}, P)$. When $p = \infty$, $L^\infty(\Omega, \mathcal{F}, P)$ denotes the space of essentially bounded random variables on $(\Omega, \mathcal{F}, P)$. We denote the extension of $L^p(\Omega, \mathcal{F}, P)$ for $p \in [1, \infty]$ to vector-valued random variables by $L^p(\Omega, \mathcal{F}, P; A)$ where A is a Banach space. Additionally, for any two Banach spaces A and B, we denote the space of bounded linear operators that map A into B by $L(A, B)$. We further denote the dual space of A by $A^* = L(A, \mathbb{R})$ and the associated dual pairing by $\langle a^*, a \rangle_{A^*,A} = a^*(a)$ for $a^* \in A$ and $a \in A$. Finally, if A is a Hilbert space, we denote the inner product on A by $\langle a', a \rangle_A$ for all a', $a \in A$.

Now, let U, V and Z be real Hilbert spaces. Here, U is the deterministic state space (i.e., the space of PDE solutions) and Z is the space of optimization variables. We will refer to $z \in Z$ as a control variable. Since controls typically must be implemented prior to observing the uncertainty in the PDE solution, we require that Z is a space of deterministic functions or vectors. We consider the following parametrized, linear PDE: for fixed $z \in Z$, find $u : \Xi \to U$ such that

$$M(u(\xi), z; \xi) := \mathcal{L}(\xi)u(\xi) + \mathcal{B}(\xi)z + \ell(\xi) = 0 \quad \forall\, \xi \in \Xi \tag{1}$$

where $\mathcal{L}(\xi) \in L(U, V^*)$ for all $\xi \in \Xi$ is the parametrized PDE operator, $\mathcal{B}(\xi) \in L(Z, V^*)$ for all $\xi \in \Xi$ is the parametrized control operator and $\ell(\xi) \in V^*$ for all $\xi \in \Xi$ is the parametrized load or force. If U, V and Z are finite-dimensional, $\mathcal{L}(\xi)$ and $\mathcal{B}(\xi)$ are matrices and $\ell(\xi)$ is a vector representing, e.g., the discretization of a PDE. We make the following assumption regarding (1).

Assumption 1 (PDE) *We assume that the operators defining (1) satisfy:*

1) The differential operator satisfies the stability conditions: $\exists\, \kappa_1 > 0$ independent of $\xi \in \Xi$ such that for all $\xi \in \Xi$

$$\inf_{u \in U\setminus\{0\}} \sup_{v \in V\setminus\{0\}} \frac{|\langle \mathcal{L}(\xi)u, v\rangle_{V^*,V}|}{\|u\|_U \|v\|_V} =: \gamma(\xi) \geq \kappa_1, \quad \forall\, \xi \in \Xi \tag{2a}$$

$$\langle \mathcal{L}(\xi)u, v\rangle_{V^*,V} = \langle \mathcal{L}(\xi)^* v, u\rangle_{U^*,U} = 0 \quad \forall\, u \in U \qquad \Longrightarrow \qquad v = 0 \tag{2b}$$

2) The control operator satisfies: $\exists\, \kappa_2 \geq 0$ independent of $\xi \in \Xi$ such that $\|\mathcal{B}(\xi)\|_{L(Z,V^*)} \leq \kappa_2$ for all $\xi \in \Xi$;

3) The force satisfies: $\exists\, \kappa_3 \geq 0$ *independent of* $\xi \in \Xi$ *such that* $\|\ell(\xi)\|_{V^*} \leq \kappa_3$ *for all* $\xi \in \Xi$.

Under Assumption 1, the Banach-Nečac-Babuška Theorem [6] ensures that a unique solution to (1) exists for all $\xi \in \Xi$ and $z \in Z$, and that the solution is bounded independent of $\xi \in \Xi$. We denote the solution of (1) by $S(\xi; z) \in U$ and assume that $S(\xi(\cdot); z)$ is strongly measurable, i.e., $S(\xi(\cdot); z) \in L^\infty(\Omega, \mathcal{F}, P; U)$ for all $z \in Z$. Note that one could achieve measurability of $S(\xi(\cdot); z)$ by imposing additional assumptions on the measurability of $\mathcal{L}(\xi(\cdot))$, $\mathcal{B}(\xi(\cdot))$ and $\ell(\xi(\cdot))$ (see, e.g., [8]). We omit these assumptions to simplify the presentation.

Turning our attention to the target optimization problem, let $\wp : Z \to \mathbb{R}$ and $G : U \times \Xi \to \mathbb{R}$, and note that $G(S(\xi(\omega); z), \xi(\omega))$ is a function of ω and hence is viewed as a random variable. We assume that $G(S(\xi; z), \xi) \in \mathcal{X} := L^p(\Omega, \mathcal{F}, P)$ for some $p \in [1, \infty]$ and seek to minimize

$$J(z) = \mathcal{R}(G(S(\xi; z), \xi)) + \wp(z) \tag{3}$$

where $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is a risk measure with the form

$$\mathcal{R}(X) = \inf_{t \in \mathbb{R}} \{t + \mathbb{E}[v(X - t)]\}. \tag{4}$$

Here, $v : \mathbb{R} \to \mathbb{R}$ is convex and satisfies $v(0) = 0$ and $v(x) > x$ for $x \neq 0$. For example, when $v(x) = (1 - \beta)^{-1}[x]^+$ with $\beta \in (0, 1)$ and $[\cdot]^+ = \max\{0, \cdot\}$, the risk measure $\mathcal{R}$ is the conditional value-at-risk (CVaR) [12],

$$\mathcal{R}(X) = \text{CVaR}_\beta[X] := \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{1 - \beta} \mathbb{E}[X - t]^+ \right\}. \tag{5}$$

It is important to point out that the objective function in (3) may be nonsmooth in z depending on the function v. Therefore, to take advantage of gradient-based optimization algorithms, we only consider differentiable v. In the context of CVaR, the associated v is not differentiable at $x = 0$. As such, one could consider a smooth approximation similar to those studied in [11].

To minimize (3), we incorporate t from the definition of $\mathcal{R}$ as an additional optimization variable and solve

$$\min_{t \in \mathbb{R},\, z \in Z} \widehat{J}(t, z) \quad \text{where} \quad \widehat{J}(t, z) := t + \mathbb{E}\left[v(G(S(\xi; z), \xi) - t)\right] + \wp(z). \tag{6}$$

Under appropriate assumptions on M, G, $\wp$ and v, the objective function $\widehat{J}$ is continuously Fréchet differentiable and the partial derivatives $\widehat{J}$ are

$$\nabla_z \widehat{J}(t, z) = \mathbb{E}\left[v'(G(S(\xi; z), \xi) - t)\mathcal{B}(\xi)^* \lambda(\xi)\right] + \nabla \wp(z) \tag{7a}$$

$$\nabla_t \widehat{J}(t, z) = 1 - \mathbb{E}\left[v'(G(S(\xi; z), \xi) - t)\right] \tag{7b}$$

where $\lambda : \Xi \to V$ solves the adjoint equation

$$\mathcal{L}(\xi)^* \lambda(\xi) = -\nabla_u G(S(\xi; z), \xi) \quad \forall\, \xi \in \Xi. \tag{8}$$

Notice that the computation of the gradient requires the solution of the state equation (1) and the solution of the adjoint equation (8) for all $\xi \in \Xi$. We denote the adjoint solution for a fixed state variable u by $\Lambda(\xi; u)$. For example, in (8), we can write $\lambda(\xi) = \Lambda(\xi; S(\xi; z))$ for $\xi \in \Xi$. To simplify notation in the subsequent sections, we denote $W := \mathbb{R} \times Z$, $w = (t, z)$ and $\widehat{J}(w) = \widehat{J}(t, z)$.

III. Risk Measures

Since the objective function $G(S(\xi; \cdot), \xi)$ is a random variable, we cannot directly minimize it. Instead, we minimize the deterministic quantity $\mathcal{R}(G(S(\xi; \cdot), \xi))$ where the functional $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is chosen to quantify the overall hazard or risk associated with the uncertain objective function $G(S(\xi; \cdot), \xi)$. In this setting, we refer to $\mathcal{R}$ as a risk measure. The risk measure $\mathcal{R}$ is said to be coherent [2] if it satisfies the following axioms:

(C1) **Convexity:** $\mathcal{R}(tX + (1 - t)X') \leq t\mathcal{R}(X) + (1 - t)\mathcal{R}(X')$ for all X, $X' \in \mathcal{X}$ and $t \in [0, 1]$;
(C2) **Monotonicity:** $\mathcal{R}(X) \leq \mathcal{R}(X')$ whenever X, $X' \in \mathcal{X}$ with $X \leq X'$ almost surely;
(C3) **Translation Equivariance:** $\mathcal{R}(X + t) = \mathcal{R}(X) + t$ for all $X \in \mathcal{X}$ and $t \in \mathbb{R}$;
(C4) **Positive Homogeneity:** $\mathcal{R}(tX) = t\mathcal{R}(X)$ for all $X \in \mathcal{X}$ and $t \geq 0$.

In the context of engineering applications, (C4) ensures that $\mathcal{R}(X)$ inherits, e.g., any change of units in X while (C3) and (C4) ensure that deterministic quantities are "riskless," i.e., $\mathcal{R}(t) = t$ for all $t \in \mathbb{R}$. In addition, (C1) and (C4) are equivalent to (C4) and $\mathcal{R}$ being subadditive, i.e., $\mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X')$ for all X, $X' \in \mathcal{X}$. It is worth pointing out that the common mean-plus-deviation risk measure of order p with $p \in [1, \infty)$

$$\mathcal{R}(X) = \mathbb{E}[X] + c\mathbb{E}[|X - \mathbb{E}[X]|^p]^{\frac{1}{p}}, \quad c \geq 0,$$

is not coherent since it does not satisfy the monotonicity property (C2).

The risk measure $\mathcal{R}$ defined by (4) is called an optimized certainty equivalent [3]. Any such $\mathcal{R}$ satisfies (C1) and (C3). Furthermore, $\mathcal{R}$ satisfies (C2) if and only if $v(x) \leq 0$ whenever $x < 0$ and satisfy (C4) if and only if v is positive homogeneous [13]. Three common risk measures arising from (4) are the mean-plus-variance

$$\mathcal{R}(X) = \mathbb{E}[X] + c\mathbb{E}[|X - \mathbb{E}[X]|^2], \quad c > 0,$$

in which case $v(x) = x + cx^2$ and $p = 2$, the conditional value-at-risk (5) in which case $v(x) = (1 - \beta)^{-1}[x]^+$ and $p = 1$, and the entropic risk

$$\mathcal{R}(X) = \log(\mathbb{E}[\exp(cX)])/c, \quad c > 0,$$

in which case $v(x) = (\exp(cx) - 1)/c$ and $p = \infty$. Note that the mean-plus-variance risk measure does not satisfy (C2) and (C4) while the entropic risk measure fails to satisfy (C4). Since v is finite valued and convex, it is continuous. Additionally, for $\mathcal{R}$ to be finite valued, we require that one of the following two properties to holds.

Assumption 2 (Growth Condition for v) *The function $v : \mathbb{R} \to \mathbb{R}$ satisfies one of the following conditions:*
(V_p) $\mathcal{X} = L^p(\Omega, \mathcal{F}, P)$ with $p \in [1, \infty)$ and there exists γ_0, $\gamma_1 \in \mathbb{R}$ with $\gamma_1 \geq 0$ such that

$$|v(x)| \leq \gamma_0 + \gamma_1 |x|^p, \quad \forall\, x \in \mathbb{R};$$

(V_∞) $\mathcal{X} = L^\infty(\Omega, \mathcal{F}, P)$ and for all $c \geq 0$ there exists $\gamma = \gamma(c) \geq 0$ such that

$$|v(x)| \leq \gamma \quad \forall\, x \in \mathbb{R},\ |x| \leq c.$$

Under Assumption 2, $v : \mathcal{X} \to L^1(\Omega, \mathcal{F}, P)$ is continuous [7, Thm. 4] and $\mathcal{R}(X)$ is finite for all $X \in \mathcal{X}$ since $\mathbb{E}[v(X)]$ is.

Since our goal is to apply derivative-based optimization algorithms to minimize (3), we further assume that v is continuously differentiable and its derivative v' satisfies one of the two following conditions.

Assumption 3 (Growth Condition for v') *The derivative $v' : \mathbb{R} \to \mathbb{R}$ satisfies one of the following conditions:*
(D_p) $\mathcal{X} = L^p(\Omega, \mathcal{F}, P)$ with $p \in (1, \infty)$ and there exists δ_0, $\delta_1 \in \mathbb{R}$ with $\delta_1 \geq 0$ such that

$$|v'(x)| \leq \delta_0 + \delta_1 |x|^{p-1}, \quad \forall\, x \in \mathbb{R};$$

(D_∞) $\mathcal{X} = L^\infty(\Omega, \mathcal{F}, P)$ and for all $c \geq 0$ there exists $\delta = \delta(c) \geq 0$ such that

$$|v'(x)| \leq \delta \quad \forall\, x \in \mathbb{R},\ |x| \leq c.$$

Under Assumption 3, $v : \mathcal{X} \to L^1(\Omega, \mathcal{F}, P)$ is continuously Fréchet differentiable and its derivative at $X \in \mathcal{X}$ is given by $v'(X)$ [7, Thm. 7]. Therefore, $\mathbb{E}[v(\cdot)]$ is also continuously Fréchet differentiable. Note that we require a *norm gap* between the domain and range of v, i.e., since the range is $L^1(\Omega, \mathcal{F}, P)$, the domain must be $L^p(\Omega, \mathcal{F}, P)$ with $p > 1$ in order to obtain differentiability.

IV. Inexact Trust Region Algorithm with Adaptive Local Reduced Bases

To overcome the potentially enormous computational cost of minimizing $\widehat{J}$, we develop an adaptive optimization framework based on two key components. The first component is the adaptive reduced basis approximation introduced in [16] to build computationally inexpensive approximations of S. Using these low-cost surrogate models, we can approximate the objective function value and its gradient in a computationally tractable manner. The second key component of our method is the inexact TR algorithm developed in [10], which prescribes accuracy requirements on the surrogate model approximations that guarantee global convergence to local minimizers of $\widehat{J}$.

A. Inexact Trust Regions

We now briefly describe the inexact TR algorithm. For a concrete statement of the algorithm as well as the rigorous convergence theory, see [10]. Given the current iterate $w_k = (t_k, z_k) \in W$, we construct a local model m_k of the objective function $s \mapsto \widehat{J}(w_k + s)$ in the region $\{s \in W : \|s_k\|_W \leq \Delta_k\}$. Here $\Delta_k > 0$ is the current TR radius and $\|\cdot\|_W$ denotes a norm on W, e.g., $\|w\|_W^2 = t^2 + \|z\|_Z^2$. We determine the trial step s_k by *approximately* solving the TR subproblem

$$\min_{s \in W} \; m_k(s) \quad \text{subject to} \quad \|s\|_W \leq \Delta_k. \tag{9}$$

We require that any approximate solution to (9) satisfies the so called *fraction of Cauchy decrease condition* (see [10] for more details). Once a step s_k is computed, we decide whether to accept or reject s_k based the ratio of the actual and predicted reduction

$$\rho_k := \frac{\text{ared}_k}{\text{pred}_k} \quad \text{where} \quad \text{ared}_k := \widehat{J}(w_k) - \widehat{J}(w_k + s_k) \quad \text{and} \quad \text{pred}_k := m_k(0) - m_k(s_k).$$

Here, ared_k denotes the *actual reduction* obtained by the step s_k while pred_k denotes the *predicted reduction* based on the model m_k.

Now, to ensure convergence of the TR algorithm from arbitrary initial guesses, we require that the model in (9) satisfies the inexact gradient condition

$$\|\nabla m_k(0) - \nabla \widehat{J}(w_k)\|_W \leq \kappa \min\{\|\nabla m_k(0)\|_W, \Delta_k\} \tag{10}$$

where $\kappa > 0$ is independent of k. Additionally, since the computation of ared_k requires the evaluation of the true objective $\widehat{J}$, which is computationally prohibitive or even impossible, we evaluate ared_k inexactly. To this end, we introduce a model $\widehat{J}_k$ that approximates $\widehat{J}$ and allows us to approximate ared_k as

$$\text{ared}_k \approx \text{cred}_k := \widehat{J}_k(w_k) - \widehat{J}_k(w_k + s_k). \tag{11}$$

Here, cred_k is the *computed reduction*. To guarantee convergence of the TR algorithm, we require that

$$|\text{ared}_k - \text{cred}_k| \leq K\theta_k \quad \text{with} \quad \theta_k^{\omega} \leq \eta \min\{\text{pred}_k, r_k\} \tag{12}$$

for some $K > 0$, fixed $\omega \in (0, 1)$ and

$$\eta < \min\{\eta_1, 1 - \eta_2\} \quad \text{and} \quad r_k \geq 0 \text{ with } \lim_{k \to \infty} r_k = 0$$

where $0 < \eta_1 < \eta_2 < 1$ are algorithmic constants and r_k is a forcing sequence. The authors in [10] prove that as long as one can control the surrogate approximation error for the objective function value and gradient to satisfy conditions (10) and (12), then the TR algorithm converges to a local minimizer for any starting point $w_0 = (t_0, z_0)$.

B. Local Reduced Bases

To construct approximations that satisfy (10) and (12), we employ the local RB method [16]. In the local RB method, we partition the parameter space into Voronoi cells, i.e., $\Xi = \cup_{k=1}^n \Xi_k$, seeded at n selected atoms ξ_k, $k = 1, \ldots, n$. Within each cell, we form a local basis for approximating the PDE solution using, e.g., full-order PDE solutions at a fixed number of proximal atoms as well as the gradient of the solution at the given seed. For example, Figure 1 shows a partition of a parameter domain $\Xi \subseteq \mathbb{R}^2$ with 2,000 Monte Carlo samples of ξ in the background. The surrogate solution at the large blue dot is computed using a basis consisting of full PDE solutions at the large solid red dots as well as the solution and gradient at the large black dot. The number of neighbors N for the local basis is usually chosen to be a fixed constant. In general, the number of neighbors is an algorithmic choice of the user, but can be chosen adaptively depending on the desired accuracy of the local approximation.

The local RB surrogate model $\widehat{S}(\xi; z)$ of $S(\xi; z)$ is given by

$$\widehat{S}(\xi; z) = \sum_{k=1}^{n} \mathbb{1}_{\Xi_k}(\xi)\, S_k(\xi; z) \tag{13}$$

where $\mathbb{1}_{\Xi_k}$ denotes the characteristic function of the set Ξ_k, i.e., $\mathbb{1}_{\Xi_k}(\xi) = 1$ if $\xi \in \Xi_k$ and $\mathbb{1}_{\Xi_k}(\xi) = 0$ otherwise, and $S_k(\cdot; z) = u_k : \Xi_k \to U_k$ is the solution of the reduced problem

$$\langle M(u_k(\xi), z; \xi), v \rangle_{V^*, V} = 0 \quad \forall\, v \in V_k, \quad \forall\, \xi \in \Xi_k.$$

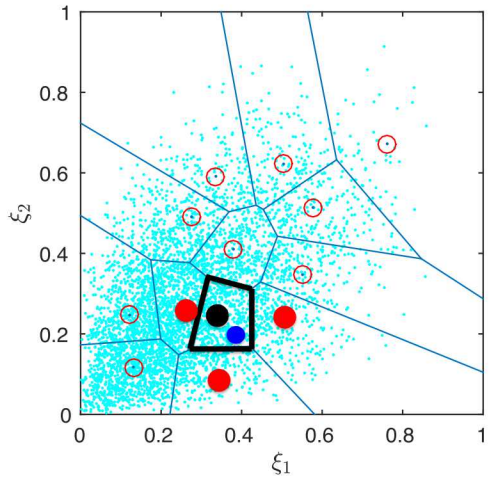

Fig. 1 The local reduced basis method with two random parameters. The surrogate solution at the large blue dot is computed using a basis consisting of the solution at the large solid red dots as well as the solution and gradient at the large black dot. We plot 2,000 Monte Carlo samples of ξ in the background.

Here, Φ_k is a "local basis" within Ξ_k, e.g.,

$$\Phi_k = \left[S(\xi_k; z), \nabla_\xi S(\xi_k; z), S(\xi_{k_1}; z), S(\xi_{k_2}; z), \ldots, S(\xi_{k_N}; z)\right],$$

$U_k = \text{span}(\Phi_k)$, and V_k is a finite-dimensional subspace of V. Since the cardinality of Φ_k is typically much smaller than the full discretization of the PDE, we often realize significant computational savings by using $\widehat{S}(\xi; z)$ as a surrogate model for $S(\xi; z)$. In addition, due to the local nature of the approximation, the evaluation cost of $\widehat{S}(\xi; z)$ at any $\xi \in \Xi$ does not increase as the number of atoms n increases.

To efficiently construct the local RB surrogate, we employ a greedy adaptive sampling procedure to select the atom set $\Theta := \{\xi_k\}_{k=1}^n$ [16]. The adaptive selection of Θ is guided by reliable a posteriori error indicators, denoted by $\epsilon^u(\xi; z)$, i.e.,

$$\|\widehat{S}(\xi; z) - S(\xi; z)\|_U \lesssim \epsilon^u(\xi; z)$$

where $x \lesssim y$ denotes "x is less than or equal to a constant times y." That is, given k atoms, the next atom ξ_{k+1} is selected from the region of Ξ where the current surrogate error is the largest. The error indicators $\epsilon^u(\xi; z)$ used in [16] are residual-based error estimates which have been used with great success in adaptive finite elements [4] and in the adaptive construction of reduced bases [14]. In fact, we have shown in [16] that the error indicator $\epsilon^u(\xi; z)$ can be further used to build more complex error indicators that are specifically targeted for the approximation of coherent risk measures evaluated at quantities of interest depending on PDE solutions. The versatility provided by the adaptive sampling procedure tailored to individual tasks is of fundamental importance for the algorithmic development in this paper.

C. Inexact Trust Regions with Local Reduced Bases

To simplify the subsequent analysis, we assume that the random inputs ξ are discretely distributed. Let $\{\xi_j\}_{j=1}^N$ denote the atoms of ξ with the associated probabilities $\{p_j\}_{j=1}^N$ with $p_1 + \ldots + p_N = 1$ and $p_j > 0$ for $j = 1, \ldots, N$. In this setting, we can then rewrite (4) as

$$\mathcal{R}(G(S(\xi; z), \xi)) = \inf_{t \in \mathbb{R}} \left\{ t + \sum_{j=1}^N p_j v(G(S(\xi_j; z), \xi_j) - t) \right\}. \tag{14}$$

For example, if $\{\xi_j\}$ are Monte Carlo samples of ξ, then $p_j = 1/N$ for $j = 1, \ldots, N$. The assumption that ξ is discretely distributed can be relaxed if we are given a discrete approximation to ξ that can be refined to control its approximation error (e.g., quadrature approximation or Monte Carlo).

As mentioned above, to obtain surrogate models that meet the accuracy requirements of the TR algorithm, we employ the adaptive local RB method. To this end, we construct two surrogate models of $S(\xi; z)$, denoted by $S_{\text{mod}}(\xi; z)$ and $S_{\text{obj}}(\xi; z)$, which we then use to construct m_k and $\widehat{J}_k$, respectively. Moreover, since (10) depends on gradient information, we construct a surrogate model for the adjoint variable $\Lambda(\xi; u)$, denoted $\Lambda_{\text{mod}}(\xi; u)$. By constructing $S_{\text{mod}}(\xi; z)$, $\Lambda_{\text{mod}}(\xi; u)$ and $S_{\text{obj}}(\xi; z)$ adaptively, we aim to achieve the desired error levels with as few high-fidelity solves as possible. To this end, we rely on the computable a posterior error indicators in the local RB method. We denote the computable a posterior error indicators for $S_{\text{mod}}(\xi; z)$, $\Lambda_{\text{mod}}(\xi; u)$ and $S_{\text{obj}}(\xi; z)$ by $\epsilon^u_{\text{mod}}(\xi; z)$, $\epsilon^\lambda_{\text{mod}}(\xi; u)$ and $\epsilon^u_{\text{obj}}(\xi; z)$, respectively; that is,

$$\begin{aligned}\|S_{\text{mod}}(\xi; z) - S(\xi; z)\|_U &\lesssim \epsilon^u_{\text{mod}}(\xi; z)\\ \|\Lambda_{\text{mod}}(\xi; u) - \Lambda(\xi; u)\|_V &\lesssim \epsilon^\lambda_{\text{mod}}(\xi; u)\\ \|S_{\text{obj}}(\xi; z) - S(\xi; z)\|_U &\lesssim \epsilon^u_{\text{obj}}(\xi; z).\end{aligned}$$

See [16] for the explicit form of these error indicators. We also require the following assumptions on G and v.

Assumption 4 (Lipschitz Continuous Derivatives) *We require the following Lipschitz continuity conditions to hold:*

1) $G(\cdot, \xi)$ is continuously Fréchet differentiable for all $\xi \in \Xi$ and there exists $K_G \geq 0$ independent of ξ such that

$$\|\nabla_u G(u, \xi) - \nabla_u G(u', \xi)\|_U \leq K_G \|u - u'\|_U \quad \forall\, u,\, u' \in U \quad \forall\, \xi \in \Xi;$$

2) There exists $K_v \geq 0$ such that

$$|v'(x) - v'(x')| \leq K_v |x - x'| \quad \forall\, x,\, x' \in \mathbb{R}.$$

Using the local RB approximation S_{mod}, one can choose, e.g., the trust-region subproblem model

$$m_k(s) = \left((t_k + \tau) + \sum_{j=1}^N p_j v(G(S_{\text{mod}}(\xi_j; z_k + \zeta); \xi_j) - (t_k + \tau)) \right) + \wp(z_k + \zeta) \tag{15}$$

where $s = (\tau, \zeta) \in W$. Since this m_k is often highly nonlinear, solving the trust-region subproblem can be computationally challenging. Instead, we typically employ the quadratic model

$$m_k(s) = \frac{1}{2}\langle H_k s, s\rangle_W + \langle g_k, s\rangle_W$$

where $H_k \in L(W, W)$ is an approximation of the Hessian of $\widehat{J}(w_k)$ and $g_k \in W$ is an approximation of the gradient of $\widehat{J}(w_k)$. We then can use efficient methods such as the truncated conjugate gradient method [5] to solve the TR subproblem (9). Since our focus is on satisfying (10), we assume H_k is provided and is bounded for all k. We then choose g_k according to (7) with S and Λ replaced by S_{mod} and Λ_{mod}.

Now, let $u(\xi) = S(\xi; z_k)$, $u_j = u(\xi_j)$, $\widehat{u}(\xi) = S_{\text{mod}}(\xi; z_k)$, $\widehat{u}_j = \widehat{u}(\xi_j)$, $\lambda(\xi) = \Lambda(\xi; u(\xi))$, $\lambda_j = \lambda(\xi_j)$, $\widehat{\lambda}(\xi) = \Lambda_{\text{mod}}(\xi; \widehat{u}(\xi))$ and $\widehat{\lambda}_j = \widehat{\lambda}(\xi_j)$. Using the explicit form of the gradient (7) and the triangle inequality, we arrive at the following bounds

$$\|\nabla_z \widehat{J}(w_k) - \nabla_z m_k(0)\|_Z \leq \sum_{j=1}^N p_j \|v'(G(u_j, \xi_j) - t_k)\mathcal{B}(\xi_j)^* \lambda_j - v'(G(\widehat{u}_j, \xi_j) - t_k)\mathcal{B}(\xi_j)^* \widehat{\lambda}_j\|_Z \tag{16}$$

and

$$|\nabla_t \widehat{J}(w_k) - \nabla_t m_k(0)| \leq \sum_{j=1}^N p_j |v'(G(u_j, \xi_j) - t_k) - v'(G(\widehat{u}_j, \xi_j) - t_k)|. \tag{17}$$

We then bound the individual samples on the right hand side of (16) by

$$\begin{aligned}\|\mathcal{B}(\xi_j)^*(v'(G(u_j, \xi_j) - t_k)\lambda_j - v'(G(\widehat{u}_j, \xi_j) - t_k)\widehat{\lambda}_j)\|_Z \leq{}& \kappa_2 |v'(G(u_j, \xi_j) - t_k) - v'(G(\widehat{u}_j, \xi_j) - t_k)| \|\lambda_j\|_V\\ &+ \kappa_2 |v'(G(\widehat{u}_j, \xi_j) - t_k)| \|\widehat{\lambda}_j - \lambda_j\|_V.\end{aligned} \tag{18}$$

Using Assumption 4 and the integral mean value theorem, we bound the first term on the right hand side of (18) and the right hand side of (17) using

$$\begin{aligned}|v'(G(u_j,\xi_j)-t_k)-v'(G(\widehat{u}_j,\xi_j)-t_k)| &\le K_v|G(u_j,\xi_j)-G(\widehat{u}_j,\xi_j)|\\ &= K_v\left|\int_0^1 \langle \nabla_u G(\widehat{u}_j+t(u_j-\widehat{u}_j),\xi_j),(u_j-\widehat{u}_j)\rangle_U \,\mathrm{d}t\right|\\ &\le \frac{1}{2}K_vK_G\|u_j-\widehat{u}_j\|_U^2 + K_v\|\nabla_u G(\widehat{u}_j,\xi_j)\|_U\|u_j-\widehat{u}_j\|_U^2\\ &\lesssim \epsilon^u_{\mathrm{mod}}(\xi_j,z_k)^2 + \|\nabla_u G(\widehat{u}_j,\xi_j)\|_U\epsilon^u_{\mathrm{mod}}(\xi_j;z_k).\end{aligned}$$

Now, let $\lambda_j^0 = \Lambda(\xi_j;\widehat{u}_j)$, then we can bound the second term on the right hand side of (18) using

$$\|\widehat{\lambda}_j-\lambda_j\|_V \le \|\widehat{\lambda}_j-\lambda_j^0\|_V + \|\lambda_j^0-\lambda_j\|_V \lesssim \|\widehat{\lambda}_j-\lambda_j^0\|_V + \|\widehat{u}_j-u_j\|_U \lesssim \epsilon^\lambda_{\mathrm{mod}}(\xi_j,\widehat{u}_j)+\epsilon^u_{\mathrm{mod}}(\xi_j,z_k).$$

By combining these bounds, we obtain the gradient error bound

$$\|\nabla\widehat{J}(w_k)-\nabla m_k(0)\|_W \lesssim \sum_{j=1}^N p_j\delta_{\mathrm{mod}}(\xi_j;w_k) =: E_{\mathrm{mod}}(w_k)$$

where

$$\delta_{\mathrm{mod}}(\xi_j;w_k) := |v'(G(\widehat{u}_j,\xi_j)-t_k)|(\epsilon^\lambda_{\mathrm{mod}}(\xi_j,\widehat{u}_j)+\epsilon^u_{\mathrm{mod}}(\xi_j,z_k)) + \epsilon^u_{\mathrm{mod}}(\xi_j,z_k)^2 + \|\nabla_u G(\widehat{u}_j,\xi_j)\|_U\epsilon^u_{\mathrm{mod}}(\xi_j;z_k).$$

and as long as $|v'(G(\widehat{u}_j,\xi_j)-t_k)|$ and $\|\nabla_u G(\widehat{u}_j,\xi_j)\|_U$ are bounded for all iterations k, we can satisfy (10) by refining our local reduced basis models S_{mod} and Λ_{mod}. The error indicator $E_{\mathrm{mod}}(z_k)$ is only small provided we approximate both forward and adjoint solutions accurately. We therefore modify the local RB method using a new set of local bases that is composed of both forward and adjoint solutions as well as their gradients. Note that even though m_k changes at each TR step, we can employ the same S_{mod} to obtain m_k for all steps, which means that S_{mod} can have a basis composed of solutions under different z_k. The error indicator will guide the refinement of S_{mod} at each step k so that (10) is always satisfied. The recycling of S_{mod} reduces the computational cost substantially since we do not need to rebuild an S_{mod} whenever z_k is updated in the optimization process. The same recycling strategy applies to Λ_{mod}. The complete algorithm for constructing S_{mod} and Λ_{mod} is listed in Algorithm 1.

Algorithm 1: Adaptive algorithm to build S_{mod} and Λ_{mod} for m_k

If $k=0$**, model initialization**:

- Let $\xi_0=\mathbb{E}[\xi]$ and build the initial surrogate models S_{mod} and Λ_{mod} based on the solution $S(\xi_0;z_0)$, its gradient $\nabla_\xi S(\xi_0;z_0)$, the adjoint solution $\Lambda(\xi_0;S(\xi_0;z_0))$ and its gradient $\nabla_\xi\Lambda(\xi_0;S(\xi_0;z_0))$.

Model refinement:

Given w_k, Δ_k and S_{mod}, which is recycled from step $k-1$,

- Build m_k with S_{mod}, evaluate $\delta_{\mathrm{mod}}(\xi_j;w_k)$ for all $j=1,\dots,N$, compute $E_{\mathrm{mod}}(w_k)$ and $\|\nabla m_k(0)\|_Z$.
- **While** $E_{\mathrm{mod}}(w_k) > \kappa\min\{\|\nabla m_k(0)\|_Z,\Delta_k\}$, **do**
  - Select $\xi_{\max}=\arg\max_{j=1,\dots,N} p_j\delta_{\mathrm{mod}}(\xi_j;w_k)$.
  - Compute $S(\xi_{\max};z_k)$, $\Lambda(\xi_{\max};S(\xi_{\max};z_k))$, $\nabla_\xi S(\xi_{\max};z_k)$ and $\nabla_\xi\Lambda(\xi_{\max};S(\xi_{\max};z_k))$.
  - Incorporate the new information at $\xi_{\max}$ into S_{mod} and Λ_{mod} using the local RB method.
  - Update m_k with S_{mod} and Λ_{mod}.
  - Update $\delta_{\mathrm{mod}}(\xi_j;w_k)$ for $j=1,\dots,N$ and $E_{\mathrm{mod}}(w_k)$, and recompute $\|\nabla m_k(0)\|_Z$.

  End

Return S_{mod} and Λ_{mod}.

Similar to the TR subproblem model m_k, we approximate the objective function by

$$\widehat{J}_k(w) = \left(t+\sum_{j=1}^N p_j v(G(S_{\mathrm{obj}}(\xi_j;z),\xi_j)-t)\right)+\wp(z) \tag{19}$$

By similar arguments as above, i.e., using Assumption 4 and multiple applications of the integral mean value theorem, we arrive at the bound

$$|\widehat{J}_k(w) - \widehat{J}(w)| \lesssim \sum_{j=1}^{N} p_j \delta_{\text{obj}}(\xi_j; w) =: E_{\text{obj}}(w) \tag{20}$$

where

$$\delta_{\text{obj}}(\xi_j; w) := \epsilon^u_{\text{obj}}(\xi_j; z)^2 + |v'(G(S_{\text{obj}}(\xi_j; z)) - t)|\epsilon^u_{\text{obj}}(\xi_j; z).$$

Therefore, we can bound the difference between the actual and computed reduction by

$$|\text{ared}_k - \text{cred}_k| \leq |\widehat{J}_k(w_k) - \widehat{J}(w_k)| + |\widehat{J}_k(w_k + s_k) - \widehat{J}(w_k + s_k)| \lesssim E_{\text{obj}}(w_k) + E_{\text{obj}}(w_k + s_k). \tag{21}$$

Again, as long as $|v'(G(S_{\text{mod}}(\xi_j; z_k), \xi_j) - t_k)|$ and $|v'(G(S_{\text{mod}}(\xi_j; z_k + \zeta_k), \xi_j) - (t_k + \tau_k))|$ are bounded for all iterates $w_k = (t_k, z_k)$ and steps $s_k = (\tau_k, \zeta_k)$, we can satisfy (12) by refining our local reduced basis model S_{obj}. The error indicator $E_{\text{obj}}(w_k)$ requires us to accurately approximate the solution to (1) at both w_k and $w_k + s_k$. This motivates us to use a local basis that is composed of solutions and gradients at both w_k and $w_k + s_k$. We employ a similar recycling scheme as above to save computations while refining S_{obj}. The complete algorithm for constructing S_{obj} is listed in Algorithm 2.

Algorithm 2: Adaptive algorithm to build S_{obj} for $\widehat{J}_k$

If $k = 0$**, model initialization**:

- Let $\xi_0 = \mathbb{E}[\xi]$ and build the initial surrogate model S_{mod} based on the solution $S(\xi_0, z_0)$, the gradient $\nabla_\xi S(\xi_0, z_0)$.

Else, model refinement:

Given w_k, s_k, pred_k, r_k and S_{obj}, which is recycled from step $k-1$,

- Compute $\gamma_k = K\,(\eta \min\{\text{pred}_k, r_k\})^{\frac{1}{\omega}}$, evaluate $\delta_{\text{obj}}(\xi_j; w_k)$, $\delta_{\text{obj}}(\xi_j; w_k + s_k)$ for $j = 1, \ldots, N$ and compute $E_{\text{obj}}(w_k)$ and $E_{\text{obj}}(w_k + s_k)$.
- **While** $E_{\text{obj}}(w_k) + E_{\text{obj}}(w_k + s_k) > \gamma_k$, **do**
  - Select $\xi_{\max} = \arg\max_{j=1,\ldots,N} p_j \left(\delta_{\text{obj}}(\xi_j; w_k) + \delta_{\text{obj}}(\xi_j; w_k + s_k)\right)$.
  - Compute $S(\xi_{\max}, z_k)$, $\nabla_\xi S(\xi_{\max}, z_k)$, $S(\xi_{\max}, z_k + \zeta_k)$, and $\nabla_\xi S(\xi_{\max}, z_k + \zeta_k)$.
  - Incorporate the new information at $\xi_{\max}$ into S_{obj} using the local RB method.
  - Update $\delta_{\text{mod}}(\xi_j; w_k)$ and $\delta_{\text{obj}}(\xi_j; w_k + s_k)$ for $j = 1, \ldots, N$, update $E_{\text{obj}}(w_k)$ and $E_{\text{obj}}(w_k + s_k)$.
  - **End**

End

Return S_{obj}.

V. Numerical Results.

The subsequent numerical example demonstrates that the above algorithms are quite efficient (in terms of the number of high-fidelity PDE solves) for constructing surrogate models that meet their respective requirements imposed by the TR algorithm. The efficiency can be largely attributed to our choice of the local bases that specifically aims to reduce the error indicators. The example also highlights the flexibility of the local RB method in terms of the choice of the local basis. We demonstrate the TR algorithm on the 1D advection-diffusion example from [16]. Let $D = (0, 1)$ and consider the weak form of the boundary value problem

$$\begin{aligned} -\nu \frac{\partial^2 u}{\partial x^2}(x, \xi) + b(x, \xi)\frac{\partial u}{\partial x}(x, \xi) &= f(x), \quad x \in X, \text{ a.s.} \\ u(0, \xi) = u(1, \xi) &= 0, \quad \text{a.s.} \end{aligned} \tag{22}$$

The diffusivity, ν, and source, f, are deterministic whereas the advection field, b, is a piecewise constant random field given by

$$b(x, \xi) = [b_1 + \xi_1]\, \mathbb{1}_{[0,0.5)}(x) + [b_2 + \xi_2]\, \mathbb{1}_{[0.5,1]}(x) \tag{23}$$

where $\mathbb{1}_S(x)$ is one if x is in the set S and is zero otherwise. Here, ξ_1 and ξ_2 are independent random variables that are uniformly distributed on the interval $[-1, 1]$. For this problem, $U = V = H^1(D)$.

We define the general objective function as

$$G(u;\xi) = \frac{1}{2}\|\mathcal{D}u - d\|_Y^2 \tag{24}$$

where Y is a Hilbert space, $\mathcal{D} \in L(U, Y)$ is an observation operator, and $d \in Y$ is a desired profile. For this problem, we set $Y = \mathbb{R}^9$ and $\mathcal{D}$ to evaluate the PDE solution u at nine equally spaced points $x \in \{0.1, 0.2, \ldots, 0.9\}$. Note that since $D \subset \mathbb{R}$, U is continuously embedded in to $C(D)$ [1] (i.e., $u \in U$ is a continuous function) and thus $\mathcal{D}$ is a bounded linear operator. The target vector d is simply $3\sin(x)$ evaluated at the same set of points. Moreover, the control penalty is

$$\wp(z) = \frac{\alpha}{2}\|z\|_Z^2, \quad \alpha > 0 \tag{25}$$

and the control is assumed to be piecewise constants, i.e., $z(x) = \sum_{i=1}^{10} z_i \mathbb{1}_{I_i}(x)$ where $I_i = (0.1(i-1), 0.1i)$, $i = 1, \ldots, 10$.

We solved the risk-averse optimization problem (6) with $\mathcal{R}$ set to a smooth approximation of CVaR at confidence level $\beta = 0.9$ [11] using the adaptive local TR/RB algorithm. We used the following TR parameters: $\eta_1 = 0.05$, $\eta_2 = 0.75$, $\gamma = 0.5$, $\omega = 0.75$, $\mu = 10$, $K = 1$ and $\Delta_0 = 3$. We set the penalty parameter to $\alpha = 0.1$ and employed $N = 5{,}000$ Monte Carlo samples to approximate ξ. Finally, we used 3 neighbors to construct the local bases Φ_k.

In Figure 2, we depict the optimal controls (left image) and the cumulative distribution functions (CDFs) with the 0.9-quantile and 0.9-CVaR of $G(u(\xi), \xi)$ (right image) obtained by the TR method (red) and by solving the full optimization problem using Monte Carlo (blue). In Figure 3, we depict the distribution of $G(u(\xi), \xi)$ over the entire

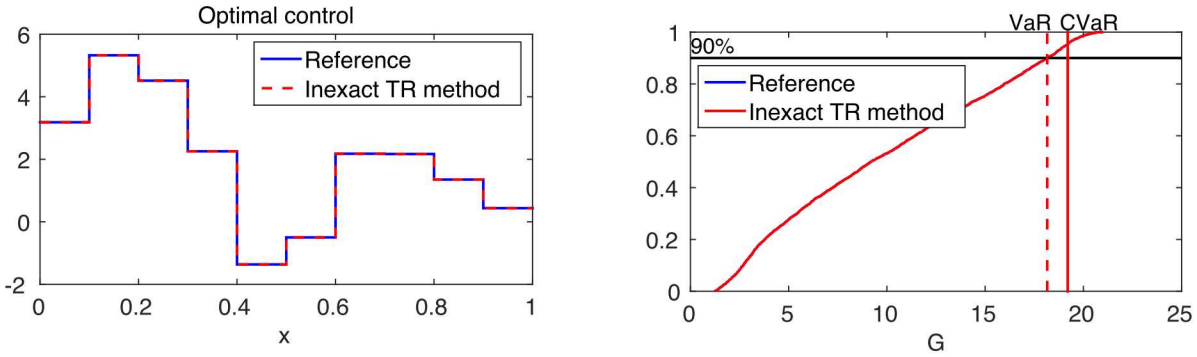

Fig. 2 The left image provides a comparison of the optimal controls computed by the proposed TR/RB method (red) and those computed by solving the full-order Monte Carlo reference problem (blue). The right image depicts the CDF of $G(u, \xi)$ evaluated at the optimal state from the TR/RB method (red) and the optimal state from the full-order Monte Carlo problem (blue). The dashed vertical lines correspond to the 0.9-quantile whereas the solid vertical lines correspond to the 0.9-CVaR. These lines (blue for the reference solution and red for the TR/RB solution) are virtually indistinguishable.

parameter space evaluated at the optimal control z_{opt} obtained using the full model $S(\xi; z_{\text{opt}})$ and the adaptive RB model $S_{\text{obj}}(\xi; z_{\text{opt}})$. The adaptive TR results are virtually indistinguishable from the reference Monte Carlo solution. In Figure 4, we plot the atoms of the surrogate models S_{mod} (black) and S_{obj} (red) which we used to construct m_k and $\widehat{J}_k$, respectively. At each atom, the surrogate model construction required six high-fidelity solves for the entire optimization process. We compare the total number of high-fidelity solves required for the reference Monte Carlo solution and the total number of high-fidelity solves required for our adaptive TR method in Figure 5. Solving the full optimization problem with Monte Carlo results in nearly 1,000,000 high-fidelity evaluations. In comparison, the total number of high-fidelity evaluations for our TR approach is less than 1,000. Finally, we plot the norm of the true gradient at each iteration of the adaptive TR algorithm in Figure 6. Notice that the norm of true gradient drops more than five orders of magnitude by the fourth iteration, indicating that the algorithm makes significant progress towards satisfying the necessary optimality conditions in few iterations and high-fidelity evaluations.

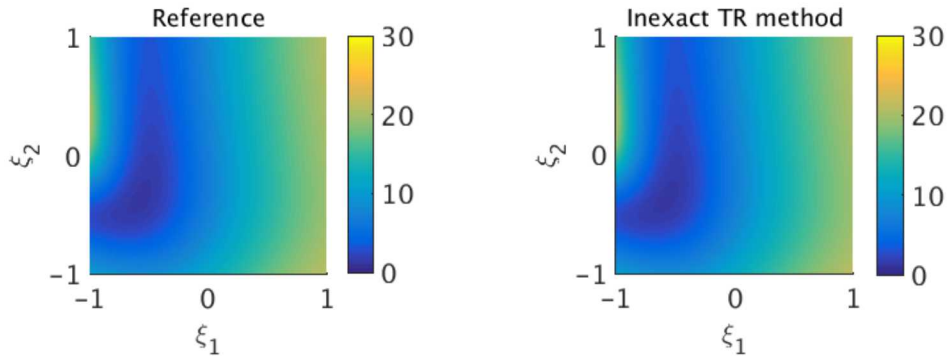

Fig. 3 The uncertain objective function $G(u, \xi)$ evaluated at the optimal states from the full-order Monte Carlo reference problem (left) and the proposed TR/RB method (right).

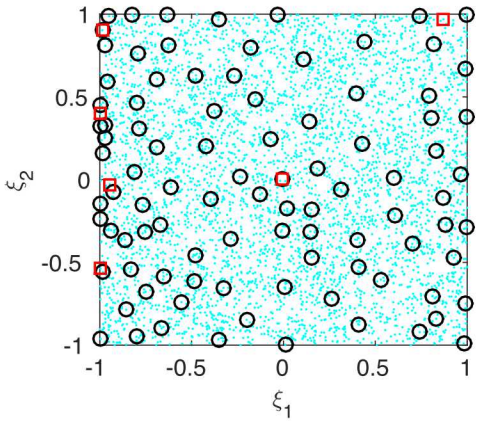

Fig. 4 The atoms of the surrogate models S_{mod} (black) and S_{obj} (red) that are used to build the models m_k and $\widehat{J}_k$, respectively. The blue dots are the 5,000 Monte Carlo samples used for the reference problem.

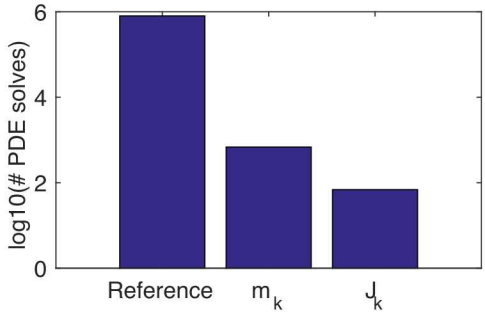

Fig. 5 A comparison of the total number of full-order PDE solves for the reference Monte Carlo approach and our proposed TR/RB approach.

VI. Conclusion

In this work, we described an approach for solving risk-averse PDE-constrained optimization problems. Our approach leverages two essential components to achieve computational efficiency: (i) we employed the local RB method [16] to build surrogate models to approximate the expensive objective value and gradient computations; and (ii) we employed an inexact TR algorithm [10] to control the surrogate approximation accuracy and achieve convergence. We derived error indicators to guide the adaptive refinement of the surrogate models in an efficient manner. Also, to ensure

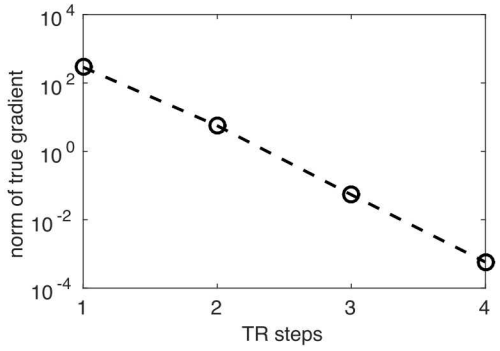

Fig. 6 The norm of the true gradient evaluated at the iterates produced by our proposed TR/RB approach.

that our surrogate models rapidly satisfy the prescribed accuracy of the TR method, we employed local bases that were specifically tailored for these error requirements.

In our numerical example, we applied the local reduced basis TR method to a risk-averse optimal control problem constrained by a 1D advection-diffusion equation with a random advection field. We demonstrated the efficiency of our method in terms of number of full PDE solves required to achieve a highly accurate solution. Our TR/RB approach obtained virtually indistinguishable optimal controls with less than 1,000 full PDE solves when solving the full problem without surrogate approximation required nearly 1,000,000 full PDE solves. This cost reduction is significant. The developed method has the potential to be applied to more complicated and realistic problems in the future.

References

[1] Adams, R. A. and Fournier, J. J.-F. *Sobolev Spaces*, second ed. Elsevier, Amsterdam, 2008.

[2] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. "Coherent measures of risk", *Mathematical Finance*, 9:203-228, 1999.

[3] Ben-Tal, A. and Teboulle, M. "An Old-New Concept of Convex Risk Measures: The Optimized Certainty Equivalent", *Mathematical Finance*, 17(3):449-476, 2007.

[4] Carstensen, C., Hoppe, R.H.W., Löbhard, C. and Eigel, M. "A review of unified a posteriori finite element error control", Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN, October 2010.

[5] Conn, A. R., Gould, N. I. M. and Toint, Ph. L. *Trust–Region Methods*. SIAM, Philadelphia, 2000.

[6] Ern, A. and Guermond, J. L. *Theory and Practice of Finite Elements*, in *Applied Mathematical Scineces*. Springer, New York, 2013.

[7] Goldberg, H., Kampowsky, W. and Tröltzsch, F. On Neymtskij operators in Lp-spaces of abstract functions. *Mathematische Nachrichten 155*, 1:127-140, 1992.

[8] Himmelberg, C. J. Measurable relations. *Fundamenta Mathematicae 87*, 1:53-72, 1975.

[9] Krokhmal, P., Palmquist, J. and Uryasev, S. "Portfolio optimization with conditional value-at-risk objective and constraints", *Journal of risk*, 4:43-68, 2002.

[10] Kouri, D.P., Heinkenschloss, M., Ridzal, D. and van Bloemen Waanders, B.G. "Inexact Objective Function Evaluations in a Trust-Region Algorithm for PDE-Constrained Optimization under Uncertainty", *SIAM Journal of Scientific Computing*, 36(6):A3011-A3029, 2014.

[11] Kouri, D.P. and Surowiec, T.M. "Risk-Averse PDE-constrained Optimization using the Conditional Value-at-Risk", *SIAM Journal of Optimization*, 26(1):365-396, 2016.

[12] Rockafellar, R. T. and Uryasev S. "Optimization of conditional value-at-risk", *The Journal of Risk*, 2:21–41, 2000.

[13] Rockafellar, R. T. and Uryasev, S. "The fundamental risk quadrangle in risk management, optimization and statistical estimation", *Surveys in Operations Research and Management Science*, 18:33–53, 2013.

[14] Rozza, G., Huynh, D. B. P. and and Patera, A. T. "Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equation", *Archives of Computational Methods in Engineering*, 15(3):229-275, 2008.

[15] Zahr, M. J. "Adaptive Model Reduction to Accelerate Optimization Problems Governed by Partial Differential Equations", Doctoral dissertation, Stanford University, 2016.

[16] Zou, Z., Kouri, D.P. and Aquino, W. "An Adaptive Sampling Approach for Solving PDEs with Uncertain Inputs and Evaluating Risk", *19th AIAA non-deterministic approaches conference*, p. 1325, 2017.