Taylor approximation for chance constrained optimization problems governed by partial differential equations with high-dimensional random parameters *

Peng Chen† and Omar Ghattas ‡

Abstract. We propose a fast and scalable optimization method to solve chance or probabilistic constrained optimization problems governed by partial differential equations (PDEs) with high-dimensional random parameters. To address the critical computational challenges of expensive PDE solution and high-dimensional uncertainty, we construct surrogates of the constraint function by Taylor approximation, which relies on efficient computation of the derivatives, low rank approximation of the Hessian, and a randomized algorithm for eigenvalue decomposition. To tackle the difficulty of the non-differentiability of the inequality chance constraint, we use a smooth approximation of the discontinuous indicator function involved in the chance constraint, and apply a penalty method to transform the inequality constrained optimization problem to an unconstrained one. Moreover, we design a gradient-based optimization scheme that gradually increases smoothing and penalty parameters to achieve convergence, for which we present an efficient computation of the gradient of the approximate cost functional by the Taylor approximation. Based on numerical experiments for a problem in optimal groundwater management, we demonstrate the accuracy of the Taylor approximation, its ability to greatly accelerate constraint evaluations, the convergence of the continuation optimization scheme, and the scalability of the proposed method in terms of the number of PDE solves with increasing random parameter dimension from one thousand to hundreds of thousands.

Key words. chance constrained optimization, Taylor approximation, randomized algorithm, smooth approximation, penalty method, Hessian

AMS subject classifications. 65C20, 65D32, 65N12, 49J20, 93E20

1. Introduction.

Large-scale simulation in computational science and engineering is often carried out not only to obtain insight about a system, but also as a basis for decision-making. When the decision variables represent the design or control or data-driven inference of model parameters of an engineered or natural system, and the system is governed by partial differential equations (PDEs), the task of determining the optimal design, optimal control, or inversion parameters leads to a PDE-constrained optimization problem. Over the past several decades, research in the field of PDE-constrained optimization has exploded, and powerful theory and algorithms are now available in the case of optimization governed by deterministic PDEs (e.g., see the monographs [44,48,61,81]). However, many PDE models are characterized by random parameters due to lack of knowledge or intrinsic variability. These include initial or boundary conditions, sources, coefficients, and geometry. In many of these cases, the uncer-

* **Funding:** This research was partially funded by the National Science Foundation, Division of Mathematical Sciences under award DMS-2012453; the Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Mathematical Multifaceted Integrated Capability Centers (MMICCS) program under award DE-SC0019303; and the Simons Foundation under award 560651.

† Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX 78712. (peng@oden.utexas.edu).

‡ Oden Institute for Computational Engineering and Sciences, Department of Mechanical Engineering, and Department of Geological Sciences, The University of Texas at Austin, Austin, TX 78712. (omar@oden.utexas.edu)

tainty arises from an (infinite-dimensional) random field, leading to high-dimensional random parameters after discretization. It is critical to incorporate this uncertainty in the optimization problem to make the optimal solution more reliable and robust. Optimization under uncertainty has become an important research area and received increasing attentions in recent years [3,5,6,8,11,21,24–26,28,31–33,35,37,39–43,45,46,49,51–59,63,65,67,73,74,78,80,82–86,88–90]. To account for the uncertainty in the optimization problem, different statistical measures of the objective function have been studied, e.g., mean, variance, conditional value-at-risk, worst case scenario, etc., [3, 39, 53, 54, 57, 74, 89]. Moreover, the treatment of chance constraints, also known as probabilistic constraints, i.e., the probability that a certain function exceeds a threshold or below a certain level, has also been investigated [32, 37, 43, 46, 73, 78, 82, 88]. Several computational challenges arise in solving optimization problems under uncertainty, especially with high-dimensional random parameters and inequality chance constraints.

The first prominent challenge is that high-fidelity discretizations of the (nonlinear) PDEs often lead to large-scale (nonlinear) algebraic systems that are extremely expensive to solve in practical applications. Therefore, only a limited number of high-fidelity PDE solves can be afforded. This challenge prevents direct application of most of the conventional numerical methods for computing statistics of the objective function, since they require a large number of evaluations of the objective function and thus the PDE solution. To tackle this challenge, multigrid, multilevel, and model reduction methods have been successfully applied to solve stochastic PDE-constrained optimization problems [5,6,10,24,26,57,67,89,90]. The multigrid discretization and multilevel statistical evaluation rely on a hierarchical discretization of the PDE model and an efficient algorithm to balance the discretization error and the number of samples required for statistical evaluation at each level. However, due to the nature of the problem, it is not always possible to use multigrid discretizations or gain computational savings by multilevel sampling because a sufficiently fine grid must be used to solve the PDE model with reasonable accuracy (as is often the case with hyperbolic or multiscale problems). Meanwhile, model reduction techniques become problematic for highly nonlinear problems that require effective affine approximation or when the solution manifold becomes high-dimensional, even if the objective function lives in a low-dimensional manifold.

The second key challenge arises in computing the statistical measures, which involves integration of the objective function with respect to (w.r.t.) the probability measure of the high-dimensional random parameters. A classical approach known as sample average approximation (SAA) or Monte Carlo quadrature is to take the average of the objective function at a set of samples randomly drawn from the probability measure of the random parameters. However, its convergence rate is only $O(M^{-1/2})$, where an often expensive PDE has to be solved for each of the M samples. The resulting deterministic optimization problem has M PDE constraints that need to be solved to evaluate the objective function, and as such is typically prohibitive to solve. In recent years, stochastic Galerkin and stochastic collocation based integration methods have been used to compute the statistical moments (e.g., mean and variance) of the objective function in stochastic optimization [10,24–26,45,49,52,55,56,74,80], provided that a suitable finite dimensional parametrization of the random parameters, such as a truncated Karhunen–Loève expansion, is available. These methods achieve fast convergence when the objective function depends smoothly on the low-dimensional parameters, but suffer from the so-called curse of dimensionality, i.e. the convergence rate quickly deteriorates as

the parameter dimension increases. More recent advances in adaptive and anisotropic sparse quadrature [19,77] and high order quasi Monte Carlo methods [36] have been shown to achieve a convergence rate of $O(M^{-s})$, with s potentially much larger than $1/2$ and independent of the nominal dimension of the random parameters, thus mitigating the curse of dimensionality. However, if the objective function is not sufficiently smooth or sufficiently anisotropic in the parameter space, the convergence of these methods becomes very slow, or worse than that of Monte Carlo.

The third critical challenge comes from the non-differentiability of the chance constraint, which involve integration of a discontinuous indicator function. The discontinuity makes the cost functional non-differentiable w.r.t. the optimization variable, so that rapidly convergent derivative based optimization methods, e.g., steepest descent or Newton methods, cannot be directly applied. Therefore, solving optimization problems with such non-differentiable constraints not only requires a large number of PDE solves at each optimization iteration, but also requires a large number of optimization iterations, especially when the optimization variable dimension is high. To address this challenge, proper smoothing techniques have been employed to approximate the indicator function by differentiable functions [18, 72, 82], which introduces smoothing errors in the cost functional and optimal solution. However considerable difficulties are still encountered in finding numerical approximations that satisfy the probability constraints in the presence of high-dimensional random parameters.

Contributions. In this work we address the above computational challenges by proposing a Taylor approximation based continuation optimization method to solve chance constrained optimization problems governed by PDEs with high-dimensional random parameters. Following our recent work [3, 28], we extend the Taylor approximation—including constant, linear, and quadratic approximations—of the objective function used to evaluate its mean, to approximation of the constraint function used to accelerate the evaluation of the chance/probability of the constraint function. In particular, a double-pass randomized algorithm is employed to solve a generalized eigenvalue problem, where the eigenvalues and eigenfunctions are used to construct the quadratic term of the Taylor approximation. A set of linearized PDEs are derived for the computation of the gradient and action of the Hessian of the objective and constraint functions w.r.t. the random parameters. To solve the optimization problem, we present a continuation BFGS algorithm, which features (1) smooth approximation of the indicator function, (2) a penalty method to transform the inequality constrained optimization problem to an unconstrained one, (3) a continuation scheme with an outer loop of increasing the smoothing and penalty parameters and an inner loop of BFGS optimization. The computation of the approximate cost functional and its gradient w.r.t. the optimization variable are presented in detail for both SAA and Taylor approximation. For the proposed method, we demonstrate (1) the accuracy of the Taylor approximations, (2) the efficiency on the surrogate acceleration, (3) the convergence of the continuation optimization algorithm, and (4) the independence of the number of PDE solves from increasing random parameter dimension. The demonstrations are carried out by numerical experiments for an example of water management in agricultural irrigation, where the PDE model is a Darcy flow equation that describes groundwater flow in the presence of an uncertain permeability field. The optimization objective is to extract water at given well locations that meets a target extraction rate, while a

chance constraint is imposed on an integrated pressure field to prevent a pressure that is low enough to lead to collapse or damage of the aquifer.

Notations. Let $\mathcal{X}$ be a Banach space and $\mathcal{X}'$ the dual space; ${}_{\mathcal{X}}\langle\cdot,\cdot\rangle_{\mathcal{X}'}$ then denotes the duality pairing between the spaces $\mathcal{X}$ and $\mathcal{X}'$. For ease of notation, we will omit specification of the subscritps $\mathcal{X}$ and $\mathcal{X}'$ and simply write $\langle\cdot,\cdot\rangle$ when the spaces can be inferred from the context without ambiguity. Given two Banach spaces $\mathcal{X}$ and $\mathcal{Y}$ and a map $f : \mathcal{X}\times\mathcal{Y}\mapsto\mathbb{R}$, $\partial_x f(x,y)\in\mathcal{X}'$ denotes the Fréchet derivative of $f(x,y)$ with respect to x evaluated at (x,y), which satisfies

$$\lim_{\tilde{x}\to 0}\frac{f(x+\tilde{x},y)-f(x,y)-\langle\tilde{x},\partial_x f(x,y)\rangle}{||\tilde{x}||_{\mathcal{X}}}=0. \tag{1.1}$$

Let $\partial_{xy}f(x,y) : \mathcal{Y}\mapsto\mathcal{X}'$ denote the Fréchet derivative of $\partial_x f(x,y)$ with respect to y evaluated at (x,y), or the second order (mixed) Fréchet derivative of $f(x,y)$ with respect to x and y evaluated at (x,y). Similarly, $\partial_{yx}f(x,y) : \mathcal{X}\mapsto\mathcal{Y}'$ denotes the Fréchet derivative of $\partial_y f(x,y)$ with respect to x evaluated at (x,y), which is the adjoint operator of $\partial_{xy}f(x,y)$ and satisfies

$$\langle\tilde{x},\partial_{xy}f\,\hat{y}\rangle := {}_{\mathcal{X}}\langle\tilde{x},\partial_{xy}f\,\hat{y}\rangle_{\mathcal{X}'} = {}_{\mathcal{Y}}\langle\hat{y},\partial_{yx}f\,\tilde{x}\rangle_{\mathcal{Y}'} =: \langle\hat{y},\partial_{yx}f\,\tilde{x}\rangle,\quad \forall\tilde{x}\in\mathcal{X},\,\hat{y}\in\mathcal{Y}, \tag{1.2}$$

where we have omitted the argument (x,y) for simplicity.

The rest of the paper is organized as follows. In Section 2 we present the general formulation of PDE and chance constrained optimization problems, which is followed by Section 3 on SAA, Taylor approximation, a randomized algorithm for low rank approximation, and computation of the gradient and Hessian action of the objective and constraint function w.r.t. the random parameters. Section 4 is devoted to the presentation of a continuation gradient-based optimization method that involves smooth approximation of the indicator function, a penalty method for the inequality constraint, a continuation scheme to increase the smoothing and penalty parameters, and the computation of the gradient of the approximate cost functional w.r.t. the optimization variable. Numerical experiments and results are reported in Section 5 for the demonstration of the accuracy, efficiency, convergence, and scalability of the proposed method. Conclusions and perspectives are drawn in Section 6.

2. Chance constrained optimization.

We consider a system to be optimized under uncertainty, which is modeled by PDE presented in an abstract (residual) form as: find $u\in\mathcal{U}$, such that

$$\mathcal{R}(u,m,z)=0\quad\text{in }\mathcal{V}', \tag{2.1}$$

where $m\in\mathcal{M}$ is an uncertain or random parameter field that lives in a separable Banach space $\mathcal{M}$, which has a probability distribution μ; $z\in\mathcal{Z}$ is an optimization variable in a separable Banach space $\mathcal{Z}$; and $\mathcal{R}(\cdot,m,z):\mathcal{U}\mapsto\mathcal{V}'$ denotes a (possibly nonlinear) operator from $\mathcal{U}$ to $\mathcal{V}'$, the dual of $\mathcal{V}$, where $\mathcal{U}$ and $\mathcal{V}$ are two separable Banach spaces. The weak form of (2.1) is given by means of duality pairing as: find $u\in\mathcal{U}$, such that

$$r(u,v,m,z) := {}_{\mathcal{V}}\langle v,\mathcal{R}(u,m,z)\rangle_{\mathcal{V}'}=0\quad\forall v\in\mathcal{V}, \tag{2.2}$$

where v is a test variable or an adjoint variable in the optimization context. By definition, $r(u, v, m, z)$ is linear with respect to (w.r.t.) v, and may be nonlinear w.r.t. u, m, and z.

By $q : \mathcal{U} \times \mathcal{M} \times \mathcal{Z} \mapsto \mathbb{R}$ and $f : \mathcal{U} \times \mathcal{M} \times \mathcal{Z} \mapsto \mathbb{R}$ we denote an objective function and a constraint function for the optimization as real-valued, continuous, and possibly nonlinear maps of u, m, z. Since u depends on m and z through (2.2), for simplicity we write $q(m, z) = q(u(m, z), m, z)$ and $f(m, z) = f(u(m, z), m, z)$ by slight abuse of notation.

For the optimization problem, we consider a cost functional

$$\mathcal{J}(z) = \mathbb{E}[q(\cdot, z)] + \mathcal{P}(z) \tag{2.3}$$

where the first term is the mean of the objective function q defined as

$$\mathbb{E}[q(\cdot, z)] = \int_{\mathcal{M}} q(m, z) d\mu(m), \tag{2.4}$$

the second term $\mathcal{P}(z)$ represents a penalty or regularization term for the optimization variable z. Moreover, we consider a chance constraint

$$P(f(\cdot, z) \geq 0) \leq \alpha \tag{2.5}$$

for a critical chance $0 < \alpha < 1$, where the probability is given by

$$P(f(\cdot, z) \geq 0) = \mathbb{E}[\mathbb{I}_{[0,\infty)}(f(\cdot, z))] = \int_{\mathcal{M}} \mathbb{I}_{[0,\infty)}(f(m, z)) d\mu(m), \tag{2.6}$$

where $\mathbb{I}_{[0,\infty)}(f(m, z))$ is an indicator function defined as

$$\mathbb{I}_{[0,\infty)}(f(m, z)) = \begin{cases} 1 & \text{if } f(m, z) \geq 0, \\ 0 & \text{if } f(m, z) < 0. \end{cases} \tag{2.7}$$

Then the PDE and chance constrained optimization problem can be formulated as

$$\min_{z \in \mathcal{Z}} \mathcal{J}(z), \text{ subject to the chance constraint (2.5).} \tag{2.8}$$

3. Taylor approximation.

We first present a sample average approximation (SAA) for the optimization problem (2.8). Then we introduce an (up to quadratic) Taylor approximation for both the objective function and the constraint function, which requires an efficient eigenvalue decomposition of the Hessian of the objective and constraint functions w.r.t. the random parameter field. We present a double-pass randomized algorithm for this task, which requires only actions of the Hessian in random directions without direct access to the entries of the Hessian matrix.

3.1. Sample average approximation.

The mean of the objective function can be evaluated by the sample average approximation (SAA)

$$\mathbb{E}[q(\cdot, z)] \approx q_M(z) := \frac{1}{M_q} \sum_{i=1}^{M_q} q(m_i, z), \tag{3.1}$$

where m_i, $i = 1, \ldots, M_q$, are independent identically distributed (i.i.d.) random samples drawn from the probability distribution μ. Similarly, the chance constraint (2.6) can be approximated by

$$P(f(\cdot, z) \geq 0) \approx f_M(z) := \frac{1}{M_f} \sum_{i=1}^{M_f} \mathbb{I}_{[0,\infty)}(f(m_i, z)), \tag{3.2}$$

where m_i, $i = 1, \ldots, M_f$, are i.i.d. random samples drawn from μ. Note that to obtain an accurate approximation q_M and especially f_M for α close to 0, a large number of samples are required due to the slow convergence (with rate $O(M^{-1/2})$) of the SAA approximation, thus making this approach computationally prohibitive if the PDE solve at one sample is expensive.

3.2. Taylor approximation. We assume that the objective function q admits the k-th order Fréchet derivative with respect to m at $\bar{m} \in \mathcal{M}$, denoted as $\nabla_m^k q(\bar{m}, z)$, for $k = 1, \ldots, K$. A K-th order Taylor expansion of the objective function q evaluated at $\bar{m} \in \mathcal{M}$ is given by

$$T_K q(m, z) = \sum_{k=0}^{K} \frac{1}{k!} \nabla_m^k q(\bar{m}, z)(m - \bar{m})^k. \tag{3.3}$$

Note that $\nabla_m^k q(\bar{m}, z) : \mathcal{M}^k \mapsto \mathbb{R}$ is a multilinear map defined in the tensor-product space

$$\mathcal{M}^k = \prod_{i=1}^{k} \mathcal{M}_i, \text{ where } \mathcal{M}_i = \mathcal{M}. \tag{3.4}$$

For the random field m, we consider a Gaussian measure $\mu = \mathcal{N}(\bar{m}, \mathcal{C})$ with mean $\bar{m}$ and covariance $\mathcal{C}$ given by a pseudodifferential operator [60, 79], see an example later in (5.7), which plays a dominant role in spatial statistics. For such a measure μ, we have the analytic expression [3]

$$\mathbb{E}[q(\cdot, z)] \approx \mathbb{E}[T_K q(\cdot, z)] = \begin{cases} q(\bar{m}, z), & K = 0, 1, \\ q(\bar{m}, z) + \frac{1}{2}\text{trace}\left(\mathcal{H}_q\right), & K = 2, \end{cases} \tag{3.5}$$

where $\text{trace}(\mathcal{H}_q)$ represents the trace of $\mathcal{H}_q = \mathcal{C}^{1/2} \nabla_m^2 q(\bar{m}, z) \mathcal{C}^{1/2}$, which is the covariance-preconditioned Hessian of the objective function q. We remark that (3.5) also holds for non Gaussian measure μ as long as the random field m has covariance $\mathcal{C}$.

Similar to (3.3), under the assumption that the constraint function f admits the k-th order Fréchet derivative with respect to m at $\bar{m} \in \mathcal{M}$ for $k = 1, \ldots, K$, we construct a K-th order Taylor expansion of the constraint function f at $\bar{m}$, denoted $T_K f$. Then we can approximate the probability $P(f(m, z) \geq 0)$ by SAA (3.2) with the Taylor approximation $T_K f$ as

$$P(f(m, z) \geq 0) \approx f_M^K(z) := \frac{1}{M_f} \sum_{i=1}^{M_f} \mathbb{I}_{[0,\infty)}(T_K f(m_i, z)). \tag{3.6}$$

The attractiveness of the Taylor approximations of the objective and constraint, (3.5) and (3.6), is that once they are constructed, no further PDE solves are required. In the next section, we shall see how these Taylor approximations can be efficiently constructed.

3.3. Low-rank approximation. To compute the trace in (3.5), we can use

$$\text{trace}(\mathcal{H}_q) = \sum_{n=1}^{\infty} \lambda_n^q, \tag{3.7}$$

where $(\lambda_n^q)_{n\geq 1}$ are the eigenvalues of $\mathcal{H}_q = \mathcal{C}^{1/2}\nabla_m^2 q(\bar{m}, z)\mathcal{C}^{1/2}$, which are equivalent to the generalized eigenvalues of $(\nabla_m^2 q(\bar{m}, z), \mathcal{C}^{-1})$. In practice, the (absolute) eigenvalues decay rapidly, $|\lambda_n^q| \to 0$ as $n \to \infty$, as proven for some model problems and demonstrated numerically for many others in forward uncertainty quantification, Bayesian inversion and experimental design, and stochastic optimization [1–3, 7, 12–17, 20, 22, 23, 27–30, 34, 38, 50, 64, 71, 87]. This is because the function q is typically not equally sensitive to all directions in the parameter space, as reflected by the decay of the eigenvalues of its Hessian $\nabla_m^2 q(\bar{m}, z)$, which measures the curvature of the function. Moreover, the covariance $\mathcal{C}$ is often assumed to be compact with its eigenvalues converging to zero [60, 79]. Given rapid decay of the eigenvalues, we can approximate the trace by the N_q largest (in absolute value) eigenvalues λ_n^q, $n = 1, \ldots, N_q$, i.e.,

$$\text{trace}(\mathcal{H}_q) \approx \sum_{n=1}^{N_q} \lambda_n^q. \tag{3.8}$$

For computational efficiency, we consider the generalized eigenvalues of $(\nabla_m^2 q(\bar{m}, z), \mathcal{C}^{-1})$ by solving the generalized eigenvalue problem: find (λ_n^q, ψ_n^q), $n = 1, \ldots, N_q$, such that

$$\nabla_m^2 q(\bar{m}, z)\psi_n^q = \lambda_n^q \mathcal{C}^{-1}\psi_n^q, \quad n = 1, \ldots, N_q, \tag{3.9}$$

with $|\lambda_1^q| \geq \cdots \geq |\lambda_{N_q}^q|$ corresponding to the N_q largest eigenvalues in absolute value and the eigenfunctions satisfying the orthonormality condition with respect to $\mathcal{C}^{-1}$, i.e.,

$$\langle \psi_n^q, \mathcal{C}^{-1}\psi_{n'}^q \rangle = \delta_{nn'}, \quad n, n' = q, \ldots, N_q. \tag{3.10}$$

We remark that the action of $\mathcal{C}^{-1}$ in a given direction ψ, i.e., $\mathcal{C}^{-1}\psi$, can be efficiently computed by solving an elliptic differential equation [60].

To compute the Taylor approximation $T_K f(m, z)$ at $\bar{m}$ for $K = 2$, we need to evaluate $\nabla_m^2 f(\bar{m}, z)(m - \bar{m})^2$. If the eigenvalues of $\mathcal{H}_f = \mathcal{C}^{1/2}\nabla_m^2 f(\bar{m}, z)\mathcal{C}^{1/2}$ decay rapidly to 0, we can approximate $\nabla_m^2 f(\bar{m}, z)(m - \bar{m})^2$ by a low-rank approximation as

$$\nabla_m^2 f(\bar{m}, z)(m - \bar{m})^2 \approx \sum_{n=1}^{N_f} \lambda_n^f \langle m - \bar{m}, \mathcal{C}^{-1}\psi_n^f \rangle^2, \tag{3.11}$$

where (λ_n^f, ψ_n^f), $n = 1, \ldots, N_f$, are the solution of the generalized eigenvalue problem

$$\nabla_m^2 f(\bar{m}, z)\psi_n^f = \lambda_n^f \mathcal{C}^{-1}\psi_n^f, \quad n = 1, \ldots, N_f, \tag{3.12}$$

with $|\lambda_1^f| \geq \cdots \geq |\lambda_{N_f}^f|$ corresponding to the N_f largest eigenvalues in absolute value and the eigenfunctions satisfying the orthonormality condition with respect to $\mathcal{C}^{-1}$, i.e.,

$$\langle \psi_n^f, \mathcal{C}^{-1}\psi_{n'}^f \rangle = \delta_{nn'}, \quad n, n' = 1, \ldots, N_f. \tag{3.13}$$

With the low-rank (LR) decomposition of the Hessian $\nabla_m^2 f(\bar{m})$ in (3.11), we define the quadratic Taylor approximation of f corresponding to (3.3) as

$$T_2^{\mathrm{LR}} f(m,z) := f(\bar{m},z) + \nabla_m f(\bar{m},z)(m-\bar{m}) + \frac{1}{2}\sum_{n=1}^{N_f} \lambda_n^f \langle m-\bar{m}, \mathcal{C}^{-1}\psi_n^f\rangle^2. \tag{3.14}$$

To solve the generalized eigenvalue problems (3.9) and (3.12) for the dominant eigenvalues, we apply a double-pass randomized algorithm [47, 76], presented in Algorithm 3.1. Here, by H and C^{-1} of size $N_h \times N_h$ each, we denote discrete approximations of the Hessians $\nabla_m^2 q$ or $\nabla_m^2 f$ and the covariance $\mathcal{C}^{-1}$, e.g., by a finite element method. In Algorithm 3.1 only the action of H and C on a given vector is required, which does not require access to the entries of H and C.

Algorithm 3.1 Double-pass randomized eigensolver for (H, C^{-1})

Input: the number of desired eigenpairs N, an oversampling factor $c \leq 10$.
Output: (Λ_N, Ψ_N) with $\Lambda_N = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ and $\Psi_N = (\psi_1, \ldots, \psi_N)$.
1. Draw a Gaussian random matrix $\Omega \in \mathbb{R}^{N_h \times (N+c)}$.
2. Compute $Y = C(H\Omega)$.
3. Compute QR-factorization $Y = QR$ such that $Q^\top C^{-1} Q = I_{N+c}$.
4. Form $T = Q^\top H Q \in \mathbb{R}^{(N+c)\times(N+c)}$ and compute eigendecomposition $T = S\Lambda S^\top$.
5. Extract $\Lambda_N = \Lambda(1:N, 1:N)$ and $\Psi_N = QS_N$ with $S_N = S(:, 1:N)$.

In the next section, we present the computation of the Hessian actions $H\Omega$ and HQ, which dominates the cost in Algorithm 3.1, by solving linearized PDEs. As observed in [**?**, 20, 28] and in the numerical results in Section 5, the advantages of Algorithm 3.1 are: (i) the error in the approximation of the eigenvalues λ_n, $n = 1, \ldots, N$, is bounded by the remaining ones λ_n, $n > N$, which is small if they decay rapidly; (ii) the computational cost is dominated by $2(N + c)$ Hessian actions, where the multiplication of C with a vector is inexpensive, e.g., for C discretized from a differential operator as in (5.7), it takes only $O(N_h)$ operations by a multigrid solver or using multigrid as preconditioner for an iterative solver; (iii) it is scalable in terms of the number of PDEs to solve, because N typically does not change when N_h increases; and (iv) computing the Hessian actions $H\Omega$ and HQ can be asynchronously parallelized.

3.4. Computation of the gradient and Hessian action. For a given optimization variable z, we first compute $u(\bar{m})$ by solving the state equation (2.2) at $\bar{m}$, which can be equivalently written as: find $u \in \mathcal{U}$ such that

$$\langle \tilde{v}, \partial_v \bar{r}\rangle = 0 \quad \forall \tilde{v} \in \mathcal{V}, \tag{3.15}$$

where for ease of notation we use $\bar{r}$ to represent the weak form (2.2) at $\bar{m}$, i.e.,

$$\bar{r} = r(u, v, \bar{m}, z). \tag{3.16}$$

Then we can evaluate the objective function $q(\bar{m}, z) = q(u(\bar{m}, z), \bar{m}, z)$ and the constraint function $f(\bar{m}, z) = f(u(\bar{m}, z), \bar{m}, z)$. The gradient and Hessian of q is computed the same way as for f, so we present only the derivation for q and then state the result for f.

We use the Lagrangian formalism to derive the expressions for the gradient $\nabla_m q(\bar{m}, z)$ evaluated at $\bar{m}$ and for the action of the Hessian $\nabla^2_m q(\bar{m}, z)$ evaluated at $\bar{m}$ in given direction $\hat{m}$, as in [48]. First, we define the Lagrangian functional

$$\mathcal{L}(u, v^q, m, z) := q(u, m, z) + r(u, v^q, m, z), \tag{3.17}$$

where the adjoint v^q is the Lagrange multiplier for the state equation in the computation w.r.t. the objective function q. In what follows, for ease of notation, we define

$$\bar{r}_q = r(u, v^q, \bar{m}, z), \text{ and } \bar{q} = q(u(\bar{m}, z), \bar{m}, z). \tag{3.18}$$

By setting the first order variation of (3.17) at $\bar{m}$ with respect to the state u to zero, we obtain the adjoint problem: find $v^q \in \mathcal{V}$, such that

$$\langle \tilde{u}, \partial_u \bar{r}_q \rangle = -\langle \tilde{u}, \partial_u \bar{q} \rangle, \quad \forall \tilde{u} \in \mathcal{U}. \tag{3.19}$$

Then, the gradient of q at $\bar{m}$ acting in direction $\tilde{m}$ is given by

$$\langle \tilde{m}, \nabla_m \bar{q} \rangle = \langle \tilde{m}, \partial_m \mathcal{L} \rangle = \langle \tilde{m}, \partial_m \bar{q} + \partial_m \bar{r}_q \rangle, \quad \forall \tilde{m} \in \mathcal{M}, \tag{3.20}$$

where u solves the state problem (3.15) and v solves the adjoint problem (3.19). Similarly, for the computation of the Hessian of q acting in direction $\hat{m}^q$, we form the Lagrangian

$$\mathcal{L}^H(u, v^q, \bar{m}, z; \hat{u}^q, \hat{v}^q, \hat{m}^q) := \langle \hat{m}^q, \partial_m \bar{q} + \partial_m \bar{r}_q \rangle + \langle \hat{v}^q, \partial_v \bar{r}_q \rangle + \langle \hat{u}^q, \partial_u \bar{r}_q + \partial_u \bar{q} \rangle, \tag{3.21}$$

where $\hat{u}^q$ and $\hat{v}^q$ denote the incremental state and incremental adjoint, respectively. By taking the variation of (3.21) with respect to the adjoint v^q and using (1.2), we obtain the incremental state problem: find $\hat{u}^q \in \mathcal{U}$ such that

$$\langle \tilde{v}, \partial_{vu} \bar{r}_q \, \hat{u}^q \rangle = -\langle \tilde{v}, \partial_{vm} \bar{r}_q \, \hat{m}^q \rangle, \quad \forall \tilde{v} \in \mathcal{V}, \tag{3.22}$$

where the derivatives $\partial_{vu} \bar{r}_q : \mathcal{U} \mapsto \mathcal{V}'$ and $\partial_{vm} \bar{r}_q : \mathcal{M} \mapsto \mathcal{V}'$ are linear operators. The incremental adjoint problem, obtained by taking variation of (3.21) with respect to the state u and using (1.2), reads: find $\hat{v}^q \in \mathcal{V}$ such that

$$\langle \tilde{u}, \partial_{uv} \bar{r}_q \, \hat{v}^q \rangle = -\langle \tilde{u}, \partial_{uu} \bar{r}_q \, \hat{u}^q + \partial_{uu} \bar{q} \, \hat{u}^q + \partial_{um} \bar{r}_q \, \hat{m}^q + \partial_{um} \bar{q} \, \hat{m}^q \rangle, \quad \forall \tilde{u} \in \mathcal{U}, \tag{3.23}$$

where $\partial_{uv} \bar{r}_q : \mathcal{V} \mapsto \mathcal{U}'$ is the adjoint of $\partial_{vu} \bar{r}_q : \mathcal{U} \mapsto \mathcal{V}'$ in the sense of (1.2). The Hessian of q at $\bar{m}$ acting in direction $\hat{m}^q$ can then be computed by taking variation of (3.21) with respect to m and using (1.2) as

$$\begin{aligned} \langle \tilde{m}, \nabla^2_m \bar{q} \, \hat{m}^q \rangle &= \langle \tilde{m}, \partial_m \mathcal{L}^H \rangle \\ &= \langle \tilde{m}, \partial_{mv} \bar{r}_q \, \hat{v}^q + \partial_{mu} \bar{r}_q \, \hat{u}^q + \partial_{mu} \bar{q} \, \hat{u}^q + \partial_{mm} \bar{r}_q \, \hat{m}^q + \partial_{mm} \bar{q} \, \hat{m}^q \rangle, \ \forall \tilde{m} \in \mathcal{M}, \end{aligned} \tag{3.24}$$

where the incremental state $\hat{u}^q$ and adjoint $\hat{v}^q$ solve (3.22) and (3.23), respectively.

Similarly, we can compute the gradient and Hessian action of f. For ease of notation we define

$$\bar{r}_f = r(u, v^f, \bar{m}, z), \text{ and } \bar{f} = f(u(\bar{m}, z), \bar{m}, z). \tag{3.25}$$

By solving the adjoint problem: find $v^f \in \mathcal{V}$, such that

$$\langle \tilde{u}, \partial_u \bar{r}_f \rangle = -\langle \tilde{u}, \partial_u \bar{f} \rangle, \quad \forall \tilde{u} \in \mathcal{U}, \tag{3.26}$$

we obtain the gradient of f at $\bar{m}$ as

$$\langle \tilde{m}, \nabla_m \bar{f} \rangle = \langle \tilde{m}, \partial_m \bar{f} + \partial_m \bar{r}_f \rangle, \quad \forall \tilde{m} \in \mathcal{M}. \tag{3.27}$$

Then by solving the incremental state problem: find $\hat{u}^f \in \mathcal{U}$ such that

$$\langle \tilde{v}, \partial_{vu} \bar{r}_f \, \hat{u}^f \rangle = -\langle \tilde{v}, \partial_{vm} \bar{r}_f \, \hat{m}^f \rangle, \quad \forall \tilde{v} \in \mathcal{V}, \tag{3.28}$$

and the incremental adjoint problem: find $\hat{v}^f \in \mathcal{V}$ such that

$$\langle \tilde{u}, \partial_{uv} \bar{r}_f \, \hat{v}^f \rangle = -\langle \tilde{u}, \partial_{uu} \bar{r}_f \, \hat{u}^f + \partial_{uu} \bar{f} \, \hat{u}^f + \partial_{um} \bar{r}_f \, \hat{m}^f + \partial_{um} \bar{f} \, \hat{m}^f \rangle, \quad \forall \tilde{u} \in \mathcal{U}, \tag{3.29}$$

we obtain the Hessian action of f at $\bar{m}$ in direction $\hat{m}^f$ as

$$\langle \tilde{m}, \nabla_m^2 \bar{f} \, \hat{m}^f \rangle = \langle \tilde{m}, \partial_{mv} \bar{r}_f \, \hat{v}^f + \partial_{mu} \bar{r}_f \, \hat{u}^f + \partial_{mu} \bar{f} \, \hat{u}^f + \partial_{mm} \bar{r}_f \, \hat{m}^f + \partial_{mm} \bar{f} \, \hat{m}^f \rangle, \quad \forall \tilde{m} \in \mathcal{M}. \tag{3.30}$$

4. Gradient-based optimization. In this section, we develop a gradient based optimization method to solve the chance constrained optimization problem (2.8). The method employs (1) a smooth approximation of the indicator function involved in the chance evaluation, (2) an exterior penalty method for the inequality chance constraint, (3) a continuation scheme to refine the smooth approximation and the penalty for inequality constraint, and (4) an approximate cost functional, with both SAA and Taylor approximations, and their gradients with respect to the optimization variable.

4.1. Smooth approximation of the indicator function. The evaluation of the probability (2.6) involves the indicator function $\mathbb{I}_{[0,\infty)}(f(m,z))$, which is discontinuous at $f(m,z) = 0$. To use a gradient-based optimization method, we consider a smooth approximation of the indicator function by a logistic function

$$\mathbb{I}_{[0,\infty)}(x) \approx \ell_\beta(x) = \frac{1}{1+e^{-2\beta x}}, \text{ with } \nabla \ell_\beta(x) = \frac{2\beta e^{-2\beta x}}{(1+e^{-2\beta x})^2} \tag{4.1}$$

where a larger β corresponds to a sharper transition at $x = 0$, as shown in Figure 1. With the definition $\mathbb{I}_{[0,\infty)}(0) = \frac{1}{2}$, we have the convergence

$$\lim_{\beta \to \infty} \ell_\beta(x) = \mathbb{I}_{[0,\infty)}(x) \text{ and } \lim_{\beta \to \infty} \nabla \ell_\beta(x) = \nabla \mathbb{I}_{[0,\infty)}(x) \text{ except at } x = 0. \tag{4.2}$$

4.2. A penalty method for inequality constraint optimization. To solve the optimization problem (2.8) with the inequality constraint (2.5), for simplicity we employ a quadratic penalty method [68] by first defining the exterior penalty function

$$\mathcal{S}_\gamma(x) := \frac{\gamma}{2} \left(\max\{0, x\}\right)^2, \text{ with } \nabla \mathcal{S}_\gamma(x) = \gamma \max\{0, x\}, \tag{4.3}$$

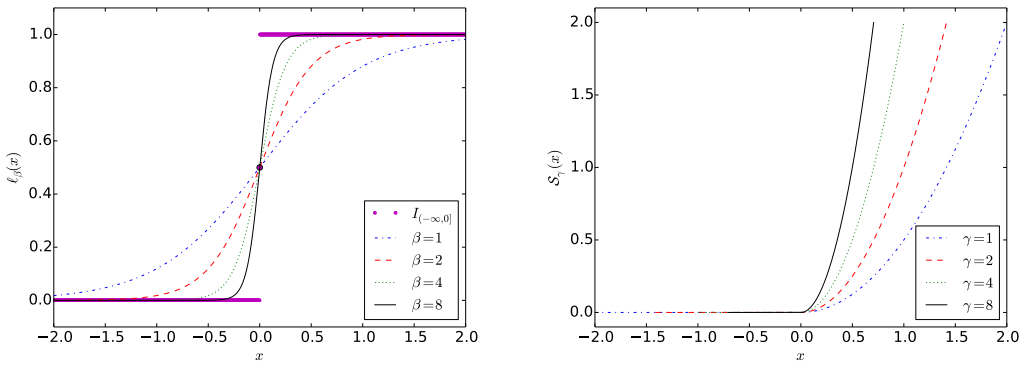

Figure 1. *Left: Smooth approximation of the discontinuous indicator function* $\mathbb{I}_{[0,\infty)}(x)$ *by a logistic function* $\ell_\beta(x) = \frac{1}{1+e^{-2\beta x}}$ *with* $\beta > 0$. *Right: Penalty function* $\mathcal{S}_\gamma(x) = \frac{\gamma}{2}(\max\{0, x\})^2$ *with* $\gamma > 0$.

for a constant $\gamma > 0$ controlling the weight of the penalty. Note that a proper reformulation technique combined with a Newton method can be used to alleviate the ill conditioning for large γ [68]. We remark that other methods, e.g., an interior-point method, or an augmented Lagrangian method could also be used [68]. Then the chance constrained optimization problem (2.8) can be approximated by the unconstrained problem

$$\min_{z\in\mathcal{Z}} \mathcal{J}(z) + \mathcal{S}_\gamma(\mathbb{E}[\ell_\beta(f)] - \alpha). \tag{4.4}$$

For general inequality constrained optimization problems, convergence of the optimization variable as $\gamma \to \infty$ by the penalty method is studied in, e.g., [68].

4.3. Adaptive BFGS optimization. Let $\mathcal{E}(z)$ denote an approximate cost functional, which is an approximation of the cost functional in (4.4) by SAA or Taylor approximations. Let $\nabla_z\mathcal{E}(z)$ denote its gradient with respect to the optimization variable, which is computed in next section. We present a continuation optimization scheme by increasing the smoothing parameter β and the penalty parameter γ in an outer loop and applying a gradient-based quasi Newton optimization algorithm, BFGS [66, 68], to solve the optimization problem (4.4) (with possible bound constraint on the optimization variable z) using the approximate cost functional $\mathcal{E}(z)$ and gradient $\nabla_z\mathcal{E}(z)$ in an inner loop. The optimization procedure is presented in Algorithm 4.1. For the continuation, we specify an initial smoothing parameter β_0 and penalty parameter γ_0, and scale them with the power parameters σ_β and σ_γ in Line 6. We stop the outer loop if the maximum number of iterations is reached or the difference between subsequent change of the optimization variable or the approximate chance, defined in (3.2) for SAA or (3.6) for the Taylor approximation of the constraint function, is smaller than a tolerance; see Line 3. The inner loop of BFGS optimization is stopped if the maximum number of iterations is reached or the gradient of the approximate cost functional is smaller than a tolerance; see Line 4. We remark that this tolerance can be set relatively large and decreased in the outer loop to improve efficiency of the continuation [9].

4.4. Computation of the approximate cost functional and its gradient. Algorithm 4.1 requires the computation of the approximate cost functional $\mathcal{E}(z)$ and its gradient $\nabla_z\mathcal{E}(z)$ at a

Algorithm 4.1 Continuation BFGS for chance constrained optimization

1: **Input**: initial value z_0, bound constraint $[z_{\min}, z_{\max}]$, smoothing parameter β_0, penalty parameter γ_0, power parameters σ_β and σ_γ, maximum number of iterations $k_{\max}$ and $l_{\max}$, and tolerances ε_{in} for the inner loop and $\varepsilon^z_{\text{out}}, \varepsilon^f_{\text{out}}$ for the outer loop.
2: Set $l = 0$, compute the approximate chance $\hat{f}_l$ at z_l, where $\hat{f}_l = f_M(z_l)$ in (3.2) for SAA or $\hat{f}_l = f^K_M(z_l)$ in (3.6) for Taylor. Set $\hat{f}_{-1} = \hat{f}_0 + 2\varepsilon^f_{\text{out}}, z_{-1} = z_0 + 2\varepsilon^z_{\text{out}}$.
3: **while** $||z_l - z_{l-1}||_{\mathcal{Z}} \geq \varepsilon^z_{\text{out}}$ or $|\hat{f}_l - \hat{f}_{l-1}| \geq \varepsilon^f_{\text{out}}$ and $l \leq l_{\max}$ **do**
4: Run an inner loop optimization by the Algorithm L-BFGS-B [66]

$$z_{l+1} = \text{L-BFGS-B}(\mathcal{E}(z), z_l, \nabla_z \mathcal{E}(z), k_{\max}, \varepsilon_{\text{in}}, [z_{\min}, z_{\max}]),$$

with objective $\mathcal{E}(z)$ and gradient $\nabla_z \mathcal{E}(z)$ given by either SAA or Taylor approximation.

5: Compute the approximate chance $\hat{f}_{l+1}$ at z_{l+1}.
6: Set $\beta_{l+1} = \sigma_\beta^{l+1}\beta_l$, $\gamma_{l+1} = \sigma_\gamma^{l+1}\gamma_l$, $l \leftarrow l + 1$.
7: **end while**
8: **return** Optimal variable $z_{\text{opt}} = z_l$.

given optimization variable z, constrained by the state PDE (2.2). In this section, we present their computation for both SAA and Taylor approximations.

4.4.1. Sample average approximation. By SAA of the mean and chance presented in Section 3.1, the cost functional in (4.4) can be approximated as

$$\begin{aligned} &\mathcal{J}(z) + \mathcal{S}_\gamma(\alpha - \mathbb{E}[\ell_\beta(f(\cdot, z))]) \\ &= \mathbb{E}[q(\cdot, z)] + \mathcal{P}(z) + \mathcal{S}_\gamma(\mathbb{E}[\ell_\beta(f(\cdot, z))] - \alpha) \\ &\approx \frac{1}{M_q}\sum_{i=1}^{M_q} q(m_i^q, z) + \mathcal{P}(z) + \mathcal{S}_\gamma\left(\frac{1}{M_f}\sum_{i=1}^{M_f} \ell_\beta(f(m_i^f, z)) - \alpha\right) =: \mathcal{E}(z), \end{aligned} \tag{4.5}$$

where M_q and M_f independent random samples are taken such that the SAA errors of the two approximate terms are balanced. A practical approach to determine M_q and M_f is to first evaluate the variances of q and $\ell_\beta(f)$ and then set the two numbers with the ratio equal to that of the variances. For simplicity, one can use the same $M_q = M_f$ random samples. Note that to compute $q(m_i^q, z) = q(u(m_i^q), m_i^q, z)$ and $f(m_i^f, z) = f(u(m_i^f), m_i^f, z)$, we need to solve the state equation (2.2) at m_i^q and m_i^f, respectively.

To compute the gradient of this approximate cost functional, we define the Lagrangian

$$\begin{aligned} &\mathcal{L}_{SAA}\big(z, \{u_i^q\}, \{u_i^f\}, \{v_i^q\}, \{v_i^f\}\big) \\ &:= \frac{1}{M_q}\sum_{i=1}^{M_q} q(m_i^q, z) + \mathcal{P}(z) + \mathcal{S}_\gamma\left(\frac{1}{M_f}\sum_{i=1}^{M_f} \ell_\beta(f(m_i^f, z)) - \alpha\right) \\ &+ \sum_{i=1}^{M_q}\langle v_i^q, \partial_v r(u_i^q, v, m_i^q, z)\rangle + \sum_{i=1}^{M_f}\langle v_i^f, \partial_v r(u_i^f, v, m_i^f, z)\rangle, \end{aligned} \tag{4.6}$$

where $\{v_i^q\}$ and $\{v_i^f\}$ are the Lagrangian multipliers, which can be computed by solving the adjoint problems obtained by setting the variation of the Lagrangian w.r.t. u_i^q to zero, i.e., for each $i = 1, \dots, M_q$, find $v_i^q \in \mathcal{V}$ such that

$$\langle \tilde{u}, \partial_{uv} r(u_i^q, v, m_i^q, z) v_i^q \rangle = -\frac{1}{M_q} \langle \tilde{u}, \partial_u q(u_i^q, m_i^q, z) \rangle, \quad \forall \tilde{u} \in \mathcal{U}, \tag{4.7}$$

where we recall that $q(u_i^q, m_i^q, z) = q(m_i^q, z)$ by a slight abuse of notation. Similarly, for each $i = 1, \dots, M_f$, find $v_i^f \in \mathcal{V}$ such that

$$\langle \tilde{u}, \partial_{uv} r(u_i^f, v, m_i^f, z) v_i^f \rangle = -\frac{1}{M_f} \langle \tilde{u}, \nabla \mathcal{S}_\gamma \nabla \ell_\beta \partial_u f(u_i^f, m_i^f, z) \rangle, \quad \forall \tilde{u} \in \mathcal{U}. \tag{4.8}$$

Then the gradient of the approximate cost functional can be evaluated as

$$\begin{aligned} \langle \tilde{z}, \nabla \mathcal{E}(z) \rangle &= \langle \tilde{z}, \partial_z \mathcal{L}_{SAA} \rangle \\ &= \frac{1}{M_q} \sum_{i=1}^{M_q} \langle \tilde{z}, \partial_z q(m_i^q, z) \rangle + \langle \tilde{z}, \nabla \mathcal{P}(z) \rangle + \frac{1}{M_f} \sum_{i=1}^{M_f} \langle \tilde{z}, \nabla \mathcal{S}_\gamma \nabla \ell_\beta \partial_z f(m_i^f, z) \rangle \\ &\quad + \sum_{i=1}^{M_q} \langle \tilde{z}, \partial_{zv} r(u_i^q, v, m_i^q, z) v_i^q \rangle + \sum_{i=1}^{M_f} \langle \tilde{z}, \partial_{zv} r(u_i^f, v, m_i^f, z) v_i^f \rangle. \end{aligned} \tag{4.9}$$

In summary, $M_q + M_f$ state PDEs and $M_q + M_f$ linearized (adjoint) PDEs have to be solved to evaluate the SAA of $\mathcal{E}(z)$ and its gradient $\nabla \mathcal{E}(z)$ at any given z.

4.4.2. Taylor approximation. For simplicity, we present only the approximate cost functional and its gradient by the quadratic Taylor approximation T_2 of the objective and constraint functions. The computation corresponding to the constant and linear approximations T_0 and T_1, which are contained within the quadratic approximation, are omitted.

Using the quadratic Taylor approximations of the objective function q and the constraint function f in (3.3), the low-rank approximations of the trace in (3.8) and the Hessian in (3.11), as well as the sample average approximation for the probability (3.2), the cost functional in the unconstrained optimization problem (4.4) becomes

$$\begin{aligned} \mathcal{J}(z) + \mathcal{S}_\gamma(\alpha - \mathbb{E}[\ell_\beta(f(\cdot, z))]) &\approx \mathbb{E}[T_2 q(\cdot, z)] + \mathcal{P}(z) + \mathcal{S}_\gamma(\mathbb{E}[\ell_\beta(T_2 f(\cdot, z))] - \alpha) \\ &\approx \bar{q} + \frac{1}{2} \sum_{n=1}^{N_q} \lambda_n^q + \mathcal{P}(z) + \mathcal{S}_\gamma \left(g_{\beta, M_f}(T_2^{\mathrm{LR}} f(\cdot, z)) \right) \\ &=: \mathcal{E}(z) \end{aligned} \tag{4.10}$$

where for simplicity we denote

$$g_{\beta, M_f}(T_2^{\mathrm{LR}} f(\cdot, z)) = \frac{1}{M_f} \sum_{i=1}^{M_f} \ell_\beta(T_2^{\mathrm{LR}} f(m_i, z)) - \alpha, \tag{4.11}$$

where the quadratic Taylor approximation with low-rank decomposition $T_2^{\mathrm{LR}} f(m_i, z)$ is given in (3.14). Note that $\bar{q}$ and λ_n^q, $n = 1, \dots, N_q$, which are part of the approximation of $\mathbb{E}[q]$, as

well as $\bar{f}$, $\nabla_m \bar{f}$, λ_n^f and ψ_n^f, $n = 1, \ldots, N_f$, which are part of the approximation of $\mathbb{E}[\ell_\beta(f)]$, all implicitly (possibly also explicitly) depend on the optimization variable z through the state equation (3.15), the adjoint equations (3.19) and (3.26), the generalized eigenvalue problems (3.9) and (3.12) with orthonormal conditions (3.10) and (3.13), the incremental state and adjoint equations (3.22) and (3.23) for the incremental state $\hat{u}^q = \hat{u}_n^q$ and adjoint $\hat{v}^q = \hat{v}_n^q$ at $\hat{m}^q = \psi_n^q$ needed by the Hessian action $\nabla_m^2 \bar{q}\, \psi_n^q$ in (3.9) through (3.24), $n = 1, \ldots, N_q$, as well as the incremental state and adjoint equations (3.28) and (3.29) for the incremental state $\hat{u}^f = \hat{u}_n^f$ and adjoint $\hat{v}^f = \hat{v}_n^f$ at $\hat{m}^f = \psi_n^f$ needed by the Hessian action $\nabla_m^2 \bar{f}\, \psi_n^f$ in (3.12) through (3.30), $n = 1, \ldots, N_f$. To derive the gradient of the approximate cost functional (4.10), we define a Lagrangian $\mathcal{L}_2$ to enforce all of the PDE constraint equations as follows:

$$
\begin{aligned}
(4.12)\quad & \mathcal{L}_2\big(z, u, v^*, v^q, (u^q)^*, \{\lambda_n^q\}, \{\psi_n^q\}, \{\hat{u}_n^q\}, \{\hat{v}_n^q\}, \{(\lambda_{nn'}^q)^*\}, \{(\psi_n^q)^*\}, \{(\hat{u}_n^q)^*\}, \{(\hat{v}_n^q)^*\}, \\
& \qquad v^f, (u^f)^*, \{\lambda_n^f\}, \{\psi_n^f\}, \{\hat{u}_n^f\}, \{\hat{v}_n^f\}, \{(\lambda_{nn'}^f)^*\}, \{(\psi_n^f)^*\}, \{(\hat{u}_n^f)^*\}, \{(\hat{v}_n^f)^*\}\big) \\
& := \bar{q} + \frac{1}{2}\sum_{n=1}^{N_q} \lambda_n^q + \mathcal{P}(z) + \mathcal{S}_\gamma\left(g_{\beta, M_f}(T_2^{\mathrm{LR}} f)\right) \\
& + \langle v^*, \partial_v \bar{r} \rangle \\
& + \langle (u^q)^*, \partial_u \bar{r}_q + \partial_u \bar{q} \rangle \\
& + \sum_{n=1}^{N_q} \langle (\psi_n^q)^*, \nabla_m^2 \bar{q}\, \psi_n^q - \lambda_n^q \mathcal{C}^{-1} \psi_n^q \rangle \\
& + \sum_{n,n'=1}^{N_q} (\lambda_{nn'}^q)^* (\langle \psi_n^q, \mathcal{C}^{-1} \psi_{n'}^q \rangle - \delta_{nn'}) \\
& + \sum_{n=1}^{N_q} \langle (\hat{v}_n^q)^*, \partial_{vu} \bar{r}_q\, \hat{u}_n^q + \partial_{vm} \bar{r}_q\, \psi_n^q \rangle \\
& + \sum_{n=1}^{N_q} \langle (\hat{u}_n^q)^*, \partial_{uv} \bar{r}_q\, \hat{v}_n^q + \partial_{uu} \bar{r}_q\, \hat{u}_n^q + \partial_{uu} \bar{q}\, \hat{u}_n^q + \partial_{um} \bar{r}_q\, \psi_n^q + \partial_{um} \bar{q}\, \psi_n^q \rangle \\
& + \langle (u^f)^*, \partial_u \bar{r}_f + \partial_u \bar{f} \rangle \\
& + \sum_{n=1}^{N_f} \langle (\psi_n^f)^*, \nabla_m^2 \bar{f}\, \psi_n^f - \lambda_n^f \mathcal{C}^{-1} \psi_n^f \rangle \\
& + \sum_{n,n'=1}^{N_f} (\lambda_{nn'}^f)^* (\langle \psi_n^f, \mathcal{C}^{-1} \psi_{n'}^f \rangle - \delta_{nn'}) \\
& + \sum_{n=1}^{N_f} \langle (\hat{v}_n^f)^*, \partial_{vu} \bar{r}_f\, \hat{u}_n^f + \partial_{vm} \bar{r}_f\, \psi_n^f \rangle
\end{aligned}
$$

$$+\sum_{n=1}^{N_f}\langle(\hat{u}_n^f)^*, \partial_{uv}\bar{r}_f\,\hat{v}_n^f + \partial_{uu}\bar{r}_f\,\hat{u}_n^f + \partial_{uu}\bar{f}\,\hat{u}_n^f + \partial_{um}\bar{r}_f\,\psi_n^f + \partial_{um}\bar{f}\,\psi_n^f\rangle. \tag{4.13}$$

We compute all the Lagrange multipliers by setting the variation of the Lagrangian $\mathcal{L}_2$ to zero. Specifically, by setting the variation of the Lagrangian $\mathcal{L}_2$ w.r.t. λ_n^q, $n = 1, \ldots, N_q$, to zero, we obtain

$$(\psi_n^q)^* = \frac{1}{2}\psi_n^q, \quad n = 1, \ldots, N_q. \tag{4.14}$$

Subsequently, for each $n = 1, \ldots, N_q$, by setting the variation of $\mathcal{L}_2$ w.r.t. $\hat{v}_n^q$ to zero, using the Hessian action (3.24) and (1.2), we have: find $(\hat{u}_n^q)^* \in \mathcal{U}$ such that

$$\langle\tilde{v}, \partial_{vu}\bar{r}_q(\hat{u}_n^q)^*\rangle = -\langle\tilde{v}, \partial_{vm}\bar{r}_q(\psi_n^q)^*\rangle, \quad \forall\tilde{v} \in \mathcal{V}, \tag{4.15}$$

which, together with (4.14) and (3.22) with $(\hat{u}^q, \hat{m}^q) = (\hat{u}_n^q, \psi_n^q)$, leads to

$$(\hat{u}_n^q)^* = \frac{1}{2}\hat{u}_n^q, \quad n = 1, \ldots, N_q. \tag{4.16}$$

Similarly, for each $n = 1, \ldots, N_q$, by setting the variation of $\mathcal{L}_2$ w.r.t. $\hat{u}_n^q$ to zero, using the Hessian action (3.24) and (1.2), we have: find $(\hat{v}_n^q)^* \in \mathcal{U}$ such that

$$\langle\tilde{u}, \partial_{uv}\bar{r}_q\,(\hat{v}_n^q)^*\rangle = -\langle\tilde{u}, \partial_{uu}\bar{r}_q\,(\hat{u}_n^q)^* + \partial_{uu}\bar{q}\,(\hat{u}_n^q)^* + \partial_{um}\bar{r}_q\,(\psi_n^q)^* + \partial_{um}\bar{q}\,(\psi_n^q)^*\rangle, \quad \forall\tilde{u} \in \mathcal{U}, \tag{4.17}$$

which, together with (4.14), (4.16), and (3.23) with $(\hat{v}^q, \hat{u}^q, \hat{m}^q) = (\hat{v}_n^q, \hat{u}_n^q, \psi_n^q)$, leads to

$$(\hat{v}_n^q)^* = \frac{1}{2}\hat{v}_n^q, \quad n = 1, \ldots, N_q. \tag{4.18}$$

Then by setting the variation of $\mathcal{L}_2$ w.r.t. v^q to zero, we obtain: find $(u^q)^* \in \mathcal{U}$ such that

$$\begin{aligned}\langle\tilde{v}, \partial_{vu}\bar{r}_q(u^q)^*\rangle = &-\sum_{n=1}^{N_q}\langle\tilde{v}, \partial_{vmu}\bar{r}_q\hat{u}_n^q(\psi_n^q)^* + \partial_{vmm}\bar{r}_q\psi_n^q(\psi_n^q)^*\rangle \\ &-\sum_{n=1}^{N_q}\langle\tilde{v}, \partial_{vuu}\bar{r}_q\hat{u}_n^q(\hat{u}_n^q)^* + \partial_{vum}\bar{r}_q\psi_n^q(\hat{u}_n^q)^*\rangle, \quad \tilde{v} \in \mathcal{V}.\end{aligned} \tag{4.19}$$

By setting the $\mathcal{L}_2$ w.r.t. λ_n^f, $n = 1, \ldots, N_f$, to zero, we obtain

$$(\psi_n^f)^* = c_n^f\psi_n^f, \quad n = 1, \ldots, N_f, \tag{4.20}$$

where the constant c_n^f is given by

$$c_n^f = \nabla\mathcal{S}_\gamma(g_{\beta,M_f}(T_2^{\mathrm{LR}}f))\frac{1}{2M_f}\sum_{i=1}^{M_f}\nabla\ell_\beta(T_2^{\mathrm{LR}}f(m_i, z))\langle m_i - \bar{m}, \mathcal{C}^{-1}\psi_n^f\rangle^2, \tag{4.21}$$

where $\nabla \mathcal{S}_\gamma$ and $\nabla \ell_\beta$ are defined in (4.2) and (4.3), respectively. Subsequently, for each $n = 1, \ldots, N_f$, by setting the variation of $\mathcal{L}_2$ w.r.t. $\hat{v}_n^f$ to zero, and using the Hessian action (3.30) and (1.2), we have: find $(\hat{u}_n^f)^* \in \mathcal{U}$ such that

$$\langle \tilde{v}, \partial_{vu} \bar{r}_f (\hat{u}_n^f)^* \rangle = -\langle \tilde{v}, \partial_{vm} \bar{r}_f (\psi_n^f)^* \rangle, \quad \forall \tilde{v} \in \mathcal{V}, \tag{4.22}$$

which, together with (4.20) and (3.28) with $(\hat{u}^f, \hat{m}^f) = (\hat{u}_n^f, \psi_n^f)$, leads to

$$(\hat{u}_n^f)^* = c_n^f \hat{u}_n^f, \quad n = 1, \ldots, N_f. \tag{4.23}$$

Similarly, for each $n = 1, \ldots, N_f$, by setting the variation of $\mathcal{L}_2$ w.r.t. $\hat{u}_n^f$ to zero, using the Hessian action (3.30) and (1.2), we have: find $(\hat{v}_n^f)^* \in \mathcal{U}$ such that

$$\langle \tilde{u}, \partial_{uv} \bar{r}_f (\hat{v}_n^f)^* \rangle = -\langle \tilde{u}, \partial_{uu} \bar{r}_f (\hat{u}_n^f)^* + \partial_{uu} \bar{q} (\hat{u}_n^f)^* + \partial_{um} \bar{r}_f (\psi_n^f)^* + \partial_{um} \bar{f} (\psi_n^f)^* \rangle, \quad \forall \tilde{u} \in \mathcal{U}, \tag{4.24}$$

which, together with (4.20), (4.23), and (3.29) with $(\hat{v}^f, \hat{u}^f, \hat{m}^f) = (\hat{v}_n^f, \hat{u}_n^f, \psi_n^f)$, leads to

$$(\hat{v}_n^f)^* = c_n^f \hat{v}_n^f, \quad n = 1, \ldots, N_f. \tag{4.25}$$

Then by setting the variation of $\mathcal{L}_2$ w.r.t. v^f to zero, we obtain: find $(u^f)^* \in \mathcal{U}$ such that

$$\begin{aligned} \langle \tilde{v}, \partial_{vu} \bar{r}_f (u^f)^* \rangle = &- \sum_{n=1}^{N_f} \langle \tilde{v}, \partial_{vmu} \bar{r}_f \hat{u}_n^f (\psi_n^f)^* + \partial_{vmm} \bar{r}_f \psi_n^f (\psi_n^f)^* \rangle \\ &- \sum_{n=1}^{N_f} \langle \tilde{v}, \partial_{vuu} \bar{r}_f \hat{u}_n^f (\hat{u}_n^f)^* + \partial_{vum} \bar{r}_f \psi_n^f (\hat{u}_n^f)^* \rangle \\ &- \langle \tilde{v}, \partial_{vm} \bar{r}_f m^f \rangle, \quad \tilde{v} \in \mathcal{V}, \end{aligned} \tag{4.26}$$

where the last term is due to the gradient (3.27), which appears in the quadratic Taylor approximation (3.14) that is used in $\mathcal{S}_\gamma(g_{\beta,M_f}(T_2^{\mathrm{LR}} f)$, with m^f given by

$$m^f = \nabla \mathcal{S}_\gamma(g_{\beta,M_f}(T_2^{\mathrm{LR}} f)) \frac{1}{M_f} \sum_{i=1}^{M_f} \nabla \ell_\beta(T_2^{\mathrm{LR}} f(m_i, z))(m_i - \bar{m}). \tag{4.27}$$

Finally, by setting the variation of $\mathcal{L}_2$ w.r.t. u to zero, we obtain $v^* \in \mathcal{V}$ by solving an equation presented in Appendix A. With all the Lagrange multipliers computed above, we can evaluate the gradient of the cost functional (4.10) as $\langle \tilde{z}, \nabla_z \mathcal{E}(z) \rangle$, which is explicitly given in Appendix A.

We remark that the constraints of the orthonormal conditions in the Lagrangian $\mathcal{L}_2$ do not explicitly depend on the optimization variable z, so that for computing the gradient $\nabla_z \mathcal{E}(z)$ we do not need the Lagrange multipliers $\{(\lambda_{nn'}^q)^*\}$ and $\{(\lambda_{nn'}^f)^*\}$, neither of which are used in computing all other Lagrange multipliers. To solve the chance constrained optimization problem under uncertainty with the unconstrained penalty formulation (4.4) and its approximation in (4.10), we apply a gradient-based BFGS algorithm, where the evaluation of the cost functional $\mathcal{E}(z)$ and its gradient $\nabla_k \mathcal{E}(z)$ at z, and their computational cost in terms of PDE solves, are summarized in Algorithm 4.2. In summary, one evaluation of the cost functional takes 1 state PDE solve and $2(N_q + c) + 2(N_f + c) + 2$ linearized PDE solves, while one evaluation of its gradient takes $2N_q + 2N_f + 3$ linearized PDE solves, where N_q and N_f are the ranks in (3.8) and (3.11), c is a small oversampling parameter in Algorithm 3.1.

Algorithm 4.2 Compute $\mathcal{E}(z)$ and its gradient $\nabla_z \mathcal{E}(z)$

Compute $\mathcal{E}(z)$:
solve the state equation (3.15) for u; // 1 state PDE
solve the linearized PDE (3.19) for v^q; // 1 linear PDE
solve the linearized PDE (3.26) for v^f; // 1 linear PDE
solve the generalized eigenvalue problems (3.9) for $(\lambda_n^q, \psi_n^q)_{n=1}^{N_q}$; // $2(N_q + c)$ linear PDEs
solve the generalized eigenvalue problems (3.12) for $(\lambda_n^f, \psi_n^f)_{n=1}^{N_f}$; // $2(N_f + c)$ linear PDEs
compute the cost functional $\mathcal{E}(z_k)$ by (4.10).
Compute $\nabla_z \mathcal{E}(z)$:
solve the linearized PDEs (3.22) and (3.23) at $(\psi_n^q)_{n=1}^{N_q}$ for $(\hat{u}_n^q, \hat{v}_n^q)_{n=1}^{N_q}$; // $2N_q$ linear PDEs
solve the linearized PDEs (3.28) and (3.29) at $(\psi_n^f)_{n=1}^{N_f}$ for $(\hat{u}_n^f, \hat{v}_n^f)_{n=1}^{N_f}$; // $2N_f$ linear PDEs
set $((\psi_n^q)^*, (\hat{u}_n^q)^*, (\hat{v}_n^q)^*)_{n=1}^{N_q}$ by (4.14), (4.16), and (4.18);
solve the linearized PDE (4.19) for $(u^q)^*$; // 1 linear PDE
set $((\psi_n^f)^*, (\hat{u}_n^f)^*, (\hat{v}_n^f)^*)_{n=1}^{N_f}$ by (4.20), (4.23), and (4.25);
solve the linearized PDE (4.26)for $(u^f)^*$; // 1 linear PDE
solve the linearized PDE (A.1) for v^*; // 1 linear PDE
compute the gradient $\nabla_z \mathcal{E}(z_k)$ by (A).

5. Numerical examples.

We consider the following PDEs that model a steady state Darcy flow,

$$\begin{aligned} \mathbf{v} + \frac{e^m}{\mu} \nabla u = 0 \quad &\text{in } D, \\ \nabla \cdot \mathbf{v} = h \quad &\text{in } D, \end{aligned} \tag{5.1}$$

where a homogeneous Dirichlet boundary condition for the pressure u is imposed along the boundary ∂D of a physical domain $D = (0,1)^2$. e^m represents a random permeability field, while μ is the fluid viscosity. For simplicity we specify $\mu = 1$ in a dimensionless setting. The source term h depends on an L-dimensional (we take $L = 25$ in the numerical test) optimization variable $z = (z_1, \ldots, z_L)$, given by

$$h(z) = -\sum_{\ell=1}^{L} z_\ell h_\ell, \tag{5.2}$$

where z_ℓ is a pointwise optimization variable with bound $z_\ell \in [z_{\min}, z_{\max}]$, where we take $z_{\min} = 0$ and $z_{\max} = 36$; h_ℓ is a smooth mollifier function defined at point $x_\ell \in D$ as

$$h_\ell = \exp\left(-\frac{1}{\varepsilon^2} ||x - x_\ell||^2\right) \tag{5.3}$$

for a positive number $\varepsilon > 0$, which we take $\varepsilon = 0.1$. The system (5.4) models steady state groundwater flow, z_ℓ represents the water extraction rate at location x_ℓ of L wells. Figure 2 illustrate the groundwater flow, where the velocity field $\mathbf{v}$ at the mean $\bar{m}$ and an optimal

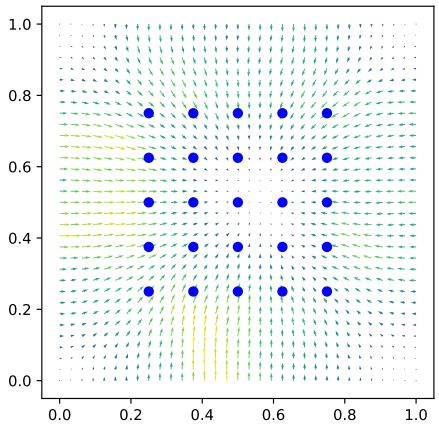

Figure 2. *Groundwater flow in physical domain* $(0,1)^2$*. Blue dots stand for the location of the extraction wells. A velocity field* $\mathbf{v}$ *is shown at the mean of the log-permeability* $m = \bar{m}$ *and optimal variable* z^**, which is obtained with quadratic approximation of the constraint function; see Figure 6 for the value of* z^**.*

extraction rate (with value given in Figure 6) is shown. Note that by eliminating the velocity field $\mathbf{v}$ from (5.1), we obtain a single equation for the pressure

$$-\nabla \cdot \left(\frac{e^m}{\mu} \nabla u \right) = h \quad \text{in } D. \tag{5.4}$$

The objective of the optimization problem is to achieve a target groundwater extraction rate $\bar{z}_\ell$ at each well, which can be represented by

$$q(z) = \frac{1}{L} \sum_{\ell=1}^{L} (z_\ell - \bar{z}_\ell)^2. \tag{5.5}$$

We use a penalty term $\mathcal{P}(z) = \frac{\eta}{2}||z||_2^2$ with $\eta = 10^{-5}$ for the optimization variable z, representing the cost of the extraction. To prevent excessive extraction leading to potential collapse of the aquifer, we consider the constraint function for the state (pressure field) u

$$f(u) = \int_{D_o} u^2(x) dx - f_c, \tag{5.6}$$

where $D_o \subset D$ is a region of interest, for which we take $D_o = (0.25, 0.75)^2$; and $f_c > 0$ is a critical value, which we take $f_c = 2$. We consider the chance or probability constraint (2.5), i.e., $P(f \geq 0) \leq \alpha$, such that the probability of f greater than or equal to zero should be less than or equal to a given value $\alpha > 0$ (we take $\alpha = 0.05$), where the probability is defined with respect to the probability distribution of the random field m.

We use a Gaussian random field $m \sim \mathcal{N}(\bar{m}, \mathcal{C})$ with mean $\bar{m}$ and covariance field $\mathcal{C}$, which is represented as the square of the inverse of an elliptic operator [**?**, 60, 79],

$$\mathcal{C} = (-\eta \nabla \cdot (\Theta \nabla) + \delta)^{-2}, \tag{5.7}$$

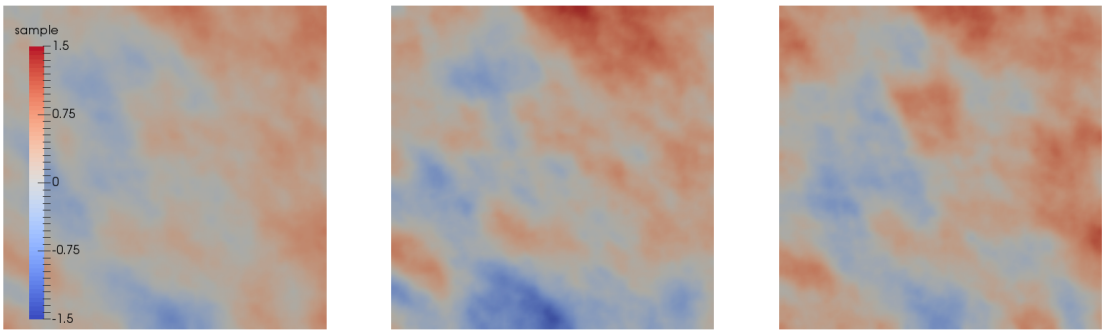

Figure 3. *Mean of the log-permeability field $\bar{m}$ (left) and two random samples (middle and right).*

∇ and $\nabla\cdot$ are gradient and divergence operator. $\delta > 0$ and $\eta > 0$ are two parameters that control the variance and correlation of the random field. The matrix $\Theta = [c_1 \sin(\theta)^2, (c_1 - c_2)\sin(\theta)\cos(\theta); (c_1 - c_2)\sin(\theta)\cos(\theta), c_2\cos(\theta)^2]$. Sampling from this Gaussian distribution is equivalent to solving the elliptic equation

$$-\eta\nabla\cdot(\Theta\nabla m) + \delta m = \dot{W} \quad \text{in } D \tag{5.8}$$

with homogeneous Neumann boundary condition along ∂D, where $\dot{W}$ is a spatial white noise with unit pointwise variance. The mean $\bar{m}$ is given in Figure 3, which is obtained as a random sample from $\mathcal{N}(0, \mathcal{C})$ for the covariance $\mathcal{C}$ in (5.7) with $\eta = 0.1$, $\delta = 10$ and $c_1 = 1, c_2 = 2, \theta = \pi/4$. This choice of these parameters leads to pointwise variance of about 0.1, and 20% of the noise-to-signal ratio as measured by the ratio between the pointwise standard deviation and the maximum of the mean field. Two random samples drawn from $\mathcal{N}(\bar{m}, \mathcal{C})$ are also shown in Figure 3.

We use a finite element method implemented in FEniCS [62] to solve all of the PDEs, with linear elements for the approximation of both the pressure field u and the parameter field m in a uniform mesh of triangles of size 32×32, which leads to 33^2 dimension for the discrete parameter. We run the optimization algorithm, Algorithm 4.2, to solve the chance constrained optimization problem (2.8), with Taylor approximation of the constraint function $T_K f$ with $K = 0, 1, 2$, and a sample average approximation (SAA) with 1024 samples. We set the tolerance for the gradient norm as $\varepsilon_{\text{in}} = 10^{-3}$ for the stopping criterion for the inner BFGS optimization loop in Algorithm 4.1 and $\varepsilon^f_{\text{out}} = \varepsilon^z_{\text{out}} = 10^{-2}$ in the outer continuation optimization loop. For the quadratic Taylor approximation, we compute 10 eigenpairs (3.12) by the randomized algorithm, Algorithm 3.1. The smoothing and penalty parameters are specified as $(\beta, \gamma) = (2^{n+2}, 10^{n+2})$ for $n = 1, 2, \cdots$. The optimization solution at step n is used as the initial guess for that at step $n+1$. After each optimization step with quadratic approximation $T_2 f$, we compute the chance $P(g > 0) = \mathbb{E}[\mathbb{I}_{[0,\infty)}(g)]$ and the smoothed value $\mathbb{E}[\ell_\beta(g)]$, by SAA approximation in (3.2) with different numbers of samples, where g represents the constraint function f or its Taylor approximations $T_K f$ for $K = 0, 1, 2$. The results are shown in Figure 4, from which we can observe: (1) the quadratic approximation $T_2 f$ yields more accurate approximation of the chance than that by the linear and constant approximations; (2) as the smoothing parameter β increases, the smooth approximation $\ell_\beta(g)$ becomes more accurate

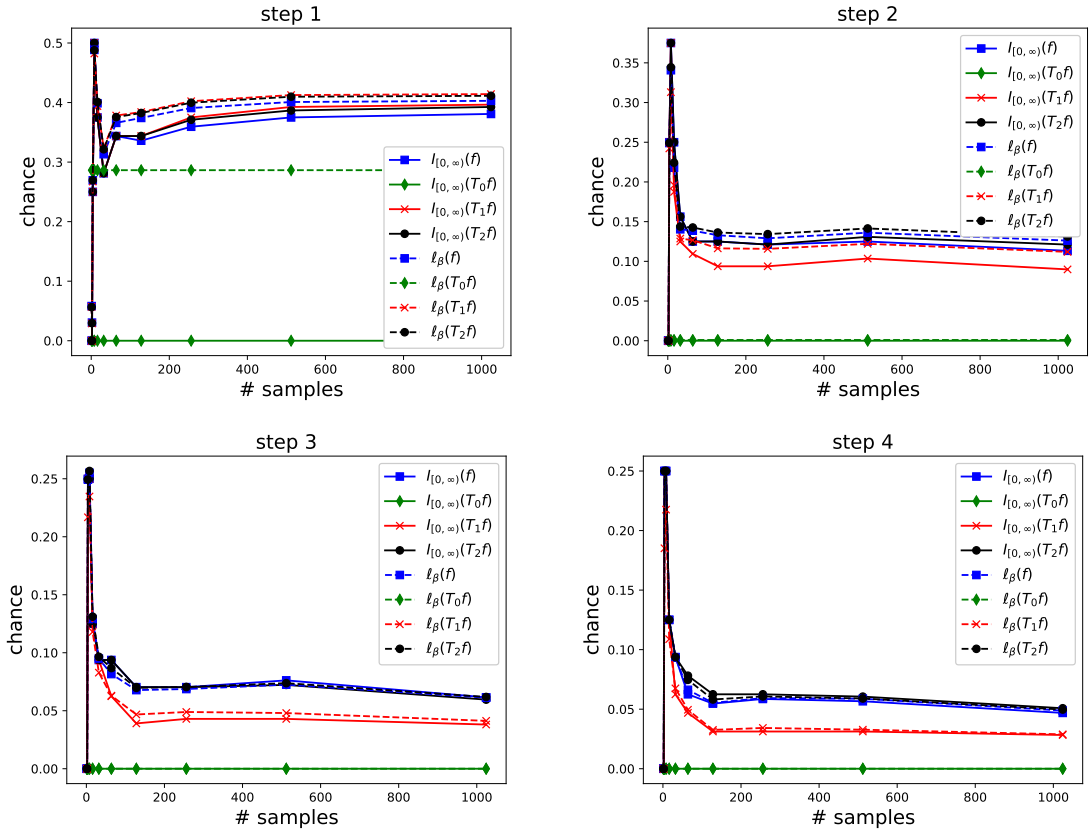

Figure 4. *The chance evaluated with different numbers of samples, by different Taylor approximations, and using different smoothing and penalty parameters* $(\beta, \gamma) = (2^{n+2}, 10^{n+2})$ *for step* $n = 1, 2, 3, 4$.

and leads to more accurate approximation of the chance for $g = f, T_K f$, $K = 0, 1, 2$; (3) as the penalty parameter γ increases, the violation of the chance $P(g > 0)$ exceeding $\alpha = 0.05$ is more strongly penalized, and the chance converges to 0.05 by the quadratic approximation $T_2 f$, while the chance by linear and constant approximations become smaller than $\alpha = 0.05$, i.e., the violation is over penalized; and (4) with an increasing number of samples for the SAA, the approximate chance becomes more accurate.

Table 1

Estimate of the chance $\hat{\ell}_\beta(g)$ *for* $g = f, T_K f$, $K = 0, 1, 2$, *by SAA with 1024 samples, and the estimation bias* $(\sqrt{MSE/1024})$ *induced by 1024 samples for SAA and estimation errors by Taylor approximations.*

step	$\hat{\ell}_\beta(f)$	SAA bias	$\lvert\hat{\ell}_\beta(f) - \hat{\ell}_\beta(T_0 f)\rvert$	$\lvert\hat{\ell}_\beta(f) - \hat{\ell}_\beta(T_1 f)\rvert$	$\lvert\hat{\ell}_\beta(f) - \hat{\ell}_\beta(T_2 f)\rvert$
0	8.25E-1	8.47E-3	1.52E-1	9.81E-3	6.14E-3
1	3.08E-1	1.08E-2	1.75E-1	1.28E-3	6.24E-3
2	1.14E-1	8.31E-3	1.14E-1	2.04E-2	2.91E-3
3	6.02E-2	6.77E-3	6.02E-2	2.26E-2	7.61E-4
4	5.00E-2	6.47E-3	4.98E-2	2.21E-2	8.60E-5

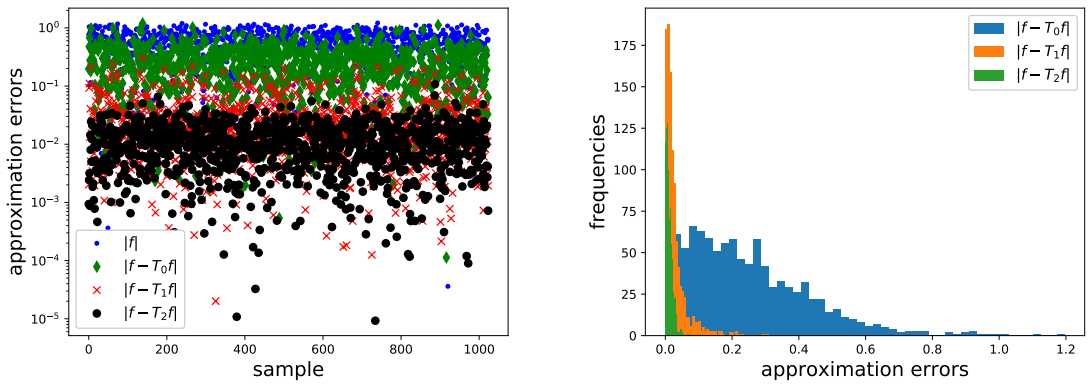

Figure 5. *Left: (absolute) values of the constraint function $|f|$ at 1024 samples and their approximation errors by Taylor approximations T_K, $K = 0, 1, 2,$. Right: histograms of the approximation errors $|f - T_K f|$.*

The observation on the accuracy of the Taylor approximations drawn from Figure 4 is further demonstrated in Figure 5, where the approximation errors for 1024 samples (at the optimal variable by the quadratic approximation) are shown on the left and their histograms are shown on the right, from which we can observe the quadratic approximation is statistically more accurate than the linear and constant approximations. Moreover, in Table 1, we report the SAA of the chance $\hat{\ell}_\beta(g)$ for $g = f, T_K f$, $K = 0, 1, 2$ with the same 1024 samples at each optimal variable z_n^* obtained by the quadratic approximation with different parameters $(\beta, \gamma) = (2^{n+2}, 10^{n+2})$ at step $n = 1, 2, 3, 4$. We take the initial guess of the optimal variable as $z_0^* = (18, \dots, 18)$ at step 0. The SAA bias for the estimate of the chance $\hat{\ell}_\beta(f)$ is computed as $\sqrt{\text{MSE}/1024}$, i.e., the square root of the mean square error of $\ell_\beta(f)$ divided by the number of samples, 1024, which is the bias from the true value induced by a finite number of samples. We can see that the quadratic approximation gives a two orders of magnitude smaller estimation error for the chance than the SAA bias with 1024 samples, while the linear and constant approximations lead to larger estimation errors. In each optimization step, 1024 state PDEs and 1024 adjoint PDEs have to be solved for the direct SAA $\hat{\ell}_\beta(f)$, while 1 state PDE, and 55 linearized PDEs (see the counts in Algorithm 4.2) are solved by the quadratic approximation. A speedup factor of $37 \approx 2048/56$ in terms of PDE solves is achieved. A higher speedup factor is achieved when (1) the number of samples is increased; and (2) the state PDE is more expensive to solve than the linearized PDEs, as is the case for nonlinear state PDEs.

The target $\bar{z} = (18, \dots, 18)$, and the optimal variables obtained by different approximations are shown in Figure 6, from which we observe that the optimal variable obtained by the quadratic approximation of the constraint function, i.e., using SAA for $\ell_\beta(T_2 f)$, is very close to that by SAA for $\ell_\beta(f)$. The distribution of the optimal variable by the quadratic approximation is shown in the right part of Figure 6, with the corresponding pressure field shown in the right part of Figure 7, whose (absolute) value in the region of interest $D_o = (0.25, 0.75)^2$ is effectively reduced from the initial state as displayed in the left part of Figure 7.

Finally, we plot the decay of the eigenvalues of (3.12) for the quadratic approximation $T_2 f$ at different optimization steps and different dimensions of the discrete random parameters in

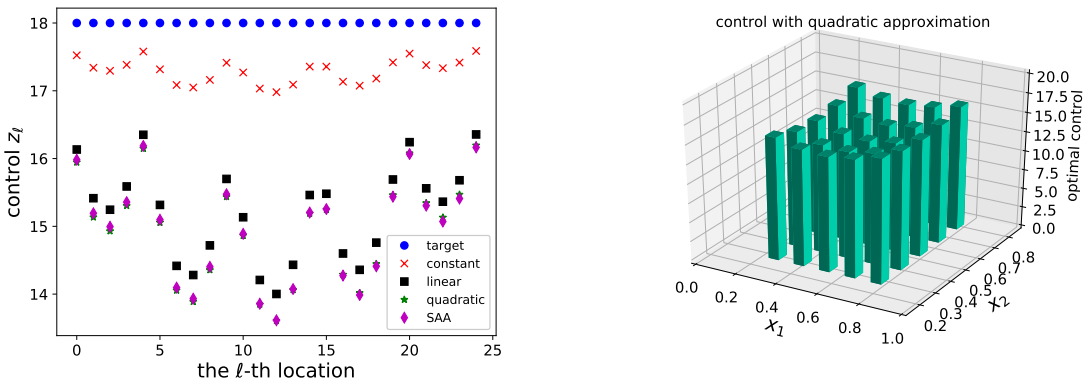

Figure 6. *Left: comparison of optimal variables z_ℓ at the ℓ-th location obtained by different approximations of the constraint function f, including Taylor approximations $T_K f$ for $K = 0$ (constant), $K = 1$ (linear), $K = 2$ (quadratic), and sample average approximation (SAA) with 1024 samples. Right: distribution of the optimal variable z^* obtained with quadratic approximation of the constraint function.*

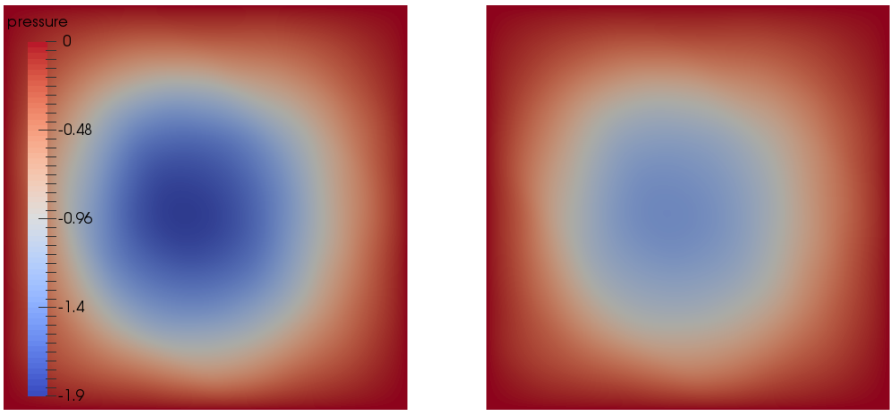

Figure 7. *Pressure field at the mean of the log-permeability $m = \bar{m}$ and the target $\bar{z}$ (left), the optimal variable z^* obtained with quadratic approximation (right); see Figure 6 for the value of z^*. The pressure (in absolute value) at the optimal variable z^* is effectively reduced from that at the target $\bar{z}$.*

Figure 8. By the similarity of the eigenvalue decay, we can conclude that the quadratic approximation is scalable with respect to the parameter dimensions in that the number of PDE solves are similar with increasing dimension. Moreover, the number of optimization iterations stays similar with increasing smoothing and penalty parameters (β, γ) as well as increasing dimension, thus demonstrating that the optimization method is also scalable for this example.

6. Conclusion. We proposed a Taylor approximation based continuation optimization method to solve chance and random PDE constrained optimization problems. We presented the derivation and efficient computation of the Taylor approximations using randomized algorithms. To address the challenges of discontinuous indicator function and inequality con-

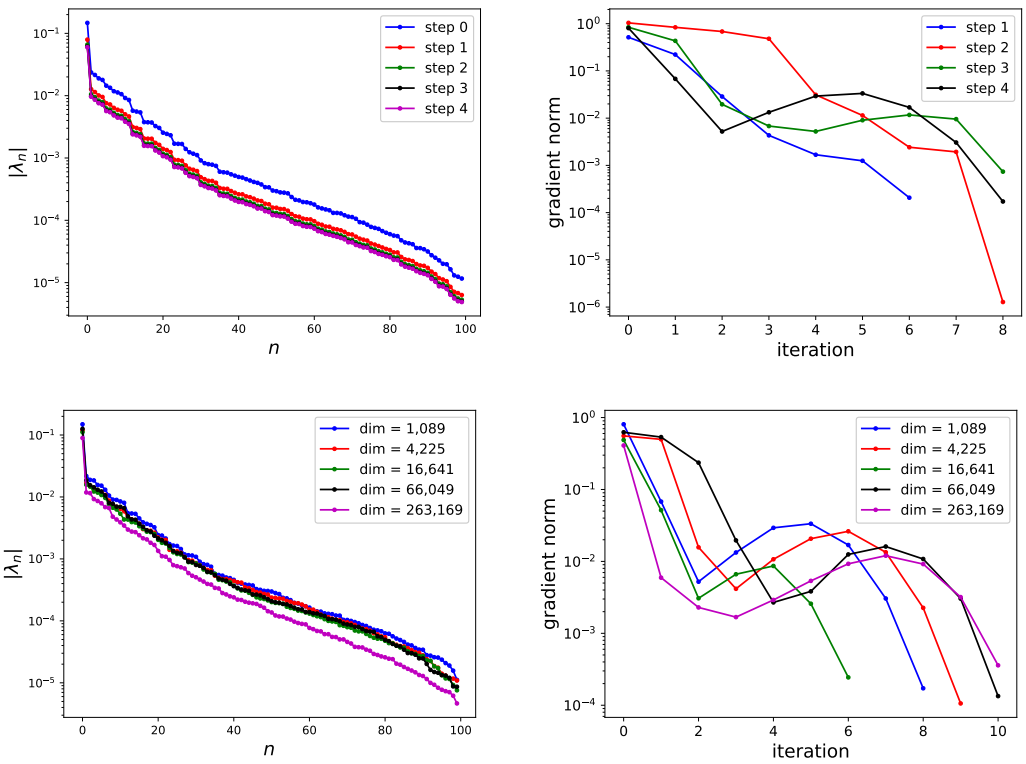

Figure 8. *Top: decay of the eigenvalues* (3.12) *for the quadratic approximation* $T_2 f$ *(left) and decay of the gradient norm of the BFGS optimization (right) at different steps (corresponding to Figure 4) with dimension* $1,089$ *for discrete* m*. Bottom: the same plot at optimal variables with different dimensions.*

straint, we employed a smooth approximation and a penalty method in a continuation BFGS optimization algorithm. We compared the accuracy of the Taylor (constant, linear, and quadratic) approximations and sample average approximation for both the chance evaluation and the optimal variable, demonstrated the acceleration by the Taylor approximation, reported the convergence of the continuation optimization algorithm in satisfying the chance constraint and minimizing the objective functional, and showed the scalability of the proposed method in that the number of PDE solves is essentially insensitive to increasing dimension of the random parameters. In particular, for the test problem the quadratic Taylor approximation achieves two orders of magnitude higher accuracy than SAA at 37X lower cost measured by the number of PDE solves.

This work motivates the following research directions: (1) since the local and quadratic Taylor approximation may become less accurate for highly nonlinear functions and large variance, higher order (beyond quadratic) and/or nonlocal Taylor approximations may improve the accuracy and efficiency of the proposed method; these rely on efficient low-rank derivative tensor decompositions which only require the tensor action [4] and/or efficient Gaussian mixture approximation of the measure μ; (2) vector or function valued chance constraint functions [37], e.g., representing pointwise pressure or the whole pressure field, require further

development of the optimization method with respect to the Taylor approximations and Lagrangian approach to computing the gradient; (3) to deal with extreme chance with a critical value $\alpha \ll 1$, importance sampling [69, 70] with Taylor approximation in the failure region rather than at the mean of the random parameters can be employed; and (4) theoretical analysis of the convergence of the Taylor approximations and the continuation optimization remain open; a promising approach is provided by the set/epigraphical convergence analysis in [75].

Appendix A. Some equations in Section 4.4.2. By setting the variation of $\mathcal{L}_2$ w.r.t. u to zero, we obtain: find $v^* \in \mathcal{V}$ such that

(A.1)
$$
\begin{aligned}
&\langle \tilde{u}, \partial_{uv}\bar{r}v^*\rangle = \\
&-\langle \tilde{u}, \partial_u\bar{q} + d^f\partial_u\bar{f} + \partial_{um}\bar{r}_f m^f + \partial_{um}\bar{f}m^f\rangle \\
&-\langle \tilde{u}, \partial_{uu}\bar{r}_q(u^q)^* + \partial_{uu}\bar{q}(u^q)^*\rangle \\
&-\sum_{n=1}^{N_q}\langle \tilde{u}, \partial_{umv}\bar{r}_q\hat{v}_n^q(\psi_n^q)^* + \partial_{umu}\bar{r}_q\hat{u}_n^q(\psi_n^q)^* + \partial_{umu}\bar{q}\hat{u}_n^q(\psi_n^q)^* + \partial_{umm}\bar{r}_q\psi_n^q(\psi_n^q)^* + \partial_{umm}\bar{q}\psi_n^q(\psi_n^q)^*\rangle \\
&-\sum_{n=1}^{N_q}\langle \tilde{u}, \partial_{uvu}\bar{r}_q\hat{u}_n^q(\hat{v}_n^q)^* + \partial_{uvm}\bar{r}_q\psi_n^q(\hat{v}_n^q)^*\rangle \\
&-\sum_{n=1}^{N_q}\langle \tilde{u}, \partial_{uuv}\bar{r}_q\hat{v}_n^q(\hat{u}_n^q)^* + \partial_{uuu}\bar{r}_q\hat{u}_n^q(\hat{u}_n^q)^* + \partial_{uuu}\bar{q}\hat{u}_n^q(\hat{u}_n^q)^* + \partial_{uum}\bar{r}_q\psi_n^q(\hat{u}_n^q)^* + \partial_{uum}\bar{q}\psi_n^q(\hat{u}_n^q)^*\rangle \\
&-\langle \tilde{u}, \partial_{uu}\bar{r}_f(u^f)^* + \partial_{uu}\bar{f}(u^f)^*\rangle \\
&-\sum_{n=1}^{N_f}\langle \tilde{u}, \partial_{umv}\bar{r}_f\hat{v}_n^f(\psi_n^f)^* + \partial_{umu}\bar{r}_f\hat{u}_n^f(\psi_n^f)^* + \partial_{umu}\bar{f}\hat{u}_n^f(\psi_n^f)^* + \partial_{umm}\bar{r}_f\psi_n^f(\psi_n^f)^* + \partial_{umm}\bar{f}\psi_n^f(\psi_n^f)^*\rangle \\
&-\sum_{n=1}^{N_f}\langle \tilde{u}, \partial_{uvu}\bar{r}_f\hat{u}_n^f(\hat{v}_n^f)^* + \partial_{uvm}\bar{r}_f\psi_n^f(\hat{v}_n^f)^*\rangle \\
&-\sum_{n=1}^{N_f}\langle \tilde{u}, \partial_{uuv}\bar{r}_f\hat{v}_n^f(\hat{u}_n^f)^* + \partial_{uuu}\bar{r}_f\hat{u}_n^f(\hat{u}_n^f)^* + \partial_{uuu}\bar{f}\hat{u}_n^f(\hat{u}_n^f)^* + \partial_{uum}\bar{r}_f\psi_n^f(\hat{u}_n^f)^* + \partial_{uum}\bar{f}\psi_n^f(\hat{u}_n^f)^*\rangle,\ \forall \tilde{u} \in \mathcal{U},
\end{aligned}
$$

where the constant d^f in the first line is given by

$$
d^f = \nabla\mathcal{S}_\gamma(g_{\beta,M_f}(T_2^{\mathrm{LR}}f))\frac{1}{M_f}\sum_{i=1}^{M_f}\nabla\ell_\beta(T_2^{\mathrm{LR}}f(m_i,z)). \tag{A.2}
$$

We evaluate the gradient of the cost functional (4.10) as

$$
\langle \tilde{z}, \nabla_z\mathcal{E}(z)\rangle = \langle \tilde{z}, \partial_z\mathcal{L}_2\rangle
$$

$$
\begin{aligned}
&= \langle \tilde{z}, \partial_z \bar{q} + \nabla_z \mathcal{P}(z) + d^f \partial_z \bar{f} + \partial_{zm} \bar{r}_f m^f + \partial_{zm} \bar{f} m^f \rangle \\
&+ \langle \tilde{z}, \partial_{zv} \bar{r} v^* \rangle \\
&+ \langle \tilde{z}, \partial_{zu} \bar{r}_q (u^q)^* + \partial_{zu} \bar{q} (u^q)^* \rangle \\
&+ \sum_{n=1}^{N_q} \langle \tilde{z}, \partial_{zmv} \bar{r}_q \hat{v}_n^q (\psi_n^q)^* + \partial_{zmu} \bar{r}_q \hat{u}_n^q (\psi_n^q)^* + \partial_{zmm} \bar{r}_q \psi_n^q (\psi_n^q)^* + \partial_{zmu} \bar{q} \hat{u}_n^q (\psi_n^q)^* + \partial_{zmm} \bar{q} \psi_n^q (\psi_n^q)^* \rangle \\
&+ \sum_{n=1}^{N_q} \langle \tilde{z}, \partial_{zvu} \bar{r}_q \hat{u}_n^q (\hat{v}_n^q)^* + \partial_{zvm} \bar{r}_q \psi_n^q (\hat{v}_n^q)^* \rangle \\
&+ \sum_{n=1}^{N_q} \langle \tilde{z}, \partial_{zuv} \bar{r}_q \hat{v}_n^q (\hat{u}_n^q)^* + \partial_{zuu} \bar{r}_q \hat{u}_n^q (\hat{u}_n^q)^* + \partial_{zum} \bar{r}_q \psi_n^q (\hat{u}_n^q)^* + \partial_{zuu} \bar{q} \hat{u}_n^q (\hat{u}_n^q)^* + \partial_{zum} \bar{q} \psi_n^q (\hat{u}_n^q)^* \rangle \\
&+ \langle \tilde{z}, \partial_{zu} \bar{r}_f (u^f)^* + \partial_{zu} \bar{f} (u^f)^* \rangle \\
&+ \sum_{n=1}^{N_f} \langle \tilde{z}, \partial_{zmv} \bar{r}_f \hat{v}_n^f (\psi_n^f)^* + \partial_{zmu} \bar{r}_f \hat{u}_n^f (\psi_n^f)^* + \partial_{zmm} \bar{r}_f \psi_n^f (\psi_n^f)^* + \partial_{zmu} \bar{f} \hat{u}_n^f (\psi_n^f)^* + \partial_{zmm} \bar{f} \psi_n^f (\psi_n^f)^* \rangle \\
&+ \sum_{n=1}^{N_f} \langle \tilde{z}, \partial_{zvu} \bar{r}_f \hat{u}_n^f (\hat{v}_n^f)^* + \partial_{zvm} \bar{r}_f \psi_n^f (\hat{v}_n^f)^* \rangle \\
&+ \sum_{n=1}^{N_f} \langle \tilde{z}, \partial_{zuv} \bar{r}_f \hat{v}_n^f (\hat{u}_n^f)^* + \partial_{zuu} \bar{r}_f \hat{u}_n^f (\hat{u}_n^f)^* + \partial_{zum} \bar{r}_f \psi_n^f (\hat{u}_n^f)^* + \partial_{zuu} \bar{f} \hat{u}_n^f (\hat{u}_n^f)^* + \partial_{zum} \bar{f} \psi_n^f (\hat{u}_n^f)^* \rangle.
\end{aligned}
$$

REFERENCES

[1] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized ℓ_0-sparsification*, SIAM Journal on Scientific Computing, 36 (2014), pp. A2122–A2148, https://doi.org/10.1137/130933381.

[2] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems*, SIAM Journal on Scientific Computing, 38 (2016), pp. A243–A272, https://doi.org/10.1137/140992564.

[3] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations*, SIAM/ASA Journal on Uncertainty Quantification, 5 (2017), pp. 1166–1192, https://doi.org/10.1137/16M106306X.

[4] N. Alger, P. Chen, and O. Ghattas, *Tensor train construction from tensor actions, with application to compression of large high order derivative tensors*, SIAM Journal on Scientific Computing, 42 (2020), pp. A3516–A3539.

[5] A. A. Ali, E. Ullmann, and M. Hinze, *Multilevel Monte Carlo analysis for optimal control of elliptic PDEs with random coefficients*, SIAM/ASA Journal on Uncertainty Quantification, 5 (2017), pp. 466–492.

[6] A. Alla, M. Hinze, P. Kolvenbach, O. Lass, and S. Ulbrich, *A certified model reduction approach for robust parameter optimization with pde constraints*, Advances in Computational Mathematics, 45 (2019), pp. 1221–1250.

[7] O. Bashir, K. Willcox, O. Ghattas, B. van Bloemen Waanders, and J. Hill, *Hessian-based model reduction for large-scale systems with initial condition inputs*, International Journal for Numerical Methods in Engineering, 73 (2008), pp. 844–868.

[8] P. Benner, A. Onwunta, and M. Stoll, *Block-diagonal preconditioning for optimal control problems*

constrained by PDEs with uncertain inputs, SIAM Journal on Matrix Analysis and Applications, 37 (2016), pp. 491–518.

[9] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, 1999.

[10] A. Borzì, *Multigrid and sparse-grid schemes for elliptic control problems with random coefficients*, Computing and Visualization in Science, 13 (2010), pp. 153–160.

[11] A. Borzì, V. Schulz, C. Schillings, and G. Von Winckel, *On the treatment of distributed uncertainties in PDE-constrained optimization*, GAMM-Mitteilungen, 33 (2010), pp. 230–246.

[12] T. Bui-Thanh, C. Burstedde, O. Ghattas, J. Martin, G. Stadler, and L. C. Wilcox, *Extreme-scale UQ for Bayesian inverse problems governed by PDEs*, in SC12: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2012.

[13] T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part I: Inverse shape scattering of acoustic waves*, Inverse Problems, 28 (2012), p. 055001, https://doi.org/10.1088/0266-5611/28/5/055001.

[14] T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part II: Inverse medium scattering of acoustic waves*, Inverse Problems, 28 (2012), p. 055002, https://doi.org/10.1088/0266-5611/28/5/055002.

[15] T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems. Part III: Inverse medium scattering of electromagnetic waves*, Inverse Problems and Imaging, 7 (2013), pp. 1139–1155.

[16] T. Bui-Thanh and O. Ghattas, *A scalable algorithm for MAP estimators in Bayesian inverse problems with Besov priors*, Inverse Problems and Imaging, 9 (2015), pp. 27–54.

[17] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler, *A computational framework for infinite-dimensional Bayesian inverse problems Part I: The linearized case, with application to global seismic inversion*, SIAM Journal on Scientific Computing, 35 (2013), pp. A2494–A2523, https://doi.org/10.1137/12089586X.

[18] C. Chen and O. L. Mangasarian, *Smoothing methods for convex inequalities and linear complementarity problems*, Mathematical programming, 71 (1995), pp. 51–69.

[19] P. Chen, *Sparse quadrature for high-dimensional integration with Gaussian measure*, ESAIM: Mathematical Modelling and Numerical Analysis, 52 (2018), pp. 631–657.

[20] P. Chen and O. Ghattas, *Hessian-based sampling for high-dimensional model reduction*, International Journal for Uncertainty Quantification, 9 (2019).

[21] P. Chen and O. Ghattas, *Sparse polynomial approximation for optimal control problems constrained by elliptic PDEs with lognormal random coefficients*, preprint, (2019). https://arxiv.org/abs/1903.05547.

[22] P. Chen and O. Ghattas, *Projected Stein variational gradient descent*, in Advances in Neural Information Processing Systems, 2020.

[23] P. Chen, M. Haberman, and O. Ghattas, *Optimal design of acoustic metamaterial cloaks under uncertainty*, Journal of Computational Physics, 431 (2021), p. 110114.

[24] P. Chen and A. Quarteroni, *Weighted reduced basis method for stochastic optimal control problems with elliptic PDE constraints*, SIAM/ASA J. Uncertainty Quantification, 2 (2014), pp. 364–396.

[25] P. Chen, A. Quarteroni, and G. Rozza, *Stochastic optimal Robin boundary control problems of advection-dominated elliptic equations*, SIAM Journal on Numerical Analysis, 51 (2013), pp. 2700–2722.

[26] P. Chen, A. Quarteroni, and G. Rozza, *Multilevel and weighted reduced basis method for stochastic optimal control problems constrained by Stokes equations*, Numerische Mathematik, 133 (2016), pp. 67–102.

[27] P. Chen, U. Villa, and O. Ghattas, *Hessian-based adaptive sparse quadrature for infinite-dimensional Bayesian inverse problems*, Computer Methods in Applied Mechanics and Engineering, 327 (2017), pp. 147–172, https://doi.org/10.1016/j.cma.2017.08.016.

[28] P. Chen, U. Villa, and O. Ghattas, *Taylor approximation and variance reduction for PDE-constrained optimal control under uncertainty*, Journal of Computational Physics, 385 (2019), pp. 163–186, https://arxiv.org/abs/1804.04301.

[29] P. Chen, K. Wu, J. Chen, T. O'Leary-Roseberry, and O. Ghattas, *Projected Stein variational Newton: A fast and scalable Bayesian inference method in high dimensions*, Advances in Neural Information Processing Systems, (2019).

[30] P. Chen, K. Wu, and O. Ghattas, *Bayesian inference of heterogeneous epidemic models: Application to COVID-19 spread accounting for long-term care facilities*, arXiv preprint arXiv:2011.01058, (2020).
[31] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz, *Shape optimization under uncertainty–a stochastic programming perspective*, SIAM Journal on Optimization, 19 (2009), pp. 1610–1632.
[32] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz, *Risk averse shape optimization*, SIAM journal on control and optimization, 49 (2011), pp. 927–947.
[33] S. Conti, M. Rumpf, R. Schultz, and S. Tölkes, *Stochastic dominance constraints in elastic shape optimization*, SIAM Journal on Control and Optimization, 56 (2018), pp. 3021–3034.
[34] B. Crestel, A. Alexanderian, G. Stadler, and O. Ghattas, *A-optimal encoding weights for nonlinear inverse problems, with application to the Helmholtz inverse problem*, Inverse Problems, 33 (2017), p. 074008, http://iopscience.iop.org/10.1088/1361-6420/aa6d8e.
[35] S. De, J. Hampton, K. Maute, and A. Doostan, *Topology optimization under uncertainty using a stochastic gradient-based approach*, arXiv preprint arXiv:1902.04562, (2019).
[36] J. Dick, Q. Le Gia, and C. Schwab, *Higher order quasi–Monte Carlo integration for holomorphic, parametric operator equations*, SIAM/ASA Journal on Uncertainty Quantification, 4 (2016), pp. 48–79.
[37] M. H. Farshbaf-Shaker, R. Henrion, and D. Hömberg, *Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization*, Set-Valued and Variational Analysis, 26 (2018), pp. 821–841.
[38] P. H. Flath, L. C. Wilcox, V. Akçelik, J. Hill, B. van Bloemen Waanders, and O. Ghattas, *Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations*, SIAM Journal on Scientific Computing, 33 (2011), pp. 407–432, https://doi.org/10.1137/090780717.
[39] S. Garreis, T. Surowiec, and M. Ulbrich, *An interior-point approach for solving risk-averse PDE-constrained optimization problems with coherent risk measures*, Preprint, submitted, Technical University of Munich, (2019).
[40] C. Geiersbach, E. Loayza-Romero, and K. Welker, *Stochastic approximation for optimization in shape spaces*, arXiv preprint arXiv:2001.10786, (2020).
[41] C. Geiersbach and T. Scarinci, *Stochastic proximal gradient methods for nonconvex problems in Hilbert spaces*, arXiv preprint arXiv:2001.01329, (2020).
[42] C. Geiersbach and W. Wollner, *A stochastic gradient method with mesh refinement for PDE constrained optimization under uncertainty*, arXiv preprint arXiv:1905.08650, (2019).
[43] A. Geletu, A. Hoffmann, P. Schmidt, and P. Li, *Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation*, ESAIM: Control, Optimisation and Calculus of Variations, 26 (2020), p. 70.
[44] M. D. Gunzburger, *Perspectives in Flow Control and Optimization*, SIAM, Philadelphia, 2003.
[45] M. D. Gunzburger, H.-C. Lee, and J. Lee, *Error estimates of stochastic optimal Neumann boundary control problems*, SIAM Journal on Numerical Analysis, 49 (2011), pp. 1532–1552, https://doi.org/10.1137/100801731, http://link.aip.org/link/?SNA/49/1532/1.
[46] S. Guo, H. Xu, and L. Zhang, *Convergence analysis for mathematical programs with distributionally robust chance constraint*, SIAM Journal on optimization, 27 (2017), pp. 784–816.
[47] N. Halko, P. G. Martinsson, and J. A. Tropp, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Review, 53 (2011), pp. 217–288.
[48] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, vol. 23 of Mathematical Modelling: Theory and Applications, Springer Netherlands, 2009, https://doi.org/10.1007/978-1-4020-8839-1, https://link.springer.com/978-1-4020-8839-1.
[49] L. S. Hou, J. Lee, and H. Manouzi, *Finite element approximations of stochastic optimal control problems constrained by stochastic elliptic PDEs*, Journal of Mathematical Analysis and Applications, 384 (2011), pp. 87–103.
[50] T. Isaac, N. Petra, G. Stadler, and O. Ghattas, *Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet*, Journal of Computational Physics, 296 (2015), pp. 348–368, https://doi.org/10.1016/j.jcp.2015.04.047.

[51] P. Kolvenbach, O. Lass, and S. Ulbrich, *An approach for robust PDE-constrained optimization with application to shape optimization of electrical engines and of dynamic elastic structures under uncertainty*, Optimization and Engineering, 19 (2018), pp. 697–731.

[52] D. Kouri, D. Heinkenschloos, M. Ridzal, and B. Van Bloemen Waanders, *A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty*, SIAM Journal on Scientific Computing, 35 (2012), pp. 1847–1879.

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

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

[55] A. Kunoth and C. Schwab, *Analytic regularity and GPC approximation for control problems constrained by linear parametric elliptic and parabolic PDEs*, SIAM Journal on Control and Optimization, 51 (2013), pp. 2442–2471.

[56] A. Kunoth and C. Schwab, *Sparse adaptive tensor Galerkin approximations of stochastic PDE-constrained control problems*, SIAM/ASA Journal on Uncertainty Quantification, 4 (2016), pp. 1034–1059.

[57] O. Lass and S. Ulbrich, *Model order reduction techniques with a posteriori error control for nonlinear robust optimization governed by partial differential equations*, SIAM Journal on Scientific Computing, 39 (2017), pp. S112–S139.

[58] C. Li and G. Stadler, *Sparse solutions in optimal control of PDEs with uncertain parameters: The linear case*, SIAM Journal on Control and Optimization, 57 (2019), pp. 633–658.

[59] J. Li, X. Wang, and K. Zhang, *An efficient alternating direction method of multipliers for optimal control problems constrained by random helmholtz equations*, Numerical Algorithms, 78 (2018), pp. 161–191.

[60] F. Lindgren, H. Rue, and J. Lindström, *An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73 (2011), pp. 423–498, https://doi.org/10.1111/j.1467-9868.2011.00777.x, http://dx.doi.org/10.1111/j.1467-9868.2011.00777.x.

[61] J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, vol. 170 of Grundlehren der mathematischen Wissenschaften, Springer-Verlag Berlin Heidelberg, 1971.

[62] A. Logg, K.-A. Mardal, and G. Wells, *Automated Solution of Differential Equations by the Finite Element Method: The FEniCS book*, vol. 84, Springer Science & Business Media, 2012.

[63] L. Ma, Q. Li, and L. Jiang, *Local–global model reduction method for stochastic optimal control problems constrained by partial differential equations*, Computer Methods in Applied Mechanics and Engineering, 339 (2018), pp. 514–541.

[64] J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas, *A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion*, SIAM Journal on Scientific Computing, 34 (2012), pp. A1460–A1487, https://doi.org/10.1137/110845598.

[65] M. Martin, F. Nobile, and P. Tsilifis, *A multilevel stochastic gradient method for pde-constrained optimal control problems with uncertain parameters*, arXiv preprint arXiv:1912.11900, (2019).

[66] J. L. Morales and J. Nocedal, *Remark on "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization"*, ACM Transactions on Mathematical Software (TOMS), 38 (2011), pp. 1–4.

[67] L. Ng and K. Willcox, *Multifidelity approaches for optimization under uncertainty*, International Journal for Numerical Methods in Engineering, 100 (2014), pp. 746–772, https://doi.org/10.1002/nme.4761.

[68] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Verlag, Berlin, Heidelberg, New York, second ed., 2006.

[69] 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.

[70] B. Peherstorfer, B. Kramer, and K. Willcox, *Multifidelity preconditioning of the cross-entropy method for rare event simulation and failure probability estimation*, SIAM/ASA Journal on Uncer-

tainty Quantification, 6 (2018), pp. 737–761.
[71] N. Petra, J. Martin, G. Stadler, and O. Ghattas, *A computational framework for infinite-dimensional Bayesian inverse problems: Part II. Stochastic Newton MCMC with application to ice sheet flow inverse problems*, SIAM Journal on Scientific Computing, 36 (2014), pp. A1525–A1555.
[72] L. Qi, D. Sun, and G. Zhou, *A new look at smoothing newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Mathematical programming, 87 (2000), pp. 1–35.
[73] L. Roald and G. Andersson, *Chance-constrained AC optimal power flow: Reformulations and efficient algorithms*, IEEE Transactions on Power Systems, 33 (2017), pp. 2906–2918.
[74] E. Rosseel and G. N. Wells, *Optimal control with stochastic PDE constraints and uncertain controls*, Computer Methods in Applied Mechanics and Engineering, 213 (2012), pp. 152–167.
[75] J. O. Royset, *Set-convergence and its application: A tutorial*, Set-Valued and Variational Analysis, (2020), pp. 1–26.
[76] A. K. Saibaba, J. Lee, and P. K. Kitanidis, *Randomized algorithms for generalized Hermitian eigenvalue problems with application to computing Karhunen–Loève expansion*, Numerical Linear Algebra with Applications, 23 (2016), pp. 314–339.
[77] C. Schillings and C. Schwab, *Sparse, adaptive Smolyak quadratures for Bayesian inverse problems*, Inverse Problems, 29 (2013), p. 065011.
[78] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on Stochastic Programming: Modeling and Theory*, Society for Industrial and Applied Mathematics, 2009.
[79] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numerica, 19 (2010), pp. 451–559, https://doi.org/10.1017/S0962492910000061.
[80] H. Tiesler, R. M. Kirby, D. Xiu, and T. Preusser, *Stochastic collocation for optimal control problems with stochastic PDE constraints*, SIAM Journal on Control and Optimization, 50 (2012), pp. 2659–2682.
[81] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications*, vol. 112, American Mathematical Society, Providence, RI, 2010.
[82] S. Uryasev, *Probabilistic constrained optimization: methodology and applications*, vol. 49, Springer Science & Business Media, 2013.
[83] W. van Ackooij, R. Henrion, and P. Pérez-Aros, *Generalized gradients for probabilistic/robust (probust) constraints*, Optimization, (2019), pp. 1–29.
[84] W. Van Ackooij and J. Malick, *Eventual convexity of probability constraints with elliptical distributions*, Mathematical Programming, 175 (2019), pp. 1–27.
[85] W. van Ackooij and P. Pèrez-Aros, *Generalized differentiation of probability functions acting on an infinite system of constraints*, SIAM Journal on Optimization, 29 (2019), pp. 2179–2210.
[86] A. Van Barel and S. Vandewalle, *Robust optimization of PDEs with random coefficients using a multilevel Monte Carlo method*, SIAM/ASA Journal on Uncertainty Quantification, 7 (2019), pp. 174–202.
[87] K. Wu, P. Chen, and O. Ghattas, *A fast and scalable computational framework for large-scale and high-dimensional Bayesian optimal experimental design*, arXiv preprint arXiv:2010.15196, (2020).
[88] B. Xu, S. E. Boyce, Y. Zhang, Q. Liu, L. Guo, and P.-A. Zhong, *Stochastic programming with a joint chance constraint model for reservoir refill operation considering flood risk*, Journal of Water Resources Planning and Management, 143 (2017), p. 04016067.
[89] H. Yang and M. Gunzburger, *Algorithms and analyses for stochastic optimization for turbofan noise reduction using parallel reduced-order modeling*, Computer Methods in Applied Mechanics and Engineering, 319 (2017), pp. 217–239.
[90] M. J. Zahr, K. T. Carlberg, and D. P. Kouri, *An efficient, globally convergent method for optimization under uncertainty using adaptive model reduction and sparse grids*, SIAM/ASA Journal on Uncertainty Quantification, 7 (2019), pp. 877–912.