TAYLOR APPROXIMATION AND VARIANCE REDUCTION FOR
PDE-CONSTRAINED OPTIMAL CONTROL UNDER
UNCERTAINTY *

PENG CHEN †, UMBERTO VILLA †, AND OMAR GHATTAS ‡

Abstract. In this work we develop a scalable computational framework for the solution of PDE-constrained optimal control problems under high/infinite-dimensional uncertainty. Specifically, we consider a mean-variance risk adverse formulation of the stochastic optimization problem and employ a Taylor expansion with respect to the uncertain parameter either to directly approximate the control objective or as a control variate for Monte Carlo variance reduction. The expressions for the mean and variance of the Taylor approximation are known analytically, although their evaluation requires to efficiently compute the trace of the (preconditioned) Hessian of the control objective. We propose to estimate such trace by solving a generalized eigenvalue problem using a randomized algorithm that only requires the action of the Hessian in a small number of random directions. Then, the computational work does not depend on the nominal dimension of the uncertain parameter but only on the effective dimension (i.e. the rank of the preconditioned Hessian), thus significantly alleviating or breaking the curse of dimensionality. Moreover, when the use of the possibly biased Taylor approximation results in large error of the optimal control function, we use such approximation as a control variate for variance reduction, which results in considerable computational savings (several orders of magnitude) compared to a simple Monte Carlo method. In summary, our approach amounts to solving an optimal control problem constrained by the original PDE and a set of linearized PDEs, which arise from the computation of the gradient and the Hessian of the control objective with respect to the uncertain parameter. We use the Lagrangian formalism to derive expression for the gradient with respect to the control variable and apply a gradient-based optimization method to solve the problem. Symbolic differentiation of the Lagrangian with respect to the state variable, the adjoint variable, the uncertain parameter and the control variable allows us to calculate (high-order) derivatives even for complex PDE models. We demonstrate the accuracy, efficiency, and scalability of the proposed computational method for two problems with infinite-dimensional uncertain parameters: a subsurface flow in a porous medium modeled as an elliptic PDE, and a turbulent jet flow modeled as Reynolds-averaged Navier–Stokes equations coupled with a nonlinear advection-diffusion equation. In particular, for the latter (and more challenging) example we show scalability of our algorithm up to one million parameters after discretization.

Key words. PDE-constrained optimal control, Taylor approximation, Monte Carlo integration, variance reduction, uncertainty quantification, curse of dimensionality, symbolic differentiation

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

1. Introduction. Optimal control or optimization problems constrained by partial differential equations (PDEs) are ubiquitous in many applications of practical relevance for science and engineering. The objective of such problems is to seek an optimal control function that minimizes a cost functional, which often consists of a control objective and a penalization term. The control objective is often given as a distance between the solution of the PDE and a desirable state or, in general, as a quantity of interest related to the solution, while the penalization term is used to limit the energy norm or to impose regularity of the control function. In practical applications, uncertainties are inevitable and can arise from various sources, such as the

---

*This work was supported by DARPA’s EQUiPS program under contract number W911NF-15-2-0121, NSF grants CBET-1508713 and ACI-1550593, and DOE grants DE-SC0010518 and DE-SC0009286.

†Institute for Computational Engineering & Sciences, The University of Texas at Austin, Austin, TX 78712 (peng@ices.utexas.edu, uvilla@ices.utexas.edu).

‡Institute for Computational Engineering & Sciences, Department of Mechanical Engineering, and Department of Geological Sciences, The University of Texas at Austin, Austin, TX 78712 (omar@ices.utexas.edu).
PDE coefficients, initial or boundary conditions, external loadings, or computational geometries. Different realizations of these uncertain parameters may lead to significant differences in the optimal control function. Mathematical theories and computational methods have been developed for several decades to deal with deterministic PDE-constrained optimal control and optimization problems [43, 29, 30, 34, 59, 13]. More recently, PDE-constrained optimal control and optimization problems under uncertainty have become important areas of research and have drawn increasing attention [14, 56, 35, 31, 52, 38, 58, 22, 42, 21, 40, 47, 23, 41, 39, 10, 2, 3].

Different formulations of the cost functional have been studied to incorporate uncertainty into the PDE-constrained optimal control or optimization problems and to obtain robust optimal control functions. A straightforward choice is to optimize the mean of the control objective, i.e., the integration of the control objective with respect to the probability measure of the uncertain parameter [14, 35, 31, 38, 21]. However, this formulation does not control the variability of the control objective. To avoid a possible and undesirable large variation of the control objective, which may represent the risk of a system failure, one can include the variance or higher moments of the control objective in the cost functional [52, 10, 58]. An alternative approach is to use the Value-at-Risk (VaR) or conditional VaR [39], which measure the (conditional) probability of the control objective surpassing a certain critical value. An extreme choice is robust optimization (also referred as min-max optimization), where one seeks for the control that yields the optimal value of the objective in the worst-case scenario, i.e. to optimize the extreme value of the control objective in the range of the uncertain parameter [12, 42].

All these formulations present similar computational challenges. The prominent one is to compute the statistics of the control objective (i.e., mean, variance, higher moments, or conditional probabilities) and involves the integration of the control objective or related quantities with respect to the probability measure of the uncertain parameter. This challenge becomes more relevant for high- or infinite-dimensional uncertain parameter space, which is often the case when the parameter is a random spatially correlated field. A straightforward approach is to apply a Monte Carlo method to compute the statistics by taking the average of the control objective or related quantities at a set of samples randomly drawn according to the probability measure of the parameter. However, the convergence rate of the Monte Carlo estimator is only $O(M^{-1/2})$, where $M$ is the number of samples. This may lead to an extremely large number of samples needed to achieve a target accuracy, and therefore it may not be feasible when the state equation is a complex nonlinear PDE problem. For small to medium scale problems, quasi Monte Carlo methods using low-discrepancy sequences improve the convergence rate to $O((\log(M))^{d}/M)$ for $d$-dimensional parameter [25]. Stochastic Galerkin and stochastic collocation methods have also been used to compute the statistical moments in the cost functional [52, 58], provided that a suitable finite dimensional parametrization of the uncertain parameter, such as a truncated Karhunen–Loève expansion, is applicable. These methods achieve fast convergence when the control objective depends smoothly on the low-dimensional parameter, but suffer from the so-called curse-of-dimensionality, i.e. the convergence rate quickly deteriorates as the dimension of the parameter increases. Recent advancement in adaptive/anisotropic sparse quadrature [57, 19] and high-order quasi Monte Carlo methods [26] have been shown to achieve a dimension-independent convergence rate $N^{-s}$ with $s$ potentially much larger than $1/2$. Here $s$ is independent of the nominal dimension of the uncertain parameters and depends only on the choice of parametriza-
tion and on the regularity of the control objective with respect to the parameter, thus breaking the curse of dimensionality. However, if the choice of the parametrization does not correctly capture the effective dimension in parameter space, the convergence of these methods may still be slow.

Another computational challenge is that, in application of practical relevance, the PDE may lead, after discretization, to possibly nonlinear large-scale systems that are extremely expensive to solve. Thus, only a limited number of high-fidelity PDE-solves can be afforded. This challenge prevents a direct application of most of the computational methods introduced above as they require a large number of evaluations of the control objective for computing its statistics. To tackle this challenge, reduced basis methods [53, 49, 33] or other model order reduction techniques [9] can be employed to exploit the low dimensionality of the solution manifold, even when the parameter lives in high dimensions. Such methods solve the high-fidelity state equation at a few carefully selected samples of the parameter, and use such high-fidelity solutions (known as snapshots) as the basis to approximate the solution at any other realization of the uncertain parameter by a (Petrov)-Galerkin projection. Solving the projected problem in the reduced basis space allows for considerable computational savings compared to solving in the full space, such as a finite-element space with a large number of degrees of freedom. These techniques have been successfully developed for solving parametric and stochastic PDE-constrained optimal control and optimization problems with several classes of PDE models [46, 36, 11, 45, 21, 61, 23, 5]. Nevertheless, challenges in using such techniques arise for highly nonlinear problems that require effective affine approximation (see [49, 33] and references therein) or when the solution manifold becomes high-dimensional, even if the effective dimensions of the manifold of the control objective is low [18, 20].

In this work, we propose a scalable computational framework that uses a Taylor expansion of the control objective with respect to the uncertain field to efficiently and accurately approximate the cost functional. This framework takes advantage both of the smooth dependence of the control objective on the uncertain parameter and of structure of the Hessian of the control objective, which exposes the low effective dimension of the nominally high- or infinite-dimensional parameter space. Such Taylor expansions can be used directly to approximate the moments of the control objective as proposed in [2] or as control variates to reduce the variance of Monte Carlo estimators, as discussed in the present work. Analytic expressions for the mean and the variance of the Taylor expansions are known; we refer to [2] for such expressions in case of the linear and quadratic approximations. Such expressions depend on the trace of the preconditioned Hessian (second order derivative of the control objective with respect to the uncertain parameter) and/or its square. While for small scale problems, it is possible to explicitly compute the Hessian operator (see [28]), this approach is unfeasible for large-scale PDE problems, where — after discretization — the Hessian is a formally large dense operator implicitly defined by its action on a given vector. Gaussian trace estimators (see [4]) only need to perform the action of the Hessian in a limited number of random directions, and have been successfully used for the solution of optimal experimental design and optimization under uncertainty problems in [1]. However, such estimators may still require a large number of Hessian action to achieve high accuracy, see [4] for probabilistic bounds on the number of Hessian actions necessary to achieve a target accuracy with high probability. In this work, we approximate the trace of the preconditioned Hessian with the sum of its dominant eigenvalues. Randomized algorithms [32, 54] are computationally efficient methods
to compute dominant eigenvalues of a large-scale implicitly defined operator whose spectra exhibit fast decay. The spectral properties of the Hessian operator have been extensively studied, and it has been observed numerically or proven analytically that for many problems the Hessian operator is either nearly low-rank or its eigenvalues exhibit fast decay [7, 27, 15, 16, 17, 24]. Thus, this approach can provide highly accurate approximation of the trace at a cost (measured in number of Hessian action calculations) much smaller than Gaussian trace estimators.

The main limitation of the Taylor approximation is that it introduces bias in the evaluation of the cost functional, which may result in large errors for the optimal control function. Specifically, the norm of the remainder of the Taylor expansion scale as $O(\text{tr}(C)^{(r+1)/2})$ where $C$ is the covariance of the uncertain parameter and $r$ is the order of the truncated Taylor expansion, e.g., $r = 1$ and 2 for linear and quadratic approximations, respectively [2]. To address this issue, we propose to use the Taylor approximation as a control variate in a variance reduced Monte Carlo estimator for the mean and variance of the exact control objective. In cases when Taylor approximations are highly correlated with the exact control objective, reduction in the computational cost can be dramatic, up to several orders of magnitude in our numerical examples.

We consider two optimal control problems to numerically illustrate the accuracy and efficiency of our computational framework. The first is an optimal source control problem with applications to subsurface flow in porous media, where the control variable is the injection rate at given wells’ locations, the state equation is an elliptic PDE with random diffusion coefficient (Darcy equations), and the objective is to drive the pressure field to a desirable state. The second is an optimal boundary control problem for a turbulent jet flow, where the control variable is the inlet velocity profile, the state equations are the Reynolds-average Navier–Stokes equations coupled with a nonlinear stochastic advection-diffusion equation, and the objective is to maximize the jet width in a cross-section at a certain distance to the inlet boundary. In both cases, the uncertainty parameter is infinite-dimensional in the continuous setting and becomes high-dimensional after discretization.

For the solution of the optimal control problems with the Taylor approximations and Monte Carlo corrections we employ a gradient-based method, specifically the limited memory BFGS method [48]. The cost functional involves the analytical expressions for the mean and variance of the Taylor approximations and, possibly, the Monte Carlo corrections; the constraint involves the PDE and a set of linearized/adjoint PDEs used in the computation of the gradient and the Hessian of the control objective with respect to the uncertain parameter. We use the Lagrangian formalism to derive the gradient of the cost functional with respect to the control variable. Specifically, we obtain the mixed high-order derivatives of the Lagrangian functional with respect to the state variable, the adjoint variable, the uncertain parameter and the control function by symbolic differentiation in a variational settings before discretization. This allows us to efficiently apply our framework to rather complex PDE models such as the turbulent jet example, which features coupling between different equations and nonlinear numerical stabilization techniques for advection dominated equation.

The rest of the paper is organized as follows: we present the basic setting for the PDE-constrained optimal control under uncertainty in Sec. 2 using a general abstract notation for the PDE. Sec. 3 is devoted to the development of the Taylor approximation and the Monte Carlo correction, followed by Sec. 4 where we derive the gradient of the cost functional with respect to the control function. Numerical experiments for the two optimal control problems are presented in Sec. 5. In the last
section, Sec. 6, we close with some conclusions and perspectives.

2. PDE-constrained optimal control under uncertainty. In this section we present the PDE-constrained optimal control problem under uncertainty in an abstract setting. To this aim, let us first introduce the notation used through the paper. Let $X$ be a separable Banach space and $X'$ the dual space, then $\langle \cdot, \cdot \rangle_{X'X}$ denotes the duality pairing between the spaces $X$ and $X'$. For ease of notation, we will omit to specify the subscripts $X$ and $X'$ and simply write $\langle \cdot, \cdot \rangle$ when the spaces can be inferred from the context without ambiguity. Given two separable Banach spaces $X$ and $Y$ and a map $f : X \times Y \to \mathbb{R}$, $\partial_x f \in X'$ denotes the Fréchet derivative of $f(x,y)$ with respect to $x \in X$ defined as

$$\langle \hat{x}, \partial_x f \rangle := \left. \frac{d}{dt} f(x + t\hat{x}, y) \right|_{t=0} \quad \forall \hat{x} \in X.$$ 

In a similar manner, the linear operator $\partial_{xy} f : Y \to X'$ denotes the second order (mixed) Fréchet derivative of $f$ with respect to $x$ and $y$. $\partial_{yx} f : X \to Y'$ is the adjoint operator of $\partial_{xy} f$ and satisfies

$$(\hat{x}, \partial_{xy} f \hat{y}) := X' \langle \hat{x}, \partial_{xy} f \hat{y} \rangle_{X'} = Y' \langle \hat{y}, \partial_{yx} f \hat{x} \rangle_{Y'} =: \langle \hat{y}, \partial_{yx} f \hat{x} \rangle, \quad \forall \hat{x} \in X, \hat{y} \in Y.$$ 

Finally, let us introduce the separable Banach spaces $U, V, M,$ and $Z$ for the state variable $u$, the adjoint variable $v$, the random parameter $m$, and the control $z$ respectively. The exact definition of these spaces is problem specific and it is given in Sec. 5 for the two application problems presented in the numerical results.

2.1. The stochastic PDE state equation and control objective. We consider a state equation modeled by the following stochastic PDE. Given a realization of the random parameter $m \in M$ (almost surely) and a deterministic control $z \in Z$, find the state $u \in U$, such that

$$(2.2) \quad \mathcal{R}(u, m, z) = 0 \quad \text{in } V',$$

where $\mathcal{R}(\cdot, m, z) : U \to V'$ denotes a (possibly nonlinear) operator from $U$ to the dual of $V$. We then define the weak form associated to the state equation (2.2) by means of duality pairing, and write

$$(2.3) \quad r(u, v, m, z) := V' \langle v, \mathcal{R}(u, m, z) \rangle_{V'} = 0 \quad \forall v \in V.$$ 

Note that the weak form $r(u, v, m, z)$ is, in general, nonlinear with respect to the state variable $u$ and linear with respect to the test variable $v$.

As control objective, we consider a real-valued, possibly nonlinear, functional $Q$ of the state variable $u \in U$,

$$(2.4) \quad Q : U \to \mathbb{R}.$$ 

For simplicity, we assume that $Q$ depends on $m$ and $z$ only implicitly through the solution of the state equation (2.3). The general case where $Q$ has an explicit dependence on $m$ and $z$ can be treated with slight modifications in the derivation of the cost functional and its gradient, as we address along the presentation in Sec. 3 and Sec. 4.
2.2. The optimal control problem under uncertainty. Because $Q$ depends implicitly on the random variable $m$ through the state variable $u$, the control objective $Q$ is also a random variable, thus leading to an optimal control problem under uncertainty. In this work we consider a mean-variance risk-averse cost functional of the form

\begin{equation}
J(z) = \mathbb{E}[Q] + \beta \text{Var}[Q] + \mathcal{P}(z).
\end{equation}

Above $\mathbb{E}[Q]$ and $\text{Var}[Q]$ are, respectively, the mean and the variance of $Q$ with respect to the probability measure $\mu$, and are defined as

\begin{equation}
\mathbb{E}[Q] := \int_{\mathcal{M}} Q(u(m,z))d\mu(m),
\end{equation}

and

\begin{equation}
\text{Var}[Q] := \int_{\mathcal{M}} (Q(u(m,z)) - \mathbb{E}[Q])^2d\mu(m) = \mathbb{E}[Q^2] - (\mathbb{E}[Q])^2.
\end{equation}

The scaling factor $\beta \geq 0$ controls the relative weight of the mean and the variance, and $\mathcal{P}(z)$ is a penalization term on $z$.

Recalling that the state variable $u$ implicitly depends on the random variable $m$ and the control variable $z$ through the solution of (2.3), we then formulate the PDE-constrained optimal control problem under uncertainty as: find $z^* \in \mathcal{Z}$ such that

\begin{equation}
z^* = \arg \min_{z \in \mathcal{Z}} J(z), \text{ subject to the problem (2.3)}.\end{equation}

To solve the stochastic optimization problem, a straightforward Monte Carlo integration method can be used to approximate the expectation and the variance in the cost functional (2.5). We briefly discuss this approach below, while in the following section we introduce more sophisticated formulations of the optimal control problem that allow for more efficient algorithms by exploiting high-order derivatives.

2.3. Monte Carlo integration. The Monte Carlo integration (or sample average approximation) of the mean reads

\begin{equation}
\mathbb{E}[Q] \approx \hat{Q} := \frac{1}{M_1} \sum_{i=1}^{M_1} Q(m_i),
\end{equation}

where $m_i, \ i = 1, \ldots, M_1$, are independent identically distributed (i.i.d.) random samples drawn from the probability measure $\mu$. Similarly, the Monte Carlo integration of the variance is given by

\begin{equation}
\text{Var}[Q] \approx \hat{V}_Q := \frac{1}{M_2} \sum_{i=1}^{M_2} Q^2(m_i) - \left(\frac{1}{M_2} \sum_{i=1}^{M_2} Q(m_i)\right)^2,
\end{equation}

where $m_i, \ i = 1, \ldots, M_2$, are i.i.d. random samples from $\mu$. Since each evaluation of $Q(m_i)$ involves solving the possibly nonlinear state equation (2.3), it is common practice to choose $M_1 = M_2 = M$ and reuse the same samples $m_i$ to estimate both the mean and variance of $Q$. However, a more careful choice of $M_1$ and $M_2$ relies on estimating the variance of $Q$ and $Q^2$ and aims at minimizing the mean square
error of \( \hat{Q} + \beta \tilde{V}_Q \). Note that a large number of samples is required to obtain a relatively accurate approximation of \( \mathbb{E}[Q] \) and \( \text{Var}[Q] \) due to the slow convergence of Monte Carlo integration, thus rendering this approach computationally prohibitive for practical applications due to the large number of (nonlinear) PDEs to be solved at each optimization step.

3. Taylor approximation and variance reduction.

3.1. Taylor approximation of the control objective. In this section we derive expressions for the Taylor expansion of the control objective \( Q(u(m, z)) \) with respect to the random variable \( m \) for a fixed value of the control variable \( z \). In what follows, we denote with the subscript \( m \) the total derivative of \( Q(u(m, z)) \) with respect to \( m \) and we assume sufficient smoothness so that all Fréchet derivatives are well-defined. For ease of notation, we also denote with \( \langle \cdot, \cdot \rangle \) the duality-pairing \( \mathcal{M}(\cdot, \cdot)_{\mathcal{M}'} \), and, with a slight abuse of notation, we write \( Q(m) \) to indicate the control objective \( Q(u(m, z)) \) evaluated at a fixed value of \( z \).

A formal Taylor expansion of the control objective evaluated at \( m_0 \in \mathcal{M} \) is given by

\[
Q(m) = Q(m_0) + \langle m - m_0, Q_m(m_0) \rangle + \frac{1}{2} \langle m - m_0, Q_{mm}(m_0) (m - m_0) \rangle + \cdots,
\]

where \( Q_m(m_0) \in \mathcal{M}' \) and \( Q_{mm}(m_0) : \mathcal{M} \to \mathcal{M}' \) are the first and the second order Fréchet derivatives of \( Q \) with respect to \( m \) evaluated at the nominal \( m_0 \in \mathcal{M} \) (see Sec. 3.2). In what follows, we refer to such derivatives as the gradient and the Hessian of \( Q \), respectively. Based on the Taylor expansion, we can define the linear approximation of \( Q \) as

\[
Q_{\text{lin}}(m) = Q(m_0) + \langle m - m_0, Q_m(m_0) \rangle,
\]

and the quadratic approximation of \( Q \) as

\[
Q_{\text{quad}}(m) = Q(m_0) + \langle m - m_0, Q_m(m_0) \rangle + \frac{1}{2} \langle m - m_0, Q_{mm}(m_0) (m - m_0) \rangle.
\]

We remark that \( Q_{\text{lin}}(m) \) and \( Q_{\text{quad}}(m) \) are polynomials with respect to \( m \), but still depend on \( z \) through the solution of the state equation \( r(u, v, m_0, z) = 0 \).

In the case of Gaussian measure \( \mu = \mathcal{N}(m_0, C) \), the linear approximation \( Q_{\text{lin}}(m) \) follows a Gaussian distribution with mean and variance given by

\[
\mathbb{E}[Q_{\text{lin}}] = Q(m_0), \quad \text{and} \quad \text{Var}[Q_{\text{lin}}] = \langle C Q_m(m_0), Q_m(m_0) \rangle;
\]

while the quadratic approximation \( Q_{\text{quad}}(m) \) follows a non-Gaussian distribution for which analytical expressions of the mean and the variance are available. In particular, we have

\[
\mathbb{E}[Q_{\text{quad}}] = \mathbb{E}[Q_{\text{lin}}] + \frac{1}{2} \text{tr}(H), \quad \text{and} \quad \text{Var}[Q_{\text{quad}}] = \text{Var}[Q_{\text{lin}}] + \frac{1}{2} \text{tr}(H^2),
\]

where \( H = C^{1/2} Q_{mm}(m_0) C^{1/2} \) is the covariance preconditioned Hessian, and \( \text{tr}(\cdot) \) denotes the trace operator, see details in [2]. Efficient algorithms to accurately estimate the trace of \( H \) and \( H^2 \) will be discussed in Sec. 3.3.

Finally, following [2], we approximate the cost functional \( J(z) \) in (2.5) by replacing \( \mathbb{E}[Q] \) and \( \text{Var}[Q] \) with the moments of the linear or quadratic Taylor expansions of the control objective.
3.2. Computation of the gradient and the Hessian of the control objective with respect to the uncertain parameter. In this section we use the Lagrangian formalism to derive the expressions for the gradient $Q_m$ and for the action of the Hessian $Q_{mm}$ in a given direction $\tilde{m}$ evaluated at $m_0$. In what follows, for ease of notation, we use the subscript 0 to indicate quantities evaluated at $m = m_0$ and define
\begin{equation}
(3.6) \quad r_0 := r(u,v,m_0,z), \text{ and } Q_0 := Q(u(m_0,z)).
\end{equation}

We then define the Lagrangian functional
\begin{equation}
(3.7) \quad \mathcal{L}(u,v,m,z) = Q(u) + r(u,v,m,z),
\end{equation}
where the adjoint variable $v$ is regarded as the Lagrangian multiplier of the state equation.

By requiring the first order variation of (3.7) with respect to the adjoint variable $v$ to vanish, we obtain the state problem: find $u \in U$, such that
\begin{equation}
(3.8) \quad \langle \hat{v}, \partial_u r_0 \rangle = 0, \quad \forall \hat{v} \in V.
\end{equation}
Similarly, by setting the first order variation of (3.7) with respect to the state variable $u$ to zero, we obtain the adjoint problem: find $v \in V$, such that
\begin{equation}
(3.9) \quad \langle \hat{u}, \partial_u r_0 \rangle = -\langle \hat{u}, \partial_u Q_0 \rangle, \quad \forall \hat{u} \in U.
\end{equation}

Then, the gradient of $Q$ at $m_0$ acting in the direction $\tilde{m} = m - m_0$ is given by
\begin{equation}
(3.10) \quad \langle \tilde{m}, Q_m(m_0) \rangle = \langle \tilde{m}, \partial_m r_0 \rangle,
\end{equation}
where $u$ solves the state problem (3.8) and $v$ solves the adjoint problem (3.9).

Similarly, for the computation of the Hessian of $Q$ acting on $\tilde{m}$, we consider the Lagrangian functional
\begin{equation}
(3.11) \quad \mathcal{L}^H(u,v,m,z; \hat{u}, \hat{v}, \hat{m}) = \langle \hat{m}, \partial_m r_0 \rangle + \langle \hat{v}, \partial_v r_0 \rangle + \langle \hat{u}, \partial_u r_0 + \partial_u Q_0 \rangle,
\end{equation}
where $\hat{u}$ and $\hat{v}$ denote the incremental state and adjoint variables, respectively. By taking variation of (3.11) with respect to the adjoint $v$, we obtain the incremental state problem: find $\hat{u} \in U$ such that
\begin{equation}
(3.12) \quad \langle \hat{v}, \partial_{uv} r_0 \hat{u} \rangle = -\langle \hat{v}, \partial_{vm} r_0 \hat{m} \rangle, \quad \forall \hat{v} \in V,
\end{equation}
where the derivatives $\partial_{uv} r_0 : U \to V'$ and $\partial_{vm} r_0 : U \to V'$ are linear operators. The incremental adjoint problem, obtained by taking variation of (3.11) with respect to the state $u$, reads: find $\hat{v} \in V$ such that
\begin{equation}
(3.13) \quad \langle \hat{u}, \partial_{uv} r_0 \hat{v} \rangle = -\langle \hat{u}, \partial_{uv} Q_0 \hat{u} + \partial_{um} r_0 \hat{m} \rangle, \quad \forall \hat{u} \in U,
\end{equation}
where $\partial_{uv} r_0 : V \to U'$ is the adjoint of $\partial_{uv} : U \to V'$ in the sense of (2.1). We remark that an extra term $-\langle \hat{u}, \partial_{um} Q_0 \hat{m} \rangle$ is added on the right hand side if $Q$ explicitly depends on $m$. The Hessian of $Q$ at $m_0$ acting on $\tilde{m}$ can then be computed by taking variation of (3.11) with respect to $m$ as
\begin{equation}
(3.14) \quad \langle \tilde{m}, Q_{mm}(m_0) \hat{m} \rangle = \langle \tilde{m}, \partial_{mm} r_0 \hat{v} + \partial_{mu} r_0 \hat{u} + \partial_{mm} Q_0 \hat{m} \rangle,
\end{equation}
where the incremental state $\hat{u}$ and adjoint $\hat{v}$ solve (3.12) and (3.13), respectively. Note that $\langle \tilde{m}, \partial_{mu} Q_0 \hat{u} + \partial_{mm} Q_0 \hat{m} \rangle$ is also added if $Q$ explicitly depends on $m$.

We summarize the computation of the gradient and Hessian of the control objective $Q$ with respect to the uncertain parameter $m$ in Algorithm 1.
Algorithm 1 Computation of the gradient and Hessian of $Q$ with respect to $m$.

**Input:** control objective $Q$, PDE $r(u,v,m,z)$, mean $m_0$, direction $\hat{m}$.

**Output:** the gradient $Q_m(m_0)$ and the Hessian $Q_{mm}(m_0)$ acting on $\hat{m}$.

1. Solve the state problem (3.8) for $u$.
2. Solve the adjoint problem (3.9) for $v$.
3. Compute the gradient $Q_m(m_0)$ by (3.10).
4. Solve the incremental state problem (3.12) for $\hat{u}$.
5. Solve the incremental adjoint problem (3.13) for $\hat{v}$.
6. Compute the Hessian $Q_{mm}(m_0)$ acting on $\hat{m}$ by (3.14).

### 3.3. Randomized algorithms to compute the trace.

The first and second moments of the quadratic approximation $Q_{\text{quad}}$ in equation (3.5), involve the trace of the covariance preconditioned Hessian $H$. $H$ is formally a large scale dense operator and it is only implicitly defined through its action in a given direction as in (3.14).

The standard approach to estimate the trace of an implicitly defined operator is based on a Monte Carlo method, where one computes the action of the operator in $N$ random directions sampled from a suitable distribution. For example, using the Gaussian trace estimator, one can approximate $\text{tr}(H)$ and $\text{tr}(H^2)$ as

\begin{equation}
\text{tr}(H) \approx \hat{T}_1(H) := \frac{1}{N} \sum_{j=1}^{N} \langle \hat{m}_j, Q_{mm}(m_0) \hat{m}_j \rangle
\end{equation}

and

\begin{equation}
\text{tr}(H^2) \approx \hat{T}_1(H^2) := \frac{1}{N} \sum_{j=1}^{N} \langle C Q_{mm}(m_0) \hat{m}_j, Q_{mm}(m_0) \hat{m}_j \rangle,
\end{equation}

where $\hat{m}_j, j = 1, \ldots, N$, are i.i.d. random samples from the Gaussian distribution $N(0, C)$. Lower bounds on the number of samples $N$ to obtain a guaranteed probabilistic error bound for the trace are discussed in [4]. Such bounds hold for any symmetric positive defined operator, regardless of size and spectral properties. However the number of samples $N$ necessary to obtain an accurate approximation may be prohibitively large and impractical: to guarantee a relative error of $10^{-3}$ in the estimation of the trace more than $10^6$ samples are need using the bounds in [4]. Improved bounds are demonstrated in [51] and show that the Gaussian estimator becomes very inefficient if the stable rank of the operator is small and that it allows for small $N$ only if the eigenvalues are all of approximately the same size. This means that the randomized Gaussian estimator is not a viable solution to estimate the trace of $H$, since it has been observed numerically or proven analytically that for many problems the Hessian operator is either nearly low-rank or its eigenvalues exhibit fast decay [7, 27, 15, 16, 17, 24].

In this work, we propose to estimate the trace of $H$ and $H^2$ by means of a randomized approximated eigendecomposition. Specifically, we approximate $\text{tr}(H)$ and $\text{tr}(H^2)$ by

\begin{equation}
\text{tr}(H) \approx \hat{T}_2(H) := \sum_{j=1}^{N} \lambda_j(H), \quad \text{and} \quad \text{tr}(H^2) \approx \hat{T}_2(H^2) := \sum_{j=1}^{N} \lambda_j^2(H),
\end{equation}
where $\lambda_j(\mathcal{H})$, $j = 1, \ldots, N$, are the dominant eigenvalues of $\mathcal{H}$ obtained by solving the generalized eigenvalue problem
\begin{equation}
\langle \phi, Q_{mm}(m_0) \psi_j \rangle = \lambda_j \langle \phi, \mathcal{C}^{-1} \psi_j \rangle, \quad \forall \phi \in \mathcal{M}, \quad j = 1, \ldots, N.
\end{equation}

Above, the eigenvector $\psi_j$, $j = 1, \ldots, N$, are $\mathcal{C}^{-1}$-orthonormal, that is
\begin{equation}
\langle \psi_j, \mathcal{C}^{-1} \psi_l \rangle = \delta_{ij}, \quad i, j = 1, \ldots, N,
\end{equation}
where $\delta_{ij}$ the Kronecker delta. Note that $N$ is independent of the stochastic dimension, and, as shown in the numerical results section, it is relatively small for $\mathcal{H}$ nearly low-rank or when its spectrum exhibits fast decay.

After suitable discretization, e.g., by finite element method, the generalized eigenvalue problem (3.18) gives rise to an algebraic eigenproblem of the form $A\psi = \lambda B\psi$ with $A, B \in \mathbb{R}^{n \times n}$ and $\psi \in \mathbb{R}^n$. Algorithm 2 summarizes the so-called double pass randomized algorithm to solve the algebraic generalized eigenvalue problem (see [32, 54] for details of the algorithms and [60] for its implementation).

**Algorithm 2** Randomized algorithm for generalized eigenvalue problem $(A, B)$

**Input:** matrix $A, B$, the number of eigenpairs $k$, an oversampling factor $p$.

**Output:** eigenpairs $(\Lambda_k, \Psi_k)$ with $\Lambda_k = \text{diag}(\lambda_1, \ldots, \lambda_k)$ and $\Psi_k = (\psi_1, \ldots, \psi_k)$.

1. Draw a Gaussian random matrix $\Omega \in \mathbb{R}^{n \times (k+p)}$.
2. Compute $Y = B^{-1}(A\Omega)$.
3. Compute $QR$-factorization $Y = QR$ such that $Q^\top BQ = I_{k+p}$.
4. Form $T = Q^\top AQ$ and compute eigendecomposition $T = \Lambda S S^\top$.
5. Extract $\Lambda_k = \Lambda(1 : k, 1 : k)$ and $\Psi_k = QS_k$ with $S_k = S(:, 1 : k)$.

There are multiple advantages in using Algorithm 2 to approximate the trace, see e.g., [32, 55]. In terms of accuracy, the approximation error is bounded by the sum of the remaining eigenvalues, so that the error is small if the eigenvalues decay fast or if the Hessian $\mathcal{H}$ has low rank, see [32, 55] for more details. In terms of computational efficiency, the $(2(k + p))$ Hessian matrix-vector products, which entail the solution of a pair of linearized state/adjoint equations (as shown in Sec. 3.2) and therefore dominate the computational cost of the algorithm, can be computed independently and, therefore, asynchronously across the random directions.

### 3.4. Monte Carlo variance reduction.

To use the moments of the truncated Taylor expansion $Q_{\text{lin}}$ and $Q_{\text{quad}}$ as approximation of $\mathbb{E}[Q]$ and $\text{Var}[Q]$ in the cost functional (2.5) introduces bias in the optimal control problem and may result in large approximation errors for the optimal control function if $Q_{\text{lin}}$ and $Q_{\text{quad}}$ are not accurate enough. In this circumstance, we can use the Taylor approximation as control variate to reduce the variance of a Monte Carlo estimator. The basic idea is to correct the moments of $Q_{\text{lin}}$ and $Q_{\text{quad}}$ by applying a Monte Carlo method to estimate the expect value (and the variance) of the difference between $Q$ and the Taylor approximation.

Specifically, recalling (3.4) and (3.2), the Monte Carlo correction for the mean of linear approximation reads
\begin{equation}
\mathbb{E}[Q] = \mathbb{E}[Q_{\text{lin}}] + \mathbb{E}[Q - Q_{\text{lin}}]
\approx \bar{Q}_{\text{lin}} := Q(m_0) + \frac{1}{M_1} \sum_{i=1}^{M_1} \left( Q(m_i) - Q(m_0) - \langle m_i - m_0, Q_{\text{lin}}(m_0) \rangle \right).
\end{equation}
Similarly, recalling that the variance is invariant with respect to deterministic translations, we have

\[
\text{Var}[Q] = \text{Var}[Q - Q(m_0)] = E[(Q - Q(m_0))^2] - (E[Q - Q(m_0)])^2
\]

(3.21)

(3.22)

(3.23)

(3.24)

and therefore the Monte Carlo correction of the variance of the linear approximation is

\[
\hat{V}^\text{lin}_Q := (CQ_m(m_0), Q_m(m_0))
\]

(3.25)

(3.26)

Note that, similar to the Monte Carlo integration, for simplicity we can use the same \( M = M_1 = M_2 \) random samples for the approximation of both the mean and the variance. A more careful choice of \( M_1 \) and \( M_2 \) relies on the variance of \( Q^{(1)} = Q - Q_{\text{lin}} \) and \( Q^{(2)} = (Q - Q_0)^2 - (Q - Q_{\text{lin}})^2 \), in order to balance the errors in the approximation of the mean and the variance. For the quadratic approximation, a combination of (3.5) and (3.3) leads to

\[
\hat{V}_Q^{\text{quad}} := (CQ_m(m_0), Q_m(m_0)) + \frac{1}{4}(\text{tr}(H))^2 + \frac{1}{2}\text{tr}(H^2)
\]

(3.27)
Note that to achieve certain approximation accuracy, the number of Monte Carlo samples depends on the variance of the integrand. If the approximations $Q_{\text{lin}}$ and $Q_{\text{quad}}$ are highly correlated with $Q$, the variances of $Q - Q_{\text{lin}}$ and $Q - Q_{\text{quad}}$ would be small so that only a small number of Monte Carlo samples are required.

4. Gradient-based optimization. In this section, we present a gradient-based optimization method to solve the problem (2.6) using the different approximations introduced in Section 3, namely the Monte Carlo integration, the first and second order Taylor expansions, and the Monte Carlo corrections of the mean $\mathbb{E}[Q]$ and variance $\text{Var}[Q]$. These approximated formulations feature different levels of accuracy and computational complexity, and could possibly be combined within a multifidelity framework to obtain an efficient computational procedure while preserving high accuracy. In this section, we use the Lagrangian formalism to derive expressions for the gradient of the approximated formulations of cost functional with respect to the control variable $z$, which are needed to perform the optimization. In the following, to distinguish the gradient of the cost function with respect to the control variable from the gradient of the control objective with respect to the uncertain parameter, we denote the former as $z$-gradient.

4.1. $z$-gradient for the Monte Carlo integration. By the Monte Carlo approximation of the mean and the variance, the cost functional becomes

\[
J_{\text{MC}}(z) = \hat{Q} + \beta \hat{V}_Q + \mathcal{P}(z),
\]

subject to the state problem (2.3) at $m_{i}$, $i = 1, \ldots, M$, where $M = M_1 = M_2$ if the same samples are used for the mean and the variance or $M = M_1 + M_2$ otherwise. To compute the gradient of (4.1) with respect to the control variable $z$ ($z$-gradient for short), we consider the Lagrangian functional

\[
\mathcal{L}_{\text{MC}}\left(\{u_i\}, \{v_i\}, z\right) = J_{\text{MC}}(z) + \sum_{i=1}^{M} r(u_i, v_i, m_i, z),
\]

where $\{v_i\}$ are the Lagrange multipliers (adjoint variables) of the state problems: given $m_i \in \mathcal{M}$ and $z \in \mathcal{Z}$, find $u_i \in \mathcal{U}$ such that

\[
r(u_i, v_i, m_i, z) = 0 \quad \forall v_i \in \mathcal{V}; \quad i = 1, \ldots, M.
\]

The $i$-th adjoint problem ($i = 1, \ldots, M$) is obtained by setting to zero the first variation of the Lagrangian functional with respect to the state variable $u_i$. The $i$-th adjoint variable $v_i$ then solves: given $m_i \in \mathcal{M}$, $z \in \mathcal{Z}$, and $u_i \in \mathcal{U}$ as the solution of (4.3), find $v_i \in \mathcal{V}$ such that

\[
\langle \tilde{u}_i, \partial_u r(u_i, v_i, m_i, z) \rangle = -\frac{1}{M_1} \langle \tilde{u}_i, \partial_u Q(m_i) \rangle - \frac{2\beta}{M_2} Q(m_i) \langle \tilde{u}_i, \partial_u Q(m_i) \rangle
\]

\[
+ \frac{2\beta}{M_2} \left( \frac{1}{M_2} \sum_{j=1}^{M_2} Q(m_j) \right) \langle \tilde{u}_i, \partial_u Q(m_i) \rangle, \quad \forall \tilde{u}_i \in \mathcal{U}.
\]

Then we can compute the $z$-gradient as

\[
\langle \hat{z}, D_z J_{\text{MC}}(z) \rangle = \langle \hat{z}, D_z \mathcal{L}_{\text{MC}}\left(\{u_i\}, \{v_i\}, z\right) \rangle
\]

\[
= \langle \hat{z}, D_z \mathcal{P}(z) \rangle + \sum_{i=1}^{M} \langle \hat{z}, D_z r(u_i, v_i, m_i, z) \rangle,
\]
where \( \{u_i\} \) and \( \{v_i\} \) denote the set of solutions of the state problem (4.3) and adjoint problem (4.4) at the set of samples \( \{m_i\} \), respectively. Therefore, each gradient evaluation requires \( M \) (nonlinear) state PDE solves and \( M \) linearized PDE solves. Note that, for the general case in which \( Q \) explicitly depends on \( z \), extra terms involving \( \langle \tilde{z}, \partial_z Q(m_i) \rangle \) need to be included in the gradient computation.

### 4.2. \( z \)-gradient for the Taylor approximation.

#### 4.2.1. \( z \)-gradient for the linear approximation.

We first consider the linear approximation (3.4) and approximate the cost functional (2.5) as

\[
J_{\text{lin}}(z) = Q_0 + \beta \langle C Q_m(m_0), Q_m(m_0) \rangle + \mathcal{P}(z),
\]

where \( Q_0 \) and the gradient \( Q_m(m_0) \) are defined in (3.6) and (3.10), respectively.

To compute the \( z \)-gradient of (4.6), we introduce the Lagrangian functional

\[
\mathcal{L}_{\text{lin}}(u,v,u^*,v^*,z) = J_{\text{lin}}(z) + \langle v^*, \partial_v r_0 \rangle + \langle u^*, \partial_u r_0 + \partial_u Q_0 \rangle,
\]

where the adjoint variables \( v^* \in \mathcal{V} \) and \( u^* \in \mathcal{U} \) represent the Lagrange multipliers of the state problem (3.8) and the adjoint problem (3.9).

By setting to zero the first order variation of the Lagrangian (4.7) with respect to \( v \), we have: find \( u^* \in \mathcal{U} \) such that

\[
\langle \hat{v}, \partial_{vu} r_0 u^* \rangle = -2\beta \langle \hat{v}, \partial_{vm} r_0 (C \partial_m r_0) \rangle, \quad \forall \hat{v} \in \mathcal{V}.
\]

while setting to zero the first order variation with respect to \( u \), we have: find \( v^* \in \mathcal{V} \) such that

\[
\langle \hat{u}, \partial_{uu} r_0 v^* \rangle = -\langle \hat{u}, \partial_u Q_0 \rangle - 2\beta \langle \hat{u}, \partial_{um} r_0 (C \partial_m r_0) \rangle - \langle \hat{u}, \partial_{uu} r_0 u^* + \partial_{uu} Q_0 u^* \rangle, \quad \forall \hat{u} \in \mathcal{U}.
\]

Finally, the \( z \)-gradient of \( J_{\text{lin}} \) in a direction \( \tilde{z} \in \mathcal{Z} \) is computed as

\[
\langle \tilde{z}, D_z J_{\text{lin}}(z) \rangle = \langle \tilde{z}, \partial_z \mathcal{L}_{\text{lin}}(u,v,u^*,v^*,z) \rangle = 2\beta \langle \tilde{z}, \partial_{zm} r_0 (C \partial_m r_0) \rangle + \langle \tilde{z}, \partial_z \mathcal{P}(z) \rangle + \langle \tilde{z}, \partial_{zu} r_0 v^* + \partial_{zu} r_0 u^* \rangle.
\]

In summary, evaluation of the cost functional with linear approximation requires the solution of the state problem (3.8) and the adjoint problem (3.9). Computation of the \( z \)-gradient requires, in addition, the solution of the two linear problems in (4.8) and (4.9).

#### 4.2.2. \( z \)-gradient for the quadratic approximation.

For the quadratic approximation (3.3), thanks to the relation (3.5) and the approximation (3.16), we obtain the approximation of the cost functional

\[
J_{\text{quad}}(z) = Q_0 + \frac{1}{2} \sum_{j=1}^N \lambda_j + \beta \left( \langle C Q_m(m_0), Q_m(m_0) \rangle + \frac{1}{2} \sum_{j=1}^N \lambda_j^2 \right) + \mathcal{P}(z),
\]

subject to the state and the adjoint problems (3.8) and (3.9) for the evaluation of \( Q \) and its gradient at \( m_0 \), the generalized eigenvalue problems (3.18) for the computation...
of the eigenvalues, as well as the incremental state and adjoint problems (3.12) and (3.13) for the Hessian action in (3.18). Correspondingly, we form the Lagrangian

\[ L_{\text{quad}} (u, v, \{\lambda_j\}, \{\psi_j\}, \{\tilde{u}_j\}, \{\tilde{v}_j\}, u^*, \{\lambda_j^*\}, \{\psi_j^*\}, \{\tilde{u}_j^*\}, \{\tilde{v}_j^*\}, z) \]

\[ = J_{\text{quad}} (z) + \langle v^*, \partial_u r_0 \rangle + \sum_{j=1}^{N} \langle \psi_j^*, (Q_{mm}(m_0) - \lambda_j C^{-1}) \psi_j \rangle \]

\[ + \sum_{j=1}^{N} \lambda_j^* ((\psi_j, C^{-1} \psi_j) - 1) \]

\[ + \sum_{j=1}^{N} \langle \tilde{v}_j^*, \partial_v r_0 \tilde{u}_j + \partial_y r_0 \psi_j \rangle \]

\[ + \sum_{j=1}^{N} \langle \tilde{u}_j^*, \partial_u r_0 \tilde{v}_j + \partial_y r_0 \tilde{u}_j + \partial_y Q_0 \tilde{u}_j + \partial_y r_0 \psi_j \rangle. \]

(4.12)

Here we assume that the dominating eigenvalues are not repeated, so that the constraints \( \langle \psi_j, C^{-1} \psi_j \rangle = 1, j = 1, \ldots, N \), are sufficient to guarantee the orthonormality \( \langle \psi_i, C^{-1} \psi_j \rangle = \delta_{ij} \) for any \( i, j = 1, \ldots, N \). In fact, for any \( i \neq j \), by definition we have

(4.13) \( \langle \psi_i, Q_{mm} \psi_j \rangle = \langle \psi_i, \lambda_i C^{-1} \psi_i \rangle \) and \( \langle \psi_i, Q_{mm} \psi_j \rangle = \langle \psi_i, \lambda_j C^{-1} \psi_j \rangle \),

which, by the symmetry of \( Q_{mm} \) and \( C^{-1} \), leads to

(4.14) \( (\lambda_i - \lambda_j) \langle \psi_i, C^{-1} \psi_j \rangle = 0 \), so that \( \langle \psi_i, C^{-1} \psi_j \rangle = 0 \) if \( \lambda_i \neq \lambda_j \).

By setting the variation of \( L_{\text{quad}} \) with respect to \( \lambda_j \) as zero, we obtain

(4.15) \( \psi_j^* = \frac{1 + 2\beta \lambda_j}{2} \psi_j, \quad j = 1, \ldots, N. \)

Subsequently, for each \( j = 1, \ldots, N \), taking the variation of \( L_{\text{quad}} \) with respect to \( \tilde{v}_j \) as zero, and using the Hessian action equation (3.14), we have: find \( \tilde{u}_j^* \in U \) such that

(4.16) \( \langle \tilde{v}, \partial_v r_0 \tilde{u}_j^* \rangle = -\langle \tilde{v}, \partial_y r_0 \psi_j^* \rangle, \quad \forall \tilde{v} \in V, \)

which is the same as the incremental forward problem (3.12). Therefore, from the result (4.15), we have

(4.17) \( \tilde{u}_j^* = \frac{1 + 2\beta \lambda_j}{2} \tilde{u}_j, \quad j = 1, \ldots, N. \)

Then, taking variation with respect to \( \tilde{u}_j \) as zero, we have: find \( \tilde{v}_j^* \in V \) such that

(4.18) \( \langle \tilde{u}, \partial_u r_0 \tilde{v}_j^* \rangle = -\langle \tilde{u}, \partial_y r_0 \tilde{u}_j^* + \partial_y Q_0 \tilde{u}_j^* + \partial_y r_0 \psi_j^* \rangle, \quad \forall \tilde{u} \in U, \)

which is the same as the incremental adjoint problem (3.13). Therefore, by (4.17) and (4.15), we obtain

(4.19) \( \tilde{v}_j^* = \frac{1 + 2\beta \lambda_j}{2} \tilde{v}_j, \quad j = 1, \ldots, N. \).
By setting to zero the variation with respect to \( v \), we have: find \( u^* \in \mathcal{U} \) such that

\[
\langle \bar{v}, \partial_{vu} r_0 u^* \rangle = -2\beta \langle \bar{v}, \partial_{vm} r_0 (C \partial_m r_0) \rangle
- \sum_{j=1}^{N} (\bar{v}, \partial_{vmu} r_0 \hat{u}_j \psi_j^* + \partial_{vmr} r_0 \psi_j \psi_j^*)
- \sum_{j=1}^{N} (\bar{v}, \partial_{vum} r_0 \hat{u}_j \psi_j^* + \partial_{vum} r_0 \psi_j \psi_j^*), \quad \forall \bar{v} \in \mathcal{V}.
\]

(4.20)

Lastly, by setting to zero the variation with respect to \( u \), we have: find \( v^* \in \mathcal{V} \) such that

\[
\langle \bar{u}, \partial_{uv} r_0 v^* \rangle =
- \langle \bar{u}, \partial_{uu} Q_0 \rangle - 2\beta \langle \bar{u}, \partial_{um} r_0 (C \partial_m r_0) \rangle
- \langle \bar{u}, \partial_{uu} r_0 u^* + \partial_{uu} Q_0 u^* \rangle
- \sum_{j=1}^{N} (\bar{u}, \partial_{uvu} r_0 \hat{v}_j \psi_j^* + \partial_{um} r_0 \hat{u}_j \psi_j^* + \partial_{uum} r_0 \psi_j \psi_j^*)
- \sum_{j=1}^{N} (\bar{u}, \partial_{vuv} r_0 \hat{v}_j \psi_j^* + \partial_{uum} r_0 \hat{u}_j \psi_j^*)
- \sum_{j=1}^{N} (\bar{u}, \partial_{uuu} r_0 \hat{v}_j \psi_j^* + \partial_{uum} r_0 \hat{u}_j \psi_j^* + \partial_{uum} Q_0 \hat{u}_j \hat{u}_j^*) \forall \bar{u} \in \mathcal{U},
\]

(4.21)

where the third order derivatives take the form of the bilinear map \( \partial_{abc} f : C \times B \rightarrow A' \). Then, the \( z \)-gradient of the cost functional in the direction \( \hat{z} \in \mathcal{Z} \) can be computed as

\[
\langle \hat{z}, D_z J_{\text{quad}}(z) \rangle = \langle \hat{z}, z L_{\text{quad}} (\text{states, adjoints, } z) \rangle,
\]

(4.22)

where states and adjoints represent the state and adjoint variables for short. Note that we do not need to solve for \( \lambda^*_j \) as the terms involving \( \lambda^*_j \) are independent of \( z \). In summary, to get these state and adjoint variables, we need to solve 1 (nonlinear) state problem and \( 2N + 3 \) linear problems (1 for \( v \), \( N \) for \( \{\hat{u}_j\} \), \( N \) for \( \{\hat{v}_j\} \), 1 for \( u^* \), and 1 for \( v^* \)), where the linear operators are the same as those in (3.12) and (3.13). Moreover, to compute the eigenvalues by the double-pass randomized algorithm 2, which requires the action of the Hessian on \( N + p \) random directions twice, we need to solve \( 4(N + p) \) linear problems.

We summarize the evaluation of the cost functional and its \( z \)-gradient with quadratic approximation in the following algorithm.

### 4.3. \( z \)-gradient for the variance reduction.

#### 4.3.1. \( z \)-gradient for variance reduction with linear approximation.

The approximation in (4.6) and (4.11) introduced by the Taylor generates bias in the evaluations of the mean \( \mathbb{E}[Q] \) and the variance \( \text{Var}[Q] \) and may lead to errors in the optimal solution. Recalling (3.20) and (3.22), we obtain a MC-corrected unbiased linear approximation for the cost functional, which is given by

\[
J_{\text{lin}}^{\text{MC}}(z) = \tilde{Q}_{\text{lin}} + \beta \tilde{V}_{Q}^{\text{lin}} + \mathcal{P}(z),
\]

(4.23)
Algorithm 3 Computation of the cost functional $J_{\text{quad}}$ and its $z$-gradient.

**Input:** cost functional $J(z)$ and PDE $r(u,v,m,z)$.

**Output:** the approximate cost functional $J_{\text{quad}}$ and its $z$-gradient $D_zJ_{\text{quad}}$.

1. Solve the state problem (3.8) for $u$ at $m_0$ and evaluate $Q_0 = Q(u(m_0))$.
2. Solve the adjoint problem (3.9) for $v$ and evaluate $Q_m(m_0)$ by (3.10).
3. Use Algorithm 1 and 2 to compute the generalized eigenpairs $(\lambda_j, \psi_j)_{j=1}^N$.
4. Compute the approximate cost functional $J_{\text{quad}}$ by (4.11).
5. Solve linearized problems (4.20) for $\hat{u}^*$.
6. Solve linearized problems (4.20) for $\hat{v}^*$.
7. Compute the $z$-gradient $D_zJ_{\text{quad}}$ by (4.22).

subject to the state problem (3.8) at $m_0$, the adjoint problem (3.9), and the state problems (2.3) at $m_i$, $i = 1, \ldots, M$, where $M = M_1 = M_2$ if the same samples are used for the mean and the variance or $M = M_1 + M_2$ otherwise. We form the corresponding Lagrangian functional as

$$L_{\text{lin}}^\text{MC}(u,v,\{u_i\},u^*,v^*,\{v_i\},z)$$

(4.24)

$$= J_{\text{lin}}^\text{MC}(z) + \langle v^*, \partial_v r_0 \rangle + \langle u^*, \partial_u r_0 + \partial_u Q_0 \rangle + \sum_{i=1}^M r(u_i,v_i,m_i,z).$$

Setting variation with respect to $v^*, v_i$, and $u^*$ to zero we obtain $u$, $u_i$, and $v$ by solving $1 + M$ state problems (2.3) at $m_i$, $i = 0, \ldots, M$, and 1 linear adjoint problem (3.9). In a similar way, setting variation with respect to $v$, $u_i$, and $u$ to zero, we obtain $u^*$, $v_i$, and $v^*$, by solving $2 + M$ linear problems. Thus, in addition, $2 + M$ linear problems need to be solved to compute the $z$-gradient

(4.25) \( \langle \tilde{z}, D_zJ_{\text{lin}}^\text{MC}(z) \rangle = \langle \tilde{z}, \partial_z L_{\text{lin}}^\text{MC}(u,v,\{u_i\},u^*,v^*,\{v_i\},z) \rangle \).$$

4.3.2. $z$-gradient for variance reduction with quadratic approximation.

As for the quadratic approximation (3.3), by the Monte Carlo corrections (3.23) and (3.25), we have an unbiased evaluation of the cost functional as

(4.26) \( J_{\text{quad}}^\text{MC}(z) = \tilde{Q}_{\text{quad}} + \beta \tilde{V}_{Q}^\text{quad} + \mathcal{P}(z) \),

subject to the state problems (2.3) and (3.8), the adjoint problem (3.9), the incremental state and adjoint problems (3.12) and (3.13). By defining the Lagrange multiplier
(u^*, v^*, \{v_i\}, \{\lambda_j^*\}, \{\psi_j^*\}, \{\hat{u}_j^*\}, \hat{v}_j^*, z), we can form the Lagrangian as

$$L_{MC}^{\text{quad}}(u, v, \{u_i\}, \{\lambda_j\}, \{\psi_j\}, \{\hat{u}_j\}, \hat{v}_j, \{\hat{v}_j\}, z, \{v^*_i\})$$

$$= J_{\text{quad}}^{MC} + \langle u^*, \partial_u r_0 \rangle + \langle v^*, \partial_u r_0 + \partial_u Q_0 \rangle + \sum_{i=1}^{M} r(u_i, v_i, m_i, z).$$

$$+ \sum_{j=1}^{N} \langle \psi_j^* \rangle (Q_{mm}(m_0) - \lambda_j^{-1}) \psi_j \rangle$$

$$+ \sum_{j=1}^{N} \lambda_j^* \langle \psi_j, 0 \rangle (\psi_j, 0) - 1)$$

(4.27)

By setting variation with respect to \( v^*, u^*, v_i, \hat{v}_j, \hat{v}_j^*, \hat{u}_j^* \) to zero, we obtain the state variables \( u, v, u_i, \hat{u}_j, \hat{v}_j, \hat{u}_j^* \) by solving 1 + \( M \) (nonlinear) state problems and 1 + 2\( N + 2M \) linear problems, \( i = 1, \ldots, M, j = 1, \ldots, N \). For the eigenpairs, 4(\( N + p \)) linear problems needs to be solved. By setting variation with respect to \( \lambda_j \) to zero, we obtain the same expression (4.15) for the adjoint variables \( \psi_j^* \). Setting variation with respect to \( \hat{v}_j, \hat{u}_j, \hat{v}_j^*, \hat{u}_j^* \) to vanish, we get the adjoint variables \( \hat{u}_j^*, \hat{v}_j^*, \hat{u}_j^*, \hat{v}_j, u^*, v^* \) by solving 2 + 2\( N + 3M \) linear problems, which are defined similarly as that for the quadratic approximation, but with the additional correction terms in both the cost functional and the constraints, thus in total 1 + \( M \) (nonlinear) state problems and 3 + 8\( N + 4p + 5M \) linear problems to evaluate the z-gradient

$$\langle \hat{z}, D_z L_{MC}^{\text{quad}}(z) \rangle = \langle \hat{z}, D_z L_{\text{MC}}^{\text{states, adjoints}, z} \rangle,$$

where states and adjoints represent the state and adjoint variables for short.

4.4. z-gradient-based optimization method. To solve the PDE-constrained optimization problem we use a gradient-based optimization method, since it only requires the ability to evaluate the cost functional (2.5) and its z-gradient. In particular, we use a BFGS method [48].

The five approximations of the optimization under uncertainty problem described in this section all differ among each other in terms of accuracy and computational cost. Table 4.4 summarizes the cost, measured in number of (nonlinear) state and linear PDE solves, to evaluate cost functional and its z-gradient for each approximation. The linear approximation in (4.6) is the cheapest to compute but also the most
Taylor approximation and variance reduction for optimal control under uncertainty

While the MC-corrected approximations in (4.23) and (4.26) are the most accurate but also the most expensive. The quadratic approximation (4.11) is, in many applications, more accurate than the linear approximation and, for highly nonlinear problems, is much cheaper to compute than the MC-corrected approximation. In fact, all the linear PDEs solve in the trace approximation share the same left-hand side \( \partial u r_0 \) (or its adjoint \( \partial u^* r_0 \)), and therefore one can amortize the cost of computing an expensive preconditioner or sparse factorization over several solutions of the linearized PDE. It is worth noting that these different approximations of the control problem can be exploited in a multifidelity framework where the optimal solution of a cheaper approximation is used as initial guess for a more accurate (and more expensive) approximation. For example, in the numerical results section the optimal solution using the MC-corrected formulation are obtained by first solving the linear approximation in (4.6), then solving the quadratic approximation in (4.11) and finally (4.23) or (4.26).

5. Numerical experiments. We consider two PDE-constrained optimal control problems under uncertainty where the uncertain parameter is the discretization of an infinite-dimensional field. These experiments aim to demonstrate the computational efficiency and accuracy of the approximation based on the Taylor expansion and their MC-corrected counterparts. The first problem considers a fluid flow in a porous medium: the control variable is the fluid injection rate into the subsurface at specific wells locations, the uncertain parameter is the permeability field (modeled as a log-normal random field), and the objective is to drive the pressure measured at the production wells to a given target distribution. The second problem seeks an optimal velocity profile on the inlet boundary of a turbulent jet flow such that the jet width in a given cross-section is maximized: the governing equations are the Reynolds-averaged Navier–Stokes equations (RANS) with an algebraic closure model and an ad-hoc nonlinear advection-diffusion equation with uncertain coefficients is introduced to model the intermittency of turbulence at the edges of the jet.

The numerical results presented in this section are obtained using hIPPYlib \([60]\), an extensible software framework for large scale PDE-based deterministic and Bayesian inverse problems. hIPPYlib builds on FEniCS [44] (a parallel finite element library) for the discretization of the PDE and on PETSc [6] for scalable and efficient linear algebra operations and solvers. Finally, we use the implementation of the L-BFGS algorithm in the Python library SciPy, specifically the \texttt{fmin_l_bfgs_b} routine (BFGS with limited memory and box constraint).

5.1. Optimal source control of fluid flow in a porous medium. We consider an optimal source control problem as presented in [2] motivated by subsurface flow in petroleum engineering, where the state problem is a linear elliptic equation
describing the fluid flow in a porous medium (Darcy’s law). By letting \( u \) be the state variable representing the pressure and \( m \sim \mathcal{N}(m_0, \mathcal{C}) \) be a Gaussian random variable with mean \( m_0 \) and covariance \( \mathcal{C} \) representing the logarithm of the permeability field, the state problems read

\[
\begin{align*}
-\nabla \cdot (e^m \nabla u) &= Fz \quad \text{in } D, \\
\mathcal{C}_{\text{approx}} \quad \mathcal{C} &= \mathcal{C} \\
e^m \nabla u \cdot n &= 0 \quad \text{on } \Gamma_N,
\end{align*}
\]

where \( D \) is an open and bounded physical domain \( D \subset \mathbb{R}^d \) (\( d = 2, 3 \)) with Lipschitz boundary \( \partial D \), and \( \Gamma_D, \Gamma_N \subset \partial D \) (\( \Gamma_D \cup \Gamma_N = \partial D \) and \( \Gamma_D \cap \Gamma_N = \emptyset \)) denote respectively the Dirichlet and Neumann portion of \( \partial D \). \( F \) is a map from the control variable \( z \) to the source term \( Fu \) defined as

\[
Fz = \sum_{i=1}^{n_c} z_i f_i,
\]

where \( f_i, i = 1, \ldots, n_c \), are the mollified Dirac functions located at the \( n_c \) injection wells, and \( z = (z_1, \ldots, z_{n_c})^\top \in \mathbb{R}^{n_c} \) is a vector of control variables that represents the fluid injection rates. Finally, we let \( \mathcal{Z} = [z_{\text{min}}, z_{\text{max}}]^{n_c} \) be set of admissible values for the control variable, where \( z_{\text{min}} \) and \( z_{\text{max}} \) denote the minimum and maximum injection rate of each well.

The weak formulation of the problem (5.1) reads: given \( m \in \mathcal{M} = L^2(D) \) almost surely and \( z \in \mathcal{Z} \), find \( u = u_0 + g \) with \( u_0 \in \mathcal{U} = \{ u \in H^1(D) : u|_{\Gamma_D} = 0 \} \) such that

\[
r(u, v, m, z) = 0, \quad \forall v \in \mathcal{V},
\]

where \( \mathcal{V} = \{ v \in H^1(D) : v|_{\Gamma_D} = 0 \} \), and

\[
r(u, v, m, z) = \int_D e^m \nabla u \cdot \nabla v \, dx - \int_D Fz \, v \, dx.
\]

The control objective measures the difference between the pressure (computed by solving the state problem (5.3)) and a target pressure distribution \( \bar{u} = (\bar{u}_1, \ldots, \bar{u}_{n_p}) \) at the locations of production wells \( x_1, \ldots, x_{n_p} \) in \( D \). Specifically, we write

\[
Q(u) = ||Bu - \bar{u}||_2^2,
\]

where \( B : \mathcal{V} \to \mathbb{R}^{n_p} \) is a pointwise observation operator at the locations \( x^1, \ldots, x^{n_p} \).

In the numerical test, we set \( D = (0, 2) \times (0, 1) \), \( \Gamma_D = \{ 0, 2 \} \times [0, 1] \), and \( \Gamma_N = (0, 2) \times \{ 0, 1 \} \). We impose the Dirichlet data \( g = 2 - x_1 \). The mean field \( m_0 \) is given as in the top part of Fig 1, in which the locations for the 20 injection wells and the 12 production wells are also shown. For the covariance \( \mathcal{C} \), we set \( \mathcal{C} = (-\alpha_1 \Delta + \alpha_2 I)^{-2} \) with \( \alpha_1 = 0.1 \) and \( \alpha_2 = 20 \). Three random samples of the Gaussian measure \( \mathcal{N}(m_0, \mathcal{C}) \) are shown in the bottom part of Fig. 1. In the cost functional (2.5), we set \( \beta = 1 \) and use the penalization \( P(z) = 10^{-5} \times ||z||_2^2 \). We use a finite element discretization in a uniform triangular mesh of size 65 \times 33 with piecewise linear element for the discrete approximation of the state \( u \), the control function \( z \), and the parameter \( m \), which results in 2145 dimensions for the parameter \( m \). We specify the desirable pressure \( \bar{u}_i = 3 - 8(x_1^2 - 1)^2 - 4(x_2^2 - 0.5)^2 \) at the locations of the production wells \( (x_1^i, x_2^i) \), \( i = 1, \ldots, 12 \), as shown in the top part of Fig. 1. Moreover, we set the lower and upper bounds \( z_{\text{min}} = 0 \) and \( z_{\text{max}} = 32 \) for the control function, i.e., \( z \in [0, 32]^{20} \).
Taylor approximation and variance reduction for optimal control under uncertainty

Fig. 1. Top: mean $m_0$ of the Gaussian random variable $m \sim \mathcal{N}(m_0,C)$. The red diamonds represent the locations of the injection wells and the black squares represent the locations of the production wells. Bottom: three samples randomly drawn from the Gaussian measure $\mathcal{N}(m_0,C)$.

We consider the five different approximations of the cost functional presented in Sec. 3, and, to solve the optimal control problem under uncertainty, we use the limited memory BFGS method with box constraint, where the $z$-gradient of the cost functional are computed as in Sec. 4. We chose $z_0 = (16, \ldots, 16)$ as initial guess for the control function. We used $M = 100$ random samples from the Gaussian measure $\mathcal{N}(m_0,C)$ to estimate the mean and variance of the control objective in both the plain Monte Carlo integration presented in Sec. 2.3 and the variance reduction in Sec. 3.4.

For the computation of the trace appearing in the quadratic approximations (3.5), (3.23), and (3.25), we use the trace estimator $\hat{T}_2(\cdot)$ (3.17) by solving the generalized eigenvalue problem (3.18) using the randomized Algorithm 2, with $k = 100$ and $p = 10$. We also test the Gaussian trace estimator $\hat{T}_1(\cdot)$ in (3.15). In Fig. 2, we plot the errors, denoted as $\text{error}_1$ and $\text{error}_2$ for the two trace approximations, respectively, against the reference value $\sum_{j=1}^k \lambda_j$ with $k = 140$ at both the initial control function $z = z_0$ and the optimal control function $z = z_{\text{quad}}^{MC}$ obtained by the variance reduction with quadratic approximation. We can observe that in both cases, the errors for $\hat{T}_2(\mathcal{H})$ decays much faster than that for $\hat{T}_1(\mathcal{H})$ as the result of fast decaying eigenvalues, especially at the optimal control function where more than two order of magnitude’s difference of accuracy is observed at $N = 100$. We remark that the slow decay of the error for $\hat{T}_1(\cdot)$ may demand a very large $N$ if relatively high accuracy of the trace estimator is required, which leads to a large number of PDE solves at each step of the optimization procedure, thus deteriorating the computational advantage of using the Taylor approximation. In the following test, we use $\hat{T}_2(\cdot)$ with $N = 100$.

The optimal control functions obtained with the different approximations are shown in Fig. 3. From the plot on the right we can observe that the Monte Carlo integration (MC), the variance reduction with linear approximation (Lin+MC), and the variance reduction with quadratic approximation (Quad+MC) result in very similar optimal control functions, while the linear (Lin) and the quadratic (Quad) approximations lead to optimal control functions that are close to each other but that differ from that of the unbiased approximations. This indicates that approximation of the cost functional using only the linear or quadratic approximation may yield a large error for the optimal control function. The number of functional and $z$-gradient evaluations required by BFGS to reduce the $\ell^\infty$ norm of the projected gradient below an absolute
Fig. 2. The decay of the generalized eigenvalues, $\lambda_+$ for positive eigenvalues and $\lambda_-$ for negative eigenvalues, and the errors, denoted as $\text{error}_1$ and $\text{error}_2$ for the trace estimators $\hat{T}_1$ and $\hat{T}_2$ with respect to the number of estimate terms $N$ in (3.15) and (3.17). Left: $z = z_0$; right: $z = z_{MC}$.

tolerance of $10^{-5}$ are $(57, 54)$ for MC, $(51, 50)$ for Lin, $(47, 46)$ for Quad, $(21, 18)$ for Lin+MC, and $(21, 18)$ for Quad+MC. We remark that for Lin+MC and Quad+MC we start the optimization from the optimal control functions $z_{\text{lin}}$ and $z_{\text{quad}}$, allowing us to save more than half of the computational cost in solving the state PDEs compared to MC starting from $z_0$, see the comparison of cost in Table 4.4.

Fig. 3. Left: optimal control function obtained as the solution of the optimization problem (2.6) with the cost functional computed by the quadratic approximation and variance reduction for the mean (3.23) and the variance (3.25). Right: optimal control functions at the 30 locations, which are obtained with the cost functional computed by the Monte Carlo integration (MC), linear approximation (Lin), quadratic approximation (Quad), linear approximation with variance reduction (Lin+MC), quadratic approximation with variance reduction (Quad+MC), respectively.

Fig. 4 shows the relative errors at 100 random samples of the linear and the quadratic approximations of $Q$ in (5.5) used for the computation of $E[Q]$, as well as of $q = (Q - Q_0)^2$ in (3.22) and (3.25) for the computation of $\text{Var}[Q]$. We note that, for both quantities, the quadratic approximation is more accurate than the linear approximation for most of the realizations. We remark that both the linear and the quadratic approximations yield relatively large errors for $q$, thus potentially leading to inaccurate approximation of the variance $\text{Var}[Q]$. This may account for the difference in the optimal control functions obtained by only using the linear or the quadratic approximations for the computational of the cost functional and its derivative, as
Taylor approximation and variance reduction for optimal control under uncertainty

shown in Fig. 3.

Since the quadratic approximation is strongly correlated to the true value, we may expect large variance reduction by using them as control variates. In fact, the correlation coefficient between the quadratic approximation and the true value is very close to one, 0.996 for $Q$ and 0.998 for $q$. On the contrary, the linear approximation is poorly correlated to the true value (-0.211 for $Q$ and 0.718 for $q$), thus leading to negligible variance reduction. In Table 2 and 3, we show the effect of the variance reduction by the linear and the quadratic approximations for the computation of the mean $E[Q]$ and the variance $\text{Var}[Q]$, respectively. This reduction is measured by the mean-square-error (MSE), which is defined (e.g., for $Q$) as

\begin{equation}
\text{MSE}(Q) = \frac{\text{Var}_p[Q]}{M},
\end{equation}

where $\text{Var}_p[Q]$ is the population variance of $Q$ at $M$ random samples.

| control | # samples | $E^{MC}[Q]$ | MSE($Q$) | MSE($Q - Q_{\text{lin}}$) | MSE($Q - Q_{\text{quad}}$) |
|---------|-----------|-------------|----------|--------------------------|--------------------------|
| $z_0$   | 10        | 2.42e+01    | 1.10e+00 | 7.94e-03                 | 1.52e-04                 |
|         | 100       | 2.54e+01    | 2.54e-01 | 4.16e-03                 | 7.40e-05                 |
| $z_{\text{MC lin}}$ | 10       | 3.01e-01    | 1.10e-03 | 1.31e-03                 | 1.69e-05                 |
|         | 100       | 2.89e-01    | 9.68e-05 | 7.49e-05                 | 9.00e-07                 |
| $z_{\text{MC quad}}$ | 10       | 3.16e-01    | 1.11e-03 | 8.02e-04                 | 1.05e-05                 |
|         | 100       | 2.98e-01    | 1.10e-04 | 9.98e-05                 | 1.12e-06                 |

In all cases, we can see that the quadratic approximation achieves more significant variance reduction than the linear approximation. This means that the number of samples needed to achieve a target MSE is significantly smaller for the quadratic approximation than for the linear. For example, at the initial control function $z_0$, the quadratic approximation allows for a speed-up (defined as the ratio between the MSE of the Monte Carlo estimator and the MSE of the correction) of order 1000X and 100X.
in the estimation of $\mathbb{E}[Q]$ and $\text{Var}[Q]$, respectively; while the linear approximation allows for a speed-up of order 100X and 10X respectively. In addition, at the optimal control function $z_{\text{lin}}^{\text{MC}}$ and $z_{\text{quad}}^{\text{MC}}$, the quadratic approximation can still achieve a $\sim$ 100X speed up for evaluation of both the mean and the variance; while the linear approximation is too poor to achieve evident variance reduction. Note that the mean and the variance are both much smaller at the optimal control than those at the initial control, confirming that the cost functional is much reduced by solving the minimization problem (2.6).

5.2. Optimal boundary control of a turbulent jet flow. We consider a turbulent jet flow modeled by a Reynolds-averaged Navier–Stokes equations coupled with a nonlinear stochastic advection-diffusion equation for the modeling of the turbulent viscosity. Specifically, the governing equations read

\begin{equation}
\begin{aligned}
-\nabla \cdot \left( (\nu + \nu_t) \left( \nabla u + \nabla u^T \right) \right) + (u \cdot \nabla) u + \nabla p = 0, & \quad \text{in } D, \\
\nabla \cdot u = 0, & \quad \text{in } D, \\
-\nabla \cdot \left( (\nu + (\gamma + e^m) \nu_{t,0}) \gamma \nabla \gamma \right) + u \cdot \nabla \gamma - \frac{1}{2} \frac{u \cdot e_1}{x_1 + b} \gamma = 0, & \quad \text{in } D,
\end{aligned}
\end{equation}

with the boundary conditions

\begin{equation}
\begin{aligned}
\sigma_n(u) \cdot \tau = 0, & \quad u \cdot n + \chi_W \phi(z) = 0, \quad \text{on } \Gamma_I, \\
\sigma_n(u) \cdot n = 0, & \quad u \cdot \tau = 0, \quad \text{on } \Gamma_O \cup \Gamma_W, \\
\sigma_n(u) \cdot \tau = 0, & \quad u \cdot n = 0, \quad \text{on } \Gamma_C, \\
\gamma - \gamma_0 = 0, & \quad \text{on } \Gamma_I \cup \Gamma_W, \\
\sigma_n^0(\gamma) \cdot n = 0, & \quad \text{on } \Gamma_O \cup \Gamma_C.
\end{aligned}
\end{equation}

Here, $D = (0, 30) \times (0, 10)$ is the computational domain (half of the physical domain for the jet flow) as shown in the left part of Fig. 5. $e_1 = (1, 0)$, $\tau$ and $n$ are the unit tangential vector and the unit normal vector pointing outside the domain $D$ along the boundary $\partial D$. $\sigma_n(u) = (\nu + \nu_t) (\nabla u + \nabla u^T) \cdot n$ and $\sigma_n^0(\gamma) = (\nu + (\gamma + e^m) \nu_{t,0}) \nabla \gamma \cdot n$ are the stress tensors. $\chi_W$ is an indicator function with $\chi_W(x) = 1$ if $x \in [0, W]$ and $\chi_W(x) = 0$ if $x \in (W, 10)$. Here we take $W = 1.5$, $u = (u_1, u_2)$ and $p$ are the time-averaged velocity and pressure. $\nu_t = \gamma \nu_{t,0}$ is the turbulent viscosity, where $\nu_{t,0}$ represents an algebraic closure model and $\gamma$ is the indicator function that is close to one near the centerline $\Gamma_C$ and zero outside the jet region. Based on a dimensional analysis of the turbulent planar jet, the turbulent viscosity along the jet centerline is

| control | # samples | $\mathbb{E}^\text{MC}[q]$ | MSE($q$) | MSE($q_{\text{lin}}$) | MSE($q_{\text{quad}}$) |
|---------|-----------|-----------------|---------|-----------------|------------------|
| $z_0$   | 10        | 1.32e+01        | 1.71e+01| 9.38e-01        | 6.05e-03         |
|         | 100       | 2.54e+01        | 1.35e+01| 2.61e+00        | 3.09e-02         |
| $z_{\text{lin}}^{\text{MC}}$ | 10        | 1.99e-02        | 1.00e-04| 9.35e-05        | 3.92e-06         |
|         | 100       | 1.65e-02        | 1.32e-05| 1.22e-05        | 2.44e-07         |
| $z_{\text{quad}}^{\text{MC}}$ | 10        | 2.38e-02        | 1.82e-04| 1.79e-04        | 8.07e-07         |
|         | 100       | 1.99e-02        | 1.66e-05| 1.65e-05        | 3.81e-07         |

Table 3: Variance reduction by the linear and the quadratic approximations at the initial control function $z_0$, and the optimal control functions $z_{\text{lin}}^{\text{MC}}$ and $z_{\text{quad}}^{\text{MC}}$. Reported are the MSE for the evaluation of the variance $\text{Var}[Q]$, where $q = (Q - Q_0)^2$, $q_{\text{lin}} = (Q_{\text{lin}} - Q_0)^2$, $q_{\text{quad}} = (Q_{\text{quad}} - Q_0)^2$.
Taylor approximation and variance reduction for optimal control under uncertainty

given by

\[ \nu_{t,0} = C \sqrt{M(x_1 + aW)^{1/2}}, \]

where \( M = \int_{\Gamma_I} ||u||^2 ds \) and \( W \) are the momentum and the jet width at the inlet boundary, \( C \) and \( a \) are two adimension parameters that need to be calibrated against either experimental or direct Navier-Stokes simulation (DNS) data. The indicator function \( \gamma \) satisfies the nonlinear stochastic advection-diffusion equation in (5.7) with the boundary data

\[ \gamma_0 = 0.5 - 0.5 \tanh \left( 5 \left( \frac{30 - x_1}{30} \right) (x_2 - 1 - 0.5x_1) \right). \]

The stochastic parameter \( m \) is a Gaussian random field with measure \( \mathcal{N}(m_0, \mathcal{C}) \). We set \( m_0 \) as a constant field and \( \mathcal{C} = (-\alpha_1 \nabla \cdot (\Theta \nabla) + \alpha_2 I)^{-2} \), where \( \alpha_1 = \alpha_2 = 0.5 \), \( \Theta = [15, 0; 0, 1] \). Two samples drawn from \( \mathcal{N}(0, \mathcal{C}) \) are shown in Fig. 6. In general, also the parameters \((C, a, m_0, b)\) are random variables whose distributions can be inferred from experimental or DNS data. Here, for simplicity, we only consider the uncertainty in \( m \) and assume deterministic values for \((C, a, m_0, b) = (0.012, 5.29, -0.47, 9.58)\).\(^3\)

---

\(^3\)Such values are the maximum a posteriori estimates obtained by solving a calibration problem where the misfit functional is the \( L^2 \) distance between the velocity of (5.7) and the DNS data provided in [37].
the boundary \( \Gamma_I \) in \( x_2 \) direction, which is represented by a B-spline function as

\[
\phi(z)(t) = \sum_{k=1}^{K} z_k B_{k,n}(t),
\]

where the coefficient vector \( z := (z_1, \ldots, z_K) \in [0,1]^K \) is the control variable and \( B_{k,n}(t) \) are B-spline basis functions recursively defined as

\[
B_{k,n}(t) = \frac{t - t_k}{t_{k+n} - t_k} B_{k,n-1}(t) + \frac{t_{k+n+1} - t}{t_{k+n+1} - t_{k+1}} B_{k+1,n-1}(t),
\]

where

\[
B_{k,0}(t) = \begin{cases} 1 & \text{if } t_k \leq t < t_{k+1}; \\ 0 & \text{otherwise}. \end{cases}
\]

Here we take \( K = 11 \) and use the B-spline of order \( n = 3 \) with 15 knots (4 repetitions at each edge) depicted in Fig. 5. A B-spline interpolation of the inflow velocity profile used as initial control is shown on the right in Fig. 5.

The control objective is the jet width \( Q \) along the cross-section \( \Gamma_0 \) at \( x_1 = 20 \) as shown in Fig. 5, and is defined as

\[
Q(u) = \frac{1}{\delta_1^0(u)} \int_{\Gamma_0} u \cdot e_1 dx_2,
\]

where \( \delta_1^0(u) = u(x^0) \cdot e_1 \) with \( x^0 = (20,0) \). The objective is to maximize the expected value of the jet width \( Q \), i.e., to minimize \( \mathbb{E}[-Q] \) subject to the state problem (5.7) and the constraint that the inlet momentum \( M \) is prescribed. Specifically, we consider the risk adverse optimization problem where the cost functional reads

\[
J(z) = \mathbb{E}[-Q] + \beta \text{Var}[Q] + \mathcal{P}(z).
\]

Here we set \( \beta = 10^3 \) to balance the mean and the variance, and define the penalization term \( \mathcal{P}(z) \) as

\[
\mathcal{P}(z) = \beta_1 \left( \int_{\Gamma_I} (\phi(z))^2 dx - M \right)^2 + \beta_2 \int_{\Gamma_I} |\nabla \phi(z)|^2 dx,
\]

where the first term penalizes violations of the inlet momentum constraint (\( \beta_1 = 10^3 \)), and the second term controls the regularity of the inlet velocity (\( \beta_2 = 1 \)).

The weak formulation of the turbulence model is given by: given the realization \( m \in M = L^2(D) \) almost surely and the control function \( z \in Z = \mathbb{R}^K \), find \( u \in V \) such that

\[
r(u,v,m,z) = 0, \quad \forall v \in V,
\]

where we denote the state and adjoint variables \( u = (u,p,\gamma) \in U \) and \( v = (v,q,\eta) \in V \), with \( U = V = (H^1(D))^2 \times L^2(D) \times H^1(D) \).

Specifically, the weak form in (5.15) consists of three terms and reads

\[
r(u,v,m,z) = \text{Model}(u,v,m) + \text{Stabilization}(u,v) + \text{Nitsche}(u,v,m,z).
\]
The first term represents the model in the weak form given by

$$\text{Model}(u, v, m) = \int_D (\nu + \nu_t) 2S(u) \cdot S(v) dx + \int_D [(u \cdot \nabla) u] \cdot v dx - \int_D p \nabla \cdot v dx$$

$$+ \int_D q \nabla u dx$$

$$+ \int_D (\nu + (\gamma + e^m) \nu_t) \nabla \gamma \cdot \nabla \eta dx + \int_D [u \cdot \nabla \gamma] \eta dx - \int_D \frac{1}{2} \frac{u \cdot e_1}{x_1 + b} \gamma dx.$$  

(5.17)

where $S(u) = (\nabla u + \nabla u^\top)/2$ denotes strain tensor.

The second term of (5.15) represents the stabilization by Galerkin Least-Squares (G-LS) method for advection-dominated problems

$$\text{Stabilization}(u, v) = \int_D \tau_1 L_1(u) \cdot D_u L_1(u)(v) dx$$

$$+ \int_D \tau_2 (\nabla \cdot u)(\nabla \cdot v) dx + \int_D \tau_3 (u \cdot \nabla \gamma)(u \cdot \nabla \eta) dx,$$

(5.18)

where $L_1(u)$ represents the strong form of the momentum equation in line 1 of (5.7), $\tau_1$, $\tau_2$ and $\tau_3$ are properly chosen stabilization parameters associated with the local Péclet number.

The third term of (5.15) represents the weak imposition of inlet velocity profile (i.e. the control function) by Nitsche’s method [8] and reads

$$\text{Nitsche}(u, v, m, z) = C_d \int_{\Gamma_o \cup \Gamma_w} h^{-1}(\nu + \nu_t)(u \cdot \tau)(v \cdot \tau) ds$$

$$- \int_{\Gamma_o \cup \Gamma_w} (\sigma_n(u) \cdot \tau)(v \cdot \tau) + (\sigma_n(v) \cdot \tau)(u \cdot \tau) ds$$

$$+ C_d \int_{\Gamma_t} h^{-1}(\nu + \nu_t)(u \cdot n + \chi W \phi(z))(v \cdot n) ds$$

$$- \int_{\Gamma_t} (\sigma_n(u) \cdot n)(v \cdot n) + (\sigma_n(v) \cdot n)(u \cdot n + \chi W \phi(z)) ds,$$

(5.19)

where the second and the fourth terms arise from integration by parts, and the first and the third terms guarantee the coercivity of the form. $C_d$ is a constant, set as $C_d = 10^5$. $h$ is a local length of the boundary edge along the Dirichlet boundaries.

We discretize the problem using the Finite Element Method with piecewise continuous elements (P2, P1, P1) for the state $u$ and P1 elements for the stochastic field $m$. We triangulate the computational domain using an anisotropic, locally refined in the jet region, mesh with 14400 triangular elements (60 elements along the $x$-direction and 120 along the $y$-direction). After discretization, we obtain a stochastic optimization problem where the dimension of the control is 11, the state is 73084, and the stochastic parameter is 7381. As in the previous example, we use a BFGS method to solve the optimization problem (2.6) with the cost functional defined in (5.13) and the constraint defined by the state problem (5.15). At each BFGS iteration, we solve the fully-coupled nonlinear state problem using Newton method with an LU factorization of the Jacobian operator. The initial guess for the control function $\phi(z)$ is shown in the left part of Fig. 8, and the corresponding state $u$ on the top part of Fig. 9. To
compute all the derivatives involved in the $z$-gradient presented in Sec. 4 we exploit the symbolic differentiation capabilities of FEniCS [44].

For the computation of the trace in the quadratic approximation we use both the Gaussian trace estimator $\hat{T}_1(\cdot)$ in (3.15) and (3.16), and the trace estimator $\hat{T}_2(\cdot)$ in (3.17) obtained by solving the generalized eigenvalue problem (3.18). Fig. 7 displays the decay of the eigenvalues as well as the approximation error of the two estimators for both the initial control $z_0$ and the optimal control $z_{MC}^{quad}$. The reference value for the trace $\text{tr}(\mathcal{H})$ was computed (with a precision up to a $10^{-8}$) as the sum of the first 140 dominant eigenvalues of $\mathcal{H}$ using the randomized Algorithm 2 with $k = 140$ and $p = 10$. It is evident that the second trace estimator achieves much faster decay of the error than the Gaussian trace estimator, with a gain in accuracy of about three orders magnitude when using $N = 100$ samples/eigenvalues. In the numerical test, we use the second trace estimator with $N = 20$, which already provides sufficiently accurate results. It is worth noticing that, in contrast, the Gaussian trace estimator would have required about $N = 10^4$ samples to achieve the same accuracy.

Fig. 7. The decay of the generalized eigenvalues, $\lambda_+$ for positive eigenvalues and $\lambda_-$ for negative eigenvalues, and the errors, denoted as $\text{error}_1$ and $\text{error}_2$ for the trace estimators $\hat{T}_1$ and $\hat{T}_2$ with respect to the number of estimate terms $N$ in (3.15) and (3.17). Left: $z = z_0$; right: $z = z_{MC}^{quad}$

Fig. 8 shows the optimal control functions $\phi(z)$ obtained by different approximations.

Fig. 8. Left: initial guess and the optimal control functions obtained as the solution of the optimization problem (2.6) with the cost functional computed by different approximations. Right: the differences between the optimal control functions obtained by different approximations and that by quadratic approximation with variance reduction.
mations presented in Section 3, namely the standard Monte Carlo approach (MC), the linear approximation (Lin), the quadratic approximation (Quad), and the linear and quadratic approximation with Monte Carlo correction (Lin+MC and Quad+MC, respectively). The plot on the left shows that these control functions are close to each other compared to the initial control function; the plot on the right part highlights the differences between the first four approximations and the reference optimal control solution obtained by quadratic approximation with variance reduction (Quad+MC) with $M = 10$ random samples for the Monte Carlo correction. As expected, the control function obtained by the linear approximation with variance reduction (Lin+MC) is the closest to that obtained by the Quad+MC, and standard MC approach is the least accurate. We remark that the differences of the optimal control functions are small because the control objective depends smoothly on the parameter and has relatively smaller variance compared to that of the parameter, as shown by the fast decay of the eigenvalues in Fig. 7, even if the state problem is rather complicated and the velocity of the turbulent jet has sharp variation in the computational domain, see Fig. 9. Also, Fig. 9 shows that the turbulent jet is more spread out with larger jet width at the optimal value of the control compared to that at the initial control.

![Fig. 9. The velocity field of the turbulent jet flow and obtained solving the state model for the initial (top) and the optimal (bottom) inlet velocity profile. Specifically, the optimal profile was obtained using the quadratic approximation with variance reduction.](image)

Fig. 10 shows the relative difference evaluated at 10 random samples for both $Q$ and $q = (Q - Q_0)^2$. It is clear that the quadratic approximation is more accurate than the linear approximation for both quantities. This allows for a drastic reduction in the variance of the Monte Carlo correction for both the linear and quadratic approximation with respect to standard Monte Carlo, as demonstrated by the MSE reported in Tables 4 and 5. In turn, such variance reduction translates to over a 100X speed up in the computation with respect to standard Monte Carlo for the same accuracy (MSE). Moreover, it is worth noting that both the linear and quadratic approximations are more accurate for $Q$ than for $q$, since $Q$ has a smaller relative variance (i.e. the ratio between the variance and mean square value) than of $q$.

To investigate the scalability of the proposed algorithm, we consider up to six levels of uniform mesh refinement: $30 \times 60, 60 \times 120, 120 \times 240, 240 \times 480, 480 \times$
Fig. 10. Relative errors at 10 random samples of the linear and the quadratic approximations of $Q$ for the computation of the mean $E[Q]$ (left) and of $q = (Q - Q_0)^2$ for the variance $\text{Var}[Q]$ (right). Here, the errors are show at the optimal control function $z = z_{\text{MC quad}}$.

Table 4

| parameter dimension | $E^{\text{MC}}[Q]$ | MSE($Q$) | MSE($Q - Q_{\text{lin}}$) | MSE($Q - Q_{\text{quad}}$) |
|---------------------|---------------------|----------|---------------------------|---------------------------|
| 1,891               | -1.71e+00           | 7.40e-06 | 2.68e-08                  | 1.81e-09                  |
| 7,381               | -1.59e+00           | 7.94e-06 | 1.57e-07                  | 1.46e-08                  |
| 29,161              | -1.44e+00           | 3.82e-06 | 7.23e-08                  | 1.66e-08                  |
| 115,921             | -1.42e+00           | 9.47e-06 | 6.91e-08                  | 3.06e-08                  |
| 462,241             | -1.41e+00           | 8.85e-06 | 9.00e-08                  | 1.78e-08                  |

Table 5

| parameter dimension | $E^{\text{MC}}[q]$ | MSE($q$) | MSE($q - q_{\text{lin}}$) | MSE($q - q_{\text{quad}}$) |
|---------------------|---------------------|----------|---------------------------|---------------------------|
| 1,891               | 8.05e-05            | 9.37e-10 | 1.76e-11                  | 8.77e-13                  |
| 7,381               | 8.13e-05            | 1.15e-09 | 8.87e-12                  | 1.48e-12                  |
| 29,161              | 5.60e-05            | 6.59e-10 | 3.39e-11                  | 4.04e-12                  |
| 115,921             | 1.07e-04            | 1.82e-09 | 7.04e-11                  | 2.08e-11                  |
| 462,241             | 8.96e-05            | 1.37e-09 | 7.78e-12                  | 1.33e-11                  |

The scalability of our approach with respect to the stochastic dimension depends
on three factors, namely the dimension independent behavior of (i) the spectral properties of the covariance preconditioned Hessian, (ii) the convergence rate of BFGS, and (iii) the variance of estimators for the Monte Carlo corrections of the objective function. The top-left plot of Fig. 11 demonstrates (i): the generalized eigenvalues of problem (3.18) at the optimal control obtained by linear approximation exhibit the same fast decay independently of the stochastic dimension (ranging from a thousand to a million), thus indicating that approximating the trace by randomized eigensolver algorithms is a scalable approach with respect to the stochastic parameter dimension.

The other three plots of Fig. 11 provides numerical evidence of the mesh independent convergence of the BFGS algorithm (ii) for this particular problem. In the case of the linear and quadratic approximations using mesh-continuation, the norm of the z-gradient drops below the prescribed tolerance within 12 iterations for all mesh resolutions except the coarsest one (which requires more iterations); in the case of the linear approximation with Monte Carlo correction, BFGS converges within 19 iterations for all mesh resolutions using the solution of the linear approximation as initial guess. This implies that the gradient-based BFGS optimization algorithm is also scalable with respect to the stochastic parameter dimension. Finally, Tables 4 and 5 indicate that the variance of the Monte Carlo corrections (iii) using Taylor approximation as control variates is independent of the stochastic parameter dimension, and, furthermore, it allows for a about 100X speed up with respect to standard Monte Carlo to achieve the same accuracy (MSE). Combining the dimension independent behavior of the three factors above, we conclude that the Taylor approximation and variance reduction algorithm is scalable with respect to the parameter dimension.

6. Conclusions. In this work we proposed a scalable computational framework to solve optimal control problem under uncertainty where the cost functional involves the expectation and the variance of a scalar control objective and the state problem is governed by a stochastic PDE with high/infinite-dimensional uncertain parameters. Specifically, we advocate the use of (low-order) Taylor expansion of the control objective with respect to the uncertain parameter either to directly approximate the moments of the control objective or as a variance reduction tool by correcting the Taylor expansion with a few Monte Carlo samples of the difference. Assuming that the uncertain parameter follows a Gaussian distribution, mean and variance of the Taylor expansions can be computed analytically. In particular, while the linear approximation of the control objective is also Gaussian, the quadratic approximation is not Gaussian and the computation of its mean and variance requires to approximate the trace of the (preconditioned) Hessian of the control objective with respect to the uncertain parameter. Randomized algorithms for the solution of generalized symmetric eigenproblems \[60, 54, 55\] allows for accurate and efficient approximation of this trace and only require computing the action of the preconditioned Hessian in a small number of random directions. As we showed in the numerical results, this approach is more efficient and accurate than the Gaussian trace estimator when the eigenvalues of the preconditioned Hessian exhibit fast decay, which is often true when the control objective smoothly depends on the uncertain parameter regardless of the complexity of the state problem. Moreover, when the possibly biased Taylor (linear and quadratic) approximation results in large error of the optimal control function, we can use it for a variance reduction method and correct the mean and the variance by Monte Carlo integration of the residual of the Taylor approximation, which was shown to achieve considerable computational cost reduction (by several orders of magnitude) compared to a simple Monte Carlo method. The complexity of our
Fig. 11. Top-left: decay of the generalized eigenvalues (in absolute value) of (3.18) with different parameter dimensions (dim) at the optimal control state obtained with linear approximation. Decay of the gradient of the optimization with respect to the number of BFGS iterations (# iter), with linear approximation (top-right), quadratic approximation (bottom-left), and linear approximation with variance reduction (bottom-right). Smaller number of dimensions are shown for the last case due to constraint on the computational time.

The method is independent of the nominal dimensions of the uncertain parameter; in fact, the computational cost is mostly driven by the effective parameter dimensions (i.e., the rank of the preconditioned Hessian) and by the (small) variance of the residual of the Taylor approximation, thus largely alleviating or breaking the curse of dimensionality. Gradient of the cost functional with respect to the control variable are derived from the Lagrangian functional of the optimal control problem. In the numerical examples, we used symbolic differentiation to express the (high-order) derivatives of the Lagrangian with respect to the state variable, the adjoint variable, the uncertain variable, as well as control variable. The Lagrangian formalism, in conjunction with symbolic differentiation and gradient-based optimization methods (such as BFGS), makes our methods applicable to general complex nonlinear PDE models, as demonstrated by the turbulent jet example.

Several extensions and further developments of the computational framework are interesting. At first, it is challenging yet promising to investigate higher-order Taylor expansion beyond the quadratic approximation for higher moments beyond the variance of the control objective. Secondly, further computational reduction can possibly be achieved by model order reduction for the state and the linearized PDEs. Thirdly, one can replace the Monte Carlo integration in the variance reduction by a sparse quadrature [19]. At last, application of the different Taylor approximations and vari-
Taylor approximation and variance reduction for optimal control under uncertainty may lead to additional computational saving.

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 $\ell_0$-sparsification. *SIAM Journal on Scientific Computing*, 36(5):A2122–A2148, 2014.

[2] Alan Alexanderian, Noemi Petra, Georg Stadler, and Omar 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(1):1166–1192, 2017.

[3] Ahmad Ahmad Ali, Elisabeth Ullmann, and Michael Hinze. Multilevel Monte Carlo analysis for optimal control of elliptic PDEs with random coefficients. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):466–492, 2017.

[4] H. Avron and S. Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *Journal of the ACM (JACM)*, 58(2):8, 2011.

[5] E. Bader, M. Kärcher, M.A. Grepl, and K. Veroy. Certified reduced basis methods for parametrized distributed elliptic optimal control problems with control constraints. *SIAM Journal on Scientific Computing*, 38(6):A3921–A3946, 2016.

[6] S. Balaj, S. Ahlyankar, M.F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W.D. Gropp, D.K. M.G. Knepley, I.C. McInnes, K. Rupp, B.F. Smith, S. Zampini, and H. Zhang. PETSc users manual. Technical Report ANL-95/11 - Revision 3.6, Argonne National Laboratory, 2015.

[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(6):844–868, 2008.

[8] Y. Bazilevs and T.J.R. Hughes. Weak imposition of Dirichlet boundary conditions in fluid mechanics. *Computers & Fluids*, 36(1):12–26, 2007.

[9] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531, 2015.

[10] 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(2):491–518, 2016.

[11] P. Benner, E. Sachs, and S. Volkwein. Model order reduction for PDE constrained optimization. In *Trends in PDE Constrained Optimization*, pages 303–326. Springer, 2014.

[12] D. Bertsimas, D.B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.

[13] A. Borzì and V. Schulz. *Computational optimization of systems governed by partial differential equations*, volume 8. SIAM, 2011.

[14] A. Borzì, V. Schulz, C. Schillings, and G. Von Winckel. On the treatment of distributed uncertainties in PDE-constrained optimization. *GAMM-Mitteilungen*, 33(2):230–246, 2010.

[15] T. Bui-Thanh and O. Ghattas. Analysis of the Hessian for inverse scattering problems: I. Inverse shape scattering of acoustic waves. *Inverse Problems*, 28(5):055001, 2012.

[16] T. Bui-Thanh and O. Ghattas. Analysis of the Hessian for inverse scattering problems. Part III: Inverse medium scattering of electromagnetic waves in three dimensions. *Inverse Problems and Imaging*, 7(4):1139–1155, 2013.

[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(6):A2494–A2523, 2013.

[18] T. Bui-Thanh, K. Willcox, O. Ghattas, and B. van Bloemen Waanders. Goal-oriented, model-constrained optimization for reduction of large-scale systems. *Journal of Computational Physics*, 224(2):880–896, 2007.

[19] P. Chen. Sparse quadrature for high-dimensional integration with Gaussian measure. *ESAIM: Mathematical Modelling and Numerical Analysis. accepted*, 2018.

[20] P. Chen and O. Ghattas. Hessian-based sampling for goal-oriented model reduction with high-dimensional parameter. *preprint*, 2017.

[21] P. Chen and A. Quarteroni. Weighted reduced basis method for stochastic optimal control problems with elliptic PDE constraints. *SIAM/ASA J. Uncertainty Quantification*, 2(1):364–396, 2014.

[22] P. Chen, A. Quarteroni, and G. Rozza. Stochastic optimal Robin boundary control problems of advection-dominated elliptic equations. *SIAM Journal on Numerical Analysis*, 51(5):2700
[23] 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(1):67–102, 2016.

[24] 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*, 327:147–172, 2017.

[25] J. Dick, F.Y. Kuo, and I.H. Sloan. High-dimensional integration—the Quasi-Monte Carlo way. *Acta Numerica*, 22:133–288, 2013.

[26] J. Dick, Q.T. Le Gia, and Ch. Schwab. Higher order quasi-Monte Carlo integration for holomorphic, parametric operator equations. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):48–79, 2016.

[27] H.P. 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(1):407–432, 2011.

[28] D. Ghate and M.B. Giles. Efficient Hessian calculation using automatic differentiation. *AIAA paper*, 4059:2007, 2007.

[29] R. Glowinski and J.L. Lions. *Exact and approximate controllability for distributed parameter systems*. Cambridge University Press, Cambridge, UK, 1996.

[30] M.D. Gunzburger. *Perspectives in flow control and optimization*, volume 5. SIAM, 2003.

[31] M.D. Gunzburger, H.C. Lee, and J. Lee. Error estimates of stochastic optimal Neumann boundary control problems. *SIAM Journal on Numerical Analysis*, 49:1532–1552, 2011.

[32] N. Halko, P-G Martinsson, and J.A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.

[33] J.S. Hesthaven, G. Rozza, and B. Stamm. Certified reduced basis methods for parametrized partial differential equations. *SpringerBriefs in Mathematics*, 2015.

[34] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*, volume 23. Springer Science & Business Media, 2008.

[35] 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(1):87–103, 2011.

[36] M. Kärcher and M.A. Grepl. A posteriori error estimation for reduced order solutions of parametrized parabolic optimal control problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(6):1615–1638, 2014.

[37] M Klein, A. Sadiki, and J. Janicka. Investigation of the influence of the Reynolds number on a plane jet using direct numerical simulation. *International Journal of Heat and Fluid Flow*, 24(6):785–794, 2003.

[38] D.P. Kouri, D. Heinkenschloos, M. Ridzal, and B.G. Van Bloemen Waanders. A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty. *SIAM Journal on Scientific Computing*, 35(4):A1847–A1879, 2012.

[39] D.P. Kouri and T.M. Surowiec. Risk-averse PDE-constrained optimization using the conditional Value-At-Risk. *SIAM Journal on Optimization*, 26(1):365–396, 2016.

[40] A. Kunoth and Ch. Schwab. Analytic regularity and GPC approximation for control problems constrained by linear parametric elliptic and parabolic PDEs. *SIAM Journal on Control and Optimization*, 51(3):2442–2471, 2013.

[41] A. Kunoth and Ch. Schwab. Sparse adaptive tensor Galerkin approximations of stochastic PDE-constrained optimal control problems. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1034–1059, 2016.

[42] T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza. Boundary control and shape optimization for the robust design of bypass anastomoses under uncertainty. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47(4):1107–1131, 2013.

[43] J.L. Lions. *Optimal control of systems governed by partial differential equations*. Springer-Verlag, Berlin, 1971.

[44] A. Logg, K.A. Mardal, and G. Wells. *Automated solution of differential equations by the finite element method: The FEniCS book*, volume 84. Springer Science & Business Media, 2012.

[45] F. Negri, A. Manzoni, and G. Rozza. Reduced basis approximation of parametrized optimal flow control problems for the Stokes equations. *Computers & Mathematics with Applications*, 69(4):319–336, 2015.

[46] F. Negri, G. Rozza, A. Manzoni, and A. Quarteroni. Reduced basis method for parametrized elliptic optimal control problems. *SIAM Journal on Scientific Computing*, 35(5):A2316–
Taylor approximation and variance reduction for optimal control under uncertainty

A2340, 2013.

[47] L.W.T. Ng and K.E. Willcox. Multifidelity approaches for optimization under uncertainty. *International Journal for numerical methods in Engineering*, 100(10):746–772, 2014.

[48] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[49] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction*, volume 92. Springer, 2015.

[50] T.D. Robinson, M.S. Eldred, K.E. Willcox, and R. Haimes. Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping. *Aiaa Journal*, 46(11):2814–2822, 2008.

[51] Farbod Roosta-Khorasani and Uri Ascher. Improved bounds on sample size for implicit matrix trace estimators. *Foundations of Computational Mathematics*, 15(5):1187–1212, 2015.

[52] E. Rosseel and G.N. Wells. Optimal control with stochastic PDE constraints and uncertain controls. *Computer Methods in Applied Mechanics and Engineering*, 213 - 216(0):152 – 167, 2012.

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

[54] 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(2):314–339, 2016.

[55] Arvind K Saibaba, Alen Alexanderian, and Ilse CF Ipsen. Randomized matrix-free trace and log-determinant estimators. *Numerische Mathematik*, pages 1–43, 2016.

[56] C. Schillings, S. Schmidt, and V. Schulz. Efficient shape optimization for certain and uncertain aerodynamic design. *Computers & Fluids*, 46(1):78–87, 2011.

[57] C. Schillings and Ch. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6), 2013.

[58] 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(5):2659–2682, 2012.

[59] F. Tröltzsch. *Optimal control of partial differential equations: theory, methods, and applications*, volume 112. American Mathematical Society, Providence, RI, 2010.

[60] U. Villa, N. Petra, and O. Ghattas. hIPPYlib: An extensible software framework for large-scale deterministic and linearized Bayesian inversion, 2016.

[61] M.J. Zahr and C. Farhat. Progressive construction of a parametric reduced-order model for PDE-constrained optimization. *International Journal for Numerical Methods in Engineering*, 102(5):1111–1135, 2015.