STOCHASTIC COLLOCATION FOR OPTIMAL CONTROL PROBLEMS WITH STOCHASTIC PDE CONSTRAINTS*

HANNE TIESLER†, ROBERT M. KIRBY‡, DONGBIN XIU§, AND TOBIAS PREUSSER†

Abstract. We discuss the use of stochastic collocation for the solution of optimal control problems which are constrained by stochastic partial differential equations (SPDE). Thereby the constraining SPDE depends on data which is not deterministic but random. Assuming a deterministic control, randomness within the states of the input data will propagate to the states of the system. For the solution of SPDEs there has recently been an increasing effort in the development of efficient numerical schemes based upon the mathematical concept of *generalized polynomial chaos*. Modal-based stochastic Galerkin and nodal-based stochastic collocation versions of this methodology exist, both of which rely on a certain level of smoothness of the solution in the random space to yield accelerated convergence rates. In this paper we apply the stochastic collocation method to develop a gradient descent as well as a sequential quadratic program (SQP) for the minimization of objective functions constrained by an SPDE. The stochastic function involves several higher-order moments of the random states of the system as well as classical regularization of the control. In particular we discuss several objective functions of tracking type. Numerical examples are presented to demonstrate the performance of our new stochastic collocation minimization approach.

Key words. stochastic collocation, optimal control, stochastic partial differential equations

AMS subject classifications. 60H15, 60H35, 35Q93

DOI. 10.1137/110835438

1. Introduction. The modeling, simulation, and optimization with partial differential equations (PDE) has become an important research field in applied mathematics. It is driven by the engineering disciplines and the desire to observe, analyze, and optimize physical phenomena without performing real physical experiments, or to predict future states of a physical system. Mostly for practically relevant simulations and optimizations, the parameters, coefficients, or boundary conditions of the PDE models are based on measurement data. Of course these measurements carry errors; thus for many applications not only the measured data but also its errors should be taken into account for further processing. In particular, for biological systems there is an intrinsic heterogeneity within and also between objects or species under consideration. This poses a particular problem if models which are built using a certain set of measurements are applied to new objects or species which are of the same type but for which measurements cannot be obtained. An example of this is the patient-specific simulation and optimization of cancer treatment therapies [1] in which the biophysical properties of the tissue are uncertain.

Uncertainty and errors in parameters, coefficients, and boundary and initial conditions are often modeled as random processes. Thereby the original differential

---

†Fraunhofer Institute for Medical Image Computing MEVIS, 28359 Bremen, Germany, and School of Engineering and Science, Jacobs University Bremen, 28759 Bremen, Germany (hanne.tiesler@mevis.fraunhofer.de, tobias.preusser@mevis.fraunhofer.de).

‡School of Computing, University of Utah, Salt Lake City, UT 84112 (kirby@cs.utah.edu).

§Department of Mathematics, Purdue University, West Lafayette, IN 47907 (dxiu@purdue.edu).

equations are transformed into stochastic ordinary or partial differential equations (SODE/SPDE). For equations involving randomness, the classical approach is the Monte Carlo (MC) method, which generates samples of the investigated states from random realizations of the input data. Although the convergence rate of the MC method is slow, it is a very robust approach, which moreover is independent of the stochastic dimensionality.

The main stochastic theoretical underpinning of our work presented herein is generally referred to as *generalized polynomial chaos* (gPC). Based on the Wiener–Hermite polynomial chaos expansion [27], gPC seeks to approximate second-order random processes by a finite linear combination of stochastic basis functions. Once one has chosen an approximation space of the random process of interest, a solution within that space can be found by solving the SPDE of interest in the weak form. Because of its analogy with the classic Galerkin method as employed in the finite element method (FEM), this methodology is often referred to as the generalized polynomial chaos–stochastic Galerkin (gPC-SG) method. It has been applied as a method for uncertainty quantification in the field of computational mechanics for a number of years and has recently seen a revival of interest [18, 8, 9, 17, 10, 15, 29, 31].

Although the SG method provides a solid mathematical framework from which one can do analysis and derive algorithms, it is not always the most computationally efficient means of solving large problems. Nor is it the case that one always has the freedom to rearchitect their currently available deterministic solver to employ gPC-SG. To address these issues, the stochastic collocation (SC) method was developed [28], combining the advantages of the MC approach (as a sampling methodology) with the idea of there being an underlying approximation representation of the stochastic process (gPC). The gPC–SC method evaluates solutions of a stochastic system at carefully chosen points within the random space to compute accurate statistics with significantly fewer solution evaluations than MC methods.

In this work, we will employ the gPC–SC approach specifically for these reasons in order to perform efficient optimization in the presence of uncertainty. In particular we will develop a gradient descent and a sequential quadratic programming (SQP) method for optimal control problems under constrained SPDEs. Our algorithms are based on the weak formulation for SPDEs, which involves the expectation of the classical weak form of the corresponding deterministic PDEs. Since the expectation of a random process is nothing other than integration of the process with respect to the corresponding probability measure, the derivation of the optimality system is analogous to the deterministic case. Moreover, for the development of the optimization algorithms we take advantage of the fact that the expectation of smooth random processes can be evaluated very conveniently with the SC method. Thus, we obtain methods which can be implemented by using components of our existing deterministic optimization code that must be interwoven with the evaluation of expectations in order to get descent directions and updates of iterates.

For a long time scientists have been working in the field of PDE constrained optimization. References and efforts are too numerous to be briefly summarized here. An overview of the subject can be found in, e.g., [5]. However, related work on the optimization and the optimal control of SPDE constrained problems is sparse. Most publications in this area discuss theoretical results. Almost no work exists on the design of efficient algorithms for the solution of the optimality systems. Øksendal [22] considers the optimal control of a quasi-linear stochastic heat equation and proves verification theorems of maximum principle type. Mortensen [19, 20] considers dynamic programming (Bellman algorithm) for SPDEs and noisy observations of the

states of the system. The regularity of the value function and the connection between viscosity solutions of the Bellman equation and the value function is studied by Nisio [21]. Cao [7] solves stochastic optimization using the MC approach. The estimation of parameters in the presence of noisy measurements has been treated with the Bayesian inference approach, which uses known information about the parameters to create a priori distribution [25, 26]. First approaches to stochastic inverse problems are presented by Narayanan and Zabaras in [3]. In [32] Zabaras and Ganepathysubramanian use a gradient descent method to solve the stochastic inverse heat equation with the polynomial chaos methodology.

Paper organization. In section 2 we give a short overview of the theory of SPDEs, the SG method, and the SC approach. For a general second-order elliptic SPDE we focus on the optimization with SPDE constraint in section 3, thereby deriving the optimality systems as well as focusing on a gradient descent and an SQP method. We also discuss particular objective functions which involve the stochastic moments of the states of the system. In section 4 we focus on an exemplary diffusion equation with uncertain diffusion coefficient and a tracking-type function. Our numerical results show the influence of the stochastic moments within the objective function. Conclusions are drawn in section 5.

2. Stochastic partial differential equations. Let $(\Omega, \mathcal{A}, \pi)$ be a complete probability space, where Ω denotes the event space, let $\mathcal{A} \subset \mathcal{P}(\Omega)$ be the σ-algebra of subsets of Ω, and let π be the probability measure. Following the theory of Wiener [27], Cameron and Martin [6], as well as Xiu and Karniadakis [30], we can represent any general second-order *random process* $X(\omega), \omega \in \Omega$, in terms of a collection of finitely many random variables $\boldsymbol{\xi} = \boldsymbol{\xi}(\omega) = (\xi_1(\omega), \ldots, \xi_N(\omega))$ with independent components. Let $\rho_i : \Gamma_i \to \mathbb{R}^+$ be the probability density functions (PDFs) of the random variables $\xi_i(\omega), \omega \in \Omega$, and let their images $\Gamma_i \equiv \xi_i(\Omega) \in \mathbb{R}$ be intervals in $\mathbb{R}$ for $i = 1, \ldots, N$. Then

$$\rho(\boldsymbol{\xi}) = \prod_{i=1}^{N} \rho_i(\xi_i) \ \ \forall \boldsymbol{\xi} \in \Gamma$$

is the joint PDF of the random vector $\boldsymbol{\xi}$ with the support $\Gamma = \prod_{i=1}^{N} \Gamma_i \subset \mathbb{R}^N$. On Γ we have the probability measure $\rho(\boldsymbol{\xi}) d\boldsymbol{\xi}$.

As commented in [30], this allows us to conduct numerical formulations in the finite dimensional (N-dimensional) random space Γ. Let us denote $L^2(\Gamma)$ as the probabilistic Hilbert space [16] in which the random processes based upon the random variables $\boldsymbol{\xi}$ reside. The inner product of this Hilbert space is given by

$$\langle X, Y \rangle = \int_{\Gamma} (X \cdot Y) \, \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} \qquad \text{for } X, Y \in L^2(\Gamma),$$

where we have exploited independence of the random variables to allow us to write the measure as product of measures in each stochastic direction. We similarly define the expectation of a random process $X \in L^2(\Gamma)$ as

$$(2.1) \qquad \mathbb{E}[X(\boldsymbol{\xi})] = \int_{\Gamma} X(\boldsymbol{\xi}) \, \rho(\boldsymbol{\xi}) d\boldsymbol{\xi},$$

and we refer to the expectation of the powers $\mathbb{E}[X^i(\boldsymbol{\xi})]$ as the ith moment of the random process.

Additionally considering a spatial domain $D \subset \mathbb{R}^d$, we define a set of random processes, which are indexed by the spatial position $x \in D$; i.e., we consider mappings

$$f : D \times \Gamma \to \mathbb{R}.$$

Such a set of processes is referred to as a *random field* [13], which can also be interpreted as a function-valued random variable, because for every $\boldsymbol{\xi} \in \Gamma$ the *realization* $f(\cdot, \boldsymbol{\xi}) : D \to \mathbb{R}$ is a real valued function on D. For a vector-space Y the class $L^2(\Gamma) \otimes Y$ denotes the space of random fields whose realizations lie in Y for a.e. (almost every) $\boldsymbol{\xi} \in \Gamma$. If Y is a Banach space, a norm on $L^2(\Gamma) \otimes Y$ is induced by $\|\|f(x, \boldsymbol{\xi})\|\|^2 = \mathbb{E}[\|f(x, \boldsymbol{\xi})\|_Y^2]$; for example, on $L^2(D) \otimes L^2(\Gamma)$ we have

$$\|\|f(\boldsymbol{\xi}, x)\|\|^2 = \mathbb{E}[\|f(\boldsymbol{\xi}, x)\|_{L^2(D)}^2] = \int_\Gamma \int_D \left(f(\boldsymbol{\xi}, x) \right)^2 dx \, \rho(\boldsymbol{\xi}) d\boldsymbol{\xi};$$

that is, $\|\|\cdot\|\|$ denotes the expected value of the L^2-norm of the function f.

In the following we consider elliptic SPDEs of the type

$$(2.2) \qquad \begin{aligned} -\mathrm{div}\,(a(\boldsymbol{\xi}, x)\nabla y(\boldsymbol{\xi}, x)) + c_0(\boldsymbol{\xi}, x)y(\boldsymbol{\xi}, x) &= \beta(\boldsymbol{\xi}, x)u(\boldsymbol{\xi}, x) &&\text{in } D \text{ a.e. } \boldsymbol{\xi} \in \Gamma, \\ y(\boldsymbol{\xi}, x) &= g(\boldsymbol{\xi}, x) &&\text{on } \partial D \text{ a.e. } \boldsymbol{\xi} \in \Gamma, \end{aligned}$$

where $a, c_0, \beta \in L^2(\Gamma) \otimes L^\infty(D)$, $g \in L^2(\Gamma) \otimes L^2(\partial D)$ are given random fields. Here and in the following we consider stochastic controls $u \in L^2(\Gamma) \otimes L^2(D)$. In some of the subsequent examples we will reduce u to a deterministic control $u \in L^2(D)$, which is also relevant for some practical applications.

The existence of solutions to (2.2) has been shown by Vage in [24] in the case when there exist constants $c_1, c_2 > 0$ such that a and c_0 fulfill

$$(\mathbb{E}[a]\nabla v, \nabla v)_{L^2(D)} \geq c_1 \|\nabla v\|_{L^2(D)}^2 \qquad \forall v \in H^1(D)$$

and

$$(\mathbb{E}[c_0]v, v)_{L^2(D)} \geq c_2 \|v\|_{L^2(D)}^2 \qquad \forall v \in L^2(D).$$

These conditions assert the coercivity of the corresponding bilinear form, and thus guarantee the existence of solutions through the Lax–Milgram lemma. For the case of a vanishing Helmholtz term in (2.2), i.e., for $c_0 \equiv 0$, Babuška, Tempone, and Zouraris [2] show the existence of solutions if the classical ellipticity condition is guaranteed almost everywhere in Γ, i.e., if

$$(a(\boldsymbol{\xi}, x)\nabla v, \nabla v)_{L^2(D)} \geq c_1 \|v\|_{H^1(D)}^2 \qquad \forall v \in H^1(D), \text{ a.e. } \Gamma.$$

Various variants and specializations of (2.2) have been studied, including different boundary conditions, as well as first-order derivatives. For an overview we refer the reader to [13].

The SPDE (2.2) is a model for the steady state of (heat) diffusion

$$(2.3) \qquad \begin{aligned} -\mathrm{div}\,(a(\boldsymbol{\xi}, x)\nabla y(\boldsymbol{\xi}, x)) &= \beta(\boldsymbol{\xi}, x)u(\boldsymbol{\xi}, x) &&\text{in } D \text{ a.e. } \boldsymbol{\xi} \in \Gamma, \\ y(\boldsymbol{\xi}, x) &= 0 &&\text{on } \partial D \text{ a.e. } \boldsymbol{\xi} \in \Gamma, \end{aligned}$$

where the diffusion tensor a represents the thermal conductivity, and the control u being the right-hand side of (2.3) models heat sources and sinks. In this model, the

diffusion tensor a and the coefficient β are uncertain. The actual form of a and β and their statistics or PDFs depend on the problem, as they can result from noisy or error-prone experimental measurements or other a priori data. We emphasize that although we work with a scalar diffusivity a, it is straightforward to transport the material presented in this paper to a symmetric and positive definite diffusion matrix (tensor). Finally, the random field $\beta \in L^2(\Gamma) \otimes L^\infty(D)$ can also be used to restrict the heat source u to a subdomain of D.

The notion of the weak solution for SPDEs of type (2.2), which have divergence form, is based on an extension of the classical theory: Test functions become random fields, and an integration over the stochastic space is done with respect to the corresponding measure. Thus the weak form involves expectations of the weak problem formulation in the physical space. Correspondingly, $y \in L^2(\Gamma) \otimes H_0^1(D)$ is a weak solution of (2.3) if for all test functions $v \in L^2(\Gamma) \otimes H_0^1(D)$,

$$(2.4) \qquad \mathbb{E}\left[\int_D a(\boldsymbol{\xi}, x) \nabla y(\boldsymbol{\xi}, x) \cdot \nabla v(\boldsymbol{\xi}, x) \, dx \right] = \mathbb{E}\left[\int_D \beta(\boldsymbol{\xi}, x) u(\boldsymbol{\xi}, x) v(\boldsymbol{\xi}, x) \, dx \right].$$

2.1. Monte Carlo simulation. A classical and very popular approach for the numerical treatment of (2.2) of the form of (2.3) is the Monte Carlo (MC) method. For the MC method, Q realizations $\boldsymbol{\xi}_j := (\xi_j^i)_{i=1}^N, j = 1, \ldots, Q$, of the vector of random variables $\boldsymbol{\xi}$ are generated. Consequently, Q deterministic problems are solved, which are obtained from (2.3) by considering the realizations of A, c_0, and β corresponding to $\boldsymbol{\xi}_j$. Finally, the statistics of the solution samples $y_j = y(\boldsymbol{\xi}_j, x)$ lead to the desired result. On the one hand, the MC approach is extremely robust and requires no assumptions on the smoothness of the underlying stochastic processes. On the other hand, the convergence is very slow and goes asymptotically with $1/\sqrt{Q}$.

2.2. The stochastic Galerkin method. The SG method utilizes the weak formulation (2.4) on finite-dimensional stochastic subspaces $P(\Gamma) \subset L^2(\Gamma)$. The approximating subspaces can be constructed by, for example, the gPC approach [11, 29, 30, 31]. Thereby any second-order stochastic process $X \in L^2(\Gamma)$ is expressed by a weighted (infinite) sum of polynomials L_i, which are functions of the vector of random variables $\boldsymbol{\xi} = (\xi^1, \ldots, \xi^N)$ of known PDF $\rho(\boldsymbol{\xi})$, and thus

$$(2.5) \qquad X(\boldsymbol{\xi}) = \sum_{i=0}^\infty \hat{X}_i L_i(\boldsymbol{\xi}), \qquad \hat{X}_i \in \mathbb{R}.$$

An approximation of a stochastic process is obtained by truncating the infinite expansion (2.5) to $Q + 1$ terms for some $Q \in \mathbb{N}$. Thus, we can write our approximation $\tilde{X}$ of $X \in L^2(\Gamma)$ as
(2.6)

$$X(\boldsymbol{\xi}) \approx \tilde{X}(\boldsymbol{\xi}) := \sum_{i=0}^Q \hat{X}_i L_i(\boldsymbol{\xi}), \quad \text{with } \hat{X}_i = \left(\int_\Gamma L_i^2(\boldsymbol{\xi}) \, \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} \right)^{-1} \int_\Gamma X(\boldsymbol{\xi}) L_i(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) \, d\boldsymbol{\xi},$$

where the coefficients $\hat{X}_i$ result from a projection of the stochastic process X onto the space spanned by the polynomials. The statistics of processes represented is evaluated as the stochastic moments of $\tilde{X}$, which can be computed by integration of the expansion (2.6) over the stochastic space (cf. (2.1)), thus involving integrals of the L_i and linear combinations of the coefficients $\hat{X}_i$.

The linear system resulting from the approximation of processes in the weak form (2.4) becomes sparse in the stochastic dimension if the L_i are orthogonal with
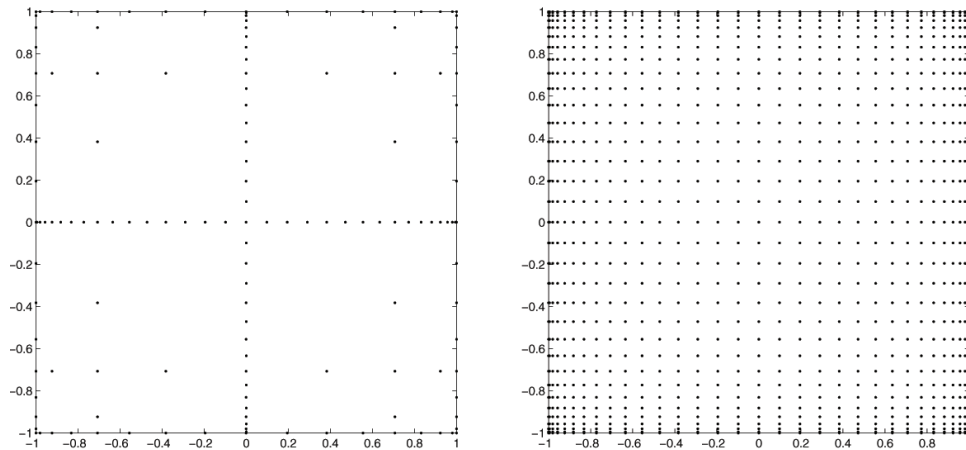
FIG. 2.1. *A comparison between a sparse grid constructed by Smolyak's algorithms (left,* 145 *grid points) and a full tensor product grid (right,* 1089 *grid points) is shown. The interpolation nodes are based on the extrema of the Chebyshev polynomials.*

respect to the corresponding measure $\rho(\boldsymbol{\xi})d\boldsymbol{\xi}$. For example, in the case of a uniformly distributed single random variable ξ, Legendre polynomials are used; if ξ is Gaussian, Hermite polynomials will be the choice, etc. [13]. Basis functions for higher stochastic dimensions result from tensor products of the single variable orthogonal polynomials.

2.3. The stochastic collocation method. An approach related to the SG method is the generalized polynomial chaos–stochastic collocation (gPC–SC) method [28], in which we employ an interpolation of the stochastic process in some polynomial space P for accomplishing integrations over the stochastic domain (as, for example, in (2.1), (2.4), and (2.6)). Analogous to classical polynomial interpolation is the use of Lagrange interpolation, which is particularly convenient: Given a set of collocation points $\{\boldsymbol{\xi}_j\}_{j=0}^Q$ and corresponding characteristic Lagrange polynomials l_i obeying $l_i(\boldsymbol{\xi}_j) = \delta_{ij}$, a process $X \in L^2(\Gamma)$ is interpolated by

$$(2.7) \qquad X(\boldsymbol{\xi}) \approx \tilde{X}(\boldsymbol{\xi}) := \sum_{j=0}^Q \hat{X}_j l_j(\boldsymbol{\xi}), \quad \text{with } \hat{X}_j = X(\boldsymbol{\xi}_j) \in \mathbb{R}.$$

Thus the interpolation involves evaluations of the stochastic process at the sampling points $\boldsymbol{\xi}_j$, and thus is much more convenient than (2.6). Moreover, the interpolation is constructed such that the residual $R(X(\boldsymbol{\xi}) - \tilde{X}(\boldsymbol{\xi}))$ between the interpolated system $\tilde{X}$ and the true process X is zero at the collocation points. Note that each collocation point $\boldsymbol{\xi}_j$ is a vector of sampling points for the N random variables, i.e., $\boldsymbol{\xi}_j = (\xi_j^1, \ldots, \xi_j^N)$.

For the choice of the collocation points $\boldsymbol{\xi}_j$ it has become popular to use points which lie on a sparse grid in the stochastic space (cf. Figure 2.1) generated by Smolyak's algorithm [28]. For high stochastic dimensions N the Smolyak sparse grids have significantly fewer points than the full tensor product grid, but the order of convergence is reduced only by some logarithmic factor. The accuracy of the collocation approach in multidimensions is often discussed in terms of the so-called *level k* of integration (a term which is related to the space of functions which is integrated exactly). The level k is the maximal degree of the polynomials used for the interpolation. The

number of nodes used in the collocation approach is approximately $\frac{2^k}{k!}N^k$ for $N \gg 1$, which is 2^k times larger than the dimension of the space of polynomials of total degree at most k. This means that the number of nodes in the Smolyak algorithm overruns the number of expansion terms in the SG method. On the other hand, the factor 2^k is independent of the stochastic dimension N, which means that the algorithm may be considered as optimal. For more details on sparse grids used in the gPC–SC approach, we refer the reader to [28] and [4]. In the following we report the level k and number of quadrature points in accordance with [28].

With the interpolation (2.7) we can approximate the expectation of the process by

$$(2.8) \qquad \mathbb{E}[X(\boldsymbol{\xi})] \approx \mathbb{E}[\tilde{X}(\boldsymbol{\xi})] = \sum_{j=0}^{Q} \omega_j \hat{X}_j, \qquad \text{with } \omega_j = \int_\Gamma l_j(\boldsymbol{\xi})\rho(\boldsymbol{\xi})d\boldsymbol{\xi}.$$

For an evaluation of the higher-order centered moments we consider $\mathbb{E}\big[\big(X(\boldsymbol{\xi}) - \mathbb{E}[X]\big)^m\big]$ for some $m \geq 2$. Although using the interpolation $\tilde{X}$ in this formula involves powers of $\tilde{X}$, it is popular to use a polynomial approximation of the same degree as $\tilde{X}$, and thus

$$(2.9) \qquad \mathbb{E}\big[\big(X(\boldsymbol{\xi}) - \mathbb{E}[X]\big)^m\big] \approx \mathbb{E}\Big[\big(\tilde{X}(\boldsymbol{\xi}) - \mathbb{E}[\tilde{X}]\big)^m\Big] \approx \sum_{j=0}^{Q} \omega_j\big(\hat{X}_j - \mathbb{E}[\tilde{X}]\big)^m.$$

Any of the above presented stochastic discretization techniques can be combined with a spatial discretization of choice for the solution of (2.2). The combination of the SG method with a FEM in the physical space D is investigated in, e.g., [10]. These discretizations involve the canonical generalizations of the above approximations of stochastic processes to random fields; i.e., a random field $f \in L^2(\Gamma) \otimes H^1(D)$ is approximated by

$$\tilde{f}(\boldsymbol{\xi}, x) = \sum_{j=0}^{Q} \hat{f}_j(x)l_j(\boldsymbol{\xi}), \quad \text{with } \hat{f}_j(x) = f(\boldsymbol{\xi}_j, x) \in H^1(D).$$

Thus, the solution of (2.2) with gPC–SC involves the evaluation of the solution y at the nodes $\boldsymbol{\xi}_j$, which corresponds to solving $Q + 1$ deterministic counterparts of (2.2) at each $\boldsymbol{\xi}$. In this sense the SC method has the same sampling character as the MC approach; however, in gPC–SC we assume smoothness of the underlying process in order to gain enhanced convergence properties. In contrast to the SG approach, the equations which have to be solved are decoupled and, furthermore, existing deterministic code can be reused easily.

Let us finally note that the collocation approach discussed above is one of multiple possible collocation methods used in the solution of SPDEs [13]. With increasing rate of convergence (and increasing assumptions on smoothness) one can use approaches based on, for example, the classical MC method, the Newton–Cotes formulas, Gaussian or Clenshaw–Curtis quadrature, or the sparse-grid spectral collocation method discussed here.

3. Optimization with SPDE constraints. We will now apply the gPC–SC methodology reviewed above to the solution of optimal control problems which are

constrained by an SPDE. To this end consider the optimal control problem

$$
\begin{aligned}
\min_{(y,u)} \ & F(y,u) \\
\text{subject to} \quad & h(y,u) = 0, \\
& u \in U_{\mathrm{ad}}
\end{aligned}
$$

(SCP)

with a convex and continuous function F and a nonempty, bounded, closed, and convex set $U_{\mathrm{ad}} \subset L^2(\Gamma) \otimes L^2(D)$. The objective function

$$
F(y,u) : L^2(\Gamma) \otimes H^1(D) \times L^2(\Gamma) \otimes L^2(D) \longrightarrow \mathbb{R}
$$

depends on the state y being a random field in $L^2(\Gamma) \otimes H^1(D)$ and the random field control $u \in U_{\mathrm{ad}} \subset L^2(\Gamma) \otimes L^2(D)$. Here we assume that the SPDE constraint denoted by $h : L^2(\Gamma) \otimes H^1(D) \times L^2(\Gamma) \otimes L^2(D) \to (L^2(\Gamma) \otimes H^1(D))^*$ is of the type (2.2) and that the condition $h(y,u) = 0$ is evaluated in a distributional sense. Since $L^2(\Gamma) \otimes H^1(D)$ as well as $L^2(\Gamma) \otimes L^2(D)$ are Hilbert spaces, the results from deterministic optimization theory (see, e.g., Chapter 2 in [23]) can be applied to the stochastic problem. Therewith, we can show that there exists an optimal solution for the problem (SCP) above.

In the following our aim is to derive the optimality system for (SCP). Note that extensions of the following derivations and algorithms to multiple constraints is straightforward. For ease of presentation we will denote the elliptic differential operator in (2.2) by $Ay(\boldsymbol{\xi}, x) := -\mathrm{div}(a(\boldsymbol{\xi}, x) \nabla y(\boldsymbol{\xi}, x))$, and for the dual pairing in $L^2(\Gamma) \otimes L^2(D)$ we will write $\langle \cdot, \cdot \rangle$.

For the definition of the Lagrange function

$$
\mathcal{L} : L^2(\Gamma) \otimes H^1(D) \times L^2(\Gamma) \otimes L^2(D) \times L^2(\Gamma) \otimes H^1(D) \longrightarrow \mathbb{R},
$$

we use a random field $\mu(\boldsymbol{\xi}, x) \in L^2(\Gamma) \otimes H^1(D)$ as the Lagrange multiplier in the weak form of (2.2). Thus, we have

(3.1)
$$
\begin{aligned}
\mathcal{L}(y,u,\mu) & := F(y,u) - \langle h(y,u), \mu \rangle \\
& = F(y,u) - \mathbb{E}\left[\int_D (Ay + c_0 y - \beta u)\,\mu\,dx - \int_{\partial D} (y - g)\mu\,dx \right].
\end{aligned}
$$

Analogous to the optimization under deterministic PDE constraints, the adjoint equation can be calculated by the derivative of the Lagrangian with respect to y,

(3.2) $\qquad \mathcal{L}_y(y,u,\mu)(w) = \langle F_y(y,u) - h_y(y,u)^* \mu, w \rangle = 0 \quad \forall w \in L^2(\Gamma) \otimes L^2(D),$

which leads to

$$
h_y(y,u)^* \mu = F_y(y,u),
$$

where * denotes the adjoint operator, and the adjoint state is μ. Moreover, for the variational inequality obtained from the derivation of the Lagrangian with respect to u, we need

$$
\begin{aligned}
\mathcal{L}_u(y,u,\mu)(v - u) & = \langle F_u(y,u) - h_u(y,u)^* \mu, v - u \rangle \\
& = \langle F_u(y,u), v - u \rangle + \mathbb{E}\left[\int_D \beta\mu(v - u)dx \right].
\end{aligned}
$$

Hence, a control $u \in L^2(\Gamma) \otimes L^2(D)$ is a solution for (SCP) if and only if u fulfills the following optimality system:

$$u \in U_{\mathrm{ad}},$$

$$\langle F_u(y, u) - h_u(y, u)^* \mu, (v - u) \rangle \geq 0 \quad \forall v \in U_{\mathrm{ad}},$$

(3.3)

$$
\begin{aligned}
Ay + c_0 y &= \beta u, & A^* \mu + c_0 \mu &= F_y(y, u) & \text{in } D \text{ a.e. } \boldsymbol{\xi} \in \Gamma, \\
y &= g, & \mu &= 0 & \text{on } \partial D \text{ a.e. } \boldsymbol{\xi} \in \Gamma.
\end{aligned}
$$

In the following two sections we describe a gradient descent method and an SQP method for the solution of (3.3), which we will discretize using the gPC–SC method introduced in section 2.3.

3.1. The gradient descent method. Gradient descent methods are among the most intuitive and simplest methods used in solving optimal control problems. Starting from an initial guess, an iteration with updates in the direction of the steepest descent of the objective function is performed until a local minimum is reached. Let us denote the iterates of the gradients with $u^{(n)}$ for $n \in \mathbb{N}$, and thus $u^{(0)}$ is the initial guess. The state and the adjoint state corresponding to the control iterate at step n are denoted by $y^{(n)}$ and $\mu^{(n)}$, respectively.

Let us assume that we have initialized the method by choosing $u^{(0)} \in L^2(\Gamma) \otimes L^2(D)$ appropriately and by setting $n = 0$. In the following we describe one iteration step of the gradient projection method for (3.3) in terms of the collocation method from section 2.3. In section 4.3 we will apply the abstract algorithm to a concrete problem of the form (2.3) as follows:

1. *State and adjoint state.* For each collocation point $\boldsymbol{\xi}_j$ solve the deterministic counterparts of the SPDEs in (3.3) to obtain the state $y_j^{(n)}$ and the adjoint state $\mu_j^{(n)}$ for the actual control $u_j^{(n)}$.

2. *Search direction.* Taking the samples $y_j^{(n)}$ and $\mu_j^{(n)}$ of the current state and adjoint state, respectively, we can use the polynomial expansion (2.7) and the quadrature formula (2.8) to evaluate the descent direction v_n, i.e., the antigradient

$$v^{(n)} := -F_u(y^{(n)}, u^{(n)}) - \beta \, \mu^{(n)} \in L^2(\Gamma) \otimes L^2(D).$$

3. *Step size control.* As in the deterministic case, the optimal step size s_n is determined by

$$F\left(y^{(n)}, P_{U_{\mathrm{ad}}}\left(u^{(n)} + s_n v^{(n)}\right)\right) = \min_{s > 0} F\left(y^{(n)}, P_{U_{\mathrm{ad}}}\left(u^{(n)} + s v^{(n)}\right)\right),$$

where $P_{U_{\mathrm{ad}}}$ is the projection onto the space of admissible controls. To actually compute the optimal step width, there exists a variety of approaches such as the bisection method or Armijo's rule.

4. *Update.* Finally, the control iterate is updated according to $u^{(n+1)} := P_{U_{\mathrm{ad}}}(u^{(n)} + s_n v^{(n)})$, and n is incremented.

As usual, this iteration is continued until $\|u^{(n+1)} - u^{(n)}\| < \varepsilon$ for an appropriately chosen $\varepsilon > 0$.

Note that in an implementation it is not necessary to store all states and adjoint states in step 1 if F_u does not depend on y. In fact, in this situation, each sample $y_i^{(n)}$ can be discarded after the right-hand side of the adjoint equation has been computed.

Moreover, the samples of μ can be discarded after they have been accumulated for the computation of the antigradient in step 2.

If a deterministic control $u \in L^2(D)$ is used, the stochasticity of the problem is reduced in the computation of the descent direction (step 2). The search direction becomes a deterministic function, as we use the following term instead of the stochastic function above:

$$v^{(n)} := -F_u(y^{(n)}, u^{(n)}) - \mathbb{E}[\beta \, \mu^{(n)}] \in L^2(D).$$

This simplifies the calculations such that we have an algorithm similar to the deterministic case. In fact this is due to the involvement of the expectation in the weak formulation. For a stochastic control as described above, the stochasticity is transferred to the descent direction through the derivative F_u, and consequently the implementation of the gradient descent is more involved.

3.2. Sequential quadratic programming. In the following we will derive a sequential quadratic program for the SPDE constrained optimal control problem (SCP). Analogous to a classical SQP method with deterministic constraints, the goal is to fulfill the *Karush–Kuhn–Tucker* (KKT) conditions and in particular the vanishing of the derivative of the Lagrangian (3.1). Thus, we are looking for points $(\bar{y}, \bar{u}, \bar{\mu})$ such that $\nabla \mathcal{L}(\bar{y}, \bar{u}, \bar{\mu}) = (\mathcal{L}_y, \mathcal{L}_u, \mathcal{L}_\mu)(\bar{y}, \bar{u}, \bar{\mu}) = 0$. In SQP we use Newton's method to find this desired root of $\nabla \mathcal{L}$. In the following we describe one step of the SQP method for (SCP).

Again let us assume that we have initialized the method appropriately by choosing $u^{(0)} \in L^2(\Gamma) \otimes L^2(D)$ and $y^{(0)}, \mu^{(0)} \in L^2(\Gamma) \otimes H^1(D)$ and by setting $n = 0$. Then one step of the SQP method is as follows:

1. *Check KKT conditions.* Check whether $\nabla \mathcal{L} = 0$ is fulfilled, i.e., whether $\nabla \mathcal{L}(y(\boldsymbol{\xi}_j), u(\boldsymbol{\xi}_j), \mu(\boldsymbol{\xi}_j)) = 0$ for every collocation point $\boldsymbol{\xi}_j$. If this is the case, then stop. This involves checking (3.2) by evaluating the expected value as described in (2.8).

2. *Solve the stochastic quadratic program.* Calculate the KKT-point $(\delta y^{(n)}, \delta u^{(n)}, \mu^{(n+1)}) \in L^2(\Gamma) \otimes H^1(D) \times L^2(\Gamma) \otimes L^2(D) \times L^2(\Gamma) \otimes H^1(D)$ as the solution of the stochastic quadratic problem
   (3.4)

$$\min_{(y,u)} \nabla F(y_n, u_n)^T \cdot \begin{pmatrix} \delta y^{(n)} \\ \delta u^{(n)} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \delta y^{(n)} \\ \delta u^{(n)} \end{pmatrix}^T \cdot \nabla^2 \mathcal{L}(y^{(n)}, u^{(n)}, \mu^{(n)}) \begin{pmatrix} \delta y^{(n)} \\ \delta u^{(n)} \end{pmatrix}$$

$$\text{subject to} \quad h(y^{(n)}, u^{(n)}) + \nabla h(y^{(n)}, u^{(n)})^T \cdot \begin{pmatrix} \delta y^{(n)} \\ \delta u^{(n)} \end{pmatrix} = 0.$$

   Here, the stochasticity appears in the optimization vector $(\delta y^{(n)}, \delta u^{(n)})^T = (y - y^{(n)}, u - u^{(n)})^T$ with functions $y, y^{(n)} \in L^2(\Gamma) \otimes H^1(D)$ and $u, u^{(n)} \in L^2(\Gamma) \otimes L^2(D)$. Using SC we can split the stochastic problem into a set of independent deterministic problems which can be solved with classical deterministic methods. In the case of a linear SPDE as treated here the constraint is given by our SPDE (2.2). This quadratic problem can be solved, for example, with an inner point method or an active set method. In this paper we use a conjugate gradient method.

3. *Choice of weight parameter.* We define the weighting function $\eta^{(n+1)} \in \mathbb{R}$ according to

$$\eta^{(n+1)} := \max \left\{ \eta^{(n)}, \mathbb{E}[\|\mu^{(n+1)}\|_\infty] + \varepsilon \right\}.$$

Thus, we use the expectation of the maximum values of μ in the physical domain to determine a real valued weight $\eta^{(n+1)} \in \mathbb{R}$ for the step size control. Another choice would be a stochastically distributed weight $\eta \in L^2(\Gamma)$, which means we use all μ instead of the expectation. In turn, the stochastic weight could be reduced to a real value by choosing finally the expectation $\mathbb{E}[\eta(\boldsymbol{\xi})]$ as the weighting function for the step size control.

4. *Step size control.* For a chosen real valued penalty function φ (see below) and some $\gamma \in (0,1), \sigma \in (0,1)$, choose the step size

$$\alpha_n = \max \big\{ \alpha = \gamma^j \, \big| \, j \in \mathbb{N}, \ \varphi(y_n + \alpha \delta_y^n, u_n + \alpha \delta_u^n; \eta)$$
$$\leq \varphi(y_n, u_n; \eta) + \sigma \alpha \nabla \varphi(y_n, u_n; \eta)(\delta_y^n, \delta_u^n) \big\}.$$

5. *Update.* The new iterates for the state and the control are given by

$$y_{n+1} := y_n + \alpha_n \delta_y^n, \quad u_{n+1} := u_n + \alpha_n \delta_u^n.$$

Finally, we increment n and continue with step 1.

As in the classical deterministic case we have various choices for the penalty function φ. However, in our setting we need to reduce the stochasticity in order to obtain real valued functions; thus our penalty functions will always involve an expectation $\mathbb{E}[h]$ of the constraint. For example, we can use the L_1-penalty function as well as the augmented Lagrangian $\mathcal{L}_a$,

$$(3.5) \qquad L_1(y, u; \eta) := F(y, u) + \eta \mathbb{E}\big[\|\mathcal{A}y + c_0 y - \beta u\|_{L^1(D)}\big],$$
$$\mathcal{L}_a(y, u, \mu; \eta) := \mathcal{L}(y, u, \mu) - \frac{\eta}{2}\|\mathcal{A}y + c_0 y - \beta u\|.$$

In our implementations we use the L_1-penalty function.

4. Optimal SPDE control problem with tracking-type objective functions. Having addressed an arbitrary objective function, let us now be more descriptive by discussing in more detail the SPDE constraint (2.3) and possible choices for objective functions. First, we discuss possible objective functions of tracking type, and in the following sections we derive the optimality system and show numerical results for two of the tracking-type functionals discussed herein.

4.1. Exemplary objective functionals.

Fitting of the expected temperature. Let us assume that we have measurements of the temperatures in the physical domain, resulting from an experiment. We interpret the measurement data $Y_1 \in L^2(D)$ as the expected temperature, to which we would like to match our calculated states as well as possible. An objective function for this task is

$$(4.1) \qquad F_E(y, u) := \frac{\lambda_1}{2}\|y - Y_1\|^2 + \sum_{k=2}^q \frac{\lambda_k}{2}\mathbb{E}\left[\left(\|y\|_{L^2(D)}^2\right)^k\right] + \mathcal{R}(u)$$

for some $k \in \mathbb{N}$, $\lambda_i \geq 0$, and where $\mathcal{R}(u)$ is a regularization of the control u, e.g., $\mathcal{R}(u) := \lambda_u \|u\|_{L^2(D)}^2$, with $0 \leq \lambda_u \in \mathbb{R}$. In this objective function the first term measures the expectation of the L^2-difference of the measured data and the calculated temperature. In the second term the higher-order moments of the calculated temperature are measured. These terms act as regularizations in the stochastic space, since they prefer temperatures with small variance or small higher-order moments, depending on the choice of the λ_k. In section 4.3 we show some numerical results for this type of objective function.

Fitting of stochastic moments. In addition we can consider the case that we have multiple experimental measurements $Y^{(k)} \in L^2(D)$, which represent the physical temperature distribution under investigation. From these samples we can evaluate the statistics of the experiments, thus leading to $Y_k \in L^2(D)$, $k = 1, \ldots, q$, which represent the expected distribution of the temperature as well as higher-order moments of the experimental temperature distribution. To these we would like to fit the statistics of the calculated temperature distribution. Thus, we define (cf. also [32])

$$(4.2) \qquad F_M(y, u) := \sum_{k=1}^{q} \frac{\lambda_k}{2} \left\| \mathbb{E}\left[y^k\right] - Y_k \right\|_{L^2(D)}^2 + \mathcal{R}(u),$$

where again $\lambda_k \geq 0$ for $k = 1, \ldots, q$ and $\mathcal{R}(u)$ is a regularization of the control u (see above). This objective function penalizes deviations of the moments of the computed temperature distribution from the ones of the measured data. Thus it can be seen as a functional that penalizes deviations of the statistics of the calculated temperature from the statistics of the measured temperature. In section 4.3 we show numerical results for this type of objective function.

Fitting of probability density function. Finally, if we have even more information about the physical process under investigation, we can also match the PDF of the calculated and the measured temperature distribution. However, to simplify the computations we consider the inverse of the cumulative distribution function (CDF), which represents the stochastic distribution in a different way and which, moreover, is defined on the unit interval. Denoting by $\Upsilon^{-1}[X] : [0, 1] \to \mathbb{R}$ the inverse CDF of a stochastic process X, we define the following for the matching of the inverse CDF of calculated and measured temperature (cf. also [32]):

$$(4.3) \qquad F_D(y, u) := \left\| \int_0^1 \left(\Upsilon^{-1}[y](\vartheta) - \Upsilon^{-1}[Y](\vartheta) \right)^2 d\vartheta \right\|_{L^2(D)}^2.$$

Thus, the objective function compares the difference of the inverse CDF of the computed temperature y and the measured temperature Y in an L^2-sense over the physical domain and the unit interval.

4.2. Solving the optimal control problem. Let us now focus on the actual solution of the minimization of one of the above discussed objective functions under the SPDE constraint (2.3) with homogeneous Dirichlet boundary data, i.e., for $g(\boldsymbol{\xi}, x) \equiv 0$. Note that the incorporation of nonhomogeneous Dirichlet boundary conditions g is obtained straightforwardly through the corresponding boundary integral on the right-hand side. In the following we use the notation $(u, v)_{L^2(\Gamma) \otimes L^2(Y)} = \mathbb{E}\left[\int_Y uv \, dx \right]$.

First, we consider the objective function (4.1). To formulate the optimality system for this optimal control problem, we use the Lagrange function
$$(4.4)$$
$$\begin{aligned}
\mathcal{L}(y, u, \mu) :=& F_E(y, u) - \left(-\text{div}(a(\boldsymbol{\xi}, x)\nabla y(\boldsymbol{\xi}, x)) - \beta(\boldsymbol{\xi}, x)u(\boldsymbol{\xi}, x), \mu(\boldsymbol{\xi}, x) \right)_{L^2(\Gamma) \otimes L^2(D)} \\
& - (y(\boldsymbol{\xi}, x), \mu(\boldsymbol{\xi}, x))_{L^2(\Gamma) \otimes L^2(\partial D)} \\
=& \frac{\lambda_1}{2} \mathbb{E}\left[\int_D \left(y(\boldsymbol{\xi}, x) - Y_1(x) \, dx \right)^2 \right] + \sum_{k=2}^{q} \frac{\lambda_k}{2} \mathbb{E}\left[\left(\int_D y(\boldsymbol{\xi}, x)^2 \, dx \right)^k \right] \\
& + \mathcal{R}(u) - \mathbb{E}\left[\int_D \left(-\text{div}(a(\boldsymbol{\xi}, x)\nabla y(\boldsymbol{\xi}, x)) - \beta(\boldsymbol{\xi}, x)u(\boldsymbol{\xi}, x) \right) \mu(\boldsymbol{\xi}, x) \, dx \right] \\
& - \mathbb{E}\left[\int_{\partial D} y(\boldsymbol{\xi}, x)\mu(\boldsymbol{\xi}, x) \cdot \nu(x) \, ds(x) \right].
\end{aligned}$$

To calculate the adjoint equation we use the formal Lagrange technique, which means that we calculate the derivative of the Lagrangian with respect to $y(\boldsymbol{\xi}, x)$. Therefore for the optimization itself we need the derivative of the objective function with respect to y in the direction of the random field δ_y,

$$\langle (F_E)_y(y, u), \delta_y(\boldsymbol{\xi}, x) \rangle = \lambda_1 \mathbb{E}\left[\int_D \left(y(\boldsymbol{\xi}, x) - Y_1(x) \right) \delta_y(\boldsymbol{\xi}, x)\, dx \right]$$
$$+ \sum_{k=2}^q k\lambda_k \, \mathbb{E}\left[\left(\int_D y(\boldsymbol{\xi}, x)^2 dx \right)^{k-1} \int_D y(\boldsymbol{\xi}, x)\delta_y(\boldsymbol{\xi}, x)\, dx \right].$$

Consequently, we get

(4.5)

$$-\mathrm{div}\left(a(\boldsymbol{\xi}, x)\nabla\mu(\boldsymbol{\xi}, x) \right) = \lambda_1 \left(y(\boldsymbol{\xi}, x) - Y_1(x) \right) + \sum_{k=2}^q k\lambda_k \left(\int_D y(\boldsymbol{\xi}, x)^2 dx \right)^{k-1} y(\boldsymbol{\xi}, x)$$
$$\text{in } D \text{ for almost all } \boldsymbol{\xi} \in \Gamma,$$

(4.6) $$\mu(\boldsymbol{\xi}, x) = 0 \qquad\qquad\qquad \text{on } \partial D \text{ for almost all } \boldsymbol{\xi} \in \Gamma$$

as the adjoint equation for our state system (2.3). Furthermore, from the derivative of $\mathcal{L}(y, u, \mu)$ with respect to the control $u(\boldsymbol{\xi}, x)$, we obtain

(4.7) $$\mathcal{L}_u(y, u, \mu)\delta_u(\boldsymbol{\xi}, x) = \mathbb{E}\left[\int_D \delta_u(\boldsymbol{\xi}, x)\beta(\boldsymbol{\xi}, x)\mu(\boldsymbol{\xi}, x)\, dx \right] + \mathcal{R}_u(u)\delta_u(\boldsymbol{\xi}, x).$$

Hence, the descent direction for the gradient descent method (cf. step 2 in section 3.1) can be calculated from the antigradient of the objective, considered as a function of u. That means the descent direction v is given by the negative gradient, which is calculated by the variational inequality (4.7),

(4.8) $$v = -\beta\mu - \mathcal{R}_u(u).$$

Here μ denotes the adjoint state. In the case of a deterministic control u, the gradient reduces again to a deterministic function, and the search direction becomes

$$v = -\mathbb{E}[\beta\mu](x) - \mathcal{R}_u(u).$$

For the SQP method we also need the second derivatives of the Hessian matrix. To this end, we obtain as the pure second derivatives of the objective function

(4.9) $$\langle (F_E)_{yy}(y(\boldsymbol{\xi}, x), u(\boldsymbol{\xi}, x)), \delta_y(\boldsymbol{\xi}, x)\delta_y(\boldsymbol{\xi}, x) \rangle$$
$$= \lambda_1\|\delta_y\|^2 + \sum_{k=2}^q 2k(k-1)\lambda_k \mathbb{E}\left[\left(\|y\|_{L^2(D)}^2 \right)^{k-2} \left(\int_D y\delta_y dx \right)^2 \right]$$
$$+ \sum_{k=2}^q k\lambda_k \mathbb{E}\left[\left(\|y\|_{L^2(D)}^2 \right)^{k-1} \|\delta_y\|_{L^2(D)}^2 \right]$$

and

$$\langle (F_E)_{uu}(y(\boldsymbol{\xi}, x), u(\boldsymbol{\xi}, x)), \delta_u(\boldsymbol{\xi}, x)\delta_u(\boldsymbol{\xi}, x) \rangle = \mathcal{R}_{uu}(u)\delta_u(\boldsymbol{\xi}, x)\delta_u(\boldsymbol{\xi}, x).$$

Moreover, the L_1-penalty function for the SQP method introduced in (3.5) becomes

$$L_1(y, u; \eta) = \frac{\lambda_1}{2} \|y - Y_1\|^2 + \sum_{k=2}^{q} \frac{\lambda_k}{2} \mathbb{E}\left[\left(\|y\|_{L^2(D)}^2\right)^k\right] + \mathcal{R}(u) + \eta \mathbb{E}\left[\|h(y, u)\|_{L^1(D)}\right],$$

and its derivative in the direction of $\delta = (\delta_y, \delta_u)^T$ is

$$
\begin{aligned}
L_1'(y, u; \eta)\delta =& \lambda_1 \mathbb{E}\left[\int_D (y(\boldsymbol{\xi}, x) - Y_1(x))\, \delta_y(\boldsymbol{\xi}, x)\, dx\right] \\
&+ \sum_{k=2}^{q} k\lambda_k \mathbb{E}\left[\left(\int_D y(\boldsymbol{\xi}, x)^2 dx\right)^{k-1} \int_D y(\boldsymbol{\xi}, x)\delta_y(\boldsymbol{\xi}, x)\, dx\right] \\
&+ \mathcal{R}_u(u)\delta_u(\boldsymbol{\xi}, x) + \eta \sum_{h>0} \nabla h(y, u)\delta - \eta \sum_{h<0} \nabla h(y, u)\delta \\
&+ \eta \sum_{h=0} |\nabla h(y, u)\delta|.
\end{aligned}
$$

For the other type of objective functions (4.2) and (4.3) we proceed analogously by formulating the Lagrangian and computing the corresponding derivatives. Indeed, the derivation is very similar for (4.2) and leads to

$$\langle (F_M)_y(y, u), \delta_y(\boldsymbol{\xi}, x)\rangle = \sum_{k=1}^{q} k\lambda_k \int_D \left(\mathbb{E}[y^k] - Y_k\right) \mathbb{E}[y^{k-1}\delta_y]\, dx.$$

Let us finally mention that for (4.3) we have

$$
\begin{aligned}
\langle (F_D)_y(\boldsymbol{\xi}, x), \delta_y(\boldsymbol{\xi}, x)\rangle =& \int_D \left[\int_0^1 \left(\Upsilon^{-1}[y](\vartheta) - \Upsilon^{-1}[Y](\vartheta)\right)^2 d\vartheta\right. \\
&\left.\times \int_0^1 \left(\Upsilon^{-1}[y](\vartheta) - \Upsilon^{-1}[Y](\vartheta)\right) \frac{\partial}{\partial y}\Upsilon^{-1}[y](\vartheta)\delta_y(\boldsymbol{\xi}, x)d\vartheta\right] dx.
\end{aligned}
$$

With the derived Lagrangian and its derivatives, we are now able to solve our optimal control problem with one of the algorithms described in section 3. For the gradient descent algorithm we need only the total derivative of the objective function, e.g., (4.8), to calculate the descent direction. However, for the SQP method we need the gradient of the Lagrangian, and the Hessian too, to set up the KKT-matrix and the corresponding right-hand side (see below).

4.3. Numerical results. In the following we first construct a test scenario and show results of the SC method applied to the forward problem (2.3) with uncertain heat conductivity. Second, we show results of the optimal control problem for the tracking-type objective functions (4.1) and (4.2). For all numerical results shown here we use a spatial discretization with the FEM on piecewise linear shape functions on the unit interval $D := (0, 1)$. The resulting linear systems of equations are solved with a classical conjugate gradient method without preconditioning.

The forward problem. Let us emphasize that, in particular for optimization with stochastic constraints and stochastic inverse problems, it is advisable to have proper control of the numerical errors and the accuracy achieved by iterative solvers. In [12] Kaipio and Somersalo discuss that limited numerical accuracies (i.e., discretization errors) can sometimes (effectively or ineffectively) be interpreted as the behavior
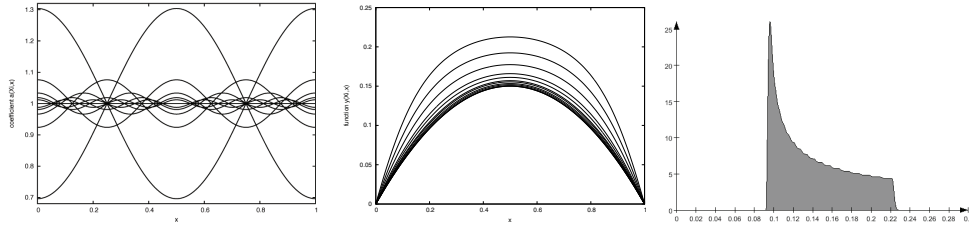
FIG. 4.1. *Left: Various realizations of the coefficient $a(\boldsymbol{\xi}, x)$ for $\kappa = 3$ and 11 collocation points for stochastic dimension $N = 5$. Middle: The state $y(\boldsymbol{\xi}, x)$ corresponding to $a(\boldsymbol{\xi}, x)$ and with a constant right-hand side $u(x)$. Right: The corresponding PDF for $y(\boldsymbol{\xi}, x)$, i.e., at the spatial point $x = 0.25$.*

of a random process. Therefore before proceeding to the optimization problem we discuss some numerical tests on the solution of the forward problem.

For reasons of simplicity we restrict our discussion to the case $\beta(\boldsymbol{\xi}, \cdot) \equiv 1$, and we use homogeneous Dirichlet boundary values only. Also, in this experiment we use a deterministic right-hand side $u(\boldsymbol{\xi}, x) = u(x)$. As discussed in section 2 we consider a vector of random variables $\boldsymbol{\xi} = (\xi^i)_{i=1}^N$, where the ξ^i are independent and uniformly distributed on $[-1, 1]$. The uncertain diffusion coefficient is defined as (cf. [28])

$$(4.10) \qquad a(\boldsymbol{\xi}, x) = 1 + \kappa \sum_{i=1}^{N} \frac{1}{i^2 \pi^2} \cos(2\pi i x) \xi^i,$$

and thus in our example the random variables ξ^i being the components of the random variable vector $\boldsymbol{\xi}$ steer different frequencies of the diffusion coefficient. Our diffusivity fulfills $\mathbb{E}[a(\boldsymbol{\xi}, \cdot)] \equiv 1$ on D and it is bounded by $1 - \frac{\kappa}{6} \leq a(\boldsymbol{\xi}, x) < 1 + \frac{\kappa}{6}$ for all $\boldsymbol{\xi}$. We have set $\kappa = 3$, the stochastic dimension $N = 5$, and the number of collocation points $Q = 10$. The spatial discretization contains $M = 129$ nodes.

We solve the problem (2.3) with the stochastic diffusion coefficient (4.10) by the SC method described in section 2.3. In Figure 4.1 we show the realizations of the diffusivity a and the state y at the collocation points, as well as an approximation [14] of the PDF of the state $y(\boldsymbol{\xi}, 0.25)$, i.e., at the spatial position $x = 0.25$. Here we have fixed the right-hand side to be $u(x) \equiv 1$.

In Figure 4.2 we show the state $y(\boldsymbol{\xi}, x)$ for the same SPDE and the same coefficient $a(\boldsymbol{\xi}, x)$ as above but with a linear right-hand side $u(x) = x$. Also shown is the corresponding variance of the state y and again an approximation of the PDF of $y(\boldsymbol{\xi}, 0.25)$, i.e., at the spatial position $x = 0.25$. For the variance graph, it might be
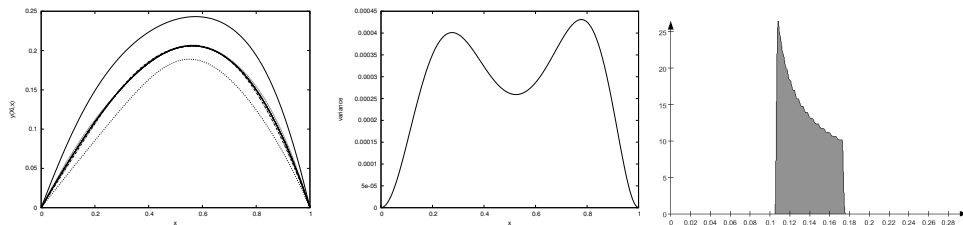

FIG. 4.2. *Left: The state $y(\boldsymbol{\xi}, x)$ corresponding to the above $a(\boldsymbol{\xi}, x)$ and with a linear right-hand side $u(x)$. Middle: The variance of the solution. Right: The corresponding PDF at the spatial point 0.25.*
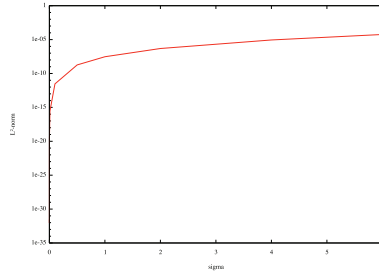
FIG. 4.3. *Convergence of the stochastic solution to the deterministic solution as $\kappa \to 0$: κ vs. the L^2-norm of the difference of the mean of the stochastic solution and deterministic solution.*

surprising at first that the highest variance in the solution is obtained where a has the smallest variance. However, this is a consequence of the global character of diffusion. At the positions with small variance in $a(\boldsymbol{\xi}, x)$ we have almost the same derivatives for all different realizations of $y(\boldsymbol{\xi}, x)$. However, the large differences in the diffusion coefficient at the other points lead to a dispersion of the function values at these points.

To further test the solver for the forward problem, we investigate the convergence of the stochastic solution as $\kappa \to 0$, i.e., as the variance of the diffusion coefficient tends to zero. Indeed, in Figure 4.3 we see the expected convergence to the function resulting from the solution of the deterministic PDE in which the diffusivity is set to $a(\boldsymbol{\xi}, \cdot) \equiv 1$.

Finally, we investigate the convergence of the stochastic solution as the number of collocation points increases. To this end we evaluate in the tensor product space $L^2(\Gamma) \otimes L^2(D)$ the quotient of solutions obtained with different numbers of collocation points, i.e.,

$$\frac{\|\mathbb{E}[y^{(k)}]\|^2_{L^2(D)}}{\|\mathbb{E}[y^{(k-1)}]\|^2_{L^2(D)}}, \quad \frac{\|\mathrm{Var}[y^{(k)}]\|^2_{L^2(D)}}{\|\mathrm{Var}[y^{(k-1)}]\|^2_{L^2(D)}}, \quad \frac{\|\|y^{(k)}\|\|}{\|\|y^{(k-1)}\|\|}, \quad k \in \mathbb{N}.$$

Here $y^{(k)}$ denotes the solution of the SPDE with (2.3), with the stochastic diffusion coefficient (4.10) computed at level k of the sparse grid approximation (cf. section 2.3). The results of the convergence test for $N = 5$ and a spatial discretization with $M = 65$ grid points are shown in Table 4.1. From these results we see a rapid convergence in the expectation, i.e., for the norms $\|\|\cdot\|\|$ and $\|\mathbb{E}[\cdot]\|$. The convergence in the variance ($\|\mathrm{Var}[\cdot]\|$) is much slower. Note that the variance $\mathrm{Var}[\cdot]$ is defined as the second-order centered moment; see (2.9) for the definition of higher-order centered moments. Further investigations on the convergence with respect to the number of collocation points are discussed in [28] as well.

TABLE 4.1
Convergence of the stochastic solution for increasing number of collocation points. The spatial discretization has $M = 65$ grid points, the stochastic dimension is $N = 5$, and the coefficient is set to $\kappa = 3$.

k	Q	$\dfrac{\|\mathbb{E}[y^{(k)}]\|^2_{L^2(D)}}{\|\mathbb{E}[y^{(k-1)}]\|^2_{L^2(D)}}$	$\dfrac{\|\mathrm{Var}[y^{(k)}]\|^2_{L^2(D)}}{\|\mathrm{Var}[y^{(k-1)}]\|^2_{L^2(D)}}$	$\dfrac{\|\|y^{(k)}\|\|}{\|\|y^{(k-1)}\|\|}$
1	10	0.999321	8250.85	0.998458
2	60	0.999979	3.95767	0.999956
3	240	1	2.24199	1
4	800	1	1.77461	1

The optimal control problem. For the following numerical results for the optimal control problem, we choose the same diffusion coefficient $a(\boldsymbol{\xi}, x)$ as before (cf. (4.10)). First, we generate the optimal state y by prescribing a fixed right-hand side (control) $u(\boldsymbol{\xi}, x) = 0.34 + 2.56x$, which does not depend on $\boldsymbol{\xi}$, and computing y through the SG method. We end up with a random field y, whose expectation and stochastic moments we use in the following as the parameters Y_k defining the objective function. Note that this control u cannot be resolved by our finite element space of piecewise linear functions, which vanish at the boundary. As before, we discretize the problem in space with the FEM with piecewise linear shape functions. We use $M = 129$ grid points in the physical space, a stochastic dimension $N = 2$, and SC at level $k = 3$, i.e., involving 29 collocation points ($Q = 28$).

In the second step of this experiment we aim at finding the optimal control that best approximates the precomputed state y in the sense of the objective functions (4.1) or (4.2). Therewith, the control is assumed to be a random field $u(\boldsymbol{\xi}, x)$. We present results for different weightings of the stochastic moments up to 10th order, i.e., $q = 10$. For the objective (4.1) we use the SQP method described in section 3.2, and for comparison we use the gradient descent method from section 3.1. The results for the objective (4.2) are generated with the gradient descent method. In all numerical experiments, we use the regularization term $\mathcal{R}(u) = \frac{1}{2}\lambda_u \mathbb{E}[\|u\|^2_{L^2(D)}]$ with $\lambda_u = 10^{-4}$.

For the objective (4.1), we can obtain the full second-order derivative in the fully discrete form, thus no longer involving the directional derivatives in direction $\delta_y(\boldsymbol{\xi}, x)$ (cf. (4.9)). For the other objective functions (4.2) and (4.3), the situation is more complicated, as the order of the integrations is reversed; i.e., the stochastic integral is taken inside the integral over the physical space. However, for the objective function (4.2) we can change the order of evaluations in the optimization algorithm accordingly such that we solve the stochastic problem at every spatial point, instead of solving the deterministic problem at every SC point. With this adaptation of the optimization algorithm, we may use the FEM as well and take advantage of symmetric matrices, as in the latter case above. A discrete version of the third objective function (4.3) has been considered by Zabaras and Ganapathysubramanian in [32]. We refer the reader to this publication, and here we do not consider numerical experiments for this type of objective function.

In Figure 4.4 we show results of the optimization with the SQP method for the first objective function (4.1) for various weightings of the stochastic moments. The particular settings of the weights λ_i in the objective functions are listed in the top part of Table 4.2. First, comparing experiments A, B, and C shows the regularizing influence of the weight of the variance. It is clearly visible from the right column in Figure 4.4 how the distribution becomes smaller (i.e., with less variance) and higher if the weight λ_2 is increased. We also observe that this is connected to a variation of the expectation of y as part of the compromise between fitting the expectation and minimizing the variance made by the optimization. Thus, adding the second moment, we obtain a smaller PDF, whereas the fitting to the expectation is not as good as without the additional moment.

Second, comparing experiments A and D shows a minor influence of the third moment. The results match the given data better, but the variance of the PDF is larger as before. Finally, comparing experiments E, F, and G demonstrates the influence of the moments of order from 4 to 10. We see that—as one would expect— taking into account these higher order moments does result in small changes to the distribution only. It can be seen that the expectation is matched slightly worse as in the latter cases.
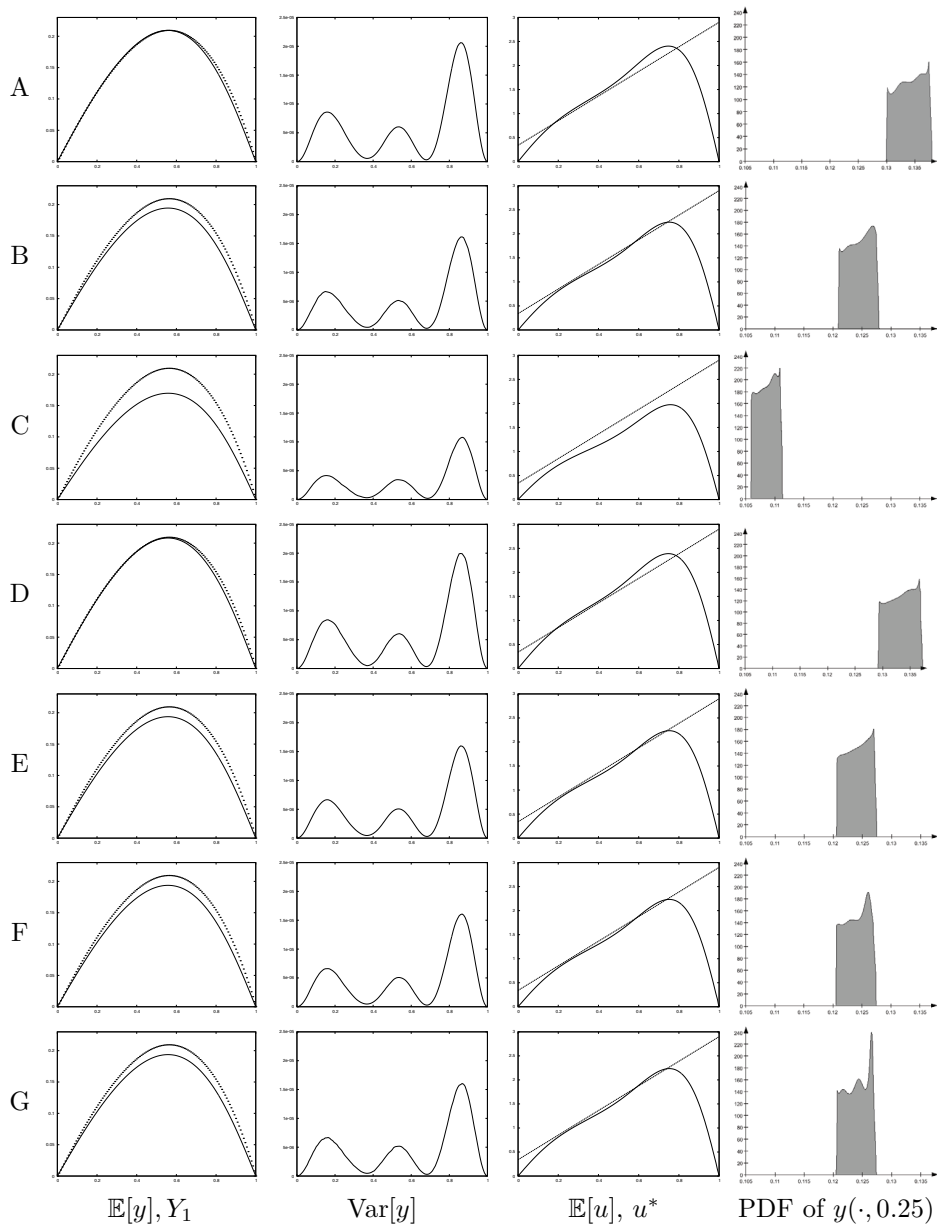
$$\mathbb{E}[y], Y_1 \qquad \mathrm{Var}[y] \qquad \mathbb{E}[u], \ u^* \qquad \text{PDF of } y(\cdot, 0.25)$$

FIG. 4.4. *Results for the minimization of* (4.1)*, solved with the SQP method described in section* 3.2. *From left to right are depicted the expectation and the variance of the state y, where the expectation is compared with the given* $Y_1(x)$ *(dotted lines), and the calculated control u, compared with the given linear right-hand side* u^*. *Finally, on the right we show the probability functions at the spatial position of the optimal state y at the spatial location* $x = 0.25$. *From top to bottom the settings of the weights vary according to the values listed in Table* 4.2. *Note that for the prescribed data we have* $Y_1(0.25) \approx 0.1347$.

TABLE 4.2

Top (A–G): *Weights used for the optimization with the objective* (4.1) *and the SQP method. The weights in A–E are additionally used for the gradient descent method with the objective* (4.1). *Bottom* (H–J): *Weights used for the optimization with the objective* (4.2) *and the gradient descent. The label given in the left column corresponds to the respective rows in Figures* 4.4, 4.5, *and* 4.6.

Label	# weights $\neq 0$	Value of nonzero weights in the objective function
A	1	$\lambda_1 = 1$
B	2	$\lambda_1 = 1, \lambda_2 = 2$
C	2	$\lambda_1 = 1, \lambda_2 = 8$
D	3	$\lambda_1 = 1, \lambda_2 = 0, \lambda_3 = 4$
E	4	$\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 4, \lambda_4 = 8$
F	6	$\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 4, \lambda_4 = 8, \lambda_5 = 16, \lambda_6 = 32$
G	10	$\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 4, \lambda_4 = 8, \lambda_5 = 16, \lambda_6 = 32, \lambda_7 = 64,$ $\lambda_8 = 128, \lambda_9 = 246, \lambda_{10} = 492$
H	1	$\lambda_1 = 1$
I	4	$\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 4, \lambda_4 = 8$
J	4	$\lambda_1 = 1, \lambda_2 = 100, \lambda_3 = 1000, \lambda_4 = 10000$

In these results we can observe that the choice of the objective, and especially of the additional moments, has an influence on the outcome of the optimization. The outcome depends on the number of used moments as well as on the used weighting factors. Since the moments are calculated by the expectation of the exponentiated states, their values decrease for higher degrees of the moments. Thus, we need significantly larger weighting factors to take higher degrees into account. However these experiments also show that we have a wealth of degrees of freedom for the design of the objective function, which allows us to construct the objective carefully corresponding to the requirements of the application under investigation.

To compare the SQP method with the gradient descent method for the optimization of the objective functional (4.1), some of the results are depicted in Figure 4.5. The weights λ_i, used for the presented results, are chosen according to the settings in rows A and E in Table 4.2. In principle, using the gradient descent requires more iterations to obtain a comparable result—approximately 40 times more iterations. However, the number of iterations can be slightly reduced by choosing a larger maximal step size, but still the gradient descent method will require more iterations. In the configurations A–E from Table 4.2 the number of required iterations is in the range of 17–110 for the SQP method and in the range of 1396–2202 for the gradient descent.

Altogether the results from the SQP method match the results from the gradient descent method. However, for the gradient descent method the calculated states show a smoother PDF than the results from the SQP method. This is based on the fact that in the gradient method the states are calculated via the constraints appropriate to the controls, whereas in the SQP method both the states and the controls are optimization variables. Therefore, the states in the gradient method need to fulfill the constraints, whereas the states in the SQP method are only as good as possible, i.e., in the present case only as smooth as possible.

The L^2-norms of the difference in the expectation of the states, in the variance of the states, and in the expectation of the control for the two methods and the configurations A–E from Table 4.2 are

$$\left\| \mathbb{E}[y_{SQP}] - \mathbb{E}[y_{GD}] \right\|_{L^2}^2 \in [1.289 \cdot 10^{-9}, 1.28 \cdot 10^{-8}],$$

$$\left\| \mathrm{Var}[y_{SQP}] - \mathrm{Var}[y_{GD}] \right\|_{L^2}^2 \in [6.989 \cdot 10^{-15}, 3.123 \cdot 10^{-13}],$$

$$\left\| \mathbb{E}[u_{SQP}] - \mathbb{E}[u_{GD}] \right\|_{L^2}^2 \in [3.772 \cdot 10^{-5}, 2.6 \cdot 10^{-4}].$$
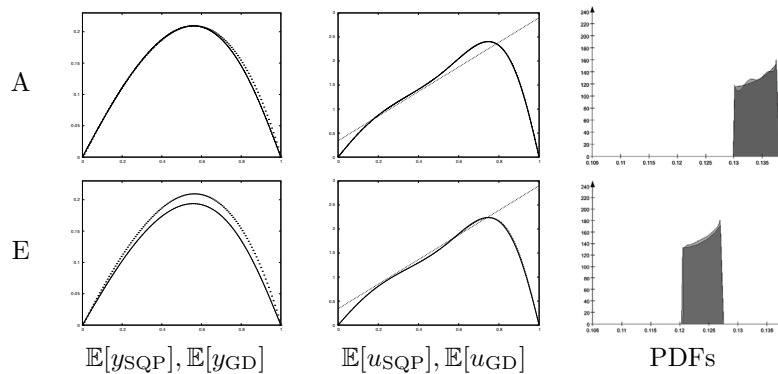
FIG. 4.5. *Comparison of the minimization of* (4.1) *with the gradient descent method and the SQP. From left to right the expectation of y, the calculated control u, and the PDF of the calculated state y at the spatial position x = 0.25 are depicted. In the graphs (left and middle) the results for the expectation value and the control are shown for the given values (dotted lines), the gradient descent (solid line), and the SQP method (dashed lines). The plots on the right show the PDF of the calculated state y from the SQP method (light grey) and from the gradient descent (dark grey). From top to bottom the settings of the weights vary according to the values listed in Table* 4.2.

The best matches between the two different methods are obtained for the optimization without any higher moments, i.e., $\lambda_i = 0, i \geq 1$ (case A).

In Figure 4.6 we show results from the minimization of the objective function (4.2). As described above, we take the stochastic moments of the precomputed state y as the functions Y_k defining the objective function (4.2). In this setting we reduce the control $u(\boldsymbol{\xi}, x)$ to a deterministic control $u(x)$. This is possible, since the control is the only optimization variable in the gradient descent method. In this case the state y and adjoint state μ are given by the constraints. The settings of the weights λ_i are shown in the bottom part of Table 4.2. Comparing the results from experiments H–J we see a better fitting of the prescribed data as more moments are taken into account and also—as expected—if the weighting for the higher-order moments is stronger. Again, from the results it is visible how the adjustment of the weights for the higher-order moments also influences the fitting of the expectation.

Comparison with Monte Carlo. In our final numerical experiments we compare the presented SQP SC method with the most prominent and robust method for stochastic computations, the MC method (cf. section 2.1). We consider the optimal control problem with the objective (4.1) as described above but without taking higher moments into account, i.e., $\lambda_i = 0$ for $i > 1$. Instead of computing the samples of the states and adjoint states at the Q collocation points, we sample at randomly distributed points in the stochastic space.

In Figure 4.7 we show the expectation and the variance of the computed optimal state as well as the optimal control resulting from the MC runs. Also, the same data is shown for the SC approach, and we see a convergence of the MC results to the SC result. In Table 4.3 we report the difference between the optimal state resulting from the SC approach y_{SC} and the optimal states resulting from the MC method with m-samples y_{MC_m}. From the table and the graphs a clear tendency of the MC solutions towards the SC solution is visible in the L^2-norm of the expectation and the variance. However, with the MC approach we need considerably more realizations to obtain a result similar to that for the collocation approach.
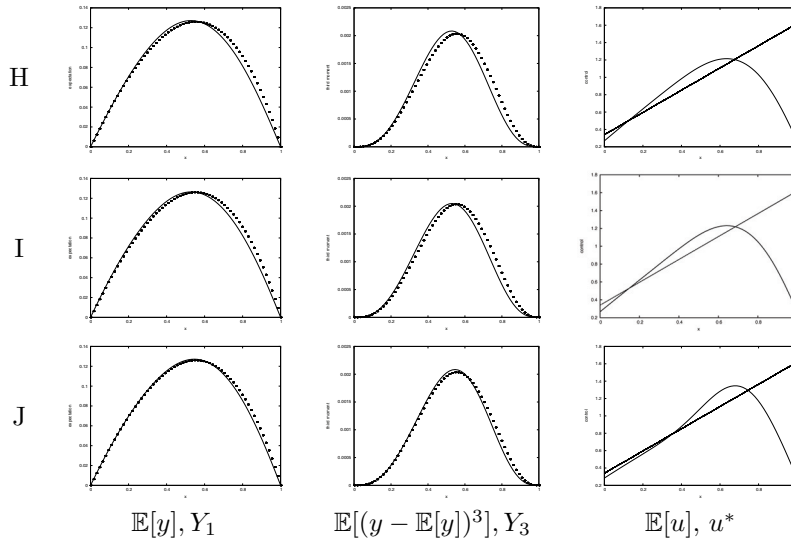
$$\mathbb{E}[y], Y_1 \qquad \mathbb{E}[(y-\mathbb{E}[y])^3], Y_3 \qquad \mathbb{E}[u], u^*$$

FIG. 4.6. *Results for the minimization of the objective function (4.2) with a gradient descent method and for up to $q = 4$ stochastic moments. From left to right the expectation of y, the third moment of the state y, as well as the calculated control u in comparison with the given linear right-hand side u^*, are shown. The expectation and the third moment of the state are each compared with the given values (dotted lines). From top to bottom the settings of the weights vary according to the values listed in Table 4.2.*
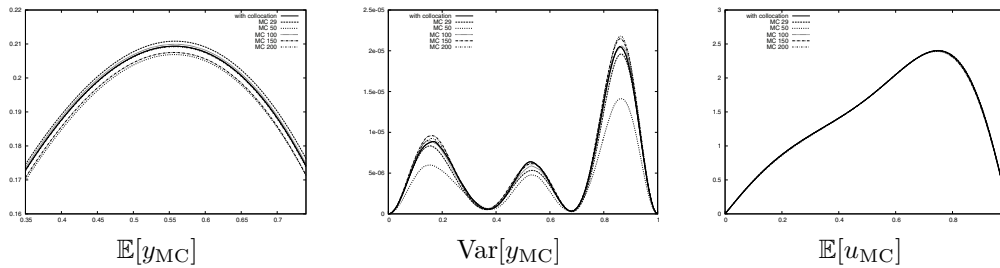

$$\mathbb{E}[y_{\mathrm{MC}}] \qquad \mathrm{Var}[y_{\mathrm{MC}}] \qquad \mathbb{E}[u_{\mathrm{MC}}]$$

FIG. 4.7. *The expectation (left) and the variance (middle) of the optimal state y as well as the expectation of the optimal control u (right) resulting from the minimization of F_E (4.2) without any higher moments ($\lambda_i = 0$ for $i > 1$). In the graphs the result from the collocation SQP method is compared with the results of MC sampling for an increasing number of sampling points (MC 29–MC 200).*

TABLE 4.3
The L^2-norm of the difference between the expectation of the solution with the collocation method and the expectation of the solution obtained with the MC method for a different number of evaluation points for the Monte Carlo method. On the right are shown the L^2-norm of the difference between the variance of the solution with the collocation method and the variance of the solution with the MC method.

m	$\|\mathbb{E}[y_{\mathrm{SC}}]-\mathbb{E}[y_{\mathrm{MC}_m}]\|_{L^2}$	$\|\mathrm{Var}[y_{\mathrm{SC}}]-\mathrm{Var}[y_{\mathrm{MC}_m}]\|_{L^2}$
29	2.96957e-06	3.11497e-13
50	5.32986e-06	5.58844e-12
100	2.92863e-07	6.3303e-14
150	4.50538e-06	1.97644e-13
200	8.78526e-08	2.08923e-13

5. Conclusions. We have presented two approaches for the treatment of optimal control problems with SPDE constraints. In the SPDE constraints we consider a stochastic control and stochastic states as well as stochastic parameter functions. Thus, the resulting states are random fields, i.e., random variables, which are indexed by a spatial coordinate. The gradient descent method as well as the SQP method discussed here are based on a spectral discretization of the constraining PDEs with the SC method. The SC method assumes that the underlying stochastic processes can be approximated by polynomials of random variables, and thus it assumes a certain smoothness in the stochastic space. The SC method can be combined with any spatial discretization technique. Here we combine it with a simple FEM incorporating piecewise linear basis functions. We have shown that both optimization approaches presented here can be implemented very conveniently and are memory effective. In fact, the final scheme results in multiple deterministic evaluations of the state or the adjoint state of the system, where the stochasticity is evaluated at different sample points.

As possible applications of the minimization techniques we have discussed three different objective functions: The first aiming at the fitting of the expectation of the stochastic state to a given (measured) function; the second aiming at the fitting of several moments of the stochastic state to prescribed data; and the third aiming at a fitting of the cumulative density function of the stochastic state. We have applied the minimization of two of the aforementioned objective functions to the problem of optimal heat source. Thereby the SPDE constraining the optimization is the stochastic analogue of the steady state heat diffusion, where the diffusion coefficient is a random field.

For this application problem we have first investigated the convergence of the SC solution for the constraining PDE. Our results show a convergence of the solutions with increasing number of collocation points, as well as a convergence to the solution of the deterministic problem as the variance of the stochastic diffusion coefficient tends to zero. Second, we have investigated the influence of the number and weighting of the stochastic moments in the objective functions for the optimal control problem. Our results show that we can steer very well the behavior of the objective function and construct it such that it matches the requirements of the application under investigation.

Moreover, we have compared the two optimization methods for an objective functional fitting the expectation and penalizing large higher-order moments. The numerical results show that the gradient descent method requires more iterations than the SQP method, especially near the optimum, which is a characteristic behavior of the gradient descent method. That means the SQP method is more efficient, as known from the deterministic optimization. But there exist further considerations when choosing the optimization method. For example, the gradient method allows us to use a deterministic control, as we have described for an objective, which fits the stochastic moments, while the constraints are still SPDEs.

Finally we have compared our new minimization technique to the most prominent and widely used technique for treating stochastic problems, the MC approach. Our numerical results show that we achieve the same results by the SC optimization and the MC optimization. However, for our application problem the MC approach needs about 10 times more evaluations of the deterministic state and adjoint state than the collocation method. This shows that under the assumption of smoothness of the underlying stochastic processes, our SC optimization algorithms outperform the classical MC approach.

Future research directions include the application of the presented methodology to real world problems in the area of optimization in medicine, such as treatment planning. Medical applications are prime examples for modeling with uncertain parameters and SPDEs due to the intrinsic variability of biological tissue. However, it is inevitable that the uncertainty and variability of the parameters are taken into account in the modeling in order to allow for robust and patient-specific optimal treatment.

REFERENCES

[1] I. ALTROGGE, T. PREUSSER, T. KRÖGER, C. BÜSKENS, P. PEREIRA, D. SCHMIDT, AND H.-O. PEITGEN, *Multiscale optimization of the probe-placement for radio-frequency ablation*, Academic Radiology, 14 (2007), pp. 1310–1324.

[2] I. BABUŠKA, R. TEMPONE, AND G. E. ZOURARIS, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825.

[3] V. A. B. NARAYANAN AND N. ZABARAS, *Stochastic inverse heat conduction using a spectral approach*, Internat. J. Numer. Methods Engrg., 60 (2004), pp. 1569–1593.

[4] V. BARTHELMANN, E. NOVAK, AND K. RITTER, *High dimensional polynomial interpolation on sparse grids*, Adv. Comput. Math., 12 (2000), pp. 273–228.

[5] L. T. BIEGLER, O. GHATTAS, M. HEINKENSCHLOSS, AND B. VAN BLOEMEN WAANDERS, EDS., *Large-Scale PDE-Constrained Optimization*, Springer, Berlin, 2003.

[6] R. H. CAMERON AND W. T. MARTIN, *The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals*, Ann. of Math., 48 (1947), pp. 385–392.

[7] Y. CAO, *Numerical Solutions for Optimal Control Problems Under SPDE Constraints*, Tech. report AFRL-SR-AR-TR-06-0475, Department of Mathematics, Florida A&M University, Tallahassee, FL, 2006.

[8] A. CHORIN, *Hermite expansions in Monte Carlo computation*, J. Comput. Phys., 8 (1971), pp. 471–482.

[9] A. CHORIN, *Gaussian fields and random flow*, J. Fluid Mech., 63 (1974), pp. 21–32.

[10] M. DEB, I. BABUŠKA, AND J. ODEN, *Solutions of stochastic partial differential equations using Galerkin finite element techniques*, Comput. Methods Appl. Mech. Eng., 190 (2001), pp. 6359–6372.

[11] R. GHANEM AND P. SPANOS, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991.

[12] J. KAIPIO AND E. SOMERSALO, *Statistical inverse problems: Discretization, model reduction and inverse crimes*, J. Comput. Appl. Math., 1098 (2007), pp. 493–504.

[13] A. KEESE, *Numerical Solution of Systems with Stochastic Uncertainties: A General Purpose Framework for Stochastic Finite Elements*, Ph.D. thesis, Technical University Braunschweig, Braunschweig, Germany, 2004.

[14] T. KRÖGER, I. ALTROGGE, O. KONRAD, R. M. KIRBY, AND T. PREUSSER., *Estimation of probability density functions for parameter sensitivity analyses*, in Proceedings of the Conference Simulation and Visualization (SimVis), Magdeburg, 2008, pp. 61–74.

[15] O. L. MAÎTRE, M. REAGAN, H. NAJM, R. GHANEM, AND O. KNIO, *A stochastic projection method for fluid flow* II: *Random process*, J. Comput. Phys., 181 (2002), pp. 9–44.

[16] P. MALLIAVIN, *Stochastic Analysis*, Springer, Berlin, 1997.

[17] F. MALTZ AND D. HITZL, *Variance reduction in Monte Carlo computations using multi-dimensional Hermite polynomials*, J. Comput. Phys., 32 (1979), pp. 345–376.

[18] W. MEECHAM AND D. JENG, *Use of the Wiener–Hermite expansion for nearly normal turbulence*, J. Fluid Mech., 32 (1968), pp. 225–249.

[19] R. E. MORTENSEN, *A priori open loop optimal control of continuous time stochastic systems*, Internat. J. Control (1), 3 (1966), pp. 113–127.

[20] R. E. MORTENSEN, *Stochastic optimal control with noisy observations*, Internat. J. Control (1), 4 (1966), pp. 455–464.

[21] M. NISIO, *Optimal control for stochastic partial differential equations and viscosity solutions of bellman equations*, Nagayo Math. J., 123 (1991), pp. 13–37.

[22] B. ØKSENDAL, *Optimal control of stochastic partial differential equations*, Stoch. Anal. Appl., 23 (2005), pp. 165–179.

[23] F. TRÖLTZSCH, *Optimal Control of Partial Differential Equations. Theory, Methods, and Applications*, Grad. Stud. Math. 112, American Mathematical Society, Providence, RI, 2010.

[24] G. VAGE, *Variational methods for pdes applied to stochastic partial differential equations*, Math. Scand., 82 (1998), pp. 113–137.

[25] J. WANG AND N. ZABARAS, *Hierarchical Bayesian models for inverse problems in heat conduction*, Inverse Problems, 21 (2005), pp. 183–206.

[26] J. WANG AND N. ZABARAS, *A Markov random field model to contamination source identification in porous media flow*, Int. J. Heat and Mass Transfer, 49 (2006), pp. 939–950.

[27] N. WIENER, *The homogeneous chaos*, Amer. J. Math., 60 (1938), pp. 897–936.

[28] D. XIU AND J. S. HESTHAVEN, *High-order collocation methods for differential equations with random inputs*, SIAM J. Sci. Comput., 27 (2005), pp. 1118–1139.

[29] D. XIU AND G. KARNIADAKIS, *Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos*, Comput. Methods Appl. Mech. Eng., 191 (2002), pp. 4927–4948.

[30] D. XIU AND G. E. KARNIADAKIS, *The Wiener–Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644.

[31] D. XIU AND G. KARNIADAKIS, *Modeling uncertainty in flow simulations via generalized polynomial chaos*, J. Comput. Phys., 187 (2003), pp. 137–167.

[32] N. ZABARAS AND B. GANAPATHYSUBRAMANIAN, *A scalable framework for the solution of stochastic inverse problems using a sparse grid collocation approach*, J. Comput. Phys., 227 (2008), pp. 4697–4735.