Multigrid Methods and Sparse-Grid Collocation Techniques for Parabolic Optimal Control Problems with Random Coefficients

A. Borzi† and G. von Winckel‡

Abstract. An efficient computational framework to solve nonlinear parabolic optimal control problems with random coefficients is presented. This framework allows us to investigate the influence of randomness or uncertainty of problem's parameters values on the control provided by the optimal control theory. The proposed framework combines space-time multigrid methods with sparse-grid collocation techniques. Theoretical and numerical results of computation of stochastic optimal control solutions and formulation of mean control functions are presented.

Key words. multigrid method, sparse grids, reaction-diffusion problems, random fields, optimal control theory

AMS subject classifications. 35K57, 60H25, 49K20, 65M55, 65C20

DOI. 10.1137/070711311

1. Introduction. An important mathematical tool for the construction of efficient control strategies for real systems is provided by optimal control theory [21, 25]. In this framework, one considers a dynamical model, a description of the control mechanism, and a criterion defining the objective that models the purpose of the control and describes the cost of its action. An optimal control problem is then formulated as the minimization of the objective under the constraint given by the modeling equations. We focus on the control of nonlinear parabolic systems with random coefficients where the nonlinearity appears as a reaction term in the system. On the one hand, choosing time-dependent reaction-diffusion systems is motivated by the many applications in, e.g., biology [24], chemistry [20], and physiology [28]. On the other hand, we intend to present a methodology for investigating PDE-based optimal control problems with random coefficients, and the choice of reaction-diffusion systems appears appropriate for the case of time-dependent problems. Optimal control functions of time-dependent systems are characterized as the solution of an optimality system consisting of evolution equations with opposite time orientation. Usually deterministic models are considered, and it is only recently that data uncertainty has boosted research [23, 33, 38] on modeling and optimization problems where the coefficients of the problem are more naturally described by random fields.

We contribute to this research work formulating optimal control problems with random coefficients and presenting an efficient multigrid methodology to solve these problems. Starting from the work in [1, 12, 26, 36] on elliptic problems with random inputs, we formulate for the first time nonlinear parabolic optimal control problems...
with random diffusion and reaction coefficients. We assume that these coefficients are modeled by random fields that can be approximated by a truncated Karhunen–Loève (KL) expansion on the probability space. With this representation, we can use the Smolyak sparse-grids algorithm [2, 10, 19] to model a high-dimensional stochastic coefficient space. Further, we suggest the use of a stochastic collocation method, where the solution of the stochastic optimal control problem is obtained solving, in the physical space, a deterministic optimality system for each point of the sparse-grid coefficient space. For the solution of the deterministic optimality system, we use the space-time multigrid schemes presented in [3, 4, 5] that provide optimal computational performance independently of the value of the optimization parameters. The combination of sparse-grids and space-time multigrid techniques results in a solution process with almost optimal computational complexity with respect to the sizes of the physical and probability grids.

In the following section, we illustrate the modeling of random fields and their approximation using the KL expansion and define the concept of solution of a stochastic PDE problem. In section 3, we formulate a class of nonlinear parabolic optimal control problems with random coefficients. We focus on optimal control problems with deterministic objectives and random time-dependent PDE constraints. In section 4, we discuss the discretization of the stochastic parameter space using sparse-grids collocation and describe the Smolyak scheme. In section 5, space-time multigrid schemes for parabolic optimal control problems are presented. This section considers two space-time smoothers whose smoothing properties and corresponding multigrid implementation are analyzed by twogrid Fourier analysis in the appendix. This analysis also provides mean values and variances of convergence factors to determine robustness of the estimates with respect to randomness of the coefficients.

Section 6 is dedicated to numerical experiments for validating the numerical performance of the space-time multigrid algorithms combined with sparse-grid collocation techniques. Typical multigrid convergence rates and robustness with respect to a large choice of optimization parameters is obtained. Results of computation of stochastic optimal control solutions are reported with a focus on the moments of the tracking ability of the optimization schemes. We investigate the construction of a new robust control obtained as the mean of the controls resulting from different realizations of diffusion and reaction fields. A conclusion section completes the exposition of our work.

2. Random coefficients. In the modeling of application systems, we acknowledge the fact that material properties, like diffusivity and reactivity, are only known at a statistical level. On the other hand, observation (or preparation) of physical configurations is subject to random noise. Both sources of randomness result in uncertainties in model and process conditions that can be accommodated in the framework described below.

To be specific, consider the following nonlinear parabolic equation

\[
\begin{align*}
  -\partial_t y + G_\delta(y) + \sigma \Delta y &= f & \quad & \text{in } Q = \Omega \times (0,T), \\
  y &= y_0 & \quad & \text{in } \Omega \times \{t = 0\}, \\
  y &= 0 & \quad & \text{on } \Sigma = \partial\Omega \times (0,T),
\end{align*}
\]

where $\Omega \subset \mathbb{R}^d$, $d \geq 1$ is the spatial dimension and $T$ is the time horizon where the reaction-diffusion process is considered. The nonlinear term $G_\delta(y)$ models the reaction kinetics for the state $y$; $\delta$ is a reaction parameter. Clearly, all coefficients in (1) can be considered subject to randomness. However, to focus our discussion, consider first
the case where only the diffusion coefficient \( \sigma \) is described by a spatially random field. That is, we assume that \( \sigma = \sigma(x, \omega) > 0 \), where \( x \in \Omega \) and \( \omega \in \mathcal{O} \), where the triple \((\mathcal{O}, \mathcal{A}, \mathcal{P})\) denotes a probability space, where \( \mathcal{O} \) is the space of elementary events, \( \mathcal{A} \) is the sigma-algebra of subsets of \( \mathcal{O} \), and \( \mathcal{P} \) the probability measure on \( \mathcal{O} \). The values of the stochastic function \( \sigma(x, \omega) \) are usually spatially correlated in a way characterized by a covariance structure. Clearly, we cannot model numerically the resulting infinite-dimensional coefficient space, and a suitable finite-dimensional approximation must be introduced. For this purpose, a common approach is to use the KL expansion \([22]\) of the random field \( \sigma(x, \omega) \) that is based on a spectral decomposition of the covariance kernel of the stochastic process. We assume that the mean and the covariance of \( \sigma(x, \omega) \) are known, respectively, as

\[
\sigma_0(x) = \mathbb{E}(\sigma)(x) := \int_{\mathcal{O}} \sigma(x, \omega) d\mathcal{P}(\omega)
\]

and

\[
C_\sigma(x, x') = \int_{\mathcal{O}} (\sigma(x, \omega) - \sigma_0(x))(\sigma(x', \omega) - \sigma_0(x')) d\mathcal{P}(\omega).
\]

We see that \( C_\sigma \) defines the kernel of a compact, positive, and self-adjoint operator. Denote with \( \lambda_j \) the real positive eigenvalues and with \( z_j(x) \) the corresponding orthonormal eigenfunctions of \( C_\sigma \) as follows:

\[
\int_{\Omega} C_\sigma(x, x') z_j(x') dx' = \lambda_j z_j(x), \quad x \in \Omega,
\]

where we assume to order the eigenvalues in decreasing order. Having the eigenpairs, we can define the following uncorrelated random variables

\[
Y_j(\omega) = \frac{1}{\sqrt{\lambda_j}} \int_{\Omega} (\sigma(x, \omega) - \sigma_0(x)) z_j(x) dx, \quad j = 1, 2, \ldots
\]

with zero mean and unit variance, i.e., \( \int_{\mathcal{O}} Y_j(\omega) Y_j(\omega) d\mathcal{P}(\omega) = \delta_{ij} \). We assume that the images \( \Gamma_j = Y_j(\mathcal{O}) \), \( j = 1, 2, \ldots \), be bounded intervals in \( \mathbb{R} \) and consider the case where the random variables \( Y_j \), with density probability \( \rho_j \), are independent.

Now, the truncated KL expansion of \( \sigma(x, \omega) \) is given by

\[
\sigma_N(x, \omega) = \sigma_0(x) + \sum_{j=1}^{N} \sqrt{\lambda_j} z_j(x) Y_j(\omega),
\]

where \( N \) denotes the number of terms in the truncation. Following \([36]\), we denote with \( \Gamma = \prod_{j=1}^{N} \Gamma_j \) and the joint probability density function \( \rho = \prod_{j=1}^{N} \rho_j \).

We see that \( \sigma_N(x, \omega) \) may provide a suitable approximation to \( \sigma(x, \omega) \) assuming the eigenvalues decay sufficiently fast and \( N \) is sufficiently large. For correlation lengths comparable to the size of the domain, a small value of \( N \) can be sufficient to obtain an accurate representation of the random field.

We consider random fields characterized by squared exponential covariance

\[
C_\sigma(x, x') = s^2 \exp \left( -\frac{|x - x'|^2}{\ell^2} \right), \quad x, x' \in [a, b],
\]
with mean $\sigma_0$ and variance $s^2$. The degree of variability of this random field can be characterized by the ratio $s/\sigma_0$, while the frequency of variation of this field is related to the ratio $L/\ell$, where $L$ is the characteristic length of the physical domain and $\ell$ represents the physical correlation length. A small $L/\ell$ corresponds to a highly correlated random field. It is clear that for $|x-x'| >> \ell$, the variables $\sigma(x,\omega)$ and $\sigma(x',\omega)$ are essentially uncorrelated. One can show that the eigenvalues have exponential decay as $\lambda \sim c_1 \ell \exp(-c_2 \ell^2)$ for some positive constants $c_1$ and $c_2$; see, e.g., [17, 26].

We consider random fields of the type given above that can be approximated by (4) with good accuracy for moderate values of $N$. Therefore, assuming finite-dimensional random fields, we can write the explicit dependence of $\sigma$ on the random variables $[Y_1,\ldots,Y_N]$ as follows:

$$\sigma(x,\omega) \approx \sigma(x,Y_1,\ldots,Y_N).$$

Further we also assume that the reaction dynamics of (1) can be modeled by a random field. We consider temporal randomness and assume the following finite-dimensional KL representation

$$\delta(t,\omega) \approx \delta(t,W_1,\ldots,W_M) = \delta_0 + \sum_{j=1}^{M} \sqrt{\mu_j} v_j(t) W_j(\omega),$$

where the $W_j$ are uncorrelated random variables with zero mean and unit variance.

In the following, $\rho$ denotes the joint probability density of $(Y_1,\ldots,Y_N,W_1,\ldots, W_M)$, with support $\Gamma = \prod_{j=1}^{N} \Gamma_Y \times \prod_{j=1}^{M} \Gamma_W$.

With the setting above and, for simplicity, assuming deterministic initial and boundary conditions, we have that the solution to (1) can be described in terms of spatial and time coordinates and the set of random variables $[Y_1,\ldots,Y_N;W_1,\ldots,W_M]$ as follows:

$$y(x,t,\omega) \approx y(x,t,Y_1,\ldots,Y_N,W_1,\ldots,W_M).$$

We see that solving a stochastic PDE problem requires us to solve for the set of all deterministic solutions corresponding to all possible events. In general, this appears to be an impossible task. However, an approximation to the stochastic solution can be obtained considering a finite-dimensional discretization of the probabilistic coefficient space.

3. Optimal control problems with random coefficients. We consider time-dependent reaction-diffusion processes controlled through source terms with the purpose of tracking a desired trajectory given by $y_d(x,t)$ or with the objective to reach a desired terminal state $y_T(x)$ at time $T$. Here, we assume that the targets are deterministic. We also assume that the theory developed in [31] for the case of elliptic problems with random coefficients is also valid in our context, that is, we suppose that the random parameter fields are sufficiently regular such that moments of the solution up to a given degree exist; see also [27].

For the ease of notation, we denote $Y = [Y_1,\ldots,Y_N]$ and $W = [W_1,\ldots,W_M]$. Also, to simplify notation, we omit the fact that all dependent variables representing the solution to the optimal control problem are functions of the event $\omega$. Thus, for the state we write $y(x,t)$ instead of $y(x,t,\omega)$ and similarly, for the control $u(x,t)$ instead of $u(x,t,\omega)$, etc.
Now, we formulate a distributed optimal control problem with random coefficients

\[
\min_{u \in L^2(Q; \mathcal{O})} J(y, u)
\]

\[
\begin{aligned}
-\partial_t y + G_{\delta(W(\omega))}(y) + \sigma(Y(\omega)) \Delta y &= f + u & \text{in } Q \times \mathcal{O}, \\
y &= y_0 & \text{in } \Omega \times \{t = 0\}, \\
y &= 0 & \text{on } \Sigma,
\end{aligned}
\]

where we assume a deterministic initial condition \(y_0 \in H^1_0(\Omega)\) and for any given event \(\omega\), the solution \(y \in H^{2,1}(Q) = \{y \in L^2(0,T; H^2(\Omega) \cap H^1_0(\Omega)) : \partial y/\partial t \in L^2(0,T; L^2(\Omega))\}\) and the control \(u \in L^2(Q)\).

To track a desired trajectory given by \(y_d(x,t) \in L^2(Q)\) or to reach a desired terminal state \(y_T(x) \in L^2(\Omega)\), we introduce the following deterministic cost functional:

\[
J(y, u) = \frac{\alpha}{2} ||y - y_d||^2_{L^2(Q)} + \frac{\beta}{2} ||y(\cdot, T) - y_T||^2_{L^2(\Omega)} + \frac{\nu}{2} ||u||^2_{L^2(Q)}.
\]

Here, \(\nu > 0\) is the weight of the cost of the control and \(\alpha \geq 0, \beta \geq 0\) are optimization parameters. For example, the case \(\alpha = 1, \beta = 0\) corresponds to tracking without terminal observation. With \(\alpha = 0, \beta = 1\), our objective is to reach a given final target configuration without any specification of the trajectory that should be followed.

The optimal control problem (5)–(6) is stochastic in the sense that for any event \(\omega\), it provides a different control. Each single event corresponds to the solution of a deterministic optimal control problem. Existence of unique solutions to the deterministic version of the optimal control problem above can be established under suitable conditions; see, e.g., [15, 25]. For \(\omega\), we have the occurrence \(\delta = \delta(t, \omega)\) and \(\sigma = \sigma(x, \omega)\). With this specification, we have that the solution to (5)–(6) is characterized by the following first-order optimality system:

\[
\begin{aligned}
-\partial_t y + G_{\delta}(y) + \sigma \Delta y &= f + u & \text{in } Q, \\
\partial_t p + G'_{\delta}(y)p + \sigma \Delta p + \alpha(y - y_d) &= 0 & \text{in } Q, \\
\nu u - p &= 0 & \text{in } Q, \\
y &= 0, p &= 0 & \text{on } \Sigma,
\end{aligned}
\]

with initial condition \(y(x, 0) = y_0(x)\) for the state variable (evolving forward in time). The terminal condition for the adjoint variable (evolving backward in time) is given by

\[
p(x, T) = \beta(y(x, T) - y_T(x)).
\]

At this point, we can introduce the so-called reduced cost functional \(\hat{J}\) given by

\[
\hat{J}(u) = J(y(u), u),
\]

where \(y(u)\) denotes the unique solution to the state equation for a given \(u\). The gradient of \(\hat{J}\) with respect to \(u\) is given by \(\nabla \hat{J}(u) = \nu u - p(u)\), where \(p(u)\) is the solution of the adjoint equation for the given \(y(u)\).

By solving optimal control problems with random coefficients, we explore the space of controls depending on the configuration of the parameters. By definition, this space represents the solution of the stochastic problem. However, a possible task is to deliver a unique control for the governing random PDE model. This control
should be at least suboptimal and provide good tracking features for all configurations of the coefficients. Therefore, the issue arises of how to define a robust deterministic control $u \in L^2(Q)$ with respect to random parameters within an uncertainty range. Clearly, it is unlikely that with different PDE-constraints, corresponding to different coefficients’ values, the objective would be optimized by the same control. What we could require is to minimize the mean of the objective while spanning the stochastic parameter space. That is, one could consider the following problem

$$
(10) \quad \min_{u \in L^2(Q)} \int_{\Omega} \hat{J}_\omega(u) \, d\mathcal{P}(\omega),
$$

where $\hat{J}_\omega(u) = \hat{J}(y_\omega(u), u)$. One can see that this formulation requires us to solve an optimization problem with infinitely many terms in the objective (the integral in $\mathcal{P}$) and infinitely many PDE-constraints (that characterize $y_\omega(u)$). Indeed, even after discretization of the stochastic space (see below), an optimization problem with a very large number of PDE-constraints must be solved: $\min_{u \in L^2(Q)} \sum_i \hat{J}_\omega(u) w_i$, with quadrature weights $w_i$. This formally results in an equivalent to a multiobjective optimization problem solved by minimizing a (given) weighted sum of objectives with as many PDE-constraints as terms in the sum.

Now, in the context of PDE-based optimization, solving (10) is an overwhelming task, and we need to search for alternative formulations. Following the discussion in [29, 30], given a functional $F$ depending on a set of random input variables $(\sigma, \delta)$, with mean $(E(\sigma), E(\delta))$ and standard deviations $(s_\sigma, s_\delta)$, we have the following second-order Taylor expansion:

$$
E(F(\sigma, \delta)) = F(E(\sigma), E(\delta)) + \frac{1}{2} \frac{\partial^2 F}{\partial \sigma^2} s_\sigma^2 + \frac{1}{2} \frac{\partial^2 F}{\partial \delta^2} s_\delta^2 + \frac{1}{2} \frac{\partial^2 F}{\partial \sigma \partial \delta} s_\sigma s_\delta + O(s^4),
$$

where the second derivatives are evaluated at $(E(\sigma), E(\delta))$. From this formula we conclude that a first-order approximation to (10) corresponds to $\min_{u \in L^2(Q)} \hat{J}(u)$, where the state equation has $E(\sigma)$ and $E(\delta)$ as diffusion and reaction coefficients, respectively. Further notice that the state equation is nonlinear and therefore, the control obtained using the mean diffusion and reaction coefficients $u_{E(\sigma)}, E(\delta)$ is not the same as the mean of the controls $E(u)$ obtained by averaging over all $(\sigma, \delta)$ configurations. Therefore, while the Taylor series expansion above shows that $u_{E(\sigma)}, E(\delta)$ is a first-order approximate solution to (10), it appears reasonable to ask whether or not $\hat{u} = E(u)$ can be also an approximate minimizer. This is not clear, since, in general, we expect

$$
J(y(u_{E(\sigma)}, E(\delta)), u_{E(\sigma)}, E(\delta)) \neq J(y(E(u)), E(u)).
$$

In the section of numerical experiments, we compare these two approximation strategies in the effort of designing a robust optimal control. It turns out that $\hat{u}$ may represent an improvement upon $u_{E(\sigma), E(\delta)}$.

4. Discretization of the probabilistic and physical spaces. We recall that the solution of a stochastic PDE problem is the set of all deterministic solutions corresponding to all possible events. A possible approximation to this set is to find a finite representation of the probability space of events. One possible approach is to represent stochastic PDE problems using the so-called generalized polynomial chaos expansion [14, 37]. In this representation, the dependent variables are considered as random processes and are written in the form $y(x, t, \omega) = \sum_{i=1}^{K} y_i(x, t) \psi_i(\xi(\omega))$, where $y_i(x, t)$ are deterministic functions and the $\psi_i$ are the Askey polynomials which
are chosen depending on the type of stochastic process $\xi(\omega)$ as input; see [37]. This approach is difficult to implement and requires us to solve very large-sized coupled systems. As the number of random input variables increases, the dimension of the approximating space and thus the computation time increase exponentially fast.

A more viable approach is by collocation in the probability space, and the temporal-spatial domain is approximated using a classical discretization scheme. In this approach, a tensor grid on the stochastic domain built upon the zeros of orthogonal polynomials is used. On this grid, we need to solve (5)–(6) for each grid point, and all statistical quantities related to these solutions are obtained by integration on this grid resulting from the Cartesian product of $\mathcal{N} = N + M$ one-dimensional grids, with $(m_1, \ldots, m_N, m_{N+1}, \ldots, m_{N+M})$ interpolation nodes in the respective dimensions. Let $m_j = 2^{j-1} + 1$ be the number of nodes for a quadrature scheme of order $j > 1$ and $m_1 = 1$. Following [26], we consider Chebyshev nodes given by

$$z_i^j = -\cos \frac{\pi(i - 1)}{m_j - 1}, i = 1, \ldots, m_j.$$ 

In addition to having an explicit formula, the Chebyshev nodes are chosen for their low Lebesgue number and because they form a nested set of nodes [12, 19, 26, 36]. In fact, if we write the set of Chebyshev nodes of order $m$ as

$$\{z^m\} = \{z_1^m, \ldots, z_m^m\},$$

then the nesting property gives that $\{z^j\} \subset \{z^{j+1}\}$.

Let $\{w^j\}$ be the set the quadrature weights in one dimension which correspond to the nodes $\{z^j\}$, the weights in multiple dimensions on a tensor product grid are written as a tensor product of one-dimensional weights $w = w^{j_1} \otimes \cdots \otimes w^{j_N}$. Thus, in particular, the mean value of a function $f : \mathbb{R}^\mathcal{N} \rightarrow \mathbb{R}$ is given by

$$E(f) = \sum_{i_1 = 1}^{m_1} \cdots \sum_{i_\mathcal{N} = 1}^{m_\mathcal{N}} f(z_{i_1}^{j_1}, \ldots, z_{i_\mathcal{N}}^{j_\mathcal{N}}) (w^{j_1} \otimes \cdots \otimes w^{j_\mathcal{N}}),$$

and the variance is $Var(f) = E((f - E(f))^2)$.

We see that the formulae above require $m_1 \times \cdots \times m_\mathcal{N}$ function evaluations, and this effort grows exponentially with the number of dimensions. To circumvent this limitation, we use a Smolyak scheme [12, 13, 26, 36] that constructs multivariate interpolation as linear combination of tensor-product formulas on a minimal number of nodes of the multidimensional space. A full $\mathcal{N}$-dimensional grid of order $\mathcal{J}$, with $j_1 = j_2 = \cdots = j_\mathcal{N} = \mathcal{J}$ in each dimension, is formed using a tensor product of the constituent one-dimensional grids and has total order $j_1 + j_2 + \cdots + j_\mathcal{N} = \mathcal{N} \mathcal{J}$. On the other hand, the sparse grid of order $\mathcal{J}$ is composed of a strict subset of full grids, where $\mathcal{J}$ is the order of the largest allowed grid and the orders of the constituent grids add up to a total order given by $\mathcal{J} + \mathcal{N} - 1$. To better explain this point, consider $\mathcal{N} = 2$ and a sparse grid of order $\mathcal{J} = 4$. This grid results in

$$\{z_1^1 \otimes z_1^1\} \cup \{z_1^2 \otimes z_1^1\} \cup \{z_1^3 \otimes z_1^2\} \cup \{z_1^4 \otimes z_1^3\}.$$ 

The resulting total number of sparse-grid points $\chi$ satisfies the following bounds [26]:

$$\mathcal{N}(2^\mathcal{J} - 1) \leq \chi \leq (2e\mathcal{N})^\mathcal{J} \min\{\mathcal{J} + 1, 2e\mathcal{N}\}.$$ 

To define the weights for sparse-grid integration, we need to define the difference weights of order $j$, which are $\eta^j = w^j - w^{j-1}$. Of course, this difference takes place on the odd-numbered nodes only. The quadrature weights on a full grid with order
Define \( j = (j_1, \ldots, j_N) \) would be the tensor product \( w = w^{j_1} \otimes \cdots \otimes w^{j_N} \). On the sparse grid of dimension \( N \) and total order \( J \), the weights are given by
\[
w^J = \sum_{i=1}^{N+J-1} \sum_{|j|=\ell} (\eta^{j_1} \otimes \cdots \otimes \eta^{j_N}).
\]
For more details on sparse grids, see [10, 19]. A MATLAB code for the setting of the sparse grids and weights described above is given in [8].

For completeness, we conclude this section illustrating the discretization of the optimality system (7) by finite differences and the backward Euler scheme. Assume \( \Omega \) is a square and \( \Omega_h \) is a uniform space mesh, where \( h \) is the mesh size. \( \Omega_h \) defines the set of interior mesh-points, \((x_i, y_j) = ((i-1)h, (j-1)h)\), \(2 \leq i, j \leq N_x\). On this mesh, \(-\Delta_h \) denotes the negative Laplacian and is approximated by the common five-point stencil. For grid functions \( v_h \) and \( w_h \) defined on \( \Omega_h \), we have the discrete \( L^2(\Omega) \)-scalar product
\[
(v_h, w_h)_{L^2(\Omega_h)} = h^2 \sum_{x \in \Omega_h} v_h(x) w_h(x),
\]
with associated norm \(|v_h| = (v_h, v_h)^{1/2}_{L^2(\Omega_h)}\). Let \( \delta t = T/N_t \) be the time step size. Define
\[
Q_{h, \delta t} = \{(x, t_m) : x \in \Omega_h, t_m = (m-1)\delta t, 1 \leq m \leq N_t + 1\}.
\]

On this grid, \( y_h^m \) and \( p_h^m \) denote grid functions at time level \( m \). The action of the time difference operator on these functions is as follows:
\[
\partial_{t^-} y_h^m = \frac{y_h^m - y_h^{m-1}}{\delta t} \quad \text{and} \quad \partial_{t^+} p_h^m = \frac{p_h^{m+1} - p_h^m}{\delta t},
\]
respectively; see [3] for details. For grid functions defined on \( Q_{h, \delta t} \), we use the discrete \( L^2(Q) \) scalar product, with norm \(||v_h, \delta t|| = (v_h, v_h)^{1/2}_{L^2(Q_{h, \delta t})}\). Later we use \( \gamma = \delta t/h^2 \).

For simplicity, we assume sufficient regularity of the data \( y_d, y_T \) such that these functions are properly approximated by their values at grid points. With this setting, the following discretization of (7) is obtained:
\[
\begin{align*}
-\partial_{t^-} y_h^m + G_{\delta}(y_h^m) + \sigma \Delta_h y_h^m &= f_h^m + u_h^m, \\
\partial_{t^+} p_h^m + G_{\delta}(y_h^m) p_h^m + \sigma \Delta_h p_h^m + \alpha(y_h^m - y_h^{m+1}) &= 0, \\
u y_h^m - p_h^m &= 0.
\end{align*}
\]
(11)

For the present finite difference setting, see [3] and the discussion in [6]. Under suitable conditions, the following estimates for the discretization error of the deterministic solution are obtained:
\[
\begin{align*}
||y_h - y|| &\leq c h^2, \\
||p_h - p|| &\leq c h^2, \text{ and } ||u_h - u|| \leq c h^2,
\end{align*}
\]
(12)
assuming there exist positive constants \( c_1 \leq c_2 \) such that \( c_1 h^2 \leq \delta t \leq c_2 h^2 \). Following \([1, 26]\) in the case of elliptic problems with random coefficients using a collocation setting, we can argue that the stochastic solution is affected by an error having three
components: 1. The error due to truncation of the KL expansion, denoted by \( e_2 \); 2. The discretization error (12), denoted by \( e_d \); 3. The interpolation error due to the collocation approach on the Smolyak sparse grid, denoted by \( e_S \). Concerning the Smolyak interpolation error, we have the following result [26] stating algebraic convergence with respect to the total number of sparse-grid collocation points:

\[
e_S \leq c \chi^{-g/(1+\log(2^N))},
\]

where \( g \) and \( c \) are positive constants that do not depend on \( \chi \).

5. **Collective smoothing multigrid schemes.** The advantage of the stochastic collocation approach on sparse grids with respect to, e.g., Monte Carlo simulation, is to greatly reduce the number of solver calls. Nevertheless, a large number of optimality system solves is required, which demands efficient solution strategies that are robust with respect to the choice of values of the optimization parameters. To meet these requirements, a space-time multigrid strategy has been proposed [3, 4, 5]. In this section, we illustrate this strategy and in the appendix, we present two-grid Fourier analysis of the resulting multigrid schemes.

Consider \( L \) grid levels indexed by \( k = 1, \ldots, L \), where \( k = L \) refers to the finest grid. The mesh of level \( k \) is denoted by \( Q_k = Q_{h_k, \delta t_k} \), where \( h_k = h_1/2^{k-1} \) and \( \delta t_k = \delta t; \) thus we choose semicoarsening in space. Any operator and variable defined on the discrete space-time cylinder \( Q_k \) is indexed by \( k \). The optimality system at level \( k \) with given initial, terminal, and boundary conditions is represented by the following nonlinear equation:

\[
A_k(w_k) = f_k, \quad w_k = (y_k, u_k, p_k).
\]

As well known [9, 16, 34, 35], the multigrid strategy combines two complementary schemes. The high-frequency components of the solution error are reduced by a smoothing iteration, denoted by \( S_k \) and defined in the following subsection, while the low-frequency error components are effectively reduced by a coarse-grid correction method as defined below.

The action of one multigrid cycle applied to (13) can be expressed in terms of a (nonlinear) multigrid iteration operator \( B_k \). Starting with an initial approximation \( w_k^{(0)} \), the result of one \( V(\nu_1, \nu_2) \)-cycle is then denoted by \( w_k = B_k(w_k^{(0)}) f_k \).

**Algorithm 1** (multigrid \( V(\nu_1, \nu_2) \)-cycle). Set \( B_1(w_1^{(0)}) \approx A_1^{-1} \) (e.g., iterating with \( S_1 \) starting with \( w_1^{(0)} \)). For \( k = 2, \ldots, L \), define \( B_k \) in terms of \( B_{k-1} \) as follows.

1. Set the starting approximation \( w_k^{(0)} \).
2. **Presmoothing.** Define \( w_k^{(l)} \) for \( l = 1, \ldots, \nu_1 \), by

\[
w_k^{(l)} = S_k(w_k^{(l-1)}, f_k).
\]

3. **Coarse-grid correction.** Set \( w_k^{(\nu_1+1)} = w_k^{(\nu_1)} + I_{k-1}^{k-1}(w_{k-1} - \hat{I}_{k-1}^{k-1} w_k^{(\nu_1)}) \), where

\[
w_{k-1} = B_{k-1}(\hat{I}_{k-1}^{k-1} w_k^{(\nu_1)}) \left[ \hat{I}_{k-1}^{k-1}(f_k - A_k(w_k^{(\nu_1)})) + A_{k-1}(\hat{I}_{k-1}^{k-1} w_k^{(\nu_1)}) \right].
\]

4. **Postsmoothing.** Define \( w_k^{(l)} \) for \( l = \nu_1 + 2, \ldots, \nu_1 + \nu_2 + 1 \), by

\[
w_k^{(l)} = S_k(w_k^{(l-1)}, f_k).
\]

5. Set \( B_k(w_k^{(0)}) f_k = w_k^{(\nu_1+\nu_2+1)} \).
In our implementation, we choose $I_{k-1}^k$ to be the full-weighted restriction operator [34] in space with no averaging in the time direction. The prolongation $I_{k-1}^k$ is defined by bilinear interpolation in space. We choose $I_{k-1}^k$ to be straight injection. We have that no interpolation in time is needed. Indeed, other choices of prolongation and restriction operators are possible; see [18].

We remark that a key component in the development of multigrid solvers for optimality systems is the design of robust collective smoothing schemes [7]. To illustrate the construction of these schemes, consider (11) in expanded form. In case of no constraints on the control, we can eliminate the control variable by means of the optimality condition $\nu u^m_h - p^m_h = 0$. We have

\begin{equation}
- [1 + 4\sigma^2] y_{ij} m + \sigma \gamma [y_{i+1,j} m + y_{i,j-1} m + y_{i,j+1} m + y_{i,j-1} m] + y_{ij} m-1
\end{equation}

+ $\delta t G_{\delta}(y_{ij} m) - \frac{\delta t}{\nu} p_{ij} m = \delta t f_{ij} m, \quad 2 \leq m \leq N_t + 1,$

\begin{equation}
- [1 + 4\sigma^2] p_{ij} m + \sigma \gamma [p_{i+1,j} m + p_{i,j-1} m + p_{i,j+1} m + p_{i,j-1} m] + p_{ij} m+1
\end{equation}

+ $\delta t G_{\delta}(y_{ij} m) + \sigma \gamma (y_{ij} m - y_{di} j m) = 0, \quad 1 \leq m \leq N_t.$

In case of terminal observation, at $t_m = T$ we have (8) in place of (15).

First, we define a pointwise smoothing scheme. Consider a collective Gauss–Seidel (–Newton) step which is applied at each space-time grid point to update $u_{ij} m = (y_{ij} m, p_{ij} m)$. For this purpose, consider (14) and (15) for the two variables $y_{ij} m$ and $p_{ij} m$ at the grid point $i j m$. We can refer to the left-hand sides of (14) and (15) as the negative of the residuals $r_y(u_{ij} m)$ and $r_p(u_{ij} m)$, respectively. A step of a collective smoothing iteration at this point consists of a local (Newton) update given by

\begin{equation}
\begin{pmatrix}
y \\
p
\end{pmatrix}^{(1)}_{ij m} = \begin{pmatrix}
y \\
p
\end{pmatrix}^{(0)}_{ij m} + \begin{pmatrix}
- (1 + 4\sigma^2) + \delta t G_{\delta} \\
\delta t (\alpha + G_{\delta}^p)
\end{pmatrix}_{ij m} \begin{pmatrix}
\delta t/\nu \\
\end{pmatrix}_{ij m}^{-1} \begin{pmatrix}
r_y \\
r_p
\end{pmatrix}^{(0)}_{ij m},
\end{equation}

where $r_y$ and $r_p$ denote the residuals at $i j m$ prior to the update. While a sweep of this smoothing iteration can be performed in any ordering of $i, j$, the problem of how to proceed along time direction arises since we need to take into account the opposite time orientation of the state equation and of the adjoint equation. For this purpose, to update the state variable, we use the first vector component of (16) marching in the forward direction, and the adjoint variable $p$ is being updated using the second component of (16) marching backward in time. In this way, a robust iteration is obtained given by the following algorithm [3, 4].

**Algorithm 2** (time-splitted collective Gauss–Seidel iteration (TS-CGS)).

1. Set the starting approximation.
2. For $m = 2, \ldots, N_t$ do
3. For $ij$ in, e.g., lexicographic order do

\begin{equation}
y^{(1)}_{ij m} = y^{(0)}_{ij m} + \begin{pmatrix}
- (1 + 4\sigma^2) + \delta t G_{\delta} \\
\delta t (\alpha + G_{\delta}^p)
\end{pmatrix}_{ij m} \begin{pmatrix}
\delta t/\nu \\
\end{pmatrix}_{ij m}^{-1} \begin{pmatrix}
r_y \\
r_p
\end{pmatrix}^{(0)}_{ij m},
\end{equation}

\begin{equation}
p_{ij,N_t-m+2}^{(1)} = p_{ij,N_t-m+2}^{(0)} + \begin{pmatrix}
- (1 + 4\sigma^2) + \delta t G_{\delta} \\
\delta t (\alpha + G_{\delta}^p)
\end{pmatrix}_{ij m} \begin{pmatrix}
\delta t/\nu \\
\end{pmatrix}_{ij m}^{-1} \begin{pmatrix}
r_y \\
r_p
\end{pmatrix}^{(0)}_{ij,N_t-m+2};
\end{equation}

4. end.
In the regime of small $\sigma$ (or $\gamma$), however, the TS-CGS iteration cannot provide robust smoothing because the coupling in the space direction becomes weak and therefore, pointwise relaxation in space is not effective in reducing the high-frequency components of the error. To overcome this problem, block-relaxation of the variables that are strongly connected must be performed. In our case, this means solving for the pairs of state and adjoint variables along the time-direction for each space coordinate.

To describe this block Gauss–Seidel procedure, consider the discrete optimality system (11) at any $i, j$ and for all time steps. For simplicity, we use the optimality condition to eliminate the control variable. Thus for each spatial grid point $i, j$, a block-tridiagonal system is obtained, where each block is a $2 \times 2$ matrix corresponding to the pair $(y, p)$. This block-tridiagonal system has the following form:

\[
M = \begin{bmatrix}
A_2 & C_2 & & \\
B_3 & A_3 & C_3 & \\
& B_4 & A_4 & C_4 \\
& & C_{N_t} & \\
& & B_{N_t+1} & A_{N_t+1}
\end{bmatrix}.
\]

Centered at $t_m$, the entries $B_m, A_m, C_m$ refer to the variables $(y, p)$ at $t_{m-1}, t_m, t_{m+1}$, respectively. The block $A_m, m = 2, \ldots, N_t$, is given by

\[
A_m = \begin{bmatrix}
-(1 + 4\sigma\gamma) + \delta t G'_\delta & -\frac{\delta \ell}{\nu} \\
\delta t (\alpha + G''_\delta p) & -(1 + 4\sigma\gamma) + \delta t G'_\delta
\end{bmatrix},
\]

where all functions within the brackets $[]$ are evaluated at $t_m$. Correspondingly, the $B_m$ and $C_m$ blocks are given by

\[
B_m = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad C_m = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.
\]

Clearly, for each time step, the variables neighboring the point $ij$ are taken as constant and contribute to the right-hand side of the system.

It remains to discuss the block $A_{N_t+1}$ for $\beta \neq 0$. At $t_m = T$, we have the terminal condition (8), which we rewrite as

\[
\beta (y^m_{N_t+1} - y^m_{Th}) - p^m_{N_t+1} = 0, \quad m = N_t + 1.
\]

Thus, the block $A_{N_t+1}$ is given by

\[
A_{N_t+1} = \begin{bmatrix}
-(1 + 4\sigma\gamma) + \delta t G'_\delta & -\frac{\delta \ell}{\nu} \\
\beta & -1
\end{bmatrix}.
\]

For each $i, j$, we have to solve a tridiagonal system $Mw = r$, where $w = (y^2_h, p^2_h, \ldots, y^N_{N_t+1}, p^N_{N_t+1})$ and $r = (r_y(w^2), r_p(w^2), \ldots, r_y(w^{N_t+1}), r_p(w^{N_t+1}))$. In particular, we have $r_y(w^{N_t+1}) = p^N_{N_t+1} - \beta (y^N_{N_t+1} - y^N_{Th})$. Block-tridiagonal systems can be solved efficiently with $O(N_t)$ effort. A block-tridiagonal solver is given in [5]. Summarizing our collective $t$-line relaxation is given by the following algorithm [4, 5].
Fig. 1. Two realizations of $\sigma(x_1, x_2, \omega)$. Left: $Y_1 = 0.8$, $Y_2 = 1.0$, $Y_3 = -0.2$, $Y_4 = 0.7$. Right: $Y_1 = -0.8$, $Y_2 = -1.0$, $Y_3 = 0.7$, $Y_4 = 0.5$.

**Algorithm 3** (time-line collective Gauss–Seidel iteration (TL-CGS)).
1. Set the starting approximation.
2. For $ij$ in, e.g., lexicographic order do
   \[
   \left( \begin{array}{c} y \\ p \end{array} \right)_{ij}^{(1)} = \left( \begin{array}{c} y \\ p \end{array} \right)_{ij}^{(0)} + M^{-1} \left( \begin{array}{c} r_y \\ r_p \end{array} \right)_{ij};
   \]
3. end.

Also in this case, $r_y$ and $r_p$ denote the residuals at $i, j$ and for all $m$ prior to the update. Since the solution in time is exact, no time splitting is required.

**6. Numerical experiments.** In this section, we validate the TS-CGS, and TL-CGS-based space-time multigrid scheme and sparse-grid collocation techniques for parabolic optimal control problems with random coefficients. We use the resulting computational tool to investigate robust control strategies obtained as the mean of the controls resulting from different realizations of diffusion and reaction fields and by using a reaction-diffusion model with average diffusion and reaction parameters. In the first part of this section, we describe problems' and algorithm's settings, then we report results of experiments.

We consider the case of a nonmonotone nonlinear reaction term given by $G_\delta(y) = \delta e^y$ that is used to model explosive combustion phenomena [6]. Depending on the choice of $\sigma$ and $\delta$, the resulting uncontrolled reaction-diffusion problem may have a stable solution or blowup at finite time.

We assume that the diffusivity and reactivity are random fields that are sufficiently well approximated by a truncated KL expansion. In the case of the diffusion coefficient, we ensure positiveness assuming, as in [1, 26], a KL expansion in the form (4) for $\log(\sigma(\omega) - \sigma_0)$. Therefore, we consider a random diffusivity coefficient given by
\[
\sigma(x_1, x_2, \omega) = \sigma_0 + \exp\{[Y_1(\omega)\cos(\pi x_2) + Y_3(\omega)\sin(\pi x_2)]e^{-1/8} \\
+ [Y_2(\omega)\cos(\pi x_1) + Y_4(\omega)\sin(\pi x_1)]e^{-1/8}\},
\]
where $\sigma_0 = 1/100$ and $Y_j \in [-1, 1], j = 1, 2, 3, 4$. This field is characterized by a squared exponential covariance typical of Gaussian processes. Two realizations of $\sigma(x_1, x_2, \omega)$ are depicted in Figure 1.

For the reaction term, we consider the following model of uncertainty:
\[
\delta(t, \omega) = \delta_0 + W(\omega)\sin(2\pi t/T),
\]
where $\delta_0 = 1$ and $W \in [-1, 1]$. For simplicity, we consider there are no background thermal sources so that $f = 0$. Notice that source terms enter linearly into the problem, while diffusivity and reactivity parameters enter bilinearly, i.e., they multiply the state variable and its derivatives and therefore, result in more complex nonlinear problem, while diffusivity and reactivity parameters enter bilinearly, i.e., they multiply the state variable and its derivatives and therefore, result in more complex nonlinear problem.

The desired target trajectory is given by

$$y_d(x, t) = (1 + t) \left(x_1 - x_1^2\right) \left(x_2 - x_2^2\right) \cos(4\pi t).$$

This is an oscillating function whose amplitude increases linearly with time. We take $y_T(x) = y_d(x, T)$.

We use one presmoothing and one postsmoothing step ($\nu_1 = \nu_2 = 1$), and $h = 1/4$ is the coarsest space mesh size. We take $\Omega = (0, 1) \times (0, 1)$ and $T = 1$. Two different grids $N_x \times N_y \times N_t$ are considered: $32 \times 32 \times 32$ and $64 \times 64 \times 64$, which result in $\gamma = 32$ and $\gamma = 64$, respectively. For the discretization of the stochastic space, we use a five-dimensional sparse grid and present results with order $J = 2$ and $J = 3$. These correspond to a total number of sparse-grid points of $\chi = 61$ and $\chi = 241$, respectively. Results of numerical experiments and results in [38] show that the first few moments can be accurately captured by a Smolyak order two interpolation scheme.

To describe the results of the experiments, we report mean values of the observed multigrid convergence factors $\eta_{obs}$, which are defined as the asymptotic value of the ratio of the norm of the dynamic residuals given by $||r_y|| + ||r_p||/\nu$ resulting from two successive multigrid cycles. The purpose of considering mean values of convergence factors is to investigate robustness of the computational performance with respect to randomness of the coefficients. In all experiments, we iterate the multigrid algorithm to a specified stopping criteria given by $||r_y|| + ||r_p||/\nu < 10^{-10}$.

The tracking ability of the space-time multigrid algorithms will be expressed in terms of the mean values of the norms of the tracking error $E(||y - y_d||)$ and of the terminal observation error $E(||y_T||).$ Other important statistical observable are the variance of the tracking error $Var \left(||y - y_d||\right) = E((||y - y_d|| - E(||y - y_d||))^2)$ and its skewness $Skew \left(||y - y_d||\right) = E((||y - y_d|| - E(||y - y_d||))^3)$.  

Next, we report results of experiments to validate the multigrid and sparse-grids approach and the optimization properties of the optimal control formulation. 

The results reported in Tables 1 and 2 in the case of tracking of trajectories illustrate efficiency and robustness of the proposed multigrid solvers. These results demonstrate usual multigrid convergence speeds which appear to be independent of the value of $\gamma$ and $\nu$, the observed convergence factor are close to that estimated by local Fourier analysis; see the appendix. Comparing Table 1 and Table 2, we see that the TS-CGS scheme and the TL-CGS scheme have similar numerical performance.

In Table 1, we can see how CPU times scale depending on the level of discretization of the physical and stochastic spaces. Notice that the CPU times scale approximately as a factor of $2^3$, that is, linearly with the number of space-time grid points. Moreover, as we move from an order-two to an order-three stochastic sparse grid, we obtain optimal complexity in CPU time. We find less improvement of the mean and of the other moments of the objective when we refine the sparse-grid mesh.

The results reported in Tables 1 and 2 also demonstrate the tracking ability of the optimal control formulation. As the value of $\nu$ decreases, smaller values of $E(||y - y_d||)$ are obtained as desired in an optimal control framework. Further numerical experiments demonstrate that the multigrid convergence behavior appears to be insensitive to the particular choice of the desired tracking function, which may not
Results for \( \alpha = 1, \beta = 0 \), with the TS-CGS multigrid scheme. Denote \( \| \Delta y \| = \| y - y_d \| \).

<table>
<thead>
<tr>
<th>( \nu )</th>
<th>( N_x \times N_y \times N_t )</th>
<th>CPU(s)</th>
<th>( \mathbb{E}(|\Delta y|) )</th>
<th>( \text{Var}(|\Delta y|) )</th>
<th>Skew((|\Delta y|))</th>
<th>( \mathbb{E}(\rho_{\text{obs}}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^{-4} )</td>
<td>32 ( \times ) 32 ( \times ) 32</td>
<td>77.23</td>
<td>3.31 \times 10^{-2}</td>
<td>4.61 \times 10^{-4}</td>
<td>2.23 \times 10^{-5}</td>
<td>0.12</td>
</tr>
<tr>
<td>( 10^{-6} )</td>
<td>32 ( \times ) 32 ( \times ) 32</td>
<td>87.89</td>
<td>1.15 \times 10^{-3}</td>
<td>5.84 \times 10^{-7}</td>
<td>5.79 \times 10^{-10}</td>
<td>0.08</td>
</tr>
<tr>
<td>( 10^{-8} )</td>
<td>32 ( \times ) 32 ( \times ) 32</td>
<td>69.75</td>
<td>3.48 \times 10^{-5}</td>
<td>6.45 \times 10^{-10}</td>
<td>2.59 \times 10^{-14}</td>
<td>0.08</td>
</tr>
</tbody>
</table>

<p>| ( \mathcal{J} = 3, \chi = 241 ) |</p>
<table>
<thead>
<tr>
<th>( \nu )</th>
<th>( N_x \times N_y \times N_t )</th>
<th>CPU(s)</th>
<th>( \mathbb{E}(|\Delta y|) )</th>
<th>( \text{Var}(|\Delta y|) )</th>
<th>Skew((|\Delta y|))</th>
<th>( \mathbb{E}(\rho_{\text{obs}}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^{-4} )</td>
<td>64 ( \times ) 64 ( \times ) 64</td>
<td>613.79</td>
<td>3.27 \times 10^{-2}</td>
<td>4.64 \times 10^{-4}</td>
<td>2.29 \times 10^{-5}</td>
<td>0.12</td>
</tr>
<tr>
<td>( 10^{-6} )</td>
<td>64 ( \times ) 64 ( \times ) 64</td>
<td>722.46</td>
<td>1.18 \times 10^{-3}</td>
<td>5.94 \times 10^{-6}</td>
<td>5.93 \times 10^{-10}</td>
<td>0.12</td>
</tr>
<tr>
<td>( 10^{-8} )</td>
<td>64 ( \times ) 64 ( \times ) 64</td>
<td>795.68</td>
<td>4.75 \times 10^{-5}</td>
<td>7.08 \times 10^{-10}</td>
<td>2.80 \times 10^{-14}</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Results for \( \alpha = 1, \beta = 0, \) and \( \mathcal{J} = 2, \chi = 61, \) with the TL-CGS multigrid scheme. Denote \( \| \Delta y \| = \| y - y_d \| \).

<table>
<thead>
<tr>
<th>( \nu )</th>
<th>( N_x \times N_y \times N_t )</th>
<th>CPU(s)</th>
<th>( \mathbb{E}(|\Delta y|) )</th>
<th>( \text{Var}(|\Delta y|) )</th>
<th>Skew((|\Delta y|))</th>
<th>( \mathbb{E}(\rho_{\text{obs}}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^{-4} )</td>
<td>32 ( \times ) 32 ( \times ) 32</td>
<td>312.89</td>
<td>3.31 \times 10^{-2}</td>
<td>4.33 \times 10^{-4}</td>
<td>9.26 \times 10^{-6}</td>
<td>0.13</td>
</tr>
<tr>
<td>( 10^{-6} )</td>
<td>32 ( \times ) 32 ( \times ) 32</td>
<td>353.51</td>
<td>1.15 \times 10^{-3}</td>
<td>6.27 \times 10^{-7}</td>
<td>1.75 \times 10^{-9}</td>
<td>0.12</td>
</tr>
<tr>
<td>( 10^{-8} )</td>
<td>32 ( \times ) 32 ( \times ) 32</td>
<td>297.82</td>
<td>3.49 \times 10^{-5}</td>
<td>4.83 \times 10^{-10}</td>
<td>2.02 \times 10^{-14}</td>
<td>0.08</td>
</tr>
<tr>
<td>( 10^{-4} )</td>
<td>64 ( \times ) 64 ( \times ) 64</td>
<td>2519.32</td>
<td>3.27 \times 10^{-2}</td>
<td>4.31 \times 10^{-4}</td>
<td>9.14 \times 10^{-6}</td>
<td>0.13</td>
</tr>
<tr>
<td>( 10^{-6} )</td>
<td>64 ( \times ) 64 ( \times ) 64</td>
<td>2867.01</td>
<td>1.18 \times 10^{-3}</td>
<td>6.37 \times 10^{-7}</td>
<td>1.80 \times 10^{-9}</td>
<td>0.13</td>
</tr>
<tr>
<td>( 10^{-8} )</td>
<td>64 ( \times ) 64 ( \times ) 64</td>
<td>3152.42</td>
<td>4.73 \times 10^{-5}</td>
<td>5.60 \times 10^{-10}</td>
<td>2.14 \times 10^{-14}</td>
<td>0.10</td>
</tr>
</tbody>
</table>

be attainable. In Tables 1 and 2, we also report the variance and the skewness of the tracking error. We see that these quantities are at least two orders of magnitude smaller than the mean value. This fact shows that the tracking ability of the optimal control formulation is less sensitive to the randomness of the parameters defining the reaction-diffusion process.

The choice of TL-CGS iteration results in a superior computational performance when only terminal observation is considered. Results for this case are reported in Table 3. Notice that, because \( \alpha = 0 \), the coupling between state and control variables is much weaker in the space-time domain, resulting in a less efficient TS-CGS-based multigrid solver. Despite of the weaker coupling, the use of TL-CGS smoothing provides a multigrid algorithm that performs equally well for tracking and for terminal observation. In Table 3, typical multigrid convergence factors are reported for the TL-CGS multigrid scheme applied to the terminal observation problem that show robustness with respect to the values of \( \gamma, \nu, \) and \( \sigma \). It is remarkable that the observed convergence factors are very close to that obtained in the case of tracking of trajectory. Nevertheless, notice that in this case, better \( \mathbb{E}(\rho_{\text{obs}}) \) are obtained for larger \( \nu \). On the other hand, even smaller values of the variance and skewness of the tracking error are obtained in the case of control with terminal observation. The results reported in the Tables 1–3 correspond to two different space-time meshes, and we see that the mean and the other moments of the tracking error are not very sensitive with respect to mesh refinement.
In previous sections, we discuss the problem of defining a unique control that provides good tracking features for all configurations of the coefficients. As a criteria to evaluate the confidence level for this control, we propose to compare, for a given value of the control parameter \( \nu \), the means and other moments of the tracking functional resulting from the optimal controls specific for each configuration (Tables 2–3) with the means and other moments of the tracking functional resulting from the proposed robust control applied to each model of the parameter space.

As possible control functions, we first consider the function \( \mathbb{E}(u) \) defined as the space-time mean function of the optimal controls corresponding to each point of the configuration space. This means computing the optimal controls for all possible parameters’ configurations modeled by the truncated KL expansion and determining the mean. Another possible way to define a unique control function is to consider average parameter values and compute the corresponding control. With these candidate control functions, we solve the forward problem corresponding to each realization of parameter values and compute the corresponding control. With these candidate control functions, for each event \( \omega \) represented on a grid point of the stochastic sparse grid, we obtain \( y_{\mathbb{E}(u)} \) and \( y_{\mathbb{E}(\sigma), \mathbb{E}(\delta)} \).

The question is then how good are \( y_{\mathbb{E}(u)} \) and \( y_{\mathbb{E}(\sigma), \mathbb{E}(\delta)} \) tracking the desired target for different parameter settings. We find that both approaches provide mean tracking errors that are close and less sensitive to the value of the weight of the cost of the control. In Table 4, we have the comparison of tracking errors for the case of tracking of trajectories with \( \nu = 10^{-6} \). The first row entry is obtained using \( \mathbb{E}(u) \) as a control function for all parameter configurations. The second row entry refers to a control obtained with average parameter values; see Figure 2. Similarly, in Table 5, we report results corresponding to terminal observation. We see that the proposed controls have similar tracking performance, although the approach with \( \mathbb{E}(u) \) performs better. Therefore, while it is true that \( J(y(u_{\mathbb{E}(\sigma), \mathbb{E}(\delta)}), u_{\mathbb{E}(\sigma), \mathbb{E}(\delta)}) \neq J(y(\mathbb{E}(u)), \mathbb{E}(u)) \), there is a difference in the tracking property of the two strategies thus motivating forthcoming investigative efforts.

### 7. Conclusions

Parabolic optimal control problems with random coefficients were formulated, and space-time multigrid methods combined with sparse-grid collocation techniques were implemented to robustly and efficiently solve these optimization problems. Multigrid schemes with typical multigrid convergence rates and

---

**Table 3**

Results for \( \alpha = 0, \beta = 1, \) and \( J = 2, \chi = 61, \) with the TL-CGS multigrid scheme. Denote \( |\Delta y| = |y - y'| \).

| \( \nu \) | \( N_x \times N_y \times N_t \) | CPU(s) | \( \mathbb{E}(|\Delta y|) \) | \( \text{Var}(|\Delta y|) \) | \( \text{Skew}(|\Delta y|) \) | \( E(\rho_{obs}) \) |
|---|---|---|---|---|---|---|
| \( 10^{-4} \) | \( 32 \times 32 \times 32 \) | 69.37 | \( 1.91 \times 10^{-4} \) | \( 1.51 \times 10^{-6} \) | \( 1.34 \times 10^{-9} \) | 0.02 |
| \( 10^{-6} \) | \( 32 \times 32 \times 32 \) | 69.21 | \( 1.90 \times 10^{-5} \) | \( 1.19 \times 10^{-10} \) | \( 1.41 \times 10^{-15} \) | 0.11 |
| \( 10^{-8} \) | \( 32 \times 32 \times 32 \) | 68.71 | \( 1.99 \times 10^{-7} \) | \( 1.91 \times 10^{-14} \) | \( 1.42 \times 10^{-21} \) | 0.11 |
| \( 10^{-4} \) | \( 64 \times 64 \times 64 \) | 552.32 | \( 2.17 \times 10^{-3} \) | \( 1.45 \times 10^{-6} \) | \( 1.91 \times 10^{-9} \) | 0.05 |
| \( 10^{-6} \) | \( 64 \times 64 \times 64 \) | 556.56 | \( 2.20 \times 10^{-5} \) | \( 1.50 \times 10^{-10} \) | \( 2.02 \times 10^{-15} \) | 0.09 |
| \( 10^{-8} \) | \( 64 \times 64 \times 64 \) | 556.31 | \( 2.20 \times 10^{-7} \) | \( 1.50 \times 10^{-14} \) | \( 2.00 \times 10^{-21} \) | 0.10 |

---

**Table 4**

Results with \( \alpha = 1, \beta = 0, \) and \( \nu = 10^{-6} \); \( N_x \times N_y \times N_t = 64 \times 64 \times 64, \chi = 241 \).

| \( y_{\mathbb{E}(u)} \) | \( E(||y - y||) \) | \( \text{Var}(||y - y||) \) | \( \text{Skew}(||y - y||) \) |
|---|---|---|---|
| \( y_{\mathbb{E}(\sigma), \mathbb{E}(\delta)} \) | \( 1.33 \times 10^{-1} \) | \( 1.20 \times 10^{-2} \) | \( 3.26 \times 10^{-4} \) |
| \( y_{\mathbb{E}(\sigma), \mathbb{E}(\delta)} \) | \( 1.43 \times 10^{-1} \) | \( 1.40 \times 10^{-2} \) | \( 4.44 \times 10^{-2} \) |
robustness with respect to a large choice of optimization parameters were discussed. These results showed sharp agreement with prediction of numerical performance obtained with stochastic twogrid Fourier analysis. Results of computation of stochastic optimal control solutions were reported with a focus on the moments of the tracking ability of the optimization schemes. The construction of robust control strategies was discussed.

This work focused on parabolic optimal control problems with linear distributed control. However, this framework can be applied to consider other classes of governing equations and different control mechanisms.

Appendix A. Fourier mode analysis of multigrid methods [9, 18, 34] for elliptic PDEs with random coefficients has been first proposed in [32] using polynomial chaos (PC) representation. The PC-multigrid approach, also considered in [11], leads to a large system of coupled deterministic PDEs, and the analysis in [32] requires us to apply local Fourier analysis to the multigrid scheme with the collective version of the local smoother applied to the coupled system. In this appendix, we illustrate a different approach, which is consistent with the stochastic collocation framework. For any given event resulting in specific values of $\delta$ and $\sigma$, we perform a twogrid Fourier analysis of the resulting deterministic problem and then compute the statistics of the local Fourier analysis estimates to obtain mean values and variance of convergence factors.

For the analysis that follows, we consider a linear case with $G_\delta(y) = \delta y$ and distributed control with the replacement $u = p/\nu$. We consider the optimality system represented by (13) with $w = (y, p)$ and assume an infinite grid and one space dimension. On the fine grid, consider the Fourier components $\phi(\mathbf{j}, \mathbf{\theta}) = e^{i\mathbf{j} \cdot \mathbf{\theta}}$, where $i$ is the imaginary unit, $\mathbf{j} = (j_x, j_t) \in \mathbb{Z} \times \mathbb{Z}$, $\mathbf{\theta} = (\theta_x, \theta_t) \in [-\pi, \pi)^2$, and $\mathbf{j} \cdot \mathbf{\theta} = j_x \theta_x + j_t \theta_t$.

In a semicoarsening setting, the frequency domain is spanned by the following two sets of frequencies:

$$\mathbf{\theta}^{(0,0)} := (\theta_x, \theta_t) \text{ and } \mathbf{\theta}^{(1,0)} := (\overline{\theta}_x, \theta_t),$$

robustness with respect to a large choice of optimization parameters were discussed. These results showed sharp agreement with prediction of numerical performance obtained with stochastic twogrid Fourier analysis. Results of computation of stochastic optimal control solutions were reported with a focus on the moments of the tracking ability of the optimization schemes. The construction of robust control strategies was discussed.

This work focused on parabolic optimal control problems with linear distributed control. However, this framework can be applied to consider other classes of governing equations and different control mechanisms.

Appendix A. Fourier mode analysis of multigrid methods [9, 18, 34] for elliptic PDEs with random coefficients has been first proposed in [32] using polynomial chaos (PC) representation. The PC-multigrid approach, also considered in [11], leads to a large system of coupled deterministic PDEs, and the analysis in [32] requires us to apply local Fourier analysis to the multigrid scheme with the collective version of the local smoother applied to the coupled system. In this appendix, we illustrate a different approach, which is consistent with the stochastic collocation framework. For any given event resulting in specific values of $\delta$ and $\sigma$, we perform a twogrid Fourier analysis of the resulting deterministic problem and then compute the statistics of the local Fourier analysis estimates to obtain mean values and variance of convergence factors.

For the analysis that follows, we consider a linear case with $G_\delta(y) = \delta y$ and distributed control with the replacement $u = p/\nu$. We consider the optimality system represented by (13) with $w = (y, p)$ and assume an infinite grid and one space dimension. On the fine grid, consider the Fourier components $\phi(\mathbf{j}, \mathbf{\theta}) = e^{i\mathbf{j} \cdot \mathbf{\theta}}$, where $i$ is the imaginary unit, $\mathbf{j} = (j_x, j_t) \in \mathbb{Z} \times \mathbb{Z}$, $\mathbf{\theta} = (\theta_x, \theta_t) \in [-\pi, \pi)^2$, and $\mathbf{j} \cdot \mathbf{\theta} = j_x \theta_x + j_t \theta_t$.

In a semicoarsening setting, the frequency domain is spanned by the following two sets of frequencies:

$$\mathbf{\theta}^{(0,0)} := (\theta_x, \theta_t) \text{ and } \mathbf{\theta}^{(1,0)} := (\overline{\theta}_x, \theta_t),$$

robustness with respect to a large choice of optimization parameters were discussed. These results showed sharp agreement with prediction of numerical performance obtained with stochastic twogrid Fourier analysis. Results of computation of stochastic optimal control solutions were reported with a focus on the moments of the tracking ability of the optimization schemes. The construction of robust control strategies was discussed.

This work focused on parabolic optimal control problems with linear distributed control. However, this framework can be applied to consider other classes of governing equations and different control mechanisms.

Appendix A. Fourier mode analysis of multigrid methods [9, 18, 34] for elliptic PDEs with random coefficients has been first proposed in [32] using polynomial chaos (PC) representation. The PC-multigrid approach, also considered in [11], leads to a large system of coupled deterministic PDEs, and the analysis in [32] requires us to apply local Fourier analysis to the multigrid scheme with the collective version of the local smoother applied to the coupled system. In this appendix, we illustrate a different approach, which is consistent with the stochastic collocation framework. For any given event resulting in specific values of $\delta$ and $\sigma$, we perform a twogrid Fourier analysis of the resulting deterministic problem and then compute the statistics of the local Fourier analysis estimates to obtain mean values and variance of convergence factors.

For the analysis that follows, we consider a linear case with $G_\delta(y) = \delta y$ and distributed control with the replacement $u = p/\nu$. We consider the optimality system represented by (13) with $w = (y, p)$ and assume an infinite grid and one space dimension. On the fine grid, consider the Fourier components $\phi(\mathbf{j}, \mathbf{\theta}) = e^{i\mathbf{j} \cdot \mathbf{\theta}}$, where $i$ is the imaginary unit, $\mathbf{j} = (j_x, j_t) \in \mathbb{Z} \times \mathbb{Z}$, $\mathbf{\theta} = (\theta_x, \theta_t) \in [-\pi, \pi)^2$, and $\mathbf{j} \cdot \mathbf{\theta} = j_x \theta_x + j_t \theta_t$.

In a semicoarsening setting, the frequency domain is spanned by the following two sets of frequencies:

$$\mathbf{\theta}^{(0,0)} := (\theta_x, \theta_t) \text{ and } \mathbf{\theta}^{(1,0)} := (\overline{\theta}_x, \theta_t),$$
where \((\theta_x, \theta_t) \in (-\pi/2, \pi/2) \times [-\pi, \pi]\) and \(\bar{\theta}_x = \theta_x - \text{sign}(\theta_x)\pi\). The components \(\phi(\cdot, \theta^\alpha)\) are called harmonics. The first harmonics \(\phi(\cdot, \theta^{(0,0)})\) represents low frequencies’ components in space. The second harmonics \(\phi(\cdot, \theta^{(1,0)})\) contains the high frequencies’ components in space direction. Both have all frequencies’ components in time direction. Using smoothers, we have that \(\phi(j, \theta^{(0,0)}) = \phi(j, \theta^{(1,0)})\) on the coarse grid.

The action of the multigrid scheme is to reduce the high-frequency error components by applying the smoothing operator \(S_k\) and to reduce the low-frequency error components by coarse grid correction given by

\[
CG_k^{k-1} = [I_k - I_k^{k-1} (A_{k-1})^{-1} I_k^{k-1} A_k].
\]

Denote with \(E_k^\theta = \text{span}[\phi_k(\cdot, \theta^\alpha) : \alpha \in \{(0,0), (1,0)\}]\). Under the assumption that all multigrid components are linear and that \((A_{k-1})^{-1}\) exists, we have a representation of the twogrid operator \(TG_k^{k-1}\) on the space \(E_k^\theta \times E_k^\theta\) by a 4 \times 4 matrix given by

\[
\hat{G}_{k}^{k-1}(\theta) = \hat{S}_{k}(\theta)^{r_2} \hat{C}_{k}^{k-1}(\theta) \hat{S}_{k}(\theta)^{r_1},
\]

where the hat denotes the Fourier symbol [34] of the given operator.

To determine the explicit form of the operator symbols given above, consider the action of the operators on the couple \((\hat{y}, \hat{p}) \in E_k^y \times E_k^p\) for a given \((\hat{j})\), where

\[
\begin{align*}
\hat{y} = & \sum_{\theta} \sum_{p=0,1} Y_{\theta}^{(p,0)} \phi_k(j, \theta^{(p,0)}) \quad \text{and} \quad \hat{\nu} = \sum_{\theta} \sum_{p=0,1} P_{\theta}^{(p,0)} \phi_k(j, \theta^{(p,0)}),
\end{align*}
\]

where \(\sum_{\theta}\) denotes formal summation in \(\theta = (\theta_x, \theta_t) \in (-\pi/2, \pi/2) \times [-\pi, \pi]\) and \(\hat{W}_\theta^\alpha = (\hat{Y}_\theta^\alpha, \hat{P}_\theta^\alpha)\) are the corresponding Fourier coefficients.

In the Fourier space, the action of one smoothing step can be expressed by \(\hat{W}_\theta^{\text{new}} = \hat{S}(\theta) \hat{W}_\theta^{\text{old}}\), where \(\hat{S}(\theta)\) is the Fourier symbol [34] of the smoothing itration. This operator applies to the two equations (14)–(15) acting on low- and high-frequency components. It has the following form:

\[
\hat{S}_{k}(\theta) = \text{diag}\{\hat{s}(\theta^{(0,0)}), \hat{s}(\theta^{(1,0)})\},
\]

where \(\hat{s}(\theta)\) is the 2 \times 2 Fourier symbol of the smoothing scheme for a generic \(\theta\).

A way to characterize the smoothing property of the operator \(S_k\) is to assume an ideal coarse grid correction which annihilates the low-frequency error components and leaves the high-frequency error components unchanged. That is, one defines the projection operator \(\overline{Q}_k^{k-1}\) on \(E_k^y \times E_k^p\) by

\[
\overline{Q}_k^{k-1}(\theta) = \begin{bmatrix}
Q_k^{k-1}(\theta) & 0 \\
0 & Q_k^{k-1}(\theta)
\end{bmatrix}, \quad \text{where} \quad Q_k^{k-1}(\theta) = \begin{cases}
\text{diag}\{0,0\} & \text{if} \ \theta = \theta^{(0,0)}; \\
\text{diag}\{1,1\} & \text{if} \ \theta = \theta^{(1,0)}.
\end{cases}
\]

In this framework, the smoothing property of \(S_k\) is defined as follows:

\[
\mu = \max\{r(\overline{Q}_k^{k-1}(\theta) \hat{S}_{k}(\theta)) : \theta \in (-\pi/2, \pi/2) \times [-\pi, \pi]\}
\]

where \(r\) is the spectral radius.

Now consider applying the TS-CGS step. We obtain

\[
\begin{pmatrix}
-(1 + 2\sigma \gamma - \delta t \delta) + \sigma \gamma e^{-i\theta_x} & -\frac{\delta t}{\nu} \\
\alpha \delta t & -(1 + 2\sigma \gamma - \delta t \delta) + \sigma \gamma e^{-i\theta_x}
\end{pmatrix}
\begin{pmatrix}
\hat{\nu}^{(\text{new})} \\
\hat{\nu}^{(\text{old})}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
-(e^{-i\theta_x} + \sigma \gamma e^{i\theta_x}) & 0 \\
0 & -(e^{i\theta_x} + \sigma \gamma e^{-i\theta_x})
\end{pmatrix}
\begin{pmatrix}
\hat{\nu}^{(\text{old})} \\
\hat{\nu}^{(\text{new})}
\end{pmatrix}.
\]
Hence

\[
\hat{s}(\theta) = \left( \frac{-(1 + 2\sigma - \delta t\delta) + \sigma \gamma e^{-i\theta x}}{\alpha \delta t} \right)^{-1} \times \left( \begin{array}{cc}
-(1 + 2\sigma - \delta t\delta) + \sigma \gamma e^{-i\theta x} & -\frac{\delta t}{\nu} \\
0 & 0
\end{array} \right) \]

Next, consider the case of TL-CGS relaxation. The Fourier symbol of the smoothing operator is given by the following \(2 \times 2\) matrix:

\[
\hat{s}(\theta) = -(A + B e^{-i\theta t} + C e^{i\theta t} + \tilde{I} e^{-i\theta t})^{-1}(\tilde{I} e^{i\theta t}),
\]

where

\[
A = \left[ \begin{array}{cc}
-(1 + 2\sigma - \delta t\delta) & -\frac{\delta t}{\nu} \\
\alpha \delta t & -(1 + 2\sigma - \delta t\delta)
\end{array} \right], \quad B_m = \left[ \begin{array}{cc} 1 & 0 \\
0 & 0 \end{array} \right], \quad \text{and} \quad C_m = \left[ \begin{array}{cc} 0 & 0 \\
0 & 1 \end{array} \right],
\]

and \(\tilde{I} = \sigma \gamma I, I\) is the \(2 \times 2\) identity matrix.

To investigate the action of the coarse-grid correction, we consider a full-weighting restriction operator whose symbol is given by

\[
\hat{\tilde{I}}_{k-1}(\theta) = \frac{1}{2} \left[ \begin{array}{cccc}
1 + \cos(\theta_x) & 0 & 1 - \cos(\theta_x) & 0 \\
0 & 1 + \cos(\theta_x) & 0 & 1 - \cos(\theta_x)
\end{array} \right].
\]

For the linear prolongation operator, we have \(\hat{\tilde{I}}_{k-1}(\theta) = \hat{\tilde{I}}_{k-1}(\theta)^T\). The symbol of the fine grid operator is

\[
\hat{A}_{k}(\theta) = \left[ \begin{array}{cccc}
\hat{a}_{y}(\theta(0,0)) & -\delta t/\nu & 0 & 0 \\
\alpha \delta t & \hat{a}_{y}(\theta(0,0)) & 0 & 0 \\
0 & 0 & \hat{a}_{y}(\theta(1,0)) & -\delta t/\nu \\
0 & 0 & \alpha \delta t & \hat{a}_{y}(\theta(1,0))
\end{array} \right],
\]

where

\[
\hat{a}_{y}(\theta) = 2\sigma \gamma \cos(\theta_x) - e^{-i\theta t} - 2\sigma \gamma + \delta t\delta - 1 \quad \text{and} \quad \hat{a}_{y}(\theta) = 2\sigma \gamma \cos(\theta_x) - e^{i\theta t} - 2\sigma \gamma - \delta t\delta - 1.
\]

The symbol of the coarse grid operator follows

\[
\hat{A}_{k-1}(\theta) = \left[ \begin{array}{cc}
\sigma \gamma \cos(2\theta_x)/2 - e^{-i\theta t} - \sigma \gamma/2 + \delta t\delta - 1 & -\delta t/\nu \\
\alpha \delta t & \sigma \gamma \cos(2\theta_x)/2 - e^{i\theta t} - \sigma \gamma/2 + \delta t\delta - 1
\end{array} \right].
\]

Notice that on the coarser grid, \(\delta t\) remains unchanged while \(\gamma \to \gamma/4\) by coarsening.

Based on the representation on \(\text{TG}^{-1}_{k}\) by a \(4 \times 4\) matrix \(\hat{\text{TG}}^{-1}_{k}(\theta)\), we can calculate the convergence factor given by

\[
\eta(\text{TG}^{-1}_{k}) = \sup(r(\hat{\text{TG}}^{-1}_{k}(\theta)) : \theta \in (-\pi/2, \pi/2) \times (-\pi, \pi)),
\]

that requires us to determine the spectral radius of a \(4 \times 4\) matrix. It results that \(\mu\) and \(\eta\) are almost independent of the value of the weight \(\nu\) and of the discretization parameter \(\gamma\) for a large range of choices of values of these parameters. For \(\sigma = 0\), no spatial coupling is present, and the TL-CGS scheme becomes an exact solver, i.e.,
The convergence factor $\eta$ for TL-CGS and TS-CGS multigrid schemes ($\nu_1 = \nu_2 = 1$); $\delta t = 1/64$, $\delta = 0$, $\sigma = 1$, $\alpha = 1$.

<table>
<thead>
<tr>
<th>$\gamma \setminus \nu$</th>
<th>TL-CGS</th>
<th>TS-CGS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$10^{-8}$</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>16</td>
<td>0.01</td>
<td>0.12</td>
</tr>
<tr>
<td>32</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td>48</td>
<td>0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>64</td>
<td>0.11</td>
<td>0.13</td>
</tr>
</tbody>
</table>

The mean $E(\eta)$ for TL-CGS and TS-CGS multigrid schemes ($\nu_1 = \nu_2 = 1$); $\delta t = 1/64$, $\alpha = 1$.

<table>
<thead>
<tr>
<th>$\gamma \setminus \nu$</th>
<th>TL-CGS</th>
<th>TS-CGS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$10^{-8}$</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>16</td>
<td>0.01</td>
<td>0.11</td>
</tr>
<tr>
<td>32</td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td>48</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td>64</td>
<td>0.09</td>
<td>0.11</td>
</tr>
</tbody>
</table>

The variance $Var(\eta) = E((\eta - E(\eta))^2)$ for TL-CGS and TS-CGS multigrid schemes ($\nu_1 = \nu_2 = 1$); $\delta t = 1/64$, $\alpha = 1$.

<table>
<thead>
<tr>
<th>$\gamma \setminus \nu$</th>
<th>TL-CGS</th>
<th>TS-CGS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$10^{-8}$</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>16</td>
<td>0.0002</td>
<td>0.0018</td>
</tr>
<tr>
<td>32</td>
<td>0.0015</td>
<td>0.0019</td>
</tr>
<tr>
<td>48</td>
<td>0.0026</td>
<td>0.0019</td>
</tr>
<tr>
<td>64</td>
<td>0.0028</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

$\mu = 0$ results. On the other hand, for moderate values of $\nu$ and corresponding to smaller values of $\sigma$, the convergence factors of the TS-CGS multigrid scheme worsen. In Table 6, local Fourier analysis quantitative estimates of the convergence factor of TL-CGS and TS-CGS multigrid schemes are given. We see that both schemes result in convergence factors that are typical of multigrid schemes for Poisson problems.

The previous estimates consider a deterministic case corresponding to an event in the stochastic space of the coefficients $\delta$ and $\sigma$ of our model. In the next step, we estimate the convergence factors of our multigrid schemes on the set of parameter values defined on a sparse grid that represents the stochastic space of the coefficients. We assume that $\sigma \in [0, 2]$ and $\delta \in [-1, 1]$ are uniformly distributed random variables. We consider a two-dimensional sparse grid of order three [19] and remark that the results below do not change considering a larger order. On this grid, we compute the mean, variance, and skewness of the distribution of $\eta$ determined by local Fourier analysis. In Table 7, we report the posterior mean value of $\eta$, denoted by $E(\eta)$, and in Table 8, we report the variance of the probability distribution of $\eta$, denoted by $Var(\eta) = E((\eta - E(\eta))^2)$. We see that $Var(\eta)$ is less sensitive to mesh and optimization parameters. In the TS-CGS cases, it increases by increasing $\nu$. Notice that in all cases, the variance of the average convergence factor is small, indicating that the computational performance of the proposed multigrid schemes is less sensitive to changes of values of the coefficients of the problem.

Skewness $E((\eta - E(\eta))^3)$ of $\eta$ distribution results to be two order of magnitude smaller than the variance. That is, the distribution is almost symmetrical.
Acknowledgments. We would like to gratefully acknowledge Andreas Klimke for support and discussion regarding sparse-grid implementation. Special thanks to Omar Ghattas, Volker Schulz, and Georg Stadler for stimulating discussions on optimization under uncertainty.

REFERENCES


