Chapter 6

PDE Optimization with Uncertainty

6.1 Introduction

The importance of PDE-constrained optimization is derived from the demand in applications where PDEs represent the constitutive models and the environment of the implementation of the results of optimization strategies is not exactly known. Often, critical modeling parameters are only known to lie in a certain region and the practical implementation of the design and control solutions is subject to a priori unknown perturbations. We illustrate recent approaches to compute solutions to PDE-constrained optimization problems which are robust with respect to the stochasticity of the application framework. This emerging field of research in PDE optimization is very important, on the one hand because it addresses the need to solve real application problems, and on the other hand because it poses new analytical and computational challenges to the PDE optimization community. While we do not provide much background material on some techniques concerning the representation and treatment of uncertainty and rely on references, we focus on the combination of the solution methodologies presented in previous chapters with these techniques.

We know that the efficient computation of the optimal solutions can be achieved only if structures in the underlying problem are exploited. Besides the structures naturally arising in PDE-constrained optimization problems, which are exploited here in the form of multigrid methods and one-shot methods, a new structure comes into focus because of the stochasticity involved. Although the exact perturbations involved in practical applications are in general unknown, usually a good estimate of the probability distribution of the perturbations is available. This knowledge is to be exploited in the treatment of uncertainty in PDE-constrained optimization.

In the following sections, we discuss the treatment of uncertainty in the case where the coefficients of the PDE models are subject to known random field perturbations. Afterwards, we consider semi-infinite robust design for aerodynamic optimization under geometric uncertainty. In these problems, the uncertainty is related to the PDE models which are deterministic and to the data which is assumed to be known. On the other hand, we shall also discuss cases where the model is stochastic and the PDE optimization framework enters in the form of a Fokker–Planck system to control the probability density function distribution of the state of the model. Further, we shall discuss a Bayesian approach to

quantify uncertainty in PDE optimization problems where the data is only known through some characterizing statistical quantities.

We start our discussion considering the following general problem formulation

$$\min\ J(y,u), \tag{6.1}$$
$$c(y,u)=0, \tag{6.2}$$

where J is a real-valued differentiable objective function and c denotes the PDE constraint. For a fixed u, we require the existence of a unique $y=y(u)$ that satisfies the PDE constraint.

In this introductory section, we discuss for ease of presentation a finite-dimensional stochastic perturbation σ "polluting" part of the data of the PDE-constrained optimization problem. This stochastic random variable σ is assumed to be a mapping

$$\sigma : \mathcal{O} \to \mathbb{R}^n,$$

where the triple $(\mathcal{O},\mathcal{A},\mathcal{P})$ denotes a probability space, $\mathcal{O}$ is the space of elementary events, and $\mathcal{A}$ is the sigma-algebra of subsets of $\mathcal{O}$ measured by the probability measure $\mathcal{P}$ on $\mathcal{O}$; see, e.g., [237]. Thus, we define the expected (mean) value of the stochastic variable σ as

$$\sigma_o := \mathbb{E}(\sigma) = \int_{\mathcal{O}} \sigma(\omega) d\mathcal{P}(\omega).$$

Now, we focus on the following three reformulations of problem (6.1)–(6.2) subject to uncertainty. We have

$$u_0 := \arg\min_u \{J(y,u)\,|\,c(y,u,\sigma_0)=0\}, \tag{6.3}$$
$$\bar{u} := \mathbb{E}(\arg\min_u \{J(y,u)\,|\,c(y,u,\sigma)=0\}), \tag{6.4}$$
$$\hat{u} := \arg\min_u \mathbb{E}(\{J(y,u)\,|\,c(y,u,\sigma)=0\}). \tag{6.5}$$

The optimal design/control u_0 corresponds to the standard approach by just performing a deterministic optimization, where the stochastic variable is substituted by its expected value. We denote with $\bar{u}$ the average optimization function obtained as the expectation of the optimization functions corresponding to all possible random configurations. With $\hat{u}$, we denote the design/control function that optimizes the expectation of the reduced objective.

Because of the arising nonlinearities in all three problem formulations, we observe in general that $u_0 \neq \bar{u} \neq \hat{u} \neq u_0$. The problem formulation (6.4) is highly modular and can be readily coupled with existing solver technologies for PDE-constrained optimization. This formulation will be investigated in the next section. The problem formulation (6.5) will be discussed in Section 6.3 for the problem of robust aerodynamic shape optimization, where also additional scalar-valued state constraints are added. In Section 6.4, we discuss the solution of (6.5) by model reduction using proper orthogonal decomposition. In this discussion, distributed uncertainty plays an important role and the Karhunen–Loève (KL) expansion is used extensively. Solution approaches for finitely many scalar uncertainties, as discussed above for ease of presentation, can be found in, e.g., [313], where problem formulation (6.5) is used. In Section 6.5, we discuss a strategy to control probability density functions of stochastic processes using the Fokker–Planck equation. The last section of this chapter is concluded with a discussion on recent techniques for Bayesian uncertainty quantification related to inverse problems governed by PDE models.

6.2 PDE Control Problems with Uncertain Coefficients

In this section, we consider nonlinear parabolic control problems with uncertain coefficients. Uncertainty may appear in nonlinear reaction terms or in the diffusion coefficient. We consider the case of distributed controls as in [44, 67]. We have the following reaction-diffusion model

$$\begin{cases} -\partial_t y + G_\delta(y) + \sigma\,\Delta y = f & \text{in } Q = \Omega \times (0,T), \\ y = y_0 & \text{in } \Omega \times \{t = 0\}, \\ y = 0 & \text{on } \Sigma = \partial\Omega \times (0,T), \end{cases} \tag{6.6}$$

where $\Omega \subset \mathbb{R}^d$, $d \geq 1$, 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. Here δ is the reaction parameter and σ is the diffusion coefficient, both described by random fields.

We have $\sigma = \sigma(x,\omega)$, where $x \in \Omega$ and $\omega \in \mathcal{O}$. 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 [237] 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 functions of $\sigma(x,\omega)$ are known, respectively, as

$$\sigma_0(x) = \mathbb{E}(\sigma)(x) := \int_{\mathcal{O}} \sigma(x,\omega)\, d\mathcal{P}(\omega) \tag{6.7}$$

and

$$C_\sigma(x,x') = \int_{\mathcal{O}} (\sigma(x,\omega) - \sigma_0(x))(\sigma(x',\omega) - \sigma_0(x'))\, d\mathcal{P}(\omega). \tag{6.8}$$

We see that C_σ defines the kernel of a compact, positive, and self-adjoint operator. Denote with λ_j the real positive eigenvalues and with $z_j(x)$ the corresponding orthonormal eigenfunctions of C_σ as follows. We have

$$\int_\Omega C_\sigma(x,x')\, z_j(x')\, dx' = \lambda_j\, z_j(x), \qquad x \in \Omega,$$

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

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

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

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

$$\sigma_N(x,\omega) = \sigma_0(x) + \sum_{j=1}^{N} \sqrt{\lambda_j}\, z_j(x)\, Y_j(\omega), \tag{6.9}$$

where N denotes the number of terms in the truncation. Following [363], we define $\Gamma = \Pi_{j=1}^{N} \Gamma_j$, and the joint probability density function with $\rho = \Pi_{j=1}^{N} \rho_j$. We see that $\sigma_N(x,\omega)$ may provide a suitable approximation to $\sigma(x,\omega)$ assuming that 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.

In particular, consider random fields characterized by squared exponential covariance

$$C_\sigma(x,x') = s^2 \exp\left(-\frac{|x-x'|^2}{\ell^2}\right), \qquad x,x' \in \Omega,$$

with mean σ_0 and variance s^2. The degree of variability of this random field can be characterized by the ratio s/σ_0, while the frequency of variation of this field is related to the ratio L/ℓ, where L is the characteristic length of the physical domain and ℓ represents the physical correlation length. We have 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., [204, 263].

We consider random fields of the type given above that can be approximated by (6.9) with good accuracy for moderate values of N. Therefore, assuming finite-dimensional random fields, we can write the explicit dependence of σ on the random variables $Y = [Y_1, \ldots, Y_N]$. We have

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

Similarly, we can assume that the reaction dynamics of (6.6) can be modeled by a KL-truncated temporal random field as follows

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

where the $W = [W_1, \ldots, W_M]$ are uncorrelated random variables with zero mean and unit variance.

With the setting above and deterministic initial and boundary conditions, we have that the solution to (6.6) 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

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

Now, we formulate a parabolic optimal control problem with random coefficients; the discussion on the elliptic case follows along the same lines of reasoning; see [44]. We have

$$\min_{y,u} J(y,u) := \frac{\alpha}{2}\|y - y_d\|^2_{L^2(Q)} + \frac{\beta}{2}\|y(\cdot,T) - y_T\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|u\|^2_{L^2(Q)}, \tag{6.10}$$

$$-\partial_t y + G_{\delta(W(\omega))}(y) + \sigma(Y(\omega))\,\Delta y = f + u \quad \text{in } Q \times \mathcal{O}, \tag{6.11}$$

where we take homogeneous boundary conditions and deterministic initial condition $y_0 \in H_0^1(\Omega)$. The objective (6.10) models the requirement to track a desired deterministic trajectory $y_d \in L^2(Q)$ and to reach a desired deterministic terminal state $y_T \in L^2(\Omega)$. Here, $\nu > 0$ is the weight of the cost of the control and $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta > 0$, are optimization

parameters. For any given event ω, we assume that the resulting random field and the nonlinearity are such that the optimal solution has the required regularity; see [268, 316] for a related discussion. To simplify notation, whenever possible in the following we omit the dependence of the variables from ω.

The optimal control problem (6.10)–(6.11) is stochastic in the sense that for any event ω, it provides a different control. Each single event corresponds to a deterministic optimal control problem with $\delta = \delta(t,\omega)$ and $\sigma = \sigma(x,\omega)$, and therefore, corresponding to each event, the discussion on the existence and uniqueness of optimal solutions and their numerical determination proceeds as in the deterministic cases discussed in Chapters 3 and 5.

We have that the solution to (6.10)–(6.11) is characterized by the following first-order optimality system

$$\begin{aligned} -\partial_t y + G_\delta(y) + \sigma\Delta y &= f + u && \text{in } Q,\\ \partial_t p + G'_\delta(y)p + \sigma\Delta p + \alpha(y - y_d) &= 0 && \text{in } Q,\\ \nu u - p &= 0 && \text{in } Q,\\ y = 0,\ p &= 0 && \text{on } \Sigma, \end{aligned} \tag{6.12}$$

with initial condition $y(x,0) = y_0(x)$ for the state variable and terminal condition for the adjoint variable given by $p(x,T) = \beta(y(x,T) - y_T(x))$.

Further, we assume that our setting is such that a unique mapping $y = y(u)$, and we recall the reduced cost functional $\hat{J}(u) = J(y(u),u)$. We have $\nabla\hat{J}(u) = \nu u - p(u)$.

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 control 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 and thus define a robust deterministic control. For this reason, we consider the following problem

$$\min_{u\in L^2(Q)} \int_{\mathcal{O}} \hat{J}(u;\omega)\, d\mathcal{P}(\omega), \tag{6.13}$$

where $\hat{J}(u;\omega) = \hat{J}(y(u;\omega),u)$. This formulation corresponds to problem (6.5) in the introduction. One can see that this formulation requires us to solve an optimization problem with infinitely many terms in the objective (the integral in $\mathcal{O}$) and infinitely many PDE constraints that characterize $y(u;\omega)$. Notice that we do not have a theory for (6.13) and we can only argue that a local optimal u exists. However, after discretization of the stochastic space, an optimization problem with a very large but finite number of PDE constraints must be solved. In this case, existence of optimal solutions can be proved with standard arguments.

Next, in an attempt to avoid solving (6.13), we explore alternative formulations, in particular formulations corresponding to (6.4). On the other hand, in Section 6.4 we discuss a proper orthogonal decomposition framework to determine robust controls with (6.13).

Following the discussion in [313, 280], given a functional F depending on a set of random input variables (σ,δ), with mean $(\mathbb{E}(\sigma),\mathbb{E}(\delta))$ and standard deviations (s_σ, s_δ), we have the following second-order Taylor expansion:

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

where the second derivatives are evaluated at $(\sigma_0, \delta_0) = (\mathbb{E}(\sigma), \mathbb{E}(\delta))$. From this formula we conclude that a first-order approximation to (6.13) corresponds to the optimal u_0, with the state equation having σ_0 and δ_0 as diffusion and reaction coefficients, respectively. Further notice that the state equation is nonlinear and that therefore the control obtained using the mean diffusion and reaction coefficients is not the same as the mean of the controls $\bar{u} = \mathbb{E}(u)$ obtained by averaging over all controls corresponding to all (σ, δ) configurations. Therefore, while the Taylor series expansion above shows that $u_0 = u_{\mathbb{E}(\sigma),\mathbb{E}(\delta)}$ is a first-order approximate solution to (6.13), it appears reasonable to ask whether or not $\bar{u}$, as defined in (6.4), can also be an approximate minimizer. This is not clear, since, in general, we expect

$$J(y(u_0), u_0) \neq J(y(\bar{u}), \bar{u}),$$

as already noted in the introduction. In Section 6.2.3, we compare these two approximation strategies in the effort of designing a robust optimal control.

6.2.1 Discretization of the Probabilistic Space

Recent research on stochastic PDE problems has focused on deterministic PDE systems with coefficients modeled by random fields [352] that can be accurately represented by a truncated KL expansion, since in this case a finite-dimensional approximation of the original infinite-dimensional stochastic space is possible. Based on the approximation of stochastic fields with the KL expansion or the polynomial chaos expansion [142], it has been possible to investigate the solution of PDE problems with random coefficients [17, 132, 263, 290, 363].

In the framework given by the generalized polynomial chaos expansion [142, 364], 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 ψ_i are the Askey polynomials which are chosen depending on the type of stochastic process $\xi(\omega)$ as input. A challenging aspect of the polynomial chaos approach is the solution of very large-sized coupled systems; see, e.g., [290].

On the other hand, a stochastic-grid collocation approach [17, 263, 363] allows us to avoid the solution of PDE problems coupled in the probability space. However, in the collocation case, a uniform Cartesian-mesh discretization of the probability parameter space may result in a curse of dimensionality, and in order to solve this problem, a Smolyak sparse-grid scheme [98, 222] has been introduced to model high-dimensional stochastic-coefficient spaces; see, e.g., [44, 67, 263]. Alternatively, Monte Carlo and quasi-Monte Carlo methods have been used to explore the probability space of PDEs with random coefficients; see, e.g., [18, 151].

In the following, we discuss the solution of (6.10)–(6.11) using a collocation approach. On the stochastic $(N+M)$-dimensional grid, the coordinates of the grid points represent the coefficients of the truncated KL expansion, while the physical space-time domain is approximated using classical discretization schemes. On the stochastic grid, we need to solve (6.10)–(6.11) 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 [263], we consider Chebyshev

nodes given by

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

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 [132, 222, 263, 363]. In fact, if we write the set of Chebyshev nodes of order j as $\{z^j\} = \{z_1^j,\ldots,z_{m_j}^j\}$, then the nesting property gives that $\{z^j\} \subset \{z^{j+1}\}$.

Let $\{w^j\}$ be the set of the quadrature weights in one dimension which correspond to the nodes $\{z^j\}$; then 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 $f : \mathbb{R}^{\mathcal{N}} \to \mathbb{R}$ is given by

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

and the variance is $Var(f) = \mathbb{E}([f - \mathbb{E}(f)]^2)$.

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, a Smolyak scheme can be used [132, 135, 263, 363] that constructs multivariate interpolation as a 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 $\{z^1 \otimes z^4\} \bigcup \{z^2 \otimes z^3\} \bigcup \{z^3 \otimes z^2\} \bigcup \{z^4 \otimes z^1\}$; see, e.g., [44, 67] for details. With χ we denote the total number of sparse-grid collocation points.

To define the weights for sparse-grid integration, we need to define the difference weights of order j, which are $\eta^j = w^j - w^{j-1}$. The quadrature weights on a full grid with order $\mathbf{j} = (j_1,\ldots,j_{\mathcal{N}})$ would be the tensor product $w = w^{j_1} \otimes \cdots \otimes w^{j_{\mathcal{N}}}$. On the sparse grid of dimension $\mathcal{N}$ and total order $\mathcal{J}$, the weights are given by

$$w^{\mathcal{J}} = \sum_{\ell=1}^{\mathcal{N}+\mathcal{J}-1} \sum_{|\mathbf{j}|=\ell} (\eta^{j_1} \otimes \cdots \otimes \eta^{j_{\mathcal{N}}}).$$

For more details on sparse grids, see [98, 222].

Following [17, 263], it is possible to estimate the global error affecting the solution of the optimal control problem with random coefficients. Let e_h be the bound of the solution error of the deterministic problem, denote with $e_{\mathcal{N}}$ the error due to truncation of the KL expansion, and let e_S denote the sparse-grid interpolation error; then $e_h + e_{\mathcal{N}} + e_S$ provides an estimate of the global error.

6.2.2 Sparse-Grid CSMG Methods

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 still required, which demands efficient solution strategies that are robust with respect to the choice of values of the optimization parameters and to the changes of the coefficients' configurations due to randomness. We have shown that these requirements are met by the CSMG strategy. Therefore in the following, we present results of experiments with optimal control problems with random coefficients where each random configuration corresponds to a sparse-grid point and the resulting deterministic optimality system is discretized in the physical space and time.

Results of numerical experiments show that the combination of sparse-grid and CSMG techniques results in a solution process with optimal computational complexity with respect to the sizes of the physical and probability grids.

6.2.3 Experiments with a Parabolic Control Problem

We consider (6.11) with a nonlinear reaction term given by $G_\delta(y) = \delta e^y$ that is used to model explosive combustion phenomena [57]. 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 [263], a KL expansion in the form (6.9) for $\log(\sigma(\omega) - \sigma_0)$. Therefore, we consider a random diffusivity coefficient given by

$$\sigma(x_1, x_2, \omega) = \sigma_0 + \exp\big\{[Y_1(\omega)\cos(\pi x_2) + Y_3(\omega)\sin(\pi x_2)]\, e^{-1/8} + [Y_2(\omega)\cos(\pi x_1) + Y_4(\omega)\sin(\pi x_1)]\, e^{-1/8}\big\},$$

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 6.1.

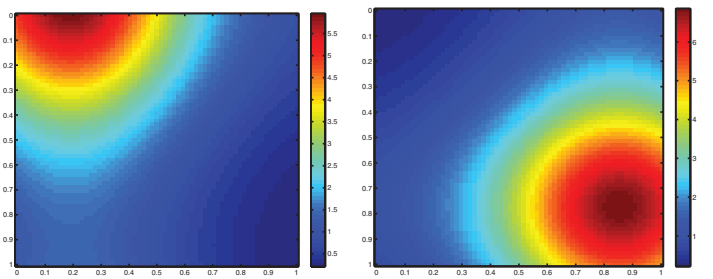

Figure 6.1. *Two realizations of* $\sigma(x_1, x_2, \omega)$*. Left:* $Y_1 = 0.8$*,* $Y_2 = 1.0$*,* $Y_3 = -0.2$*,* $Y_4 = 0.7$*. Right:* $Y_1 = -0.8$*,* $Y_2 = -1.0$*,* $Y_3 = 0.7$*,* $Y_4 = 0.5$*. This figure first appeared in A. Borzì and G. von Winckel, Multigrid methods and sparse-grid collocation techniques for parabolic optimal control problems with random coefficients, SIAM J. Sci. Comput.,* 31(3) (2009), 2172–2192.

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 $f = 0$. The desired target trajectory is given by the following oscillating

function with increasing amplitude

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

We take $y_T(x) = y_d(x,T)$.

We use one pre- and postsmoothing, and the domain $\Omega = (0,1)^2$ is discretized by a hierarchy of meshes and $h = 1/4$ is the coarsest space-mesh size. We take $T = 1$ and $\delta t = T/N_t$. 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 orders $\mathcal{J} = 2$ and $\mathcal{J} = 3$. These correspond to a total number of sparse-grid points of $\chi = 61$ and $\chi = 241$, respectively.

To describe the results of the experiments, we report the observed multigrid convergence factors ρ_{obs} and CPU times. The tracking ability of the sparse-grid multigrid algorithm is given in terms of the mean values of the norms of the tracking error $\mathbb{E}(\|y - y_d\|)$ and of the terminal observation error $\mathbb{E}(|y - y_T|)$. Other important statistical observable are the variance Var of the tracking error and its skewness $Skew$.

The results are reported in Table 6.1 in the case of tracking of trajectories. They illustrate efficiency and robustness of the proposed multigrid solvers. Usual multigrid convergence speeds are obtained which appear to be independent of the mesh size and of ν. Notice that the CPU times scale linearly as a factor of 2^3 as the space-time mesh is refined. Moreover, as we move from an order-two to an order-three stochastic sparse grid we also obtain optimal complexity in CPU time. The results in Table 6.1 also demonstrate the tracking ability of the optimal control formulation. As the value of ν decreases, smaller values of $\mathbb{E}(\|y - y_d\|)$ are obtained as desired in an optimal control framework.

Table 6.1. *Results for $\alpha = 1$, $\beta = 0$, with the CSMG multigrid scheme. Denote $\|\Delta y\| = \|y - y_d\|$.*

$\mathcal{J} = 2$, $\chi = 61$						
ν	$N_x \times N_y \times N_t$	CPU(s)	$\mathbb{E}(\|\Delta y\|)$	$Var(\|\Delta y\|)$	$Skew(\|\Delta y\|)$	$\mathbb{E}(\rho_{obs})$
10^{-6}	$32 \times 32 \times 32$	87.89	$1.15\,10^{-3}$	$5.84\,10^{-7}$	$5.79\,10^{-10}$	0.08
10^{-8}	$32 \times 32 \times 32$	69.75	$3.48\,10^{-5}$	$6.45\,10^{-10}$	$2.59\,10^{-14}$	0.08
10^{-6}	$64 \times 64 \times 64$	722.46	$1.18\,10^{-3}$	$5.94\,10^{-6}$	$5.93\,10^{-10}$	0.12
10^{-8}	$64 \times 64 \times 64$	795.68	$4.75\,10^{-5}$	$7.08\,10^{-10}$	$2.80\,10^{-14}$	0.12
$\mathcal{J} = 3$, $\chi = 241$						
ν	$N_x \times N_y \times N_t$	CPU(s)	$\mathbb{E}(\|\Delta y\|)$	$Var(\|\Delta y\|)$	$Skew(\|\Delta y\|)$	$\mathbb{E}(\rho_{obs})$
10^{-6}	$32 \times 32 \times 32$	353.51	$1.15\,10^{-3}$	$6.27\,10^{-7}$	$1.75\,10^{-9}$	0.12
10^{-8}	$32 \times 32 \times 32$	297.82	$3.49\,10^{-5}$	$4.83\,10^{-10}$	$2.02\,10^{-14}$	0.08
10^{-6}	$64 \times 64 \times 64$	2867.01	$1.18\,10^{-3}$	$6.37\,10^{-7}$	$1.80\,10^{-9}$	0.13
10^{-8}	$64 \times 64 \times 64$	3152.42	$4.73\,10^{-5}$	$5.60\,10^{-10}$	$2.14\,10^{-14}$	0.10

Once a (finite) set of controls approximating the solution of the stochastic PDE optimal control problem is obtained, the question arises of how to construct a unique control that provides good tracking features for all configurations of the coefficients. As a possible control function, we consider the function $\bar{u} = \mathbb{E}(u)$ defined as the mean function of the

optimal controls corresponding to each point of the discretized configuration space. Alternatively, we may consider the control function obtained using average parameter values. We denote the corresponding state functions as $y_{\bar{u}}$ and y_{σ_0,δ_0}, respectively. In Table 6.2 we present a comparison of the performance of the two controls. We see that the proposed controls have similar tracking performance, although the approach with $\bar{u}$ performs slightly better.

Table 6.2. *Results with* $\nu = 10^{-6}$*;* $N_x \times N_y \times N_t = 64 \times 64 \times 64$*,* $\chi = 241$*.*

		$\mathbb{E}(\|y - y_d\|)$	$Var(\|y - y_d\|)$	$Skew(\|y - y_d\|)$
$\alpha = 1$	$y_{\bar{u}}$	$1.33\,10^{-1}$	$1.20\,10^{-2}$	$3.26\,10^{-4}$
$\beta = 0$	y_{σ_0,δ_0}	$1.43\,10^{-1}$	$1.40\,10^{-2}$	$4.44\,10^{-2}$
$\alpha = 0$	$y_{\bar{u}}$	$2.86\,10^{-1}$	$5.79\,10^{-4}$	$3.99\,10^{-5}$
$\beta = 1$	y_{σ_0,δ_0}	$5.25\,10^{-1}$	$2.46\,10^{-6}$	$-1.80\,10^{-9}$

Similar results are obtained with the sparse-grid multigrid scheme applied to elliptic control problems with random coefficients [44].

6.3 Aerodynamic Design under Geometric Uncertainty

Uncertainty arises in flight conditions like the angle of incidence, the velocity (Mach number) of the plane, the density of the air, or the Reynolds number. The formulation of robust optimization problems as well as their numerical treatment is studied in [313]. Here we focus on the aircraft geometry itself as an uncertainty source [65, 300]. The real shape may vary from the planned shape due to manufacturing tolerances, temporary factors like icing, or material fatigue. Since there are so many factors having effects on the shape, this uncertainty has to be considered in the optimization problem in order to produce a design which is robust to small perturbations of the shape itself. In the literature, there are only a few papers on this topic investigating the influence of variations of the profile; see, e.g., [156, 238]. The general approach that we discuss is again to apply a KL expansion to approximate the infinite-dimensional probability space of geometrical uncertainty. To better overcome the curse of dimensionality, an adaptively refined sparse grid is used in order to compute statistics of the solution.

6.3.1 Modeling Geometric Uncertainty

The geometrical uncertainty depends on the geometry itself, so it is modeled as a Gaussian random field $\sigma : \Gamma \times \mathcal{O} \to \mathbb{R}$, defined on a probability space $(\mathcal{O}, \mathcal{A}, \mathcal{P})$ and on the shape of the airfoil Γ. In each point x of the shape Γ, the uncertainty is described by a normally distributed random variable $\sigma(x, \cdot)$. Additionally, the second order statistics, the mean value, and the covariance function are given to fully describe the random field. We have

$$\begin{aligned} \sigma_0(x) &= 0, \\ C_\sigma(x, y) &= b^2 \exp\left(-\frac{\mathrm{arc}(x, y)^2}{l^2}\right) \end{aligned} \tag{6.14}$$

for all $x, y \in \Gamma$. The function arc $\Gamma \times \Gamma \to \mathbb{R}^+$ measures the shortest distance between two points on the curved geometry Γ. The parameter l determines how quickly the covariance falls off and b controls the magnitude of the bumps. A squared exponential covariance function is chosen, since the resulting perturbed geometry is smooth due to the smoothness of the random field.

Then, a perturbed geometry is given as

$$v(x,\omega) = x + \sigma(x,\omega)\,\vec{n}(x), \tag{6.15}$$

where $\vec{n}$ is the unit vector in x normal to the profile Γ. The approximation of the random field σ is again performed by applying the KL expansion as in Section 6.2.

6.3.2 Semi-infinite Robust Design

The usual single setpoint aerodynamic shape optimization problem can be described in the following rather abstract form of an optimization problem with equality and inequality constraints

$$\min J(y,p), \tag{6.16}$$
$$c(y,p), = 0, \tag{6.17}$$
$$h(y,p) \geq 0. \tag{6.18}$$

Now, we think of (6.17) as the discretized outer flow equation around, e.g., an airfoil described by the geometry parameters $p \in \mathbb{R}^{n_p}$. Rather than more general investigations on aerodynamic shape optimization as in [123], we assume that a finite parametrization of the shape to be optimized is given. The vector y is the state vector (velocities, pressure, ...) of the flow model (6.17), and we assume that (6.17) can be solved uniquely for y for all reasonable geometries p. The objective in (6.16) typically is the drag to be minimized. The restriction (6.18) typically denotes lift or pitching moment requirements. To make the discussion here simpler, we assume a scalar-valued restriction, i.e., $h(y,p) \in \mathbb{R}$. The generalization of the discussions below to more than one restriction is straightforward. Here, we treat the angle of attack as a fixed parameter which is not adjusted to reach the required lift (cf., e.g., [233, 205, 270]). The general deterministic problem formulation (6.16)–(6.18) is influenced by stochastic perturbations. In the literature, min-max formulations and also chance constrained formulations can be found. According to our experience the semi-infinite formulation discussed below has been most successful.

Consider the following robust version of the nonlinear programming problem written in the form of a semi-infinite optimization problem

$$\min \int_{\mathcal{O}} J(y(\cdot,\omega), p, \sigma(\cdot,\omega))\, d\mathcal{P}(\omega), \tag{6.19}$$
$$c(y(\cdot,\omega), p, \sigma(\cdot,\omega)) = 0 \quad \forall \omega \in \mathcal{O}, \tag{6.20}$$
$$h(y(\cdot,\omega), p, \sigma(\cdot,\omega)) \geq 0 \quad \forall \omega \in \mathcal{O}. \tag{6.21}$$

This definition of robustness can also be found in [205, 234]. Semi-infinite optimization problems have been treated directly so far only for rather small and weakly nonlinear problems; see, e.g., [128]. For the numerical treatment of complicated design tasks, one has to

approximate the integral in the objective (6.19) resulting in the form of a multiple setpoint problem for the set points $\{\sigma_i\}_{i=1}^N$. We have

$$\min \sum_{i=1}^{N} w_i \, J(y_i, p, \sigma_i), \tag{6.22}$$

$$c(y_i, p, \sigma_i) = 0 \quad \forall i \in \{1, \ldots, N\}, \tag{6.23}$$

$$h(y_i, p, \sigma_i) \geq 0 \quad \forall i \in \{1, \ldots, N\}, \tag{6.24}$$

where w_i denote the quadrature weights. The approximation and discretization of the probability space using a goal-oriented KL expansion and an adaptive sparse-grid strategy in order to formulate the introduced multiple setpoint problem (6.22)–(6.24) are discussed in the following in more detail.

6.3.3 The Use of a Goal-Oriented KL Basis

In order to reduce the computational effort, the orthogonal basis functions $\{z_i\}$ of the KL expansion (cf. (6.9)) will be chosen goal-oriented; that is, the individual impact of the eigenvectors on the target functional will be taken into account. This method is well established in the model reduction methods of dynamic systems and the adaptive mesh refinement (cf. [28]). The idea is to develop an error indicator for the individual eigenvectors reflecting the influence on the drag. We discuss the use of sensitivity information to capture the local sensitivities of the drag with respect to the eigenvectors

$$\eta_i := \frac{\mathrm{d}J}{\mathrm{d}z_i} = -\lambda^\top \frac{\partial c}{\partial z_i} + \frac{\partial J}{\partial z_i} \qquad \forall i = 1, \ldots, d, \tag{6.25}$$

where λ solves the adjoint equation. The adjoint equation is independent of i; hence it has to be solved only once and the indicator η_i is numerically cheap to evaluate. Now, the reduced basis $\{\hat{z}_i\}$ can be automatically selected, and the eigenvector z_i with a large value η_i has to be kept in the reduced basis, whereas a small value indicates that the basis vector can be rejected from the basis.

6.3.4 Adaptive Sparse Grids for High-Dimensional Integration

We consider a dimension adaptive sparse-grid strategy in order to further reduce the number of grid points but conserve the approximation quality. First, a generalization of sparse grids will be introduced which allows us to weight the dimensions according to their importance on the target functional (cf. [141, 134, 97, 222]). The original sparse grid of order $\mathcal{J}$ combines all the incremental functions, defined as $\eta^i = w^{i+1} - w^i$, where w^j are the one-dimensional quadrature formulas of order j, which sum up to order $\mathcal{J}$. Further, [141] and [133] suggest allowing a more general index set which can then be adaptively chosen with respect to the importance of each dimension. An index set I is called *admissible* if for all $\mathbf{i} \in I$, we have

$$\mathbf{i} - e_j \in I \qquad \forall 1 \leq j \leq d, \, i_j > 1,$$

where $e_j \in \mathbb{R}^d$ is the jth unit vector. The generalized definition of sparse grids including the original sparse grid and the full tensor grid definition is then given as

$$S(\mathcal{J},d)(J) = \sum_{\mathbf{i}\in I} \left(\eta^{i_1} \otimes \cdots \otimes \eta^{i_d}\right)(J). \tag{6.26}$$

The algorithm computing the dimension adaptive sparse grid starts with the coarsest sparse grid, which means $I = \{(0,\ldots,0)\}$, and adds new indexes such that (1) the new index set remains admissible; and (2) the approximation error is reduced (cf. [141, 133]). The term $\eta_{\mathbf{i}} = \left(\eta^{i_1} \otimes \cdots \otimes \eta^{i_d}\right)(J)$ indicates the reduction in the approximated integral for each new added index, so that we directly use $\eta_{\mathbf{i}}$ as an error indicator for the adaptivity. The main advantage of the dimension adaptive refinement is the fact that the quadrature formulas can be chosen problem dependent. Considering geometric uncertainty described by a Gaussian random field, the Gauss–Hermite formulas are an appropriate choice for the quadrature. Since the nesting is a favorable feature constructing the sparse grid, only the Gauss–Hermite quadratures of orders $1, 3, 7, 15$ are taken into account in the numerical results.

6.3.5 Numerically Computed Robust Aerodynamic Designs

We present results of numerical experiments using a numerical solution strategy which is based on the one-shot approach as outlined in Section 4.6 and in [183, 182, 312]. In the experiments that follow, an active set strategy has been employed, which uses the worst violation of the lift constraint as the active set in each iteration. The robust optimization under shape uncertainty of a transonic RAE2822 profile in Euler flow is considered. We use the flow solver TAU provided by DLR which allows the computation of gradients by the adjoint approach. The TAU code is a CFD software package for the prediction of viscous and inviscid flows about complex geometries from the low subsonic to the hypersonic flow regime employing hybrid unstructured grids. The profile is described by 129 surface grid points and the airfoil is parametrized by 21 Hicks–Henne functions. The geometric uncertainty is characterized by a Gaussian random field with $\mathbb{E}(\sigma)(x) = 0$ and $C_\sigma(x,y) = (0.005)^2 \cdot \exp(-\frac{\text{arc}(x,y)^2}{(0.1)^2})$ for all $x, y \in \Gamma$. The distribution of the eigenvalues using the KL expansion of the given random field is shown in Figure 6.2.

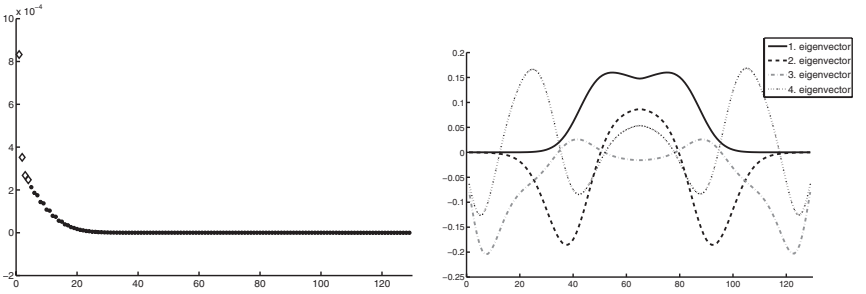

Figure 6.2. *Distribution of the eigenvalues and first four eigenvectors of the given random field σ. Reprinted with permission from A. Borzì, V. Schulz, C. Schillings, and G. von Winckel, On the treatment of distributed uncertainties in PDE-constrained optimization, GAMM-Mitteilungen,* 33(2) (2010), 230–246.

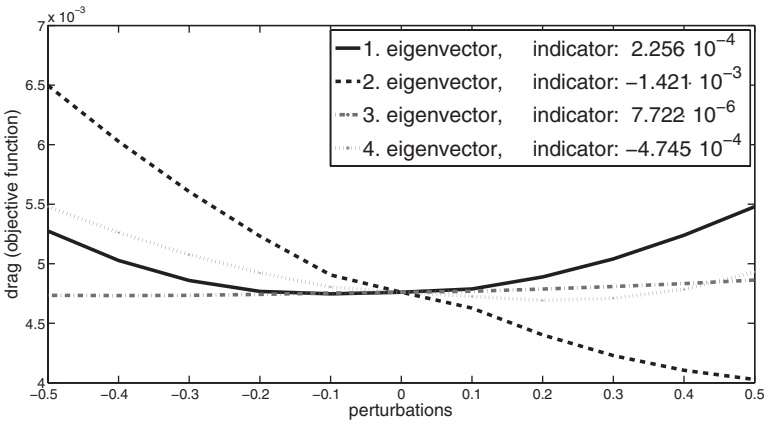

Figure 6.3. *Drag performance of the first four eigenvectors on the target functional. Reprinted with permission from A. Borzì, V. Schulz, C. Schillings, and G. von Winckel, On the treatment of distributed uncertainties in PDE-constrained optimization, GAMM-Mitteilungen,* 33(2) (2010), 230–246.

As stated before, the eigenvalues exponentially converge towards zero. For the numerical results below, we have considered only the first four eigenvalues and eigenvectors to represent the random field σ of the perturbations. The corresponding eigenvectors are shown in Figure 6.2. To further reduce the computational effort, we investigate the influence of the individual eigenvectors in order to reject those eigenvectors from the reduced basis which have no impact on the target functional. As Figure 6.3 shows, the third eigenvector has no impact on the objective function; hence it can be rejected from the KL basis and the dimension of the integral is reduced. This behavior is also reflected by the introduced indicator. The use of Gauss–Hermite formulas with an adaptive sparse-grid strategy results in a grid consisting of 21 points for a given error tolerance of 10^{-5}. Compared to a full tensor grid approach, a reduction by factor 10 can be observed; see Figure 6.4.

Figure 6.5 compares the results of the robust optimization and of the single setpoint optimization, i.e., without considering any uncertainty in the optimization.

The drag and lift performance is plotted against the 21 perturbed geometries, and the dashed line in Figure 6.5 indicates the mean value of the drag. The robust optimization improves the mean value of the target functional and also leads at the same time to a better lift performance over the whole range of perturbations, whereas the single setpoint optimization is infeasible in more than the half of the considered grid points. Summing it all up, it can be said that the robust optimization leads to a better lift-to-drag ratio than the single setpoint optimization and the resulting profile is more robust against small perturbations of the shape itself.

6.4 A Proper Orthogonal Decomposition Framework to Determine Robust Controls

In this section, we discuss a strategy for the fast computation of robust controls for PDE models with random coefficients [68]. This strategy is formulated in the framework of Section 6.2.1 and may be applied equally well to the case of robust shape design.

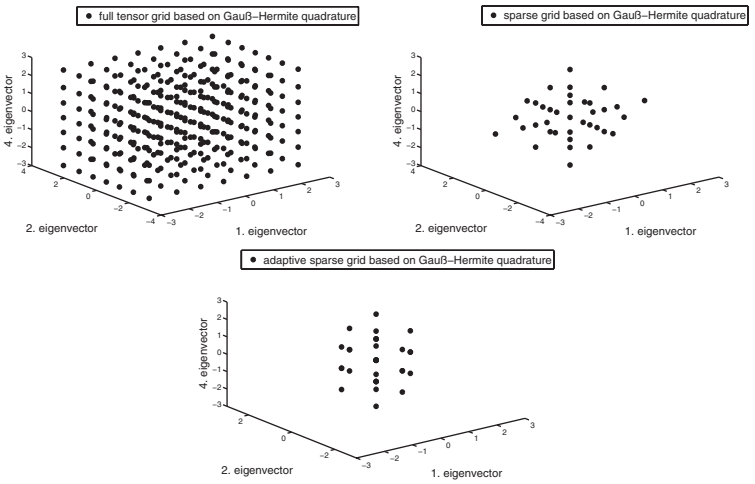

Figure 6.4. *Full tensor grid with* 343 *grid points, sparse grid with* 37 *grid points, and dimension adaptive sparse grid with* 21 *grid points. Reprinted with permission from A. Borzì, V. Schulz, C. Schillings, and G. von Winckel, On the treatment of distributed uncertainties in PDE-constrained optimization, GAMM-Mitteilungen,* 33(2) (2010), 230–246.

A robust control is defined as the control function that minimizes the expectation value of the objective over all coefficient configurations. However, a straightforward application of the full adjoint method on this problem results in a very large optimality system. In contrast, a fast method can be implemented where the expectation value of the objective is minimized with respect to a reduced proper orthogonal decomposition (POD) basis of the space of controls.

To illustrate this POD-based method, consider the following PDE optimization problem

$$\min J(y,u), \tag{6.27}$$

$$c(y,u,\omega) = 0, \tag{6.28}$$

where the PDE constraint has coefficients that are subject to randomness.

Now, assume an $\mathcal{N}$-dimensional and order $\mathcal{J}$ sparse-grid discretization of the probability space of the events $\omega \in \mathcal{O}$. Denote with χ the total number of sparse-grid points. We have that ω_i, $i = 1,\ldots,\chi$, represent the grid points of the sparse grids that correspond to different events in the probability space. For each of these points we solve (6.27)–(6.28), thus obtaining the optimal control u_i corresponding to ω_i.

In the following, we discuss a framework that, by exploring the collection of controls, u_i, $i = 1,\ldots,\chi$, allows us to determine a unique robust control for the governing random PDE model. We define this control as the minimizer $\hat{u}$ of the mean of the objective on the stochastic parameter space; recall the discussion in Sections 6.2 and 6.3. That is, we consider the following problem

$$\min_u \int_{\mathcal{O}} \hat{J}(u;\omega)\, d\mathcal{P}(\omega), \tag{6.29}$$

where $\hat{J}(u;\omega) = J(y(u;\omega),u)$ denotes the reduced objective corresponding to (6.27)–(6.28).

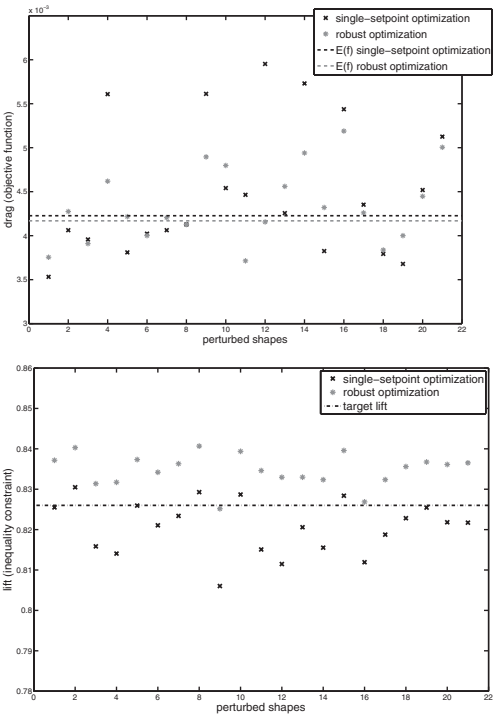

Figure 6.5. *Drag and lift performance of the* 21 *perturbed geometries. Reprinted with permission from A. Borzì, V. Schulz, C. Schillings, and G. von Winckel, On the treatment of distributed uncertainties in PDE-constrained optimization, GAMM-Mitteilungen,* 33(2) (2010), 230–246.

Now, after discretization of the stochastic space, an optimization problem with a very large number of PDE constraints must be solved. We have

$$\min_u \sum_{i=1}^{\chi} w_i \, \hat{J}(u;\omega_i), \tag{6.30}$$

where the w_i's denote the sparse-grid quadrature weights with $\sum_{i=1}^{\chi} w_i = 1$. This problem can be written in expanded form as follows

$$\begin{aligned} \min_u \sum_{i=1}^{\chi} w_i \, J(y_i,u), \\ c(y_1,u,\omega_1) = 0, \\ c(y_2,u,\omega_2) = 0, \\ \vdots \\ c(y_\chi,u,\omega_\chi) = 0, \end{aligned} \tag{6.31}$$

where $y_i = y(u;\omega_i)$, $i = 1,\ldots,\chi$.

To characterize the solution to (6.31), we define the Lagrange function

$$L(y_1, y_2, \ldots, y_\chi, p_1, p_2, \ldots, p_\chi, u) = \sum_{i=1}^{\chi} w_i\, J(y_i, u) + \sum_{i=1}^{\chi} w_i\, (c(y_i, u, \omega_i), p_i).$$

We have that solving (6.31) results in the following optimality system consisting of a very large system of state and adjoint equations coupled through an optimality condition. We have

$$\begin{aligned} c(y_1, u, \omega_1) = 0, \quad & c_y^*(y_1, u, \omega_1)\, p_1 = -J_y(y_1, u), \\ c(y_2, u, \omega_2) = 0, \quad & c_y^*(y_2, u, \omega_2)\, p_2 = -J_y(y_2, u), \\ \ldots & \\ c(y_\chi, u, \omega_\chi) = 0, \quad & c_y^*(y_\chi, u, \omega_\chi)\, p_\chi = -J_y(y_\chi, u), \\ & \sum_{i=1}^{\chi} w_i \left(c_u^*(y_i, u, \omega_i)\, p_i + J_u(y_i, u)\right) = 0. \end{aligned} \tag{6.32}$$

We call the solution of this problem the full solution.

The POD-based framework is different [68]. The sparse-grid collection of controls is computed and using POD, a reduced POD basis of this space is constructed. Then, the minimization problem (6.30) is solved in the span of the POD functions.

6.4.1 POD Analysis of the Control Space

Consider a positive random field σ representing a coefficient of our PDE governing model and assume that σ can be well approximated with a truncated KL expansion which allows a sparse-grid representation. Correspondingly, we obtain χ sparse-grid points, each of which represents a coefficient configuration function σ_i, $i = 1, 2, \ldots, \chi$.

Let $u_i \in U$, $i = 1, \ldots, \chi$, represent an input collection of the controls corresponding to different coefficient configurations. The entire collection can be written as follows

$$\mathbb{U} = [u_1, \ldots, u_\chi] \in U^\chi.$$

A reduced-order model of this collection is obtained with the following procedure.

Assume that each member of the collection can be written in terms of an nth order Galerkin finite element basis $\{\varphi_j\}_{j=1}^n$ and U is the corresponding finite element control space. Therefore, based on this representation, a finite-dimensional approximation of $\mathbb{U}$ results in an $n \times \chi$ matrix $\mathcal{U}$ where its m-column, $\mathcal{U}_{\cdot,m}$, contains the finite element coefficients of the mth member of the collection of controls. That is,

$$u_i = \sum_{j=1}^{n} \mathcal{U}_{ji}\, \varphi_j.$$

Now consider the problem of approximating all coordinate vectors $\mathcal{U}_{\cdot,m}$ simultaneously by a single normalized vector $\Psi_\cdot$ as well as possible. We assume that this vector contains the finite element coefficients of a function $\psi \in U$ that satisfies the following

optimization problem

$$\max \sum_{i=1}^{\chi} |(u_i, \psi)_U|^2 \qquad \text{s.t. } \|\psi\|_U = 1. \tag{6.33}$$

The finite-dimensional approximation of this problem is obtained assuming that we can represent $\psi = \sum_{j=1}^{n} \Psi_j \varphi_j$ and denote with M the positive definite mass matrix with elements $M_{ij} = (\varphi_i, \varphi_j)_U$. With this setting the first-order optimality condition for the optimization problem (6.33) results in the following eigenvalue problem

$$\mathcal{U}\mathcal{U}^\top M \Psi = \rho \Psi.$$

This eigenvalue problem can be put in a more elegant form defining $\hat{\mathcal{U}} = M^{1/2}\mathcal{U}$ and $v = M^{1/2}\Psi$. This yields

$$\hat{\mathcal{U}}\hat{\mathcal{U}}^\top v = \mu^2 v. \tag{6.34}$$

Now, consider the normalized eigenvector v_1 of (6.34) corresponding to the largest eigenvalue. The first POD coordinate basis function is given by $\Psi_1 = M^{-1/2} v_1$. This is a column vector that defines the first POD basis function as follows

$$\psi_1 = \sum_{j=1}^{n} \Psi_{j1} \varphi_j.$$

The second POD basis function is found by minimizing (6.33) subject to the additional constraint $(\psi_1, \psi_2)_U = 0$. One obtains the coefficients $\Psi_2 = M^{-1/2} v_2$, where v_2 is the normalized eigenvector in (6.34) corresponding to the second largest eigenvalue. This process is repeated until all nonzero eigenvalues are obtained. If the matrix $\mathcal{U}$ has rank $\ell \leq \chi$, then we obtain a POD basis given by

$$\hat{\Psi} = [\psi_1, \ldots, \psi_\ell].$$

The advantage of the POD framework is that a much smaller number of POD basis functions can approximate the input collection in the case of rapidly decaying eigenvalues. In fact, given a fraction $\delta < 1$, there exists $m \ll \ell$ such that

$$\frac{\sum_{i=1}^{m} \mu_i^2}{\sum_{i=1}^{\ell} \mu_i^2} > \delta.$$

Then m gives the number of POD basis functions that guarantee an approximation with a mean square error less than $(1-\delta)\sum_{i=1}^{\ell} \mu_i^2$. That is, the first m POD basis functions retain 100δ percent of the structure of the input collection. Therefore, we use a truncated POD basis given by

$$\Psi = [\psi_1, \ldots, \psi_m].$$

Now, it becomes convenient to pursue the minimization of the expectation of the reduced objective in the truncated POD space. We have

$$\min_u \sum_{i=1}^{\chi} w_i \hat{J}(u; \omega_i),$$

where we use the representation

$$u = \sum_{i=1}^{m} q_i \psi_i = \Psi q,$$

where $q = (q_1, q_2, \ldots, q_m)^\top$.

6.4.2 A Robust Control for Elliptic Control Problems

In this section, we illustrate the POD framework [68] to determine robust controls in the case of a distributed elliptic control problem with random coefficients. Consider the following optimal control problem

$$\min J(y,u) := \frac{1}{2}\int_\Omega [(y-z)^2 + \gamma u^2]\,dx, \tag{6.35}$$

$$\{-\partial_x^2 + \sigma\}\, y = u \qquad \text{in } \Omega, \tag{6.36}$$

$$y = 0 \qquad \text{on } \partial\Omega, \tag{6.37}$$

where $\Omega = (0,1)$ and $z(x) = x^2 \sin(\pi x)$. We assume that for any configuration of the positive random field $\sigma \in L^\infty(\Omega)$, the state equation has a unique solution.

Now, for each $\sigma_i(x)$ coefficient function, we compute the corresponding optimal control u_i so that we form a collection $\mathcal{U} = \text{span}\{u_1, u_2, \ldots, u_\chi\}$. In the following experiments, the functions σ_i are piecewise constant in Ω. They are defined as follows. We choose the value of the sparse-grid dimension $\mathcal{N}$ and consider a subdivision of Ω in $\mathcal{N}$ subintervals. On the ℓth subinterval, the constant value of σ_i is the value of the ℓth coordinate of the ith sparse-grid point. That is, we have KL basis functions that are piecewise constant with random amplitudes on equally sized subregions.

Next, assume that a POD analysis of the space of controls has been performed and it results that this space is well approximated by retaining the first m largest eigenvalues and corresponding eigenvectors. Further, denote with S_i the self-adjoint elliptic operator as in (6.36)–(6.37) with σ_i as coefficient. With this setting, the expectation of the reduced cost functional, the corresponding gradient, and the reduced Hessian are given by

$$\hat{J}(u) = \frac{1}{2}\int_\Omega \left[\gamma u^2 + \sum_{i=1}^{\chi} w_i (S_i^{-1}u - z)^2\right] dx,$$

$$\nabla \hat{J}(u) = \gamma\, u + \sum_{i=1}^{\chi} w_i\, S_i^{-1}(S_i^{-1}u - z),$$

$$\nabla^2 \hat{J}(u) = \gamma\, I + \sum_{i=1}^{\chi} w_i\, S_i^{-2}.$$

If the dimensionality of the KL expansion is large and one chooses a relatively high-order sparse-grid quadrature, there will be many solution operators S_i to invert, which can be quite expensive, as we will need a Krylov solver to compute the Newton step.

We have that the optimal control for a given σ_i satisfies the equation

$$[\gamma I + S_i^{-2}]\, u_i = -S_i^{-1} z.$$

The optimal control for the average reduced objective satisfies the following full problem

$$\left(\gamma I + \sum_{i=1}^{\chi} w_i \, S_i^{-2}\right) u = \sum_{i=1}^{\chi} w_i \, S_i^{-1} \, z. \tag{6.38}$$

Notice that in (6.38), in order to apply the Hessian, we must solve $2\,\chi$ boundary value problems.

We perform a POD decomposition on the set of control functions in order to obtain an orthogonal set of functions $[\psi_1, \psi_2, \ldots, \psi_m]$, where $m \ll \chi$. Therefore, we consider the control problem (6.38) using the POD expansion $u = \Psi\, q$. Further, we multiply both sides of (6.38) by $\Psi^\top$. We obtain

$$\Psi^\top \left(\gamma I + \sum_{i=1}^{\chi} w_i \, S_i^{-2}\right) \Psi\, q = \Psi^\top \sum_{i=1}^{\chi} w_i \, S_i^{-1} \, z.$$

Rearranging the Ψ and S_i^{-1} terms, we have

$$\left(\gamma I + \sum_{i=1}^{\chi} w_i \, (S_i^{-1}\Psi)^\top (S_i^{-1}\Psi)\right) q = \sum_{i=1}^{\chi} w_i \, (S_i^{-1}\Psi)^\top z.$$

Now, define $P_i = S_i^{-1}\,\Psi$ so that we obtain

$$\left(\gamma I + \sum_{i=1}^{\chi} w_i \, P_i^\top \, P_i\right) q = \sum_{i=1}^{\chi} w_i \, P_i^\top \, z. \tag{6.39}$$

We notice that in (6.39), we have to solve offline the χ boundary value problems $S_i\, P_i = \Psi$, $i = 1, \ldots, \chi$, with m right-hand sides each to obtain all the $P_i \in \mathbb{R}^{n\times m}$ matrices needed.

In fact, solving (6.38) requires us to solve 2χ boundary value problems, times the number of, e.g., conjugate gradient iterations to solve the Newton problem of size n. On the other hand, solving (6.39) requires the solution of $m\,\chi$ boundary value problems only once before the outer iteration starts.

In Figure 6.6, we compare the full solution obtained with (6.38) and with (6.39). We can see rapid changes of the control corresponding to the jumps in the reaction coefficient ($\mathcal{N} = 3$ on top and $\mathcal{N} = 5$ on bottom). In particular, we see that the POD robust control has larger gradients at the jump interfaces, which results in better minimizers. In Table 6.3, we see that the computational time of the POD method is an order of magnitude smaller than that of the full method. The CPU time to compute the input collection of controls and the POD basis results of approximately the same order of the POD solution process. All results with the POD method are obtained with $m = 10$ POD basis functions. The full problem (6.38) has been solved using a preconditioned conjugate gradient method.

6.5 Optimal Control of Probability Density Functions of Stochastic Processes

We remark that the discussion on the simulation and control of PDE models with data uncertainty focuses on deterministic models and the aim is to determine unique robust solutions to be used in application. On the other hand, a statistically meaningful solution would

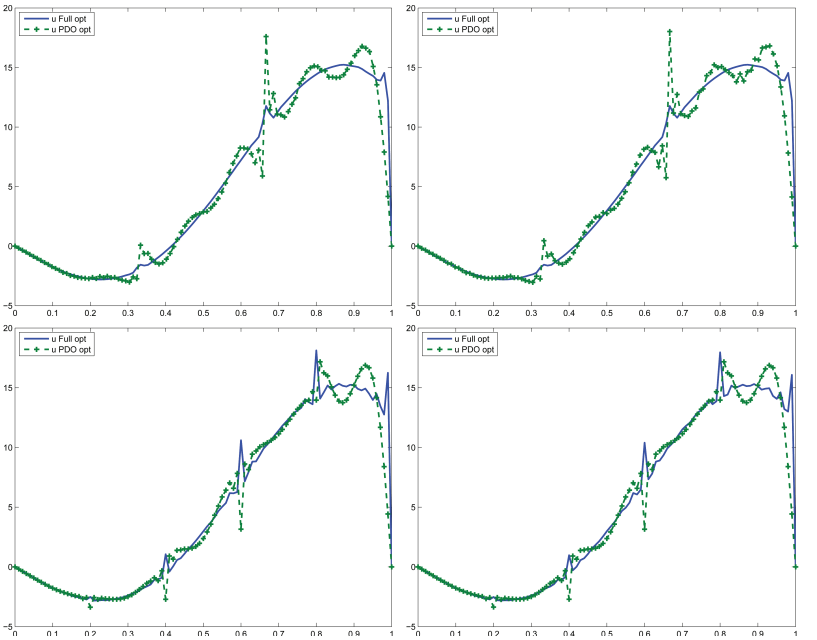

Figure 6.6. *Optimal controls computed with the full approach and the POD approach. Top: $\mathcal{N} = 3$; bottom: $\mathcal{N} = 5$. Left: $\mathcal{J} = 3$; right: $\mathcal{J} = 5$. From left to right and from top to bottom $\chi = \{69, 441, 241, 2433\}$, $m = 10$. Reprinted with permission from A. Borzì and G. von Winckel, A POD framework to determine robust controls in PDE optimization, Computing and Visualization in Science.*

Table 6.3. *Results with full and POD methods; $n = 63$, $m = 10$. CPU times include setup.*

$\mathcal{N}$	$\mathcal{J}$	χ	$\hat{J}_{pod}$	CPUsec	$\hat{J}_{full}$	CPUsec
3	3	69	$3.560\,10^{-2}$	$2.4\,10^{-2}$	$3.657\,10^{-2}$	34.2
3	5	441	$3.560\,10^{-2}$	$1.8\,10^{-1}$	$3.657\,10^{-2}$	221.6
5	3	241	$3.604\,10^{-2}$	$8.6\,10^{-2}$	$3.725\,10^{-2}$	225.6
5	5	2433	$3.605\,10^{-2}$	3.3	$3.725\,10^{-2}$	2291.4

require us to compute mean values and higher statistical moments depending on space and time. This is a very challenging problem setting especially in the infinite-dimensional functional framework where PDEs are defined. However, in the finite-dimensional case the problem of determining and controlling directly the probability density function (PDF) of a stochastic system is possible using the Fokker–Planck equation [288].

In this section, we discuss a Fokker–Planck framework for the formulation of an optimal control strategy for stochastic processes [7]. In this case, the governing model is a finite-dimensional Itô stochastic differential equation [144] and the PDE-based opti-

mization theory enters this framework to define a control strategy of the PDF of the given stochastic process. In this framework, the control objectives are defined based on the PDF and the optimal control is obtained as the minimizer of the objective under the constraint given by the Fokker–Planck equation [288] that models the evolution of the PDF.

We focus on stochastic processes described by the following model

$$\begin{cases} dX_t = b(X_t,t;u)\,dt + \sigma(X_t,t)\,dW_t, \\ X_{t_0} = X_0, \end{cases} \tag{6.40}$$

where the state variable X_t is subject to deterministic infinitesimal increments of the first addend and to random increments proportional to a Wiener process dW_t having zero mean and unit variance. We have that $\sigma(X_t,t) > 0$ is the function of variance of the stochastic process and $b(X_t,t;u)$ is the drift term including the control mechanism.

We consider the action of a control $u = u(t)$ that allows us to drive the random process to approximate a desired evolution trajectory. In deterministic dynamics, the optimal control is achieved by finding the control law u that minimizes a given cost functional $J(X,u)$. In the present nondeterministic case, the state evolution X_t is random and represents an outcome of a probability space, so that a direct insertion of a stochastic process into a deterministic cost functional will result in a random variable. Therefore, when dealing with stochastic optimal control, usually the average of the cost function is considered. In particular, we have

$$J(X,u) = \mathbb{E}\left[\int_{t_0}^{t_1} L(t,X_t,u(t))\,dt + \Psi[X_T]\right]. \tag{6.41}$$

This is a Bolza-type cost functional in the finite horizon case, and it is supposed that the controller knows the state of the system at each instant of time (complete observations).

The common methodology to find an optimal controller of random processes consists in reformulating the problem from stochastic to deterministic. This is a reasonable approach when we consider the problem from a statistical point of view, with the perspective to find out the collective behavior of the process. In fact, the average of the functional of the process X_t is omnipresent in almost all stochastic optimal control problems considered in the scientific literature.

However, while the average of a stochastic process represents only a piece of information, the state of a stochastic process can be completely characterized in many cases by the shape of its statistical distribution, which is represented by the PDF. Therefore, a control methodology defined on the PDF would provide an accurate and flexible control strategy that could accommodate a wide class of objectives. For this reason, in [7, 129, 212, 215, 358] PDF control schemes were proposed, where the cost functional depends on the PDF of the stochastic state variable. In this way, a deterministic objective results and no average is needed.

Now, the important step is to recognize that the evolution of the PDF associated with the stochastic process (6.40) is characterized as the solution of the Fokker–Planck (also Fokker–Planck–Kolmogorov) equation; see, e.g., [279, 288]. This is a PDE of parabolic type with Cauchy data given by the initial PDF distribution. Therefore, the formulation of objectives in terms of the PDF and the use of the Fokker–Planck equation provide a consistent framework to formulate an optimal control strategy of stochastic processes.

In this framework, we consider a stochastic process in a time interval, with a given initial PDF and the objective of approximating a desired final PDF target with the actual

PDF of the state variable. We assume that the control is a constant function in the time interval to be determined by the optimal control scheme once and for all the evolution of the process in the time interval. The cost functional consists of a terminal-time tracking objective and the control cost. The resulting open-loop optimal control problem is formulated as the problem of finding a controller that minimizes this cost function within the time interval under the constraint provided by the Fokker–Planck equation. In [7], this control strategy is applied to a sequence of time subintervals to construct a fast closed-loop control scheme of the stochastic process based on the receding-horizon (RH) model predictive control (MPC) approach [244, 250]. Notice that the RH-MPC approach does not optimize a true performance index; see [155] for a method to quantify the performance degradation. Nevertheless, these schemes provide robust controllers that apply equally well to linear and nonlinear models and allow us to accommodate different control and state constraints [155, 207]. For this reason, RH-MPC schemes are among the most widely used control techniques in process control.

6.5.1 A Fokker–Planck Optimal Control Formulation

Denote with $f(x,t)$ the PDF function of finding the process at x at time t. Further, let $\hat{f}(x,t;y,s)$ denote the transition density probability distribution for the stochastic process to move from y at time s to x at time t, which means that $\hat{f}(x,s;y,s) = \delta(x-y)$. Both $f(x,t)$ and $\hat{f}(x,t;y,s)$ are nonnegative functions and the following holds

$$\int_\Omega f(x,t)\,dx = 1 \quad \forall t \geq s. \tag{6.42}$$

This is the conservation condition. If $\rho(y,s)$ is the given initial density probability of the process at time s, then we have that the density probability of the process at time $t > s$ is given by the following

$$f(x,t) = \int_\Omega \hat{f}(x,t;y,s)\rho(y,s)\,dy. \tag{6.43}$$

Notice that ρ should be nonnegative and normalized: $\int_\Omega \rho(y,s)dy = 1$.

First, we consider the stochastic process in the time interval (t_k, t_{k+1}), and domain Ω is chosen large enough such that $f(x,t) = 0$ on $\Sigma = \Omega \times (t_k, t_{k+1})$. We assume to know the initial value of the process at time t_k, in the sense that we give the probability density $\rho(x,s)$ at time $s = t_k$. Our problem is to determine a control u such that starting with an initial distribution ρ the process evolves such that a desired target probability density $f_d(x,t)$ at time $t = t_{k+1}$ is matched as close as possible. For this purpose, we consider the following control problem in $Q = \Omega \times (t_k, t_{k+1})$. We have

$$\min J(f,u) := \frac{1}{2}\|f(\cdot,t_{k+1}) - f_d(\cdot,t_{k+1})\|^2_{L^2(\Omega)} + \frac{\nu}{2}|u|^2, \tag{6.44}$$

$$\partial_t \hat{f}(x,t;y,t_k) - \frac{1}{2}\partial_x^2(\sigma(x,t)^2\hat{f}(x,t;y,t_k)) + \partial_x(b(x,t;u)\hat{f}(x,t;y,t_k)) = 0, \tag{6.45}$$

$$\hat{f}(x,t_k;y,t_k) = \delta(x-y), \tag{6.46}$$

where (6.45) is the Fokker–Planck equation, i.e., the *forward Kolmogorov* equation, for the transition density probability distribution $\hat{f}(x,t;y,t_k)$ of the stochastic process X_t, and

$y \in \Omega$. Notice that the set of equations (6.45)–(6.46) can be interpreted as the fundamental solution of the stochastic process such that for any given ρ the corresponding f is obtained. This problem is as difficult to compute as the Green function of a PDE. However, in our case we assume that the initial distribution ρ is given, and hence we can reformulate (6.44)–(6.46) as follows

$$\min J(f,u) := \frac{1}{2}\|f(\cdot,t_{k+1}) - f_d(\cdot,t_{k+1})\|^2_{L^2(\Omega)} + \frac{\nu}{2}|u|^2, \tag{6.47}$$

$$\partial_t f(x,t) - \frac{1}{2}\partial_x^2(\sigma(x,t)^2 f(x,t)) + \partial_x(b(x,t;u)f(x,t)) = 0, \tag{6.48}$$

$$f(x,t_k) = \rho(x), \tag{6.49}$$

where we dropped s in the initial distribution. We consider solutions of the Fokker–Planck equation that are sufficiently regular [144], i.e., $f(x,t)$ has continuous first derivative in time, and continuous second derivative in space, jointly to the conservation condition.

6.5.2 An RH-MPC Scheme

The optimal control problem (6.47)–(6.49) is formulated for a stochastic process in the time interval (t_k, t_{k+1}) with a given terminal observation. Now, we consider the problem of controlling the PDF of a stochastic process to track a given sequence of desired PDFs in time. Let $(0,T)$ be the time interval where this process is considered. We assume time windows of size $\Delta t = T/N$, with N a positive integer. Let $t_k = k\Delta t$, $k = 0,1,\ldots,N$. At time t_0, we have a given initial PDF denoted with ρ and with $f_d(\cdot,t_k)$, $k = 1,\ldots,N$, we denote the sequence of desired PDFs. Our scheme starts at time t_0 and solves the minimization problem $\min_u J(f(u),u)$ defined in the interval (t_0,t_1). Then, with the PDF f resulting at $t = t_1$ solving the optimal control problem, we define the initial PDF for the subsequent optimization problem defined in the interval (t_1,t_2). This procedure is repeated by receding the time horizon until the last time window is reached. This is an instance of the class of RH-MPC schemes [250, 244] that is widely used in engineering applications to design closed-loop algorithms. One important aspect of this approach is that it can be applied to infinite-dimensional evolution systems [207], which is the case of the Fokker–Planck model.

The RH-MPC procedure is summarized in the following algorithm.

ALGORITHM 6.1. RH-MPC control.
Set $k = 0$; assign the initial PDF, $f(x,t_k) = \rho(x)$ and the targets $f_d(\cdot,t_k)$, $k = 0,\ldots,N-1$;

1. In (t_k,t_{k+1}), solve $\min_u J(f(u),u)$.
2. With the optimal solution u compute $f(\cdot,t_{k+1})$.
3. Assign this PDF as the initial condition for the Fokker–Planck problem in the next time window.
4. If $t_{k+1} < T$, set $k := k+1$, go to 1. and repeat.
5. End.

In the following, we report results of numerical experiments where we compute the control u that solves the control problem corresponding to a representative stochastic process. The classical problem of a particle immersed in a viscous fluid and subject to random Brownian fluctuations due to interaction with other particles is modeled by the Ornstein–Uhlenbeck process. We focus on the control of the PDF of this process. In this model, we have $b(X_t,t;u) = -\gamma X_t + u$, $\sigma(X_t,t) = \sigma$, where X_t represents the velocity of the particle and u is the momentum induced by an external force field defining the control mechanism.

In this case, the solution to the Fokker–Planck equation is well known to be a Gaussian distribution with mean

$$\mu(t;y,t_k;u) = u/\gamma + (y - u/\gamma)e^{-\gamma(t-t_k)}$$

and variance

$$\bar{\sigma}^2(t,t_k) = \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma(t-t_k)}).$$

Therefore the solution is as follows

$$\hat{f}(x,t;y,t_k;u) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2(t,t_k)}} \exp\left(-\frac{(x-\mu(t;y,t_k;u))^2}{2\bar{\sigma}^2(t,t_k)}\right). \tag{6.50}$$

This solution defines a mapping $\hat{f} = \hat{f}(u)$. Now, assuming an initial distribution ρ at $t = t_k$, and having $\hat{f}(x,t_{k+1};y,t_k)$, the final distribution $f(x,t_{k+1})$ is given by integration as defined in (6.43). This procedure provides a mapping $f = f(u)$, and thus the following reduced cost functional is obtained

$$J(f(u),u) = \frac{1}{2}\|(f(u))(\cdot,t_{k+1}) - f_d(\cdot,t_{k+1})\|^2_{L^2(\Omega)} + \frac{\nu}{2}|u|^2.$$

As a result, one obtains the following one-parameter optimization problem

$$\min_u J(f(u),u).$$

This problem can be solved efficiently by a bisection minimization procedure [265].

In Figure 6.7, we show the optimal solution for the Ornstein–Uhlenbeck process. This solution is obtained solving (6.48) with an appropriate conservative numerical scheme [7], and these results are almost identical to those obtained from the direct quadrature of (6.43) with (6.50). In this experiment, the parameters are $\gamma = 1$, $\sigma = 0.8$, $\nu = 0.1$. The initial distribution (black dotted) is a Gaussian with zero mean and variance 0.1; the target is also Gaussian with mean value following the law $2\sin(\pi t_k/5)$ and variance 0.2. The optimal distribution (solid line) is calculated at time windows of $\Delta t = 0.5$ until time $T = 5$. Other successful applications of this methodology are given in [7], where also a geometric-Brownian process with additive drift control and a controlled Shiryaev process are considered.

We remark that the complexity of this Fokker–Planck approach increases exponentially as the dimensionality of the random process X_t increases. Also the idea of solving the reduced minimization problem $\min_u J(f(u),u)$ by bisection may be too inefficient when increasing the dimensionality of the process and of the control space. In all these cases, it

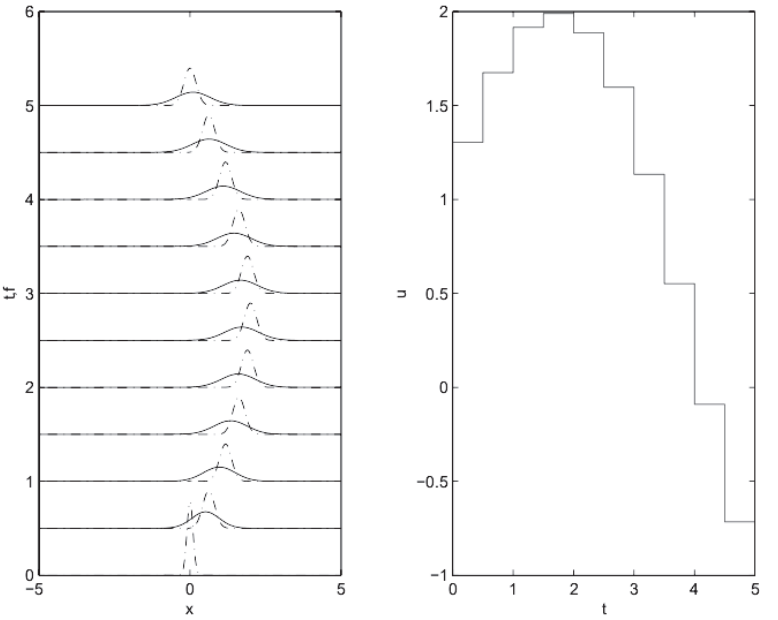

Figure 6.7. *The optimal control of the Ornstein–Uhlenbeck process. Left-hand side: computed PDF (solid line) and desired PDF (dotted-dashed line) at different time windows. Right-hand side: the optimal control function. Reprinted with permission from M. Annunziato and A. Borzì, Optimal control of probability density functions of stochastic processes, Math. Model. Anal.,* 15(4) (2010), 393–407.

becomes necessary to obtain gradient information by solving the following Fokker–Planck optimality system

$$\begin{cases} \partial_t f(x,t) - \frac{1}{2}\partial_x^2(\sigma(x,t)^2 f(x,t)) + \partial_x(b(x,t;u)f(x,t)) &= 0, \\ f(x,t_k) &= f_k(x), \\ -\partial_t p(x,t) - \frac{1}{2}\sigma(x,t)^2\partial_x^2 p(x,t) + b(x,t;u)\partial_x p(x,t) &= 0, \\ p(x,t_{k+1}) &= f_d(x,t_{k+1}) - f(x,t_{k+1}), \\ \nu u + \left(\partial_x(\frac{\partial b}{\partial u} f), p\right) &= 0. \end{cases}$$

6.6 Bayesian Uncertainty Quantification

Central to this section is the concept of the state of information over a parameter set or a parameter function. We assume to describe such a state of information as a probability density over the parameter space. Therefore also the results corresponding to such parameters and the information on the physical correlations between observable and model parameters are all described using probability densities. Here, model parameters play the role of the optimization function in the sense that the mathematical link between observable and model parameters is established through the formulation of an optimization problem. Moreover, we consider models defined by PDEs. Notice that the Bayesian approach has mainly been

investigated with inverse problems [336] and it is only recently that this framework started to be extended to inverse problems governed by PDE models; see [127, 247, 336].

6.6.1 Statistical Inverse Problems

In the statistical approach to inverse problems all variables included in the model are modeled as random variables. The randomness describes our degree of information and uncertainty concerning model realizations, and these values are coded in the probability distributions. In this framework, the solution of an inverse problem is represented by posterior probability distributions.

Assume that a quantity $y \in \mathbb{R}^q$ represents (part of) the state of the system under consideration and it is also the quantity that we measure (data) in order to determine another quantity $x \in \mathbb{R}^n$ which is not directly accessible to measurements. This assumption requires the existence of a governing model that describes how these two quantities are related. The model itself may be inaccurate and it may contain parameters that are not well known. Furthermore, the measured quantity y may be affected by noise. Therefore, within this framework it is reasonable to consider the following functional relationship

$$y = F(x, \epsilon), \tag{6.51}$$

where $F : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^q$ is the model function and $\epsilon \in \mathbb{R}^k$ is the vector containing all model uncertainty as well as the measurement noise. One possible approach to treat the presence of noise by appropriately chosen calibration measurements is to determine those components of ϵ that do not change from measurement to measurement. The noise, however, may be different from one instant to the other, and regularization methods are applied to cope with this part.

The statistical formulation of the inverse problem represented by (6.51) leads to the following

$$Y = F(X, E).$$

This is a parameter-to-observable map linking together the three random variables X, Y, and E, and consequently their probability distributions.

We call the directly observable random variable Y the measurement, and any of its realization $y_{obs} \in Y$ is the actual measurement producing the data. The nonobservable random variable X is called the unknown model. Here, $x \in X$ is a realization of X. In Bayesian theory, it is assumed that any information on the model which is available a priori can be coded into a probability density $\pi_{prior}(x)$ called the prior density. It expresses what we know about the model prior to the measurement. Further, the probability density $\pi_{noise}(\epsilon)$ describes the modeling error and the observation noise. We also introduce the likelihood function $\pi(y|x)$, which describes the relationship between the observable y and the model parameter x. It is the statistical representation of the (forward) model.

To be more specific, we illustrate the following few examples of PDF functions related to the present discussion. We do not discuss the appropriate normalization by constants and give the PDFs in terms of proportionality.

In the case where we have no a priori information on x, the $\pi_{prior}(x)$ equals a normalized constant function. If $x \in X$ is a realization of a Gaussian random field X with

given mean x_{prior}, we have

$$\pi_{prior}(x) \propto \exp\left(-\frac{1}{2}(x - x_{prior})^T C_X^{-1} (x - x_{prior})\right),$$

where C_X is the covariance matrix of X.

Concerning the modeling error and the measuring data error, for simplicity assume an additive noise such that the following model can be considered: $y = f(x) + \epsilon$. Then the theoretical model uncertainty described by a Gaussian probability density with covariance matrix C_M results in the following

$$\pi(y|x) \propto \exp\left(-\frac{1}{2}(y - f(x))^T C_M^{-1} (y - f(x))\right).$$

On the other hand, if there is no theoretical modeling uncertainty, we have

$$\pi(y|x) \propto \delta(y - f(x)),$$

where δ is the Dirac delta. In this latter case, we can only have measurement errors and we obtain

$$\pi_{noise}(\epsilon) = \pi_{noise}(y_{obs} - f(x)).$$

Now, we state Bayes' theorem that is central to the present discussion.

Theorem 6.1 (Bayes' theorem). *Given two events A and B, we have that*

$$P(B|A)P(A) = P(A|B)P(B).$$

Assume we know the (unconditional) probabilities $P(A)$ and $P(B)$ and the conditional probability $P(A|B)$ for the effect A given the cause B. If the effect A is observed, the Bayes formula gives the probability $P(B|A)$ for B being the cause. In particular, if A and B are independent, we have

$$P(A \cup B) = P(A)P(B);$$

then $P(B|A) = P(B)$ and $P(A|B) = P(A)$ and the statement becomes obvious. This latter remark means that the conditional probabilities equal the unconditional ones, and hence the term "independent" for A and B.

In our context, in terms of probability densities, Bayes' theorem states that the prior probability density, the likelihood function, and the data can be combined to form the posterior probability density on the model parameters X as follows

$$\pi_{post}(x) := \pi(x|y_{obs}) = \pi_{prior}(x)\pi(y_{obs}|x)/\pi(y_{obs}) \propto \pi_{prior}(x)\pi(y_{obs}|x). \qquad (6.52)$$

By definition, $\pi_{post}(x)$ is the solution of the inverse problem in the statistical sense.

6.6.2 A Fast Scheme for Large-Scale Linear Inverse Problems

We consider the case that the prior probability density of X and the probability density of error E are both Gaussian. With this setting, we discuss a methodology for the fast solution

of large-scale linear inverse problems presented in [127]. This methodology applies also in the case of high-dimensional parameter spaces, which arise when an infinite-dimensional parameter field is discretized.

Consider the following probability densities

$$\pi_{prior}(x) \propto \exp\left(-\frac{1}{2}(x - x_{prior})^T C_X^{-1} (x - x_{prior})\right)$$

and

$$\pi_{noise}(\epsilon) \propto \exp\left(-\frac{1}{2}(\epsilon - \epsilon_o)^T C_E^{-1} (\epsilon - \epsilon_o)\right),$$

where ϵ_o represents the mean error.

We insert these distributions in (6.52) assuming no modeling uncertainty and obtain

$$\pi_{post}(x) \propto \exp\left(-\frac{1}{2}\|x - x_{prior}\|^2_{C_X^{-1}} - \frac{1}{2}\|y_{obs} - f(x) - \epsilon_o\|^2_{C_E^{-1}}\right). \tag{6.53}$$

Notice that since $f(x)$ may be nonlinear in x, the posterior probability density may not be Gaussian. If we assume linearity of this map, $f(x) = A\,x$, then $\pi_{post}(x)$ is Gaussian, with mean x_{post} corresponding to the maximum point of $\pi_{post}(x)$, or equivalently

$$x_{post} = \underset{x}{\operatorname{argmin}} \left(\frac{1}{2}\|x - x_{prior}\|^2_{C_X^{-1}} + \frac{1}{2}\|y_{obs} - A\,x - \epsilon_o\|^2_{C_E^{-1}}\right). \tag{6.54}$$

We remark that this is similar to a regularized least-squares problem. It is clearly possible to see that the inverse covariance matrix of the posterior probability density of model parameters is given by the Hessian of the least-squares objectives. Therefore, we have

$$C_{post}^{-1} = A^T C_E^{-1} A + C_X^{-1}, \tag{6.55}$$

where $H_{misfit} = A^T C_E^{-1} A$ is the Hessian of the data misfit and $H_{prior} = C_X^{-1}$ is the Hessian of the prior. We denote $H_{post} = H_{misfit} + H_{prior}$.

This means that we have a complete knowledge of the solution to the inverse problem if we are able to estimate the mean x_{post} and the covariance matrix C_{post}, since then

$$\pi_{post}(x) \propto \exp\left(-\frac{1}{2}\|x - x_{post}\|^2_{C_{post}^{-1}}\right).$$

As discussed in [127], the task of determining the mean of the posterior density of the parameters can be found readily by solving (6.54) using state-of-the-art techniques from large-scale PDE-constrained optimization. However, computing the posterior covariance matrix of the parameters using (6.55) is a challenge. To solve this problem, in [127] a low-rank spectral approximation of the prior-preconditioned Hessian of the data misfit is computed at a cost that is a small multiple (independent of the parameter dimension) of the cost of solving the underlying forward PDEs. Then, with this low-rank approximation, the final expression for approximation of the posterior covariance is easily obtained. This approach is illustrated in the following. For all details see [127].

The first step in the development of this approach is to recognize that typically, H_{misfit} is spectrally similar to a discrete compact operator with the eigenvalues decaying rapidly to zero. This fact is used to construct a low-rank approximation of H_{misfit} by

retaining only a certain number of the largest eigenvalues and corresponding eigenvectors as follows.

Denote with λ_i and v_i the eigenvalues and eigenvectors of the prior-preconditioned Hessian of the data misfit given by

$$\tilde{H}_{misfit} = C_X^{1/2} \left(A^T C_E^{-1} A \right) C_X^{1/2}.$$

Further, let Λ be the diagonal matrix of the eigenvalues λ_i and let V be the matrix of the corresponding normalized eigenvectors, such that $\tilde{H}_{misfit} = V \Lambda V^T$.

Now, when the λ_i decay rapidly to zero, one can retain the first r largest eigenvalues and obtain the following low-rank approximation

$$\tilde{H}_{misfit} \approx V_r \Lambda_r V_r^T,$$

where Λ_r and V_r denote the truncated eigenvalue and eigenvector matrices. Specifically, the r columns of V_r correspond to the r eigenvectors of the first r largest eigenvalues with which Λ_r is constructed.

Next, an important step in [127] is to use the Sherman–Morrison–Woodbury formula to write the following expansion

$$\left(I + \tilde{H}_{misfit}\right)^{-1} = I - V_r D_r V_r^T + \mathcal{O}\left(\sum_{i=r+1}^{n} \frac{\lambda_i}{\lambda_i + 1} \right),$$

where $D_r = \text{diag}\left(\frac{\lambda_i}{\lambda_i+1}\right) \in \mathbb{R}^{r \times r}$. Notice that this expression also provides the estimate of the error due to truncation. For rapidly decaying eigenvalues the rest sum quickly approaches zero by increasing the order of truncation.

With this preparation, we are now ready to construct a low-rank approximation to C_{post}. We have

$$\begin{aligned} C_{post} &= \left(A^T C_E^{-1} A + C_X^{-1} \right)^{-1} = C_X^{1/2} \left(A^T C_E^{-1} A + I \right)^{-1} C_X^{1/2} \\ &\approx C_X^{1/2} \left(I - V_r D_r V_r^T \right) C_X^{1/2} \\ &= C_X - C_X^{1/2} V_r D_r V_r^T C_X^{1/2}. \end{aligned} \tag{6.56}$$

The importance of this formula lies in the fact that usually the choice of a small r is reasonable and independent of problem size. This is particularly true when the prior is of smoothing type.

As shown in [127] one cannot explicitly construct $\tilde{H}_{misfit}$ and apply a truncated singular value decomposition algorithm to obtain Λ_r and V_r. The viable approach is to use a matrix-free Lanczos method to find the dominant eigenvalues and corresponding eigenvectors. The application of the Lanczos scheme requires only a matrix-vector product with $\tilde{H}_{misfit}$ at each iteration. By definition of $\tilde{H}_{misfit}$, this process can be broken down in a few steps: multiplications with C_X, solving for C_E, and action of A and A^T on vectors. The last involve forward and adjoint PDE solutions, which dominate the cost of the computation when one considers a PDE forward model.

6.6.3 An Inverse Parabolic Problem

In this section, we illustrate the Bayesian framework in the case of an inverse parabolic problem. Our purpose is also to provide a clear link to the PDE-based deterministic optimization framework considered in previous sections.

Consider the following parabolic inverse problem:

$$\begin{aligned} \min_{y_0 \in L^2(\Omega)} J(y, y_0) &:= \frac{\beta_{noise}}{2} \|y(\cdot, T) - y_{obs}\|^2_{L^2(\Omega)} + \frac{\beta_{prior}}{2} \|y_0\|^2_{L^2(\Omega)}, \\ -\partial_t y + \Delta y &= 0 \qquad \text{in } Q = \Omega \times (0, T), \\ y &= y_0 \qquad \text{in } \Omega \times \{t = 0\}. \end{aligned} \tag{6.57}$$

That is, find the initial configuration y_0 that represents the unknown distributed parameter x, based on the observations at the final time T that provide the data y_{obs}.

Notice that the discretized functional $J(y, y_0)$ provides the minus exponent of the noise and prior PDEs, respectively. We have

$$C_X^{-1} = \beta_{prior}\, h^2\, I \qquad \text{and} \qquad C_E^{-1} = \beta_{noise}\, h^2\, I,$$

where h is the space mesh size. Clearly, $(y_0)_{prior} = 0$.

We denote with A the discrete parabolic operator such that $y_T = A\, y_0$, where y_T is the solution of the discretized parabolic problem in (6.57) with y_0 as initial condition. Therefore, discretization of the minimization problem (6.57) results in the following posterior mean:

$$(y_0)_{post} = \operatorname*{argmin}_{y_0} \left(\frac{h^2}{2}\, \beta_{prior}\, y_0^T\, y_0 + \frac{h^2}{2}\, \beta_{noise}\, (y_{obs} - A\, y_0)^T\, (y_{obs} - A\, y_0) \right).$$

The choice of prior and noise distributions given above result in the following:

$$\tilde{H}_{misfit} = \frac{\beta_{noise}}{\beta_{prior}}\, A^T\, A,$$

where A is the forward operator and A^T is the adjoint operator. On a square domain and homogeneous Dirichlet boundary conditions, one can easily verify that the eigenvalues λ_i of $\tilde{H}_{misfit}$ are linearly proportional to the ratio $\beta_{noise}/\beta_{prior}$ and decay exponentially; see [127].