SIAM J. MATRIX ANAL. APPL.
Vol. 33, No. 1, pp. 195–209

A POSTERIORI ERROR ANALYSIS OF PARAMETERIZED LINEAR SYSTEMS USING SPECTRAL METHODS*

T. BUTLER†, P. CONSTANTINE‡, AND T. WILDEY§

Abstract. We develop computable a posteriori error estimates for the pointwise evaluation of linear functionals of a solution to a parameterized linear system of equations. These error estimates are based on a variational analysis applied to polynomial spectral methods for forward and adjoint problems. We also use this error estimate to define an *improved linear functional* and we prove that this improved functional converges at a much faster rate than the original linear functional given a pointwise convergence assumption on the forward and adjoint solutions. The advantage of this method is that we are able to use low order spectral representations for the forward and adjoint systems to cheaply produce linear functionals with the accuracy of a higher order spectral representation. The method presented in this paper also applies to the case where only the convergence of the spectral approximation to the adjoint solution is guaranteed. We present numerical examples showing that the error in this improved functional is often orders of magnitude smaller. We also demonstrate that in higher dimensions, the computational cost required to achieve a given accuracy is much lower using the improved linear functional.

Key words. a posteriori error analysis, adjoint problem, spectral methods, parameterized linear systems

AMS subject classifications. 65D30, 65F99, 65C99, 65B99

DOI. 10.1137/110840522

1. Introduction.

Parameterized linear systems arise naturally in electronic circuit design [33], applications of PageRank [6, 11], and dynamical systems [13]. In addition, such systems are frequently encountered in computational methods for stochastic partial differential equations once the differential operator is discretized in the physical and/or time domains. In recent years, the use of spectral methods has become an increasingly popular way to approximate response surfaces and to reduce the computational burden of Monte Carlo sampling for such systems. Several methods have been developed along these lines: stochastic Galerkin [26, 12, 3, 40, 29, 28, 25], stochastic collocation [2, 39, 41, 23, 22, 34], and nonintrusive spectral projection [1, 36], just to name a few. A theoretical comparison of some of these methods for parameterized linear systems is given in [10].

Given a parameterized linear system, all these methods use a truncated spectral expansion which introduces approximation error. The general appeal of these methods can be attributed to the exponential convergence of certain moments of the solution if a proper spectral basis is chosen. In many cases, however, the goal of the simulation is not to compute moments of the solution but to compute distributions or probabilities associated with certain functionals of the solution. For example, statistical inference

*Received by the editors July 12, 2011; accepted for publication (in revised form) by D. B. Szyld November 20, 2011; published electronically March 13, 2012.
http://www.siam.org/journals/simax/33-1/84052.html

†Institute for Computational Engineering and Sciences (ICES), University of Texas at Austin, Austin, TX 78712 (tbutler@ices.utexas.edu).

‡Department of Mechanical Engineering, Stanford University, Stanford, CA 94305 (paul.constantine@stanford.edu).

§Optimization and Uncertainty Quantification Department, Sandia National Labs, Albuquerque, NM 87185 (tmwilde@sandia.gov). Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

problems using Bayesian methods require accurate and efficient estimates of distributions or probabilities. For such problems, the moments of the spectral representation are useful only if the output distribution happens to have a particularly simple form, such as Gaussian. In [31, 30], the computational efficiency of the inference problem was dramatically improved by sampling the spectral representation rather than the full model. While this approach is very appealing in terms of the computational cost, the reliability of the predictions relies on the pointwise accuracy of the spectral representation. This accuracy may be lacking for the low order spectral methods which are commonly used for high dimensional parameterized systems.

Meanwhile, computational modeling is becoming increasingly reliant on a posteriori error estimates to provide a measure of reliability on the numerical predictions. This methodology has been developed for a variety of methods and is widely accepted in the analysis of discretization error for partial differential equations [4, 15, 21]. The adjoint-based (dual-weighted residual) method is motivated by the observation that often the goal of a simulation is to compute a small number of linear functionals of the solution, such as the average value in a region or the drag on an object, rather than controlling the error in a global norm. This method has been successfully extended to estimate numerical errors due to operator splittings [16], operator decomposition for multiscale/multiphysics applications [9, 19, 20], adaptive sampling algorithms [17, 18], and inverse sensitivity analysis [5, 8]. It was also used in [32] to estimate the error in moments of linear functionals for the stochastic Galerkin approximation of a partial differential equation. In [7], the present authors used adjoint-based analysis to provide pointwise estimates for a stochastic Galerkin approximation of a partial differential equation including spatial and temporal errors. The spectral representation error was not included in the estimate, but a sufficiently high order spectral method was used so that this component was negligible.

The goal of this paper is to extend the methodology developed in [7] to parameterized linear systems and to show that the error estimate can include an accurate approximation of the spectral representation error. We also use this estimate to define an *improved linear functional* and we show that the pointwise accuracy of this improved linear functional is asymptotically more accurate than the original linear functional, even if a low order approximation of the adjoint is used. We summarize the impact of this result in the following remark:

> *If we solve the forward problem with a low order spectral method and solve the adjoint problem with another low order spectral method, then the error in the improved functional value is roughly the same as if we had solved the forward problem with a much higher order spectral method. If the parameter space is high dimensional, the computational savings of this approach may be significant.*

By focusing on parameterized linear systems rather than discretized partial differential equations, we are able to present the *improved linear functional* and the corresponding error analysis without worrying about spatial or temporal discretization errors. The extension of this work to estimate the total error in spectral approximations for discretized stochastic partial differential equations is the subject of a forthcoming article. The utilization of an a posteriori error estimate as a corrector to define an improved linear functional has been explored in other contexts [27, 35]. However, the error estimate typically requires a projection of the discrete approximations to the forward and adjoint problems into higher order spaces. Such techniques do not apply to parameterized linear systems, and the approach in this paper does not require any additional projections.

This paper is organized as follows. In section 2, we define the parameterized linear systems. In section 3, we briefly described the spectral Galerkin and the nonintrusive spectral projection algorithms for the approximation of the parameterized linear system and discuss a priori error bounds. In section 4, we develop an a posteriori error analysis for the pointwise values of linear functionals of the solution. In section 5, we define the *improved linear functional* and prove a pointwise error bound. A comparison of the computational cost of this approach versus solving the forward problem with a higher order method is made in section 6. Numerical results demonstrating the accuracy and efficiency of the improved linear functional are presented in section 7. We make our concluding remarks in section 8.

2. Parameterized linear systems. Let $\Omega = \Omega_1 \otimes \cdots \otimes \Omega_d$, where each Ω_i may be bounded or unbounded with elements denoted by $s = (s_1, \ldots, s_d)$. Let $w(s) = w_1(s_1) \cdots w_d(s_d)$ be a positive, separable weight function satisfying

$$\int_\Omega p(s)w(s)\, ds < \infty \tag{2.1}$$

for any polynomial $p(s)$, and $\int_\Omega w(s)\, ds = 1$. In a probabilistic context, $w(s)$ is a probability density function on Ω.

Let $x(s) \in \mathbb{R}^n$ satisfy the linear system of equations

$$A(s)x(s) = b(s), \qquad s \in \Omega, \tag{2.2}$$

for a bounded $A(s) \in \mathbb{R}^{n\times n}$ and $b(s) \in \mathbb{R}^n$. We assume that (2.2) is well-posed for all $s \in \Omega$, i.e., $x(s) = A^{-1}(s)b(s)$ for any $s \in \Omega$. Often, we are interested in accurately computing some functional of the solution, $g(x(s))$, rather than accurately computing the entire solution $x(s)$. In this paper, we take $g(x(s))$ to be a linear functional, but the extension to nonlinear functionals is also possible.

If the elements of $A(s)$ and $b(s)$ are analytic in a region containing Ω, then the components of $x(s)$ will also be analytic. For such cases, polynomial approximation will rapidly converge pointwise. However, for situations where the problem data is discontinuous with respect to the parameters, the solution will also be discontinuous, and polynomial approximations will not converge pointwise due to Gibbs phenomona. The analysis in section 4 holds regardless of whether the solution is smooth, and we present numerical results in section 7 confirming this.

3. Spectral approximation methods. We focus our attention on two of the most popular spectral approximation methods, namely, the spectral Galerkin method and the pseudospectral projection method. The results in this paper can easily be extended to virtually any approximation method where a global surrogate model is constructed and pointwise values are sampled throughout the parameter domain, Ω.

3.1. Background. Before reviewing the spectral Galerkin and pseudospectral methods, we set up some notation for working with multivariate polynomials. We employ the standard multi-index notation: let $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ be a multi-index, and define the basis polynomial

$$\pi_\alpha(s) = \pi_{\alpha_1}(s_1) \cdots \pi_{\alpha_d}(s_d). \tag{3.1}$$

The polynomial $\pi_{\alpha_i}(s_i)$ is the orthonormal polynomial of degree α_i, where the orthogonality is defined with respect to the weight function $w_i(s_i)$. Then for $\alpha, \beta \in \mathbb{N}^d$,

$$\int_\Omega \pi_\alpha(s)\pi_\beta(s)w(s)\, ds = \begin{cases} 1, & \alpha = \beta, \\ 0 & \text{otherwise}, \end{cases} \tag{3.2}$$

where equality between multi-indicies means componentwise equality. For more on multivariate polynomials, see [24, 14].

We define the discrete inner product,

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i, \qquad x, y \in \mathbb{R}^n,$$

and the discrete space,

$$l^2(\mathbb{R}^n) := \left\{x \in \mathbb{R}^n, \|x(s)\|_{l^2(\mathbb{R}^n)} < \infty\right\}$$

with the corresponding norm,

$$\|x\|_{l^2(\mathbb{R}^n)} = \langle x, x\rangle^{1/2}.$$

In addition, we define

$$L^2(\Omega; l^2(\mathbb{R}^n)) := \left\{x(s) : \|x(s)\|_{L^2(\Omega; l^2(\mathbb{R}^n)} < \infty\right\},$$
$$L^\infty(\Omega; l^2(\mathbb{R}^n)) := \left\{x(s) : \|x(s)\|_{L^\infty(\Omega; l^2(\mathbb{R}^n)} < \infty\right\}$$

with the following norms:

$$\|x(s)\|_{L^2(\Omega; l^2(\mathbb{R}^n))} := \left(\int_\Omega \|x(s)\|^2_{l^2(\mathbb{R}^n)} w(s)\ ds\right)^{1/2},$$
$$\|x(s)\|_{L^\infty(\Omega; l^2(\mathbb{R}^n))} := \sup_{s\in\Omega} \|x(s)\|_{l^2(\mathbb{R}^n)}.$$

We assume that $x(s) \in L^2(\Omega; l^2(\mathbb{R}^n)) \cap L^\infty(\Omega; l^2(\mathbb{R}^n))$.

In addition to the previously defined norms, we also will use

$$\|z(s)\|_{L^\infty(\Omega)} := \sup_{s\in\Omega} |z(s)|$$

for any sufficiently smooth scalar function $z : \mathbb{R} \to \mathbb{R}$.

3.2. Spectral Galerkin. The spectral Galerkin method approximates each component of the parameterized vector $x(s)$ with a finite degree polynomial. We can define the space of polynomials with a set of multi-indices. Let $\mathcal{I}_N$ be a set of multi-indices parameterized by N. For example, the most common set for a spectral Galerkin approximation is

$$\mathcal{I}_N = \{\alpha \in \mathbb{N}^d \,:\, \alpha_1 + \cdots + \alpha_d \le N\}. \tag{3.3}$$

We let $\mathbb{P}^N$ denote the space of polynomials defined by

$$\mathbb{P}^N = \text{span}\{\pi_\alpha(s) \,:\, \alpha \in \mathcal{I}_N\}. \tag{3.4}$$

This particular choice for $\mathcal{I}_N$ corresponds to the full polynomials of degree at most N. The dimension of $\mathbb{P}^N$, which we denote by $|\mathbb{P}^N|$, is $|\mathbb{P}^N| = \binom{d+N}{N}$.

The spectral Galerkin approximation seeks a vector of polynomials $X_N(s)$ with each component in $\mathbb{P}^N$ such that

$$\int_\Omega A(s)X_N(s)\pi_\alpha(s)w(s)\ ds = \int_\Omega b(s)\pi_\alpha(s)w(s)\ ds, \quad \alpha \in \mathcal{I}_N. \tag{3.5}$$

Galerkin methods yield the convenient orthogonality relation,

$$\int_\Omega A(s)e(s)\pi_\alpha(s)w(s)\ ds = \int_\Omega R(s)\pi_\alpha(s)w(s)\ ds = 0 \quad \forall\, \pi_\alpha(s) \in \mathbb{P}^N, \tag{3.6}$$

where $e(s) = x(s) - X_N(s)$ and $R(s) = b(s) - A(s)X_N(s)$.

In [10], convergence of the spectral Galerkin approximation for the parameterized linear system (2.2) in the $L^2(\Omega; l^2(\mathbb{R}^n))$ norm is shown, i.e.,

$$\|e(s)\|_{L^2(\Omega;l^2(\mathbb{R}^n))} \le C\rho^{-N},$$

for some $\rho > 1$ provided $x(s)$ is analytic in a region containing Ω.

Note that computing the spectral Galerkin approximation requires solving a large coupled linear system. The linear system has a well-defined block structure and may be sparse depending on the choice of basis and the dependence of $A(s)$ on s. However, the method is usually deemed *intrusive* or *embedded* because special algorithms need to be developed to apply stochastic Galerkin to an existing code.

3.3. Pseudospectral projection. If we assume that each element of the solution $x(s)$ of the parameterized matrix equation (2.2) is analytic in a region containing Ω, then we can write the convergent Fourier expansion in vector notation,

$$x(s) = \sum_{\alpha\in\mathbb{N}^d} x_\alpha \pi_\alpha(s), \tag{3.7}$$

where the equality is in the L^2 sense and, by othogonality,

$$x_\alpha = \int_\Omega x(s)\pi_\alpha(s)w(s)\ ds. \tag{3.8}$$

Unfortunately, the coefficients x_α depend on $x(s)$, which is unknown. The pseudospectral method approximates these coefficients using a quadrature rule, i.e.,

$$x_{\alpha,m} = \sum_{j=1}^{m} x(s_j)\pi_\alpha(s_j)w_j,$$

where $\{s_j\}$ and $\{w_j\}$ are the integration points and weights, respectively. Let $\mathbb{P}^N$ be defined as in the previous section. The approximate solution $X_N(s)$ with components in $\mathbb{P}^N$ is given by

$$X_N(s) = \sum_{\alpha\in\mathbb{I}_N} x_{\alpha,m}\pi_\alpha(s). \tag{3.9}$$

In general, the number of collocation points need not be related to N, but it was shown in [10] that a proper choice of Gaussian quadrature rule gives precisely the same solution as the spectral collocation method, and both methods converge geometrically in the $L^2(\Omega; l^2(\mathbb{R}^n))$ norm,

$$\|e(s)\|_{L^2(\Omega;l^2(\mathbb{R}^n))} \le C\rho^{-N},$$

if $x(s)$ is analytic.

The advantage in using pseudospectral over spectral Galerkin is that pseudospectral is nonintrusive, meaning that m linear systems of size $n\times n$ must be solved. These linear systems are independent, which leads to a simple parallel implementation. However, if the quadrature points are chosen to be tensor products of one-dimensional (1D) Gaussian rules, then m increases exponentially with the dimension of Ω.

3.4. Pointwise error bounds. If we assume the solution $x(s)$ is analytic over Ω, then the spectral approximation converges pointwise. Such results were derived in [37] to sparse chaos approximations of stochastic partial differential equations. Pointwise convergence of stochastic Galerkin methods was also used in [38] to obtain convergence with respect to a non-Gaussian measure.

Our objective is not to prove $L^\infty(\Omega; l^2(\mathbb{R}^n))$ rates of convergence of spectral methods. We merely aim to show that given a particular pointwise error for the forward problem, we can define an improved linear functional with a much smaller error in many cases. To this end, we assume that the $L^\infty(\Omega; l^2(\mathbb{R}^n))$ estimate

$$\|e(s)\|_{L^\infty(\Omega; l^2(\mathbb{R}^n))} \leq \epsilon_1(N) \tag{3.10}$$

holds for some $\epsilon_1(N) \geq 0$. Ideally, we would have $\epsilon_1(N) \downarrow 0$ monotonically as $N \to \infty$, but the main result in this paper, Theorem 5.2, will hold regardless of the form of $\epsilon_1(N)$.

4. A posteriori error analysis. Often we are interested in some quantity of interest given by a linear functional, $g(x(s))$, where g depends on s only through $x(s)$. According to the Reisz representation theorem, there exists a unique $\psi \in \mathbb{R}^n$ such that $g(x(s)) = \langle \psi, x(s) \rangle$ for any $s \in \Omega$. Let $\phi(s) \in \mathbb{R}^n$ satisfy the adjoint system

$$A^T(s)\phi(s) = \psi, \qquad s \in \Omega. \tag{4.1}$$

The assumptions made on the forward problem (2.2) also imply that there exists a unique $\phi(s) = (A(s))^{-T}\psi$ for any $s \in \Omega$.

Recall that the spectral approximation is frequently used as a surrogate model and is sampled many times to compute probabilities or densities of a quantity of interest. In this scenario, we are more interested in the error in a quantity of interest for a given $s \in \Omega$. Let $X_N(s)$ denote the spectral approximation and define $e(s) = x(s) - X_N(s)$ and $R(s) = b(s) - A(s)X_N(s)$. Since we assume that both (2.2) and (4.1) hold for any $s \in \Omega$, we can derive the following error representation:

$$\begin{aligned} \langle \psi, e(s) \rangle &= \left\langle A^T(s)\phi(s), e(s) \right\rangle \\ &= \langle \phi(s), A(s)e(s) \rangle \\ &= \langle \phi(s), R(s) \rangle . \end{aligned}$$

If we know $\phi(s)$, then the pointwise error can be computed exactly regardless of the accuracy in the spectral approximation of the forward problem.

Previous adjoint-based a posteriori error analysis for the spectral Galerkin approximation computed an integral of the error in the linear functional over Ω [32]. This approach could then take advantage of the orthogonality relation (3.6). In our approach, there is no orthogonality relation because we are evaluating the error representation at a point, $s \in \Omega$, rather than integrating over the domain. This implies that we can use a low order approximation of the adjoint to compute the error estimate. This observation is critically important to the remainder of this paper. One of the difficulties in using adjoint-based error estimates is the need to avoid orthogonality. This usually requires solving the adjoint with a higher order method than was used for the forward problem. As we shall see in sections 5 and 7, we are able to produce accurate error estimates and improved linear functionals using very low order adjoint approximations.

5. An improved linear functional. Let Φ_M be any Mth order approximation of ϕ defined over Ω. We are particularly interested in the case where Φ_M is computed using an Mth order spectral approximation method. Including the approximation of the adjoint solution into the error representation gives

$$\langle \psi, e(s) \rangle = \langle \Phi_M(s), R(s) \rangle + \langle \phi(s) - \Phi_M(s), A(s)e(s) \rangle . \tag{5.1}$$

The second term on the right-hand side is a truncation error and is clearly higher order since it involves the product of $e(s)$ and $\phi(s) - \Phi_M(s)$. To take advantage of this high order term, we use the error estimate to define an *improved linear functional*.

DEFINITION 5.1. *The* improved linear functional *is given by*

$$l(s) = \langle \psi, X_N(s) \rangle + \langle \Phi_M(s), R(s) \rangle ,$$

which is fully computable given X_N and Φ_M.

We recognize the improved linear functional as the functional of the spectral approximation plus an estimate of the error.

THEOREM 5.2. *Assume that the pointwise errors for the forward and adjoint approximations satisfy*

$$\|e(s)\|_{L^\infty(\Omega;l^2(\mathbb{R}^n))} \leq \epsilon_1(N) \tag{5.2}$$

and

$$\|\phi(s) - \Phi_M(s)\|_{L^\infty(\Omega;l^2(\mathbb{R}^n))} \leq \epsilon_2(M) \tag{5.3}$$

respectively, where $\epsilon_1(N), \epsilon_2(M) \geq 0$. Then the pointwise error in the improved functional satisfies

$$\| \langle \psi, x(s) \rangle - l(s) \|_{L^\infty(\Omega)} \leq C\epsilon_1(N)\epsilon_2(M),$$

where $C > 0$ depends only on $A(s)$.

Proof. Using (5.1), we easily see that the error in $l(s)$ is given by

$$\langle \psi, x(s) \rangle - l(s) = \langle \phi(s) - \Phi_M(s), A(s)e(s) \rangle .$$

We take the supremum over Ω and bound

$$\begin{aligned} \| \langle \psi, x(s) \rangle - l(s) \|_{L^\infty(\Omega)} &= \| \langle \phi(s) - \Phi_M(s), A(s)e(s) \rangle \|_{L^\infty(\Omega)} \\ &\leq C\|e(s)\|_{L^\infty(\Omega;l^2(\mathbb{R}^n))}\|\phi(s) - \Phi_M(s)\|_{L^\infty(\Omega;l^2(\mathbb{R}^n))} \\ &\leq C\epsilon_1(N)\epsilon_2(M) \end{aligned}$$

using a Cauchy–Schwarz inequality and the assumptions (5.2) and (5.3). ∎

To illustrate the consequences of Theorem 5.2, consider the following examples:

- Suppose $\epsilon_1(N) = \rho^{-N}$ and $\epsilon_2(M) = \rho^{-M}$. Therefore, we have
$$\epsilon_1(N)\epsilon_2(M) \approx \epsilon_1(N + M).$$
This is the case most often encountered if $x(s)$ and $\phi(s)$ are smooth.
- Suppose $\epsilon_1(N) = N^{-\alpha}$ and $\epsilon_2(M) = M^{-\beta}$ for some $\alpha, \beta > 0$. Then
$$\epsilon_1(N)\epsilon_2(M) = N^{-\alpha}M^{-\beta},$$
which is often smaller than
$$\epsilon_1(N + M) = (N + M)^{-\alpha}.$$
In this case, the error in the improved functional may be much better than the error computed by solving the forward problem with a higher order method.

- Suppose $\epsilon_1(N) = \rho^{-N}$ and $\epsilon_2(M) = M^{-\beta}$, meaning that the adjoint problem is more difficult to resolve than the forward problem. Then we may have

$$\epsilon_1(N)\epsilon_2(M) \gg \epsilon_1(N+M),$$

  and the error in the improved functional may not be smaller than the error computed by solving the forward problem with a higher order method.

In most cases, it is equally difficult to solve the forward and the adjoint problems. Sometimes the adjoint is even easier to solve. For example, if A does not depend on s, then the adjoint solution will not depend on s and $\epsilon_2(M)$ will be zero for $M \geq 1$. This particular case is well known and a single adjoint solve can be used to estimate sensitivity of the forward solution to variations in the data.

Remark 5.1. There are numerous examples where spectral approximation methods fail, e.g., when $x(s)$ is discontinuous. Our approach is designed for the case where the spectral approximations of the forward and the adjoint problems converge, but it also applies to the case where only the spectral approximation of the adjoint problem converges. Therefore we believe that this approach extends the applicability of the spectral approximation methods.

6. Comparison of the computational cost. In the previous section, we compared the accuracy of the improved linear functional and the linear functional computed using a higher order method. However, it is important to pay careful attention to the computational cost required to achieve this accuracy. For example, solving two lower order systems is often much easier than solving one high order system, especially if the stochastic space is high dimensional. Indeed, there may be scenarios where increasing the order of the forward problem may be virtually impossible, but solving an adjoint problem of the same (or lower) order is feasible.

To compare the computational cost using the improved functional value with the cost in using a higher order method for the forward problem, we assume that $\epsilon_1(N)\epsilon_2(M) \approx \epsilon_1(N+M)$. A comparison of the cost can be made for other cases as well. For simplicity, we set $M = N$, i.e., we solve the forward and adjoint problems with the same order approximation and compare the cost in solving two order N problems (forward and adjoint) with solving one order $2N$ forward problem.

It is well known that an Nth order spectral Galerkin method requires the solution of an $n|\mathbb{P}^N| \times n|\mathbb{P}^N|$ linear system, where $|\mathbb{P}^N|$ is defined as in section 3.1. For simplicity, we assume that the linear system can be solved in $\mathcal{O}((n|\mathbb{P}^N|)^\alpha)$ operations for some $\alpha \geq 1$. Achieving $\alpha = 1$ may be difficult in general, but as we will see, assuming a less optimal linear solver only increases the computational efficiency of the improved linear functional.

Example 1. If $d = 1$ and $\alpha = 1$, then $|\mathbb{P}^N| = N + 1$ and $2|\mathbb{P}^N| \approx |\mathbb{P}^{2N}|$. However, if $\alpha > 1$, then $2|\mathbb{P}^N| \ll |\mathbb{P}^{2N}|$ for large N and it becomes much cheaper to solve two order N problems than one order $2N$ problem.

Example 2. If $d = 10$, $\alpha = 1$, and $N = 4$, then $|\mathbb{P}^N| = 1001$, while $|\mathbb{P}^{2N}| = 43758$. Therefore, the improved functional is approximately 22 times cheaper to compute than the higher order functional. However, if $\alpha = 2$, then the improved functional is approximately 22^2 times cheaper to compute.

As mentioned in section 3, an Nth order pseudospectral method using tensor product quadrature rules requires solving m independent linear systems of size $n \times n$. To maintain a consistent notation, we assume $m = m_1^d$, where m_1 is the number of quadrature points in one dimension and is proportional to N. This assumption is merely for the ease of presentation and is not necessary in practice. For simplicity,

we assume that solving the $n \times n$ linear system may be achieved in $\mathcal{O}(n^\beta)$ operations although the value of β is not as important in this case.

Example 3. If $d = 1$, then the cost in solving the forward problem of order N is $m \approx N$ and the cost in solving the forward problem of order $2N$ is $m = 2N$. Therefore the cost is nearly the same regardless of β.

Example 4. If $d = 10$, then the cost in solving the forward problem of order $N = 4$ is $m \approx 4^{10}$, while the cost in solving the forward problem of order $N = 8$ is $m = 8^{10}$. Therefore, the improved functional is approximately 512 times easier to compute than the higher order functional.

7. Numerical results. In this section we present three numerical results to verify the error estimates and the computational efficiencies. The first example is a simple 1D problem used to verify the rate of convergence of the pointwise error in both the original and the improved linear functionals. The second example demonstrates that accurate estimates can be obtained using the improved functional even if the spectral approximation of the forward problem fails, provided the spectral approximation of the adjoint problem is accurate. Finally, the third example comes from a finite element discretization of an elliptic partial differential equation with a random permeability tensor and demonstrates the potential gains in computational efficiency for high dimensional problems.

7.1. A 2×2 example. The following example was investigated in [10] to demonstrate the $L^2(\Omega; l^2(\mathbb{R}^n))$ convergence of the spectral Galerkin method. Let $\epsilon > 0$, and consider the following parameterized matrix equation:

$$\begin{bmatrix} 1+\epsilon & s \\ s & 1 \end{bmatrix} \begin{bmatrix} x_1(s) \\ x_2(s) \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

with $s \in \Omega = [-1, 1]$ and the exact solution

$$x_1(s) = \frac{2-s}{1+\epsilon-s^2}, \qquad x_2(s) = \frac{1+\epsilon-2s}{1+\epsilon-s^2}.$$

As $\epsilon \to 0$, the singularity moves closer to Ω and the spectral approximation converges at a slower rate. We take $w(s) = 1/2$ and use the Legendre basis for the spectral Galerkin approximation and the Gauss–Legendre integration points for the pseudospectral method. We take as our quantity of interest the error in the first component of the solution so that $\psi = [1, 0]^T$ in the adjoint.

To compare the pointwise accuracy of the linear functional and the improved linear functional, we estimate the $L^\infty(\Omega)$ norm of each by randomly sampling s at 1000 points throughout Ω. In Figure 7.1, we plot the resulting $L^\infty(\Omega)$ errors for $\epsilon = 0.8$ (left) and $\epsilon = 0.2$ (right) using the spectral Galerkin approximation. As noted in [10], the spectral Galerkin and the pseudospectral spectral projection methods give the same solution for this example, so we only present one set of results. The error in the linear functional for $1 \le N \le 5$ is given by the top curve. Each of the other curves corresponds to the error in the improved linear functional for a certain order adjoint approximation. We see that even a first order adjoint approximation improves the pointwise error in the linear functional as predicted by Theorem 5.2.

In practice, it is often preferable to solve the forward and adjoint problems using the same order method. In Figure 7.2, we plot only the error in the functional value and the error in the improved functional value computed using the same order method for both the forward and adjoint problems. In this example, the error in the improved

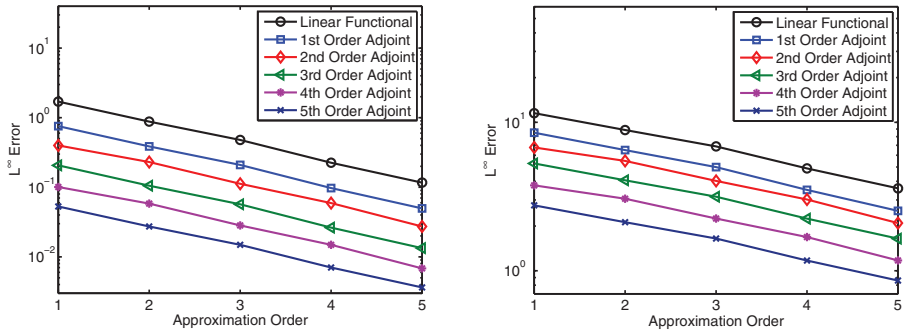

FIG. 7.1. $L^\infty(\Omega)$ *errors in the quantity of interest for* $\epsilon = 0.8$ *(left) and* $\epsilon = 0.2$ *(right). The top line (in black) is the error in the linear functional. The other lines correspond to the* $L^\infty(\Omega)$ *error in the improved functional using different order adjoint approximation.*

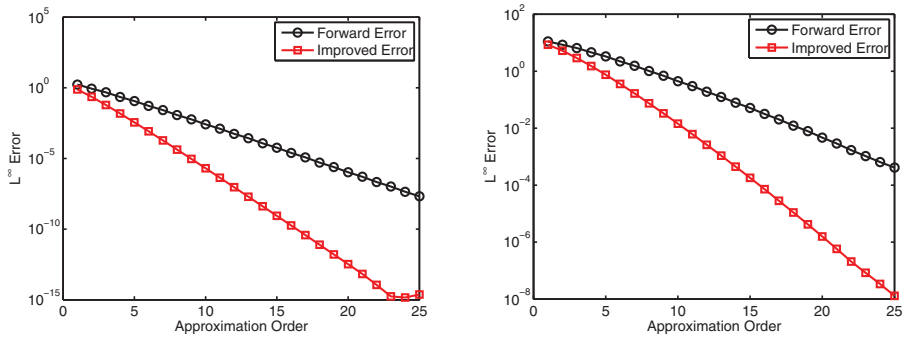

FIG. 7.2. $L^\infty(\Omega)$ *errors in the quantity of interest for* $\epsilon = 0.8$ *(left) and* $\epsilon = 0.2$ *(right). The black line gives the error in the linear functional, while the red line gives the error in the improved functional value computed by solving the forward and adjoint problems with the same order method.*

functional value using an order N method for the forward problem and an order M method for the adjoint problem is nearly the same as using an order $N + M$ method for the forward problem.

Next, we verify that the original linear functional converges pointwise at the same rate as the $L^\infty(\Omega; l^2(\mathbb{R}^n))$ error for the forward solution and that the improved linear functional converges pointwise at the product of the $L^\infty(\Omega; l^2(\mathbb{R}^n))$ rates for the forward and adjoint solutions. In Figure 7.3, we plot the $L^\infty(\Omega; l^2(\mathbb{R}^n))$ errors in the forward and adjoint solutions for each choice of ϵ. We estimate the slopes of the lines corresponding to both forward and adjoint solutions to be -0.327 for $\epsilon = 0.8$ and -0.169 for $\epsilon = 0.2$. From Figure 7.2, we estimate the slope of the line corresponding to the original linear functional to be -0.327 for $\epsilon = 0.8$ and -0.169 for $\epsilon = 0.2$. In this example, the rate of pointwise convergence of the original linear functional is almost exactly the same as the $L^\infty(\Omega; l^2(\mathbb{R}^n))$ convergence rate for the forward solution. Meanwhile, the slope of the line corresponding to the improved linear functional is -0.654 for $\epsilon = 0.8$ (before reaching machine precision) and -0.338 for $\epsilon = 0.2$. Thus, the rate of pointwise convergence of the improved linear functional is almost exactly the same as the *product* of the $L^\infty(\Omega; l^2(\mathbb{R}^n))$ convergence rates for the forward and adjoint solutions.

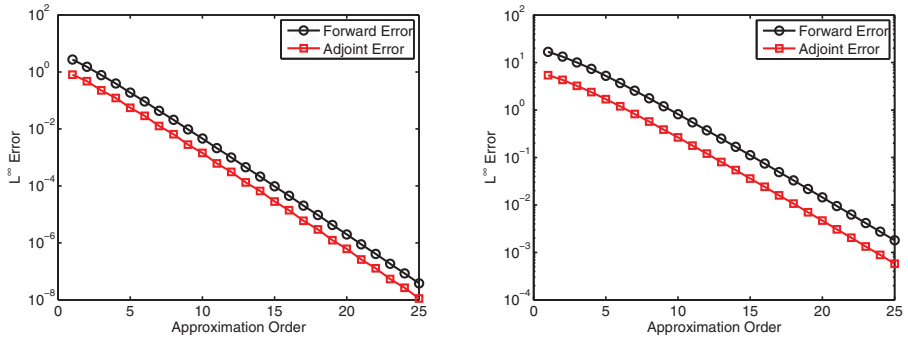

FIG. 7.3. $L^\infty(\Omega; l^2(\mathbb{R}^n))$ *errors in the forward and adjoint solutions for* $\epsilon = 0.8$ *(left) and* $\epsilon = 0.2$ *(right).*

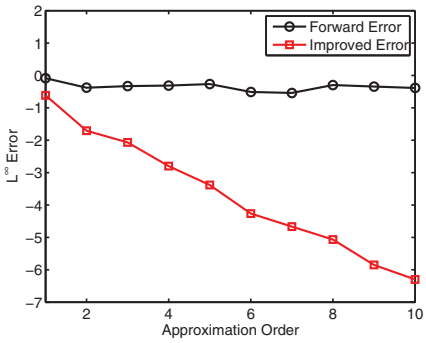

FIG. 7.4. $L^\infty(\Omega)$ *errors in the original linear functional and the improved linear functional using a pseudospectral approximation.*

7.2. A discontinuous example. Consider the parameterized linear system

$$\begin{bmatrix} 2 & -s_1 \\ -s_2 & 1 \end{bmatrix} \begin{bmatrix} x_1(s) \\ x_2(s) \end{bmatrix} = \begin{bmatrix} 1 \\ \lceil s_3 - 1/3 \rceil \end{bmatrix},$$

where $\lceil \cdot \rceil$ is the ceiling operator with $s_i \in [-1, 1]$ and $w(s) = 1/|\Omega|$. Note that the presence of the ceiling operator causes the solution (and all linear functionals of the solution) to be discontinuous at $s_3 = -2/3$ and $s_3 = 1/3$. We expect that any spectral approximation will exhibit the Gibbs phenomena and while the solution may converge in $L^2(l^2)$, it will not converge in $L^\infty(\Omega; l^2(\mathbb{R}^n))$. However, the discontinuity in the solution is entirely due to the discontinuity in the right-hand side. We are interested in $x_1(s)$, so we set $\psi = [1, 0]^T$. The adjoint problem,

$$\begin{bmatrix} 2 & -s_2 \\ -s_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1(s) \\ \phi_2(s) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

does not involve s_3 and therefore $\phi_1(s)$ and $\phi_2(s)$ will be smooth.

In Figure 7.4, we plot the error in the functional value and the error in the improved functional value computed using the same order pseudospectral method for both forward and adjoint problems. We see that the original functional value does not converge at all, while the improved functional converges at a reasonable rate. Next,

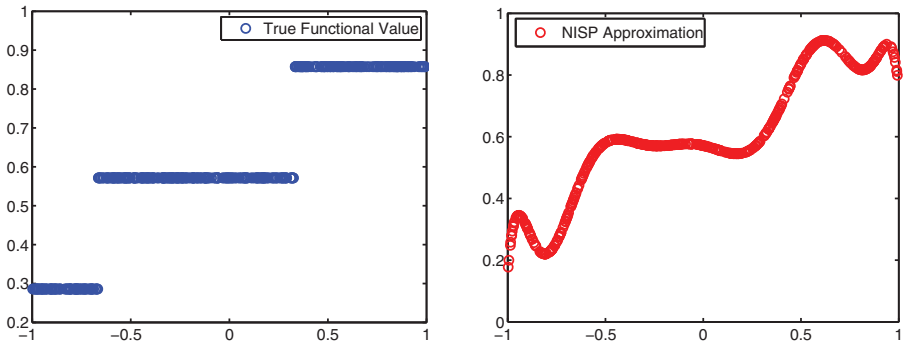

FIG. 7.5. *The true functional value and the order* 10 *pseudospectral approximation along the line* $s_1 = s_2 = 1/2$, $-1 \le s_3 \le 1$.

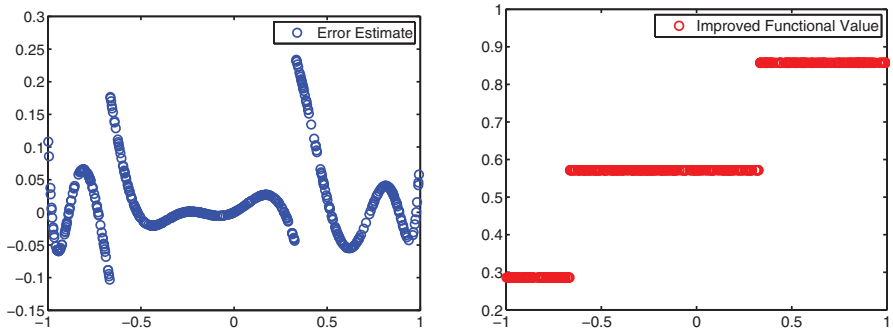

FIG. 7.6. *A posteriori error estimate (left) and the improved functional value (right) using an order* 10 *pseudospectral approximation of the forward and adjoint problems.*

we choose 500 random samples along the hypersurface (line) defined by

$$s_1 = s_2 = 1/2, \qquad -1 \le s_3 \le 1.$$

In Figure 7.5, we plot the true functional value and the order 10 pseudospectral approximation along this line. The pseudospectral approximation uses global polynomials and oscillates rapidly near $s_3 = -2/3$ and $s_3 = 1/3$. In Figure 7.6, we plot the error estimate computed for the order 10 pseudospectral approximation using an order 10 pseudospectral approximation of the adjoint problem as well as the improved functional value. Clearly, the a posteriori error estimate detects the discontinuity and corrects the functional value appropriately.

7.3. A higher dimensional example based on a discretized partial differential equation. Consider the following linear partial differential equation:

$$\begin{cases} -\nabla \cdot (K(x,y,s)\nabla u) = f(x,y), & (x,y) \in D, \\ u = 0, & (x,y) \in \Gamma_D, \\ K(x,y,s)\nabla u \cdot \mathbf{n} = 0, & (x,y) \in \partial D \backslash \Gamma_D, \end{cases} \tag{7.1}$$

where $D = [0,1] \times [0,1]$ and Γ_D is the line defined by $\{x = 0, 0 \le y \le 1\}$. The tensor $K(x,y,s)$ is parameterized as

$$K(x,y,s) = 15\mathbb{I} + \sum_{k=1}^{6} s_k \left(\sin(k\pi x) + \cos(k\pi y)\right) \mathbb{I}$$

with $s_i \in [-1,1]$ for $i = 1, 2, \ldots, 8$ and $w(s) = 1/|\Omega|$.

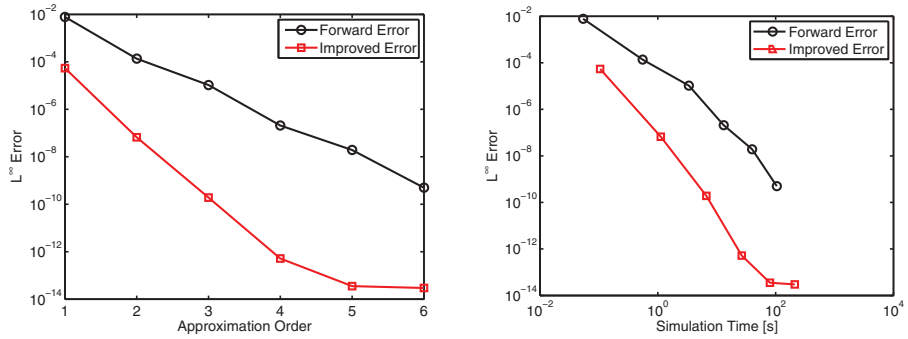

FIG. 7.7. *$L^\infty(\Omega)$ errors in the quantity of interest (left) and the simulation time required to achieve a given accuracy (right) using a pseudospectral approximation.*

We discretize (7.1) using continuous piecewise linear finite elements on a uniform triangulation of D consisting of 200 triangles and 121 degrees of freedom. The resulting parameterized linear system is of the form

$$\left(A_0 + \sum_{k=1}^{6} s_k A_k\right) x(s) = b. \tag{7.2}$$

Our quantity of interest is the solution at $(x, y) = (0.8, 0.6)$, so we define

$$\psi(x, y) = \frac{400}{\pi} \exp(-400(x - 0.8)^2 - 400(y - 0.6)^2)$$

and compute the projection onto the finite element space to obtain the right-hand side for the discrete adjoint equation,

$$\left(A_0 + \sum_{k=1}^{6} s_k A_k\right)^T \phi(s) = \psi.$$

We use the pseudospectral method to compute the approximation of $x(s)$ over Ω.

In Figure 7.7, we see that the error in the improved linear functional is much better than the original linear functional. The simulation time for the forward error depends only on the number of linear systems to solve, while the simulation time for the improved linear functional includes the construction of the pseudospectral approximation of the adjoint. We see that the computational cost to achieve a given accuracy is much lower for the improved method since the difference in solving an Nth order problem and a $2N$th order problem is much larger due to the dimension of Ω. For example, solving the forward and adjoint problems with $N = 2$ gives an improved functional value with the same accuracy as the original functional computed by solving the forward problem with $N = 4$, but the computational cost is an order of magnitude smaller.

We emphasize that the errors reported in Figure 7.7 are the pointwise errors in solving the *discrete* system (7.2) and do not include the finite element discretization errors over D. These errors can also be incorporated into the estimate but are not included here to keep the focus on parameterized linear systems. The a posteriori error analysis for discretized partial differential equations will be the subject of future work.

8. Conclusions. In this paper, we have developed a fully computable a posteriori error estimate for quantities of interest for parameterized linear systems. We also introduced an improved linear functional that converges pointwise at a much faster rate than the linear functional of the forward problem. In addition, we demonstrated that the computational cost required to achieve a given accuracy is far smaller for the improved linear functional, especially in higher parametric dimensions. In fact, we showed that solving both the forward and adjoint problems with an order N approximation and computing the improved linear functional can be as accurate as using an order $2N$ method for the forward problem but with a much lower computational cost.

While the analysis in this paper is relatively straightforward, the potential impact of the pointwise a posteriori error estimate and the improved linear functional warrants future investigation. Work is in progress to extend this error analysis and improved estimation of both linear and nonlinear functionals to approximations of stochastic partial differential equations. The error estimates can also be applied to reduction techniques for higher dimensional problems, such as sparse grid and probabilistic collocation. Finally, we are pursuing the implications of these error estimates on Bayesian inference problems.

REFERENCES

[1] S. Acharjee and N. Zabaras, *A non-intrusive stochastic Galerkin approach for modeling uncertainty propagation in deformation processes*, Comput. & Structures, 85 (2007), pp. 244–254.

[2] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034.

[3] I. Babuska, R. Tempone, and G. E. Zouraris, *Galerkin finite element approximations of stochastic differential equations*, SIAM J. Numer. Anal., 42 (2004), pp. 800–825.

[4] W. Bangerth and R. Rannacher, *Adaptive Finite Element Methods for Differential Equations*, Birkhäuser, Basel, 2003.

[5] J. Breidt, T. Butler, and D. Estep, *A measure-theoretic computational method for inverse sensitivity problems* I: *Method and analysis*, SIAM J. Numer. Anal., 49 (2011), pp. 1836–1859.

[6] C. Brezinski and M. Redivo-Zaglia, *The PageRank vector: Properties, computation, approximation, and acceleration*, SIAM. J. Matrix Anal. Appl., 28 (2006), pp. 551–575.

[7] T. Butler, C. Dawson, and T. Wildey, *A posteriori error analysis of stochastic differential equations using polynomial chaos expansions*, SIAM J. Sci. Comput., 33 (2011), pp. 1267–1291.

[8] T. Butler, D. Estep, and J. Sandelin, *A measure-theoretic computational method for inverse sensitivity problems* II: *A posterior error analysis*, SIAM J. Numer. Anal., 50 (2012), pp. 22–45.

[9] V. Carey, D. Estep, and S. Tavener, *A posteriori analysis and adaptive error control for multiscale operator decomposition solution of elliptic systems* I: *Triangular systems*, SIAM J. Numer. Anal., 47 (2009), pp. 740–761.

[10] P. G. Constantine, D. F. Gleich, and G. Iaccarino, *Spectral methods for parameterized matrix equations*, SIAM. J. Matrix Anal. Appl., 31 (2010), pp. 2681–2699.

[11] P. G. Constantine and D. F. Gleich, *Random Alpha PageRank*, Internet Math., 6 (2009), pp. 189–236.

[12] M. K. Deb, I. Babuška, and J. T. Oden, *Solution of stochastic partial differential equations using Galerkin finite element techniques*, Comput. Methods Appl. Mech. Engrg., 190 (2001), pp. 6359–6372.

[13] L. Dieci and L. Lopez, *Lyapunov exponents of systems evolving on quadratic groups*, SIAM. J. Matrix Anal. Appl., 24 (2003), pp. 1175–1185.

[14] C. F. Dunkl and Y. Xu, *Orthogonal Polynomials in Several Variables*, Cambridge University Press, Cambridge, UK, 2001.

[15] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Computational Differential Equations*, Cambridge University Press, Cambridge, UK, 1996.

[16] D. Estep, V. Ginting, J. Shadid, and S. Tavener, *An a posteriori-a priori analysis of multiscale operator splitting*, SIAM J. Numer. Anal., 46 (2008), pp. 1116–1146.
[17] D. Estep and D. Neckels, *Fast and reliable methods for determining the evolution of uncertain parameters in differential equations*, J. Comput. Phys., 213 (2006), pp. 530–556.
[18] D. Estep and D. Neckels, *Fast methods for determining the evolution of uncertain parameters in reaction-diffusion equations*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 3967–3979.
[19] D. Estep, S. Tavener, and T. Wildey, *A posteriori analysis and improved accuracy for an operator decomposition solution of a conjugate heat transfer problem*, SIAM J. Numer. Anal., 46 (2008), pp. 2068–2089.
[20] D. Estep, S. Tavener, and T. Wildey, *A posteriori error estimation and adaptive mesh refinement for a multiscale operator decomposition approach to fluid-solid heat transfer*, J. Comput. Phys., 229 (2010), pp. 4143–4158.
[21] D. Estep, *A posteriori error bounds and global error control for approximation of ordinary differential equations*, SIAM J. Numer. Anal., 32 (1995), pp. 1–48.
[22] B. Ganapathysubramanian and N. Zabaras, *Sparse grid collocation methods for stochastic natural convection problems*, J. Comput. Phys., 225 (2007), pp. 652–685.
[23] B. Ganis, H. Klie, M. F. Wheeler, T. Wildey, I. Yotov, and D. Zhang, *Stochastic collocation and mixed finite elements for flow in porous media*, Comput. Methods Appl. Mech. Engrg., 197 (2008), pp. 3547–3559.
[24] W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*, Clarendon Press, Oxford, UK, 2004.
[25] R. Ghanem and J. Red-Horse, *Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach*, Phys. D., 133 (1999), pp. 137–144.
[26] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 2002.
[27] M. B. Giles and E. Suli, *Adjoint Methods for PDEs: A Posteriori Error Analysis and Postprocessing by Duality*, Acta Numer., Cambridge University Press, Cambridge, UK, 2002, pp. 105–158.
[28] O. Le Maître, R. Ghanem, O. Knio, and H. Najm, *Multi-resolution analysis of Wiener-type propagation schemes*, J. Comput. Phys., 197 (2004), pp. 502–531.
[29] O. Le Maître, R. Ghanem, O. Knio, and H. Najm, *Uncertainty propagation using Wiener-Haar expansions*, J. Comput. Phys., 197 (2004), pp. 28–57.
[30] Y. M. Marzouk and H. N. Najm, *Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems*, J. Comput. Phys., 228 (2009), pp. 1862–1902.
[31] Y. Marzouk, H. Najm, and L. Rahn, *Stochastic spectral methods for efficient Bayesian solution of inverse problems*, J. Comput. Phys., 224 (2007), pp. 560–586.
[32] L. Mathelin and O. P. Le Maître, *Dual-based error analysis for uncertainty quantification in a chemical system*, PAMM, 7 (2007), pp. 2010007–2010008.
[33] K. Meerbergen and Z. Bai, *The Lanczos method for parameterized symmetric linear systems with multiple right-hand sides*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1642–1662.
[34] F. Nobile, R. Tempone, and C. G. Webster, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2309–2345.
[35] N. A. Pierce and M. B. Giles, *Adjoint recovery of superconvergent functionals from PDE approximations*, SIAM Rev., 42 (2000), pp. 247–264.
[36] M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio, *Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection*, Combustion and Flame, 132 (2003), pp. 545–555.
[37] R. A. Todor and C. Schwab, *Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients*, IMA J. Numer. Anal., 27 (2007), pp. 232–261.
[38] X. Wan and G. Karniadakis, *Solving elliptic problems with non-Gaussian spatially-dependent random coefficients*, Comput. Methods Appl. Mech. Engrg, 198 (2009), pp. 1985–1995.
[39] D. Xiu and J. S. Hesthaven, *High-order collocation methods for differential equations with random inputs*, SIAM J. Sci. Comput., 27 (2005), pp. 1118–1139.
[40] D. Xiu and G. Karniadakis, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., 24 (2002), pp. 619–644.
[41] D. Xiu, *Efficient collocational approach for parametric uncertainty analysis*, Commun. Comput. Phys., 2 (2007), pp. 293–309.