On the Convergence of Stochastic Finite Elements

John DeLaurentis and Irene Moshesh

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.
On the Convergence of Stochastic Finite Elements

John M. DeLaurentis
Computational Math/Algorithms Department

Irene Moshesh
Computational Math/Algorithms Department

Sandia National Laboratories
PO Box 5800
Albuquerque, NM 87185-0529

ABSTRACT

We investigate the rate of convergence of stochastic basis elements to the solution of a stochastic operator equation. As in deterministic finite elements, the solution may be approximately represented as the linear combination of basis elements. In the stochastic case, however, the solution belongs to a Hilbert space of functions defined on a cross product domain endowed with the product of a deterministic and probabilistic measure. We show that if the dimension of the stochastic space is n, and the desired accuracy is of order $\varepsilon$, the number of stochastic elements required to achieve this level of precision, in the Galerkin method, is on the order of $\ln \varepsilon |n|$. 
ACKNOWLEDGEMENTS

This work was funded by the US DOE Office of Nonproliferation & National Security, Office of Research and Development, under the Advanced Radar System (ARS) project. This effort is directed by Randy Bell of DOE NN-20.
CONTENTS

1. Introduction ........................................................................................................ 7
2. Symmetric, Semi-bounded Operators ............................................................ 8
3. Numerical Results .......................................................................................... 10
4. Summary ........................................................................................................ 12
References ......................................................................................................... 13
Distribution ........................................................................................................ 14
This page left intentionally blank.
1 Introduction

We consider the stochastic equation where the operator has been split into a deterministic part \( L \) and a random part \( \Pi \),

\[(L + \Pi)u = f.\] (1.1)

The domain of \( L + \Pi \) is a dense subset of the Hilbert space \( L^2(D \otimes \Omega, d\mu \otimes dP) \), more specifically, the domain is a dense linear span of independent functions of the form \( \varphi(x)\psi(\xi), x \in D, \xi \in \Omega \), where \( \varphi \) and \( \psi \) are measurable and \( ||\varphi\psi|| < \infty \) (\( ||\varphi\psi|| \equiv \int \varphi\psi d\mu dP \)). (Here and in the following we will use boldface letters to denote vectors, in particular, \( x = (x_1, \ldots, x_d) \) and \( \xi = (\xi_1, \ldots, \xi_n) \).) We will assume that \( L \) and \( L + \Pi \) are symmetric ((\( Lu, u \)) = (\( u, Lu \)) and ((\( L + \pi \))u, u) = (\( u, (L + \pi)u \)) where (\( \varphi, \psi \)) denotes the inner product on the Hilbert space), and positive-bounded-below ((\( Lu, u \)) \geq c^2(u, u) = c^2 \| u \|^2, for some \( c > 0 \); similar for \( L + \pi \)).

We show that to achieve a precision of level \( \epsilon \), assuming that sufficiently many of the deterministic elements \( \varphi \) have been chosen, the number of stochastic elements, \( \psi \), using the Galerkin method, is on the order of \( \ln(\epsilon) \ln \). By the expression "a precision of level \( \epsilon \)," we mean that \( \| u_n - u \| \leq \epsilon \), where \( u_n \) is the approximate solution given by the Galerkin method. This assertion implies that the computational cost, due to the stochastic component, is relatively inexpensive, provided the dimension of the probability space is not too large.

The next section summarizes the requisite theorems from Hilbert space theory. In the third section we state and prove our main results. The fourth section presents some numerical examples, and demonstrates the rapid convergence of the stochastic finite elements in the low dimensional case. Aside from the practical interest in the rate of convergence, this analysis may provide some insight into the underlying structure of the random processes that arise in engineering.

2 Symmetric, Semi-bounded Operators

Many problems in mathematical physics can be stated in variational form as the function \( u \) that minimizes the functional

\[ F(u) = (Au, u) - 2(f, u), \] (2.1)

where \( A \) is a symmetric, positive-bounded-below transformation with domain \( D_A \subseteq H \) and range contained in the real Hilbert space \( H \). (By positive-bounded-below we mean that (\( Au, u \)) \geq b^2(u, u) = b^2 \| u \|^2 \) for some \( b > 0 \).) The minimum of the functional \( F \) may be derived using the associated inner product defined on \( D_A \), namely,

\[(u, v)_A \equiv (Au, v), \] (2.2)

and \( \| u \|_A^2 = (u, u)_A \). We complete the space, denoted \( H_A \), by adjoining the limit of the Cauchy sequences in \( D_A \); in this case, the limit can be identified with elements in \( H \), see Riesz-Nagy [5].
The operator $A$ may be extended to a self-adjoint transformation $\overline{A}$ such that, if $f \in H$,
\[ \overline{A}u_0 = f, \] (2.3)
for some $u_0 \in H_A$, where $\overline{A}u = Au$ for all $u \in D_A$. The existence of the operator $\overline{A}$ and the solution to equation (2.3) are derived by applying the Riesz representation theorem to the linear functional $\Lambda_f(u) = (f, u)$. That is, since $\Lambda_f$ is a linear functional on $H_A$, there exists a unique $u_0 \in H_A$ such that $\Lambda_f(u_0) = (u_0, u)_A$ and $\| u_0 \| \leq \| f \|$. We set $u_0 = \overline{A}^{-1}f$, and observe that $b \| \overline{A}^{-1}f \| = (u_0, u)_A \leq \| f \|$, so that $\| \overline{A}^{-1}f \| \leq b^{-1} \| f \|$. It follows that the inverse of $\overline{A}$, denoted $\overline{A}^{-1}$, exists and is bounded; $\| \overline{A}^{-1} \| \leq b^{-1}$. Also, it can be shown that the operator $\overline{A}^{-1}$ is self-adjoint, see Riesz-Nagy [5].

The function that minimizes $F$ is given by the solution to equation (2.3), that is,
\[ F(u) = (\overline{A}u, u) - 2(f, u) = (u, u)_A - 2(u_0, u)_A = \| u - u_0 \|_A^2 + \| u_0 \|_A^2, \] (2.4)
where $\| u \|_A$ is the energy norm defined by $\| u \|_A^2 = (\overline{A}u, u)$. It follows that $F(u)$ has a unique minimum at $u = u_0$ and the minimum value is $- \| u_0 \|_A$, see Mikhlin [4].

## 3 Stochastic Finite Elements

A mathematical representation for a problem involving random spatial fluctuations is given by the operator equation
\[ Au = [L_0(x) + \sum_{i=1}^{n} \beta_i \xi_i L_i(x)]u = f, \] (3.1)
where $A = L + \Pi$, $L = L_0$, $\Pi = \sum \beta_i \xi_i L_i$, the $L_i(x)$'s are deterministic differential operators, $u = u(x, \xi)$, $f = f(x, \xi)$, and the $\xi_i$'s are zero-mean, uncorrelated random variables. Invoking the central limit theorem we are led to consider Gaussian variables as possible candidates for the $\xi_i$'s; however, since Gaussian deviates can assume arbitrarily large negative values the coefficients may become negative. This problem, however, is avoided if we assume that the variates $\xi_i$ are truncated. Gaussian variables also have the advantage that uncorrelated variates are actually independent. In the following, we assume that the $\xi_i$'s are independent, identically distributed (not necessarily Gaussian), bounded ($P(\xi_i > \gamma) = 0$ for some $\gamma > 0$), zero-mean random variables. Additionally, we assume that the solution satisfies a deterministic boundary condition $\sum(x)u(x, \xi) = 0$, for $x \in \partial D$.

The form of the stochastic operator in equation (3.1) was derived by replacing the coefficients of the corresponding deterministic differential operator with random coefficients. In turn, the Karhunen-Loeve expansion was used to express the random coefficients as a finite orthonormal expansion involving the variables $\xi_i$, see Ghanem and Spanos [2], Ghanem and Red-Horse [3].
Let us define the domain $D_A$ of $A$ as the linear span of functions of the form $\varphi(x)\psi(\xi)$ where $\varphi$ is sufficiently differentiable to belong to the domain of the differential operators in expression (3.1), satisfies the boundary condition $\sum(x)\varphi(x) = 0$, and $\psi(\xi)$ is a polynomial in the stochastic variables $\xi_i$. The inner product on $H$ is defined by

$$\langle u, v \rangle \equiv \int \int uv d\mu dP,$$

(3.2)

where $u = u(x, \xi)$ and $v = v(x, \xi)$ are mean square integrable with respect to $d\mu(x) \otimes dP(\xi)$, that is, $u, v \in H = L^2(D \otimes \Omega, d\mu \otimes dP)$. Here, the symbol $\mu$ denotes Lebesgue measure, $P(\xi)$ equals the joint distribution of the random variables $\xi_i$, $i = 1, \ldots, n$, $dP(\xi) = \prod_{i=1}^{n} p(\xi_i) d\xi_i$, and $p(\xi_i)$ is the probability density of $\xi_i$. (We use angled brackets to emphasize that the underlying measure is a product involving a probabilistic measure.) If $A$ is symmetric, positive-bounded-below, $f \in H$, and $D_A$ is dense in $H$, then the transformation $A$ has an extension such that $\tilde{A}u_0 = f$ has a solution (see the preceding section).

As we have seen, the problem of solving equation (3.1) may be replaced by the problem of finding the function that minimizes the functional

$$F(u) \equiv \langle \tilde{A}u, u \rangle - 2\langle f, u \rangle = \langle u, u \rangle_A - 2\langle u_0, u \rangle_A = \| u - u_0 \|^2_A - \| u_0 \|^2_A,$$

(3.3)

where $\langle u, v \rangle_A \equiv \langle \tilde{A}u, v \rangle$, $\| u \|^2_A \equiv \langle u, u \rangle_A \geq b^2 \| u \|^2$, $\tilde{A}$ extends $A$, and $\tilde{A}u_0 = f$. Let us restrict the domain of $A$ to a finite set of linearly independent basis elements, $\varphi_j(x) \psi_k(\xi)$, $1 \leq j \leq M$, $1 \leq k \leq N$, and pose the problem of finding the linear combination, $u_a(x, \xi) = \sum a_{j,k} \varphi(x)\psi(\xi)$, that minimizes $F(u)$. It can be shown that (for the restricted domain) the coefficients of the minimizing function satisfy a system of algebraic equations,

$$\langle \tilde{A}u_a, \varphi_{j'}\psi_{k'} \rangle = \langle f, \varphi_{j'}\psi_{k'} \rangle,$$

(3.4a)

or,

$$\sum_{j=1}^{M} \sum_{k=1}^{N} a_{j,k} \langle \tilde{A}(\varphi_j\psi_k), \varphi_{j'}\psi_{k'} \rangle = \langle f, \varphi_{j'}\psi_{k'} \rangle,$$

(3.4b)

where $1 \leq j' \leq M$, $1 \leq k' \leq N$, (the variational argument is essentially the same as in the deterministic case, see Ghanem and Spanos [2] or Fletcher [1]). The technique that yields the approximate solution given by equation (3.4) is known as the Galerkin method.

We may assume that the basis elements have been orthogonalized with respect to the inner product $(u, v)$. More specifically, we suppose that the $\varphi_j$'s are a subset of the orthogonal basis $\{\varphi_j(x)\}$, $1 \leq j < \infty$, where $\langle \varphi_j, \varphi_j \rangle = \delta_{j,j}$; here $\delta_{j,j}$ denotes the Kronecker delta function and $\langle \varphi_j, \varphi_j \rangle \equiv \int_D \varphi_j(x)\varphi_j(x)d\mu(x)$. Also, we assume that the $\psi_k$'s are a subset of the orthogonal basis of the form $\{\psi_k(\xi) = \prod_{i=1}^{n} \psi_k_i(\xi_i)\}$, where $k = (k_1, \ldots, k_n)$ and $\psi_k(\xi) = \sum a_{j,k} \varphi(x)\psi(\xi)$ equals the orthonormal polynomial in $\xi_i$ of degree $k_i$ over the space $L^2([-\gamma, \gamma], p(\xi_i)d\xi_i)$; that is, $(\psi_k, \psi_{k'}) = \delta_{k, k'}$, for $(\psi_k, \psi_{k'}) = \int_{-\gamma}^{\gamma} \psi_k(\xi)\psi_{k'}(\xi)p(\xi_i)d\xi_i$.

We have
\((\psi_k, \psi_k) \equiv \int \psi_k(\xi)\psi_k(\xi) dP(\xi) = \delta_{k,k'}\); where \(\delta_{k,k'} = 1\) if \(k = k'\), otherwise \(\delta_{k,k'} = 0\).

Our goal is to estimate the number of stochastic elements required to produce the estimate \(u_a\) satisfying \(\| u_a - u_0 \| \leq \epsilon\), assuming that the deterministic elements can be chosen to provide a desired level of precision. Towards this end, we introduce an equivalent (under certain assumptions on the operators \(L_i\)) problem. Applying \(L_0^{-1}\) to both sides of equation (3.1), we obtain,

\[
(I + \sum_{i=1}^{n} \beta_i \xi_i L_0^{-1} L_i)u = g, \tag{3.5}
\]

where \(x \in D\), \(I\) is the identity operator, \(g(x, \xi) = L_0^{-1} f(x, \xi)\), and \(\sum(x)u(x, \xi) = 0\), for \(x \in \partial D\). We assume that both, \(L_0\) and \(I + \sum_{i=1}^{n} \beta_i \xi_i L_0^{-1} L_i\), are symmetric, positive-bounded-below. Additionally, we suppose that \(T_\xi \equiv -\sum_{i=1}^{n} \beta_i \xi_i L_0^{-1} L_i\) is a contraction; that is, \(\| T_\xi \| \leq \eta < 1\), for some \(\eta\). Also, we assume that \(L_0^{-1}\) and \(T_\xi\) commute. It follows that the series \(\sum_{i=0}^{\infty} \| T_\xi^{(i)} g \|\) converges and we may write

\[
u_0 = (I - T_\xi)^{-1} g = \sum_{i=0}^{\infty} T_\xi^{(i)} g, \tag{3.6}\]

where \(u_0 = u_0(x, \xi)\) is the solution to equation (3.5) and \(T_\xi^0 \equiv I\). In particular, expression (3.6) asserts that \(u_0\) satisfies equation (3.1),

\[
\mathcal{A}u_0 = L_0(I - T_\xi)u_0 = L_0(I - T_\xi) \sum_{i=0}^{\infty} T_\xi^i g = L_0 g = f. \tag{3.7}
\]

It follows that we need only estimate the number of basis elements required to approximate the solution \(u_0\) given by equation (3.5).

To estimate the rate of convergence of the stochastic elements we need an assumption on the rate at which the orthogonal expansion converges to \(f\). For a given \(\delta > 0\), we assume that \(f(x, \xi)\) may be approximated by a function \(f_m(x, \xi)\) such that the maximum degree of the polynomials in \(\xi_i\) of \(f_m\) do not exceed \(m\), where \(m = m(\delta) = O(|\ln(\delta)|), \delta \downarrow 0\) such that \(\| f_m - f \| < \delta\). The expansion for \(f(x, \xi)\) converges exponentially fast in \(\xi\), that is,

\[
f = f_m + B_m = \sum_{k: \text{degree } \xi_i \leq m} \sum_{j=1}^{\infty} \langle f, \varphi_j \psi_k \rangle \lambda_j \varphi_j \psi_k + B_m, \tag{3.8}\]

where \(\| B_m \| \leq \delta = O(\exp(-m))\), as \(m \to \infty\). Also, we assume that for \(\epsilon > 0\) and a given function \(h(x) = \sum_{j=1}^{\infty} a_j \varphi_j(x)\), there exists a function \(M = M(h, \delta)\), such that,

\[
h(x) = \sum_{j=1}^{M} a_j \varphi_j(x) + C(h, \delta), \tag{3.9}\]

where \(\| C(h, \epsilon) \| \leq \delta\). We are ready to state and prove our main result.
Theorem Let us assume that \( T_\xi \) is a contraction (\( \| T_\xi \| \leq \eta < 1 \)), that \( T_\xi \) commutes with \( \overline{L}_0^{-1} (\| L_0^{-1} \| \leq c^{-1}) \), and that the expansion for \( f \) converges exponentially fast in \( \xi \). For a given \( \varepsilon > 0 \), we choose the finite basis \( \{ \varphi_j \psi_k \} \), \( 1 \leq j \leq M, k \in J = \{ (k_1, \ldots, k_n) : k_i < K \} \), \( K = l + m, m = m(\delta_1) \), with \( l \) chosen so that \( \eta(1 - \eta)^{-1} c^{-1} \| f \| \leq \delta_1 \), \( \delta_1 = \frac{1}{4}(1 + \frac{1}{(1-\eta)c})^{-1}(\frac{\| f \|}{1-\eta})^{-1} e b \), and \( M = \max_{k \in J} \{ M(h_k, \delta_2) \} \), where \( h_k(x) = \sum_{j=1}^{\infty} a_{j,k} \varphi_j(x) \), \( a_{j,k} = \langle S_l, \varphi_j \psi_k \rangle \), \( S_l = \sum_{s=0}^{l-1} T_s^* f_m \), with \( \delta_2 = \frac{1}{4}K^{-n/2}(\frac{\| f \|}{1-\eta})^{-1} e b \). For the Galerkin approximation, \( u_\alpha \), we obtain

\[
\| u_0 - u_\alpha \| \leq b^{-1} \| v_\alpha - u_\alpha \| \leq \varepsilon,
\]

where \( v_\alpha = \sum_{k \in J} \sum_{j=1}^{M} (S_l, \varphi_j \psi_k) \varphi_j \psi_k, \) and \( \| u \| \leq b \| u \| \). The basis contains \( MN = O(M \ln \varepsilon/n) \) elements; the number of stochastic elements is \( O(\ln \varepsilon/\ln \eta^n) = O(\ln \varepsilon/n) \), as \( \varepsilon \downarrow 0 \).

Proof To demonstrate the accuracy of \( u_\alpha \), it is sufficient to find an element \( v_\alpha \), belonging to the linear span of the finite basis, such that \( b^{-1} \| v_\alpha - u_\alpha \| \leq \varepsilon \). To see this, we use the fact, \( \| u \| = (A u, u)^{1/2} \geq b(u, u) = b \| u \| \), and the inequality,

\[
\| u_\alpha - u_\alpha \| \leq b^{-1} \| v_\alpha - u_\alpha \| = b^{-1} (F(u_\alpha) - F(u_0))^{1/2},
\]

\[
\leq b^{-1} (F(u_\alpha) - F(u_0))^{1/2} = b^{-1} \| v_\alpha - u_\alpha \|_A.
\]

The second inequality follows from the fact that \( u_\alpha \) minimizes \( F(u) \) over the finite basis. It follows that to establish the required accuracy for the Galerkin method we need only produce a \( v_\alpha \) satisfying \( \| v_\alpha - u_\alpha \|_A \leq \varepsilon \).

We may replace the right-hand side of equation (3.5) with \( g_m = \overline{L}_0^{-1} f_m \), that is, from equation (3.6) we obtain,

\[
u_0 = (I-T_\xi)^{-1} \overline{L}_0^{-1} f = (I-T_\xi)^{-1} (\overline{L}_0^{-1} f_m + \overline{L}_0^{-1} B_m) = (I-T_\xi)^{-1} g_m + B,
\]

where \( \| B \| \leq (1-\eta)^{-1} c^{-1} \| B_m \| \leq (1-\eta)^{-1} c^{-1} \delta_1 \). Now, we choose \( l > 0 \) so large that \( \eta(1-\eta)^{-1} c^{-1} \| f \| \leq \delta_1 \); the first term on the right-hand side of expression (3.12) may be written,

\[
(I-T_\xi)^{-1} \overline{L}_0^{-1} f_m = \sum_{s=0}^{l-1} T_s^* f_m + R = \overline{L}_0^{-1} S_l + R,
\]

where \( S_l = \sum_{s=0}^{l-1} T_s^* f_m \) and \( R = \sum_{s=0}^{\infty} T_s^* \overline{L}_0^{-1} f_m \). Here, we use the fact that \( T_\xi \) and \( \overline{L}_0^{-1} \) commute. The remainder is bounded by \( \| R \| \leq \frac{\eta'}{(1-\eta)c} \| f_m \| \leq \frac{\eta'}{(1-\eta)c} \| f \| \leq \delta_1 \). We have shown

\[
u_0 = \overline{L}_0^{-1} S_l + B + R,
\]

where \( \| B \| + \| R \| \leq \delta_1 (1 + (1-\eta)^{-1} c^{-1}) \).

11
The multinomials in $S_1 = \sum_{s=0}^{l-1} T^s \psi_m$ have degree less than $K = l + m$; $T^s$ contributes multinomials of degree less than $l$ and $\psi_m$ has degree at most $m$. It follows that the multinomials in $S_1$ may be expressed in terms of the basis elements $\psi_k(\xi), k \in J$; that is,
\begin{equation}
S_1 = \sum_{s=0}^{l-1} T^s \psi_m = \sum_{k} \sum_{j=1}^{\infty} \langle S_1, \varphi_j \psi_k \rangle \varphi_j \psi_k.
\end{equation}

Next, setting $S_{1,M} = \sum_{k \in J} \sum_{j=1}^{M} \langle S_1, \varphi_j \psi_k \rangle \varphi_j \psi_k$ and using the assumption that for each $k \in J, \| \sum_{j=M+1}^{\infty} \langle S_1, \varphi_j \psi_k \rangle \varphi_j \|_2^2 \leq \delta_2^2$, we obtain
\begin{equation}
\| S_1 - S_{1,M} \|^2 = \sum_{k \in J} \sum_{j=M+1}^{\infty} \langle S_1, \varphi_j \psi_k \rangle \varphi_j \|_2^2 \leq K^n \delta_2^2.
\end{equation}

Here, we also use that the cardinality of $J$ equals $K^n$. Setting $v_a = \bar{T}_0^{-1} S_{1,M}$ and using inequality (3.16), we obtain for the first term on the right-hand side of expression (3.14),
\begin{equation}
\| \bar{T}_0^{-1} S_1 - v_0 \|^2 = \| \bar{T}_0^{-1} (S_1 - S_{1,M}) \|^2 \leq c^{-2} K^n \delta_2^2.
\end{equation}

Now, from expressions (3.14) and (3.17), we obtain
\begin{equation}
\| u_0 - v_a \| \leq \| \bar{T}_0^{-1} S_1 - v_0 \| + \| B \| + \| R \| \leq \frac{K^{n/2}}{c} \delta_2 + \delta_1 (1 + \frac{1}{(1-\eta)c}) = \frac{1}{2} (\| f \|)^{-1} eb,
\end{equation}
where $\delta_1 = \frac{1}{2} (1 + \frac{c}{1-\eta})^{-1} (\| f \|)^{-1} eb$ and $\delta_2 = \frac{1}{2} (1 + \frac{1}{(1-\eta)c})^{-1} eb$.

To obtain an estimate for $\| u_0 - v_a \|_A$ we need a bound on $\| Av_a \|$, for this we use the identity $\bar{T}_0^{-1} = \bar{L}_0(I - T_\xi)\bar{L}_0^{-1} = I - T_\xi$, so that,
\begin{equation}
\| Av_a \| \leq \| \bar{T}_0^{-1} S_{1,M} \| \leq \| (I - T_\xi) S_{1,M} \| \leq 1 + \eta \frac{1}{1-\eta} \| f \|.
\end{equation}

Using the Cauchy inequality $\| u \|_A \leq \| Au \|_A \| u \|$, together with inequalities (3.18) and (3.19), we obtain
\begin{equation}
\| u_0 - v_a \|_A \leq \| A(u_0 - v_a) \| \leq (1 + \frac{1}{1-\eta}) \| f \| \frac{1}{2} (\| f \|)^{-1} eb = eb.
\end{equation}

as desired. This demonstrates that, for this basis, the Galerkin method provides the desired accuracy.

The number of basis elements in this set is given by $MN = MK^n$, where $N = K^n$ is the cardinality of the index set $J$. We recall that $K = l + m$ and that $l$ was chosen so that $\eta (1 - \eta)^{-1} \| f \| \leq \delta_1 = O(e)$, more precisely,
\begin{equation}
l = \left\lfloor \frac{\log \delta_1 + \log(1 - \eta) + \log c + \log \| f \|}{\log \eta} \right\rfloor + \theta = \left\lfloor \frac{\log \delta_1}{\log \eta} \right\rfloor (1 + O(\frac{1}{\log \delta_1})),
\end{equation}

(3.19)
where $0 \leq \theta < 1$. Now, since $m = O(\ln \delta_1) = O(\ln \epsilon)$, as $\epsilon \downarrow 0$, we have

$$K^n = (1 + m)^n = \left( \frac{\ln \delta_1}{\ln \eta} \right)^n \left( 1 + O\left( \frac{m}{\ln \delta_1} \right) \right)^n = O\left( \frac{\ln \epsilon}{\ln \eta} \right)^n, \quad (3.20)$$

as $\epsilon \downarrow 0$. We have arrived at the desired conclusion, namely, the number of stochastic elements are $O((\ln(\epsilon)/\ln(\eta))^n) = O((\ln(\epsilon))^n)$, as $\epsilon \downarrow 0$, where $\epsilon$ is the desired level of precision and $n$ is the dimension of the stochastic space.

The key step in the proof is the representation of the solution as a partial sum of the resolvent $\sum T^k f$. In turn, this function may be expressed as polynomials, of low degree, in the variables $\xi_i$.

4 Numerical Results

We consider a set of stochastic operator problems governed by the equation

$$-(\alpha + \beta \xi_1) \frac{\partial^2}{\partial x^2} u(x, \xi_1) = f(x), \quad (4.1)$$

for $f(x) = 1$, $f(x) = x$ and $f(x) = x + 1$ with boundary condition

$$u(0) = u(1) = 0. \quad (4.2)$$

We define $\xi_1$ to be a truncated Gaussian variable with mean zero. By a truncated Gaussian random variable we mean that

$$P(\xi_\gamma < x) = \frac{C_\gamma}{\sqrt{2\pi}} \int_{-\gamma}^x e^{-\theta^2} d\theta \quad (4.3)$$

for $-\gamma < x < \gamma$ $P(|\xi_\gamma| \geq \gamma) = 0$ and,

$$\frac{C_\gamma}{\sqrt{2\pi}} \int_{-\gamma}^\gamma e^{-\theta^2} d\theta = 1. \quad (4.4)$$

The parameters $\alpha$, $\beta$ and $\gamma$ are chosen so that

$$\frac{\beta}{\alpha} \gamma < 1, \quad (4.5)$$

for $\alpha, \beta, \gamma > 0$. We note that $\alpha, \beta$ and $\gamma$ are not unique. The set of problems described by equation (4.1) has the exact solution

$$u(x, \xi_\gamma) = \frac{1}{\alpha} \sum_{i=0}^\infty (-\frac{\beta \xi_\gamma}{\alpha})^i [(1 - x) \int_0^z y f(y) dy + x \int_x^1 (1 - y) f(y) dy]. \quad (4.6)$$

Using the Stochastic-Galerkin method, an approximate solution is introduced as

$$u_\alpha (x, \xi_\gamma) = \sum_{j=1}^M \sum_{k=0}^N a_{j,k} \varphi_j(x) \psi_k(\xi_\gamma), \quad (4.7)$$
where the $\varphi_j(x) = x - x^{j+1}$'s are chosen so that they satisfy the boundary condition (4.2). Also, we set $\psi_k(\xi) = \xi^k$, and we define the inner product by

$$\langle h, g \rangle \equiv \frac{C_6}{\sqrt{2\pi}} \int_0^1 \int_{-\gamma}^\gamma h(x, \xi)g(x, \xi)e^{-\xi^2}d\xi dx$$

$$= \int_0^1 E[h(x, \xi)g(x, \xi)]dx,$$

where $E[.]$ denotes the expected value. Evaluation of the inner product

$$\langle Au_0, \varphi_j(x)\psi_k(\xi) \rangle = \langle f, \varphi_j(x)\psi_k(\xi) \rangle,$$

where $A \equiv -(\alpha + \beta \xi)\frac{\partial^2}{\partial \xi^2}$, for the case $f = 1$, produces the matrix equation

$$\begin{pmatrix}
  a_{1,0} & \ldots & a_{1,N} \\
a_{2,0} & \ldots & a_{2,N} \\
\vdots & \ddots & \vdots \\
a_{i,0} & \ldots & a_{i,N} \\
\vdots & \ddots & \vdots \\
a_{M,0} & \ldots & a_{M,N}
\end{pmatrix}
\begin{pmatrix}
  E[-(\alpha + \beta \xi)\psi_0 \psi_k] \\
  E[-(\alpha + \beta \xi)\psi_1 \psi_k] \\
  \vdots \\
  E[-(\alpha + \beta \xi)\psi_{i-1} \psi_k] \\
  \vdots \\
  E[-(\alpha + \beta \xi)\psi_N \psi_k]
\end{pmatrix}
= \frac{i}{2(i + 2)} E[\xi_k^k]
$$

for fixed $i$ and $k$ where $i \in \{1, \ldots, M\}$ and $k \in \{0, \ldots, N\}$. If we let $M = 3$ and $N = 2$, we obtain the matrix equation

$$\begin{pmatrix}
  3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
  3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\
  \alpha & \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta & \alpha \\
  \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta \\
  \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta \\
  \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta \\
  \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta \\
  \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta \\
  \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta \\
  \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta & \alpha & \beta \\
\end{pmatrix}
\begin{pmatrix}
  a_{1,0} \\
  a_{1,1} \\
  a_{1,2} \\
  a_{2,0} \\
  a_{2,1} \\
  a_{2,2} \\
  a_{3,0} \\
  a_{3,1} \\
  a_{3,2}
\end{pmatrix}
= \begin{pmatrix}
  1/6 \\
  1/4 \\
  3/10 \\
  0 \\
  0 \\
  0 \\
  1/6 \\
  1/4 \\
  3/10
\end{pmatrix}
$$

Here, we use the fact that $E[\xi_1^2] \approx 1$ and $E[\xi_1^4] \approx \frac{3}{2}$, because $\xi$ is approximately Gaussian for $\gamma$ sufficiently large. In particular, this means that the right-hand side of the preceding matrix equation is only approximately equal to $(b_{1,0}, b_{2,0}, b_{3,0}, \ldots, b_{1,2}, b_{2,2}, b_{3,2})^t$, where $b_{i,k} = \frac{1}{2(i + 2)} E[\xi_k^i]$, $1 \leq i \leq 3$, $0 \leq k \leq 2$. Solving for the coefficients $a_{i,k}$, we obtain

$$a_{1,0} = \frac{1}{2\alpha}; a_{1,1} = \frac{-\beta}{2\alpha^2 - 3\beta^2}; a_{1,2} = \frac{\beta^2}{\alpha(2\alpha^2 - 3\beta^2)};$$

$$a_{2,0} = 0; a_{2,1} = 0; a_{2,2} = 0;$$

$$a_{3,0} = 0; a_{3,1} = 0; a_{3,2} = 0.$$
Therefore, using equation (4.7) we have
\[ u_a(x, \xi\gamma) = \left( \frac{1}{2\alpha} - \frac{\beta \xi\gamma}{2\alpha^2 - 3\beta^2} + \frac{\beta^2 \xi^2\gamma}{\alpha(2\alpha^2 - 3\beta^2)} \right) (x - x^2). \] (4.11)

It follows that,
\[ \text{Var}[u_a(x, \xi\gamma)] = \left( \frac{\beta^2}{(2\alpha^2 - 3\beta^2)^2} + \frac{\beta^4}{2\alpha^2(2\alpha^2 - 3\beta^2)^2} \right) (x - x^2)^2. \] (4.12)

The first three terms of the solution given by equation (4.6), when \( f = 1 \), are
\[ u(x, \xi\gamma) = \left( \frac{1}{2\alpha} - \frac{\beta \xi\gamma}{2\alpha^2} + \frac{\beta^2 \xi^2\gamma}{2\alpha^3} \right) (x - x^2). \] (4.13)

Therefore,
\[ \text{Var}[u(x, \xi\gamma)] = \left( \frac{\beta^2}{4\alpha^4} + \frac{\beta^4}{8\alpha^6} \right) (x - x^2)^2. \] (4.14)

We note that for \( \beta \ll \alpha \), \( u_a(x, \xi\gamma) \) is close to \( u(x, \xi\gamma) \). This is illustrated in figure 1, where the square root of the variance of \( u_a(x, \xi\gamma) \) versus the square root of the variance of \( u(x, \xi\gamma) \) for \( \alpha = 1 \), \( \beta = 1/200 \) and \( \gamma = 100 \) is plotted.
Similarly, for $f = x$, we have the matrix equation

$$
\begin{pmatrix}
  a_{1,0} & \ldots & a_{1,N} \\
  a_{2,0} & \ldots & a_{2,N} \\
  \vdots & \ddots & \vdots \\
  a_{M,0} & \ldots & a_{M,N}
\end{pmatrix}
\begin{pmatrix}
  E[-(\alpha + \beta \xi_i)\psi_0\psi_k] \\
  E[-(\alpha + \beta \xi_i)\psi_1\psi_k] \\
  \vdots \\
  E[-(\alpha + \beta \xi_i)\psi_{M-1}\psi_k]
\end{pmatrix}
= \begin{pmatrix}
  \frac{1}{12} \\
  \frac{1}{6} \\
  \vdots \\
  \frac{1}{14}
\end{pmatrix}
$$

for fixed $i$ and $k$ where $i \in \{1, \ldots, M\}$ and $k \in \{0, \ldots, N\}$. If we let $M=3$ and $N=2$, we obtain the matrix equation

$$
\begin{pmatrix}
  \frac{2}{15} & \frac{1}{12} & \frac{1}{6} & \frac{2}{15} \\
  \frac{1}{12} & \frac{1}{6} & \frac{10}{14} & \frac{5}{14} \\
  \frac{1}{6} & \frac{10}{14} & \frac{2}{15} & \frac{1}{6} \\
  \frac{2}{15} & \frac{5}{14} & \frac{1}{6} & \frac{1}{6}
\end{pmatrix}
\begin{pmatrix}
  a_{1,0} \\
  a_{1,1} \\
  a_{1,2} \\
  a_{1,3}
\end{pmatrix}
= \begin{pmatrix}
  1/12 \\
  1/6 \\
  0 \\
  1/12
\end{pmatrix}
$$

Solving for the coefficients $a_{i,k}$, we obtain

$$
a_{1,0} = 0; a_{1,1} = 0; a_{1,2} = 0; a_{2,0} = \frac{1}{6\alpha}; a_{2,1} = \frac{\beta}{9\beta^2 - 6\alpha^2}; a_{2,2} = \frac{-\beta}{\alpha(9\beta^2 - 6\alpha^2)}; a_{3,0} = 0; a_{3,1} = 0; a_{3,2} = 0.
$$

Therefore, using equation (4.7) we have

$$
\begin{aligned}
\varphi_1(x, \xi) &= \left( \frac{1}{6\alpha} - \frac{\beta \xi_1}{6\alpha^2 - 9\beta^2} + \frac{\beta^2 \xi_1^2}{\alpha(6\alpha^2 - 9\beta^2)} \right) (x - x^3). \\
\end{aligned}
$$

(4.15)

It follows that,

$$
\text{Var}[\varphi_1(x, \xi)] = \left( \frac{\beta^2}{(6\alpha^2 - 9\beta^2)^2} + \frac{\beta^4}{2\alpha^2(6\alpha^2 - 9\beta^2)^2} \right) (x - x^3)^2.
$$

(4.16)

The first three terms of the solution given by equation (4.6), when $f = x$, are

$$
\varphi_1(x, \xi) = \left( \frac{1}{6\alpha} - \frac{\beta \xi_1}{6\alpha^2} + \frac{\beta^2 \xi_1^2}{6\alpha^3} \right) (x - x^3).
$$

(4.17)

Therefore,

$$
\text{Var}[\varphi_1(x, \xi)] = \left( \frac{\beta^2}{36\alpha^4} + \frac{\beta^4}{72\alpha^6} \right) (x - x^3)^2.
$$

(4.18)
We note that for $\beta \ll \alpha$, $u_a(x, \xi)$ is close to $u(x, \xi)$. This is illustrated in figure 2, where the square root of the variance of $u_0(x, \xi)$ versus the square root of the variance of $u(x, \xi)$ for $\alpha = 1$, $\beta = 1/200$ and $\gamma = 100$ is plotted.

![Standard Error Plot for two deterministic terms and three stochastic terms](image)

For $f = x + 1$, $u_a(x, \xi)$ is equal to the sum of $u_a(x, \xi)$ for $f = 1$ and $u_a(x, \xi)$ for $f = x$. Therefore, for $f = x + 1$, $u_a(x, \xi) = u(x, \xi)$.

$$u_a(x, \xi) = \frac{2x}{3\alpha} - \frac{x^2}{2\alpha} - \frac{x^3}{6\alpha} - \frac{4\beta x \xi}{6\alpha^2 - 9\beta^2} + \frac{\beta^2 x^2 \xi}{2\alpha^2 - 3\beta^2} + \frac{\beta^2 x^3 \xi}{6\alpha^2 - 9\beta^2} + \frac{4\beta^2 x^2 \xi^2}{\alpha(6\alpha^2 - 9\beta^2)} - \frac{\beta^2 x^2 \xi^2}{\alpha(2\alpha^2 - 3\beta^2)} - \frac{\beta^2 x^3 \xi^2}{\alpha(6\alpha^2 - 9\beta^2)}. \quad (4.19)$$

The first three terms of the solution given by equation (4.6), when $f = x + 1$, are

$$u(x, \xi) = \frac{2x}{3\alpha} - \frac{x^2}{2\alpha} - \frac{x^3}{6\alpha} - \frac{2\beta x \xi}{3\alpha^2} + \frac{\beta^2 x^2 \xi}{2\alpha^2} + \frac{\beta^2 x^3 \xi}{6\alpha^2} + \frac{2\beta^2 x^2 \xi^2}{3\alpha^3} - \frac{\beta^2 x^2 \xi^2}{2\alpha^3} - \frac{\beta^2 x^3 \xi^2}{6\alpha^3}. \quad (4.20)$$
Similarly to the $f = x$ case, when $f = x + 1$, the graph of the square root of the variance of $u_\alpha(x, \xi_\gamma)$ and the graph of the square root of the variance of $u(x, \xi_\gamma)$ for $\alpha = 1$, $\beta = 1/200$ and $\gamma = 100$ are identical up to the thousandth place, for $\beta \ll \alpha$.

5 Summary

The analysis and numerical examples indicate that the increase in cost, due to the stochastic component, is relatively modest, provided the dimension of the probability space is not too large. The approach used, here, is to treat the combined deterministic and stochastic problem in a manner analogous to the deterministic case alone, namely, the elements are basis functions for the entire cross product space, and the inner products are taken with respect to a cross product measure. We have used polynomials for the stochastic finite elements, and the evaluation of the inner product requires only that the degree of the polynomials be less than the highest known moment of the underlying distribution. The numerical examples show the rapid convergence of the stochastic components in the one dimensional case, in fact, the approximate solutions have almost "converged" using only three stochastic elements. Further, the analysis demonstrates that, for operators satisfying the contraction and commutativity assumptions, the stochastic elements converge rapidly in the low dimensional case, more precisely, $O(|\ln \epsilon|^n)$ (n is the dimension of the probability space) elements provide $O(\delta)$ accuracy, as $\epsilon \downarrow 0$. In other words, setting $\epsilon = \exp(-\kappa)$, we have shown, for $O(\kappa^n)$ stochastic elements, the error decreases exponentially fast, that is, the error is $O(\exp(-\kappa))$, as $\kappa \to \infty$.

References


## INTERNAL DISTRIBUTION

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Code</th>
<th>Name</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MS 0857</td>
<td>J.Red-Horse</td>
<td>09211</td>
</tr>
<tr>
<td>1</td>
<td>MS 0819</td>
<td>T.G. Trucano</td>
<td>09211</td>
</tr>
<tr>
<td>1</td>
<td>MS 1110</td>
<td>D.E. Womble</td>
<td>09214</td>
</tr>
<tr>
<td>20</td>
<td>MS 1110</td>
<td>J.M. DeLaurentis</td>
<td>09214</td>
</tr>
<tr>
<td>1</td>
<td>MS 9018</td>
<td>Central Technical Files</td>
<td>8945-1</td>
</tr>
<tr>
<td>2</td>
<td>MS 0899</td>
<td>Technical Library</td>
<td>9616</td>
</tr>
<tr>
<td>1</td>
<td>MS 0612</td>
<td>Review &amp; Approval Desk</td>
<td>9612</td>
</tr>
<tr>
<td></td>
<td></td>
<td>for DOE/OSTI</td>
<td></td>
</tr>
</tbody>
</table>