BLOCK-DIAGONAL PRECONDITIONING FOR SPECTRAL STOCHASTIC FINITE ELEMENT SYSTEMS *

CATHERINE E. POWELL † AND HOWARD C. ELMAN ‡

Abstract. Deterministic models of fluid flow and the transport of chemicals in flows in heterogeneous porous media incorporate PDEs whose material parameters are assumed to be known exactly. To tackle more realistic stochastic flow problems, it is fitting to represent the permeability coefficients as random fields with prescribed statistics. Traditionally, large numbers of deterministic problems are solved in a Monte Carlo framework and the solutions averaged to obtain statistical properties of the solution variables. Alternatively, so-called stochastic finite element methods (SFEMs) discretise the probabilistic dimension of the PDE directly leading to a single structured linear system. The latter approach is becoming extremely popular but its computational cost is still perceived to be problematic as this system is orders of magnitude larger than for the corresponding deterministic problem. A simple block-diagonal preconditioning strategy, incorporating only the mean component of the random field coefficient and based on incomplete factorisations has been employed in the literature, and observed to be robust, for problems of moderate variance, but without theoretical analysis. We solve the stochastic Darcy flow problem in primal formulation via the spectral SFEM and focus on its efficient iterative solution. To achieve optimal computational complexity, we base our block-diagonal preconditioner on algebraic multigrid. In addition, we provide new theoretical eigenvalue bounds for the preconditioned system matrix and highlight the dependence of the iteration counts on all the SFEM parameters.

Keywords finite elements, stochastic finite elements, fast solvers, preconditioning, multigrid

1. Introduction. Fluid flow and the transport of chemicals in flows in heterogeneous porous media are modelled mathematically using partial differential equations (PDEs). In deterministic modelling, inputs such as material properties, boundary conditions and source terms are assumed to be known explicitly. Such assumptions lead to tractable computations. However, simulations based on such over-simplifications cannot be used in practice to quantify the probability of an unfavourable event such as, say, a chemical being transported at a lethal level of concentration in groundwater. If the input variables for the system being studied are subject to uncertainty, then it is fitting to represent them as random fields. Solutions to the resulting stochastic PDEs are then necessarily also random fields. Strategic decision making cannot be made without some form of uncertainty quantification. Typically, a few moments of the solution variables are required or the probability distribution of a particular quantity of interest.

We focus on the case of uncertainty in material properties. The simplest and most commonly employed way of dealing with this is via the Monte Carlo Method (MCM). Large numbers of realisations of the random system inputs are generated and each resulting deterministic problem is solved using the available numerical methods and solvers. Results are post-processed to determine the desired statistical properties of the solution variables. Care must be taken, however, to ensure that enough realisations are generated so that the probability space is sampled appropriately. The exact number of trials required depends on the problem at hand but hundreds of thousands of experiments are not untypical for realistic flow problems with large variance. Easy access to parallel computers makes this feasible in the 21st century. Quasi Monte Carlo methods and variance reduction techniques can be used to reduce the overall number of trials, making this technology even more competitive. However, minimising the computational cost of solving each deterministic problem is still a crucial and non-trivial step.

An alternative approach, pioneered in [17], couples a Karhunen-Loève expansion of the random field coefficients in the stochastic PDE with a traditional finite element discretisation on the spatial domain. The stochastic dimension of the problem is discretised directly. The advantage of this so-called stochastic finite element method (SFEM) is that a *single* linear system needs to be solved. However, this is orders of magnitude larger than the subproblems solved in the MCM. The components of the discrete solution are coefficients of a probabilistic expansion of the solution variables which can

easily be post-processed to recover the mean, variance and probability distribution of quantities of interest. SFEMs are becoming increasingly popular but their computational cost is perceived to be high. The linear systems in question are, however, highly structured and researchers have been slow to take up the challenge of solving them efficiently. Initial attempts were made in [16] and [27]. More recently, fast and efficient linear algebra for alternative SFEMs (see [3], [1], [2]) has been proposed by linear algebra specialists (see [5] [6] [7]) and fast solvers and parallel computer architectures have been exploited by the authors of [19], [20], [21] and [22].

We focus on the numerical solution of the steady-state diffusion problem which, in its deterministic formulation, is written as

$$
\begin{aligned}
-\nabla \cdot K \nabla u &= f, & &\text{in } D \subset \mathbb{R}^d, \\
u &= g & &\text{on } \partial D_D \neq \emptyset, \\
K \nabla u \cdot \vec{n} &= 0 & &\text{on } \partial D_N = \partial D \backslash \partial D_D.
\end{aligned}
\tag{1.1}
$$

The boundary-value problem (1.1) is the primal formulation of the standard second-order elliptic problem and provides a simplified model for single-phase flow in a saturated porous medium (e.g. see [29], [9].) In that physical setting, u and $\vec{q} = K \nabla u$ are the residual pressure and velocity field, respectively. K is a prescribed scalar function or a $d \times d$ symmetric and uniformly positive definite tensor, representing permeability. Since the permeability coefficients of a heterogeneous porous medium can never, in reality, be known at every point in space, we consider, here, the case where $K = K(\boldsymbol{x}, \omega)$ is a random field. We assume only statistical properties of K.

To make these notions precise, let $(\Omega, \mathcal{B}, P)$ denote a probability space where Ω, $\mathcal{B}$, and P are the set of random events, the minimal σ-algebra of the subsets of Ω and an appropriate probability measure, respectively. Then $K(\boldsymbol{x}, \omega) : D \times \Omega \to \mathbb{R}$. For a fixed spatial location $\boldsymbol{x} \in D$, $K(\cdot, \omega)$ is a random variable whilst for a fixed realisation $\omega \in \Omega$, $K(\boldsymbol{x}, \cdot)$ is a spatial function in $\boldsymbol{x}$. The stochastic problem then reads, find a random field $u(\boldsymbol{x}, \omega) : D \times \Omega \to \mathbb{R}$ such that P-almost surely $(P-\text{a.s.})$

$$
\begin{aligned}
-\nabla \cdot K(\boldsymbol{x}, \omega) \nabla u(\boldsymbol{x}, \omega) = f(\boldsymbol{x}) & & \boldsymbol{x} \in D, \\
u(\boldsymbol{x}, \omega) = g(\boldsymbol{x}), & & \boldsymbol{x} \in \partial D_D, \\
K(\boldsymbol{x}, \omega) \nabla u(\boldsymbol{x}, \omega) \cdot \vec{n} = 0, & & \boldsymbol{x} \in \partial D_N = \partial D \backslash \partial D_D,
\end{aligned}
\tag{1.2}
$$

where $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are suitable deterministic functions. The source term f can also be treated as a random field in a straight-forward manner (see [6], [3]) but we shall not consider that case.

1.1. Overview. The focus of this work is the design of fast solvers for (1.2). In section 2 we summarise the classical spectral SFEM discretisation (see [17]) and discuss some modelling issues that affect the spectral properties of the resulting linear systems and ultimately the solver performance. We highlight the structure and algebraic properties of the resulting linear system and implement the block-diagonal preconditioning scheme advocated in [16]. For the subproblems, however, we replace the traditional incomplete factorisation schemes used in [16] and [27] with a black-box algebraic multigrid (AMG) solver. We also compare the computational effort required with that of traditional Monte Carlo methods. Our main contributions, namely the derivation of key properties of the finite element matrices and eigenvalue bounds for the preconditioned system matrices, are presented in section 3. The bounds are shown to be tight for test problems commonly used in the literature. Numerical results are presented in section 4.

2. Spectral SFEM for the steady-state diffusion problem. SFEMs can be divided into two categories: non-spectral and spectral methods. The former (see [3], [6]) achieve a prescribed accuracy via polynomial approximation of a fixed degree on an increasingly fine partition of a probability range space. The latter require no such formal partition and error is reduced by increasing the degree, p, of polynomial approximation. The methods can be further subcategorised according to the choice of stochastic basis functions (orthogonal, [12], doubly-orthogonal [2] etc.) We focus on the classical spectral method [17] and employ a standard orthogonal polynomial chaos basis. The main advantage is that the dimension of this space grows more slowly than for other choices (see [2] or [3] for alternatives). However, the stochastic terms are fully coupled and this is much more challenging for

solvers. For notational convenience, we illustrate the derivation of the spectral SFEM equations for the case of homogeneous Dirichlet boundary conditions only. This derivation is completely standard and full details can be found in [17], [3], [1], [2].

If the coefficient $K(\boldsymbol{x}, \omega)$ is bounded and strictly positive, that is,

$$0 < k_1 \leq K(\boldsymbol{x}, \omega) \leq k_2 < +\infty, \quad \text{a.e. in } D \times \Omega, \tag{2.1}$$

then (1.2) can be cast in weak form in the usual way, and existing theory (i.e. the classical Lax-Milgram lemma) can be used to establish existence and uniqueness of a solution. In the present stochastic setting, the idea is to seek a weak solution in a Hilbert space $H = H_0^1(D) \otimes L^2(\Omega)$, consisting of tensor products of deterministic functions defined on the spatial domain and stochastic functions defined on the probability space.

In order to set up variational problems, some notation is first required. Let X be a real random variable belonging to $(\Omega, \mathcal{B}, P)$ and assume that there exists a density function $\rho : \mathbb{R}^d \to \mathbb{R}$ such that the expected value can be expressed via the integral,

$$\langle X \rangle = \int_{\mathbb{R}} x \rho(x) dx.$$

If $\langle X \rangle < \infty$ then we say $X \in L^1(\Omega)$. The space $L^2(D) \otimes L^2(\Omega) = \{v(\boldsymbol{x}, \omega) : D \times \Omega \to \mathbb{R} \mid \| v \| < \infty \}$, consists of random functions with finite second moment where the norm $\| \cdot \|$ is defined via,

$$\| v(\boldsymbol{x}, \omega) \|^2 = \left\langle \int_D v^2(\boldsymbol{x}, \omega) \, d\boldsymbol{x} \right\rangle. \tag{2.2}$$

Next, we define $V = \{v(\mathbf{x}, \omega) : D \times \Omega \to \mathbb{R} \mid \| v \|_V < \infty, \ v|_{\partial D \times \Omega} = 0\}$, where the 'stochastic energy' norm $\| \cdot \|_V$ is defined via,

$$\| v(\boldsymbol{x}, \omega) \|_V^2 = \left\langle \int_D K(\boldsymbol{x}, \omega) |\nabla v(\boldsymbol{x}, \omega)|^2 \, d\boldsymbol{x} \right\rangle. \tag{2.3}$$

If condition (2.1) holds, then the norm is well-defined and it can be shown that there exists a unique $u = u(\boldsymbol{x}, \omega) \in V$ satisfying the continuous variational problem,

$$\left\langle \int_D K(\boldsymbol{x}, \omega) \nabla u(\boldsymbol{x}, \omega) \cdot \nabla v(\boldsymbol{x}, \omega) \, d\boldsymbol{x} \right\rangle = \left\langle \int_D f(\boldsymbol{x}) v(\boldsymbol{x}, \omega) \, d\boldsymbol{x} \right\rangle \quad \forall \, v(\boldsymbol{x}, \omega) \in V. \tag{2.4}$$

To convert the stochastic problem (2.4) into a deterministic one, we require a finite set of random variables, $\{\xi_1(\omega), \ldots, \xi_M(\omega)\}$ that represent appropriately and sufficiently, the stochastic variability of $K(\boldsymbol{x}, \omega)$. One possibility is to approximate $K(\boldsymbol{x}, \omega)$ by a truncated Karhunen-Loève (KL) expansion, a linear combination of a finite set of *uncorrelated* random variables. We discuss this in the next section. After formally replacing $K(\boldsymbol{x}, \omega)$ by $K_M(\boldsymbol{x}, \boldsymbol{\xi})$, it can be shown that the corresponding solution also has finite stochastic dimension and the variational problem (2.4) can be re-stated as, find $u = u(\boldsymbol{x}, \boldsymbol{\xi}) \in W$ satisfying,

$$\int_\Gamma \rho(\boldsymbol{\xi}) \int_D K_M(\boldsymbol{x}, \boldsymbol{\xi}) \nabla u(\boldsymbol{x}, \boldsymbol{\xi}) \cdot \nabla w(\boldsymbol{x}, \boldsymbol{\xi}) \, d\boldsymbol{x} \, d\boldsymbol{\xi} = \int_\Gamma \rho(\boldsymbol{\xi}) \int_D f(\boldsymbol{x}) w(\boldsymbol{x}, \boldsymbol{\xi}) \, d\boldsymbol{x} \, d\boldsymbol{\xi} \tag{2.5}$$

$\forall \, w(\boldsymbol{x}, \boldsymbol{\xi}) \in W$. Here, $\rho(\boldsymbol{\xi})$ denotes the joint probability density function of the random variables and $\Gamma = \Gamma_1 \times \ldots \times \Gamma_M$ is the joint image of the random vector $\boldsymbol{\xi}$. A key point is that if the random variables are mutually *independent* then the density function is separable, i.e. $\rho(\boldsymbol{\xi}) = \rho_1(\xi_1) \cdot \rho_2(\xi_2) \cdots \rho_M(\xi_M)$ and the integrals in (2.5) simplify greatly. Many authors favour Gaussian random variables since uncorrelated Gaussian random variables are independent (see [17], [7]). However, they do not have bounded images and so (2.1) cannot be guaranteed. Other authors simply introduce independence as a modelling assumption (see [3], [6], [1]), which may not be physically justified either, in order to work with uniform random variables which have bounded images.

Formally, the space W differs from V since the definition of the norm induced by the inner-product in (2.5) is defined in terms of the density $\rho(\boldsymbol{\xi})$. Hence, to make (2.5) understood, we define,

$$W = H_0^1(D) \otimes L^2(\Gamma) = \left\{ w(\boldsymbol{x}, \boldsymbol{\xi}) \in L^2(D \times \Gamma) \mid \, \| w(\boldsymbol{x}, \boldsymbol{\xi}) \|_W < \infty \text{ and } w|_{\partial D \times \Gamma} = 0 \right\}, \qquad (2.6)$$

and the energy norm,

$$\| w(\boldsymbol{x}, \boldsymbol{\xi}) \|_W^2 = \int_\Gamma \rho(\boldsymbol{\xi}) \int_D K_M(\boldsymbol{x}, \boldsymbol{\xi}) \left| \nabla w(\boldsymbol{x}, \boldsymbol{\xi}) \right|^2 d\boldsymbol{x} \, d\boldsymbol{\xi}.$$

Representing the stochastic behaviour of $K(\boldsymbol{x}, \omega)$ by a finite set of random variables (a form of model-order reduction), can be viewed as the first step in the discretisation process. To obtain a fully discrete version of (2.5), we now need a finite-dimensional subspace $W_h \subset W = H_0^1(D) \otimes L^2(\Gamma)$. The key idea of the SFEM is to discretise the deterministic space $H_0^1(D)$ and the stochastic space $L^2(\Gamma)$ separately. Hence, given bases,

$$X_h = \mathtt{span}\,\{\phi_i(\boldsymbol{x})\}_{i=1}^{Nx} \subset H_0^1(D), \qquad S = \mathtt{span}\,\{\psi_j(\boldsymbol{\xi})\}_{j=1}^{N_\xi} \subset L^2(\Gamma), \qquad (2.7)$$

which may be chosen independently of one another, we define,

$$W^h = X_h \otimes S = \left\{ v(\boldsymbol{x}, \boldsymbol{\xi}) \in L^2(D \times \Gamma) \mid v(\boldsymbol{x}, \boldsymbol{\xi}) \in \mathtt{span}\,\{\phi(\boldsymbol{x})\psi(\boldsymbol{\xi}), \phi \in X_h, \, \psi \in S\} \right\}. \qquad (2.8)$$

We choose the basis for the deterministic space by defining the functions $\phi_i(\boldsymbol{x})$ to be the standard hat functions associated with piecewise linear (or bilinear) approximation associated with a partition T_h of the spatial domain D into triangles (or rectangles). Different classes of SFEMs are distinguished by their choices for S. In [6], [7], [3], [2], tensor products of piecewise polynomials on the subdomains Γ_i are employed. In this approach, the polynomial degree is fixed and approximation is improved by refining the partition of Γ. The classical, so-called spectral SFEM (see [17], [23], [10]) employs global polynomials of total degree p in M random variables ξ_i on Γ. In this approach, there is no partition of Γ and approximation is improved by increasing the polynomial degree. We shall adopt the latter method. Convergence and approximation properties are discussed in [2].

When the underlying random variables are Gaussian, the spectral approach uses a basis of multi-dimensional Hermite polynomials of total degree p, termed the 'polynomial chaos' (see [33]). The use of Hermite polynomials ensures that the corresponding basis functions are orthogonal with respect to the Gaussian probability measure. This leads to sparse linear systems, a crucial property which *must* be exploited for fast solution schemes. If alternative distributions are used to model the input random field then appropriate stochastic basis functions should be used to ensure orthogonality with respect to the probability measure they induce (see [34]).

For example, consider the case of Gaussian random variables with $M = 2$ and $p = 3$. The stochastic basis is the set of two-dimensional Hermite polynomials (the product of a one-dimensional Hermite polynomial in ξ_1 and a one-dimensional polynomial in ξ_2) of degree less than or equal to three. Each basis function is associated with a multi-index of length two, $\alpha = (\alpha_1, \alpha_2)$, where the components represent the degrees of polynomials in ξ_1 and ξ_2. Since the total degree of the polynomial is three, we have the possibilities $\alpha = (0,0), (1,0), (2,0), (3,0), (0,1), (1,1), (2,1), (0,2), (1,2), (0,3)$. Given that the one-dimensional Hermite polynomials of degrees $0, 1, 2, 3$ are $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, and $H_3(x) = x^3 - 3x$ we obtain

$$S = \mathtt{span}\,\{\psi_j(\boldsymbol{\xi})\}_{j=1}^{10} = \left\{ 1, \xi_1, \xi_1^2 - 1, \xi_1^3 - 3\xi_1, \xi_2, \xi_1\xi_2, \left(\xi_1^2 - 1\right)\xi_2, \xi_2^2 - 1, \left(\xi_2^2 - 1\right)\xi_1, \xi_2^3 - 3\xi_2 \right\}.$$

Note that the dimension of this space is,

$$N_\xi = 1 + \sum_{s=1}^p \frac{1}{s!} \prod_{r=0}^{s-1} (M + r) = \frac{(M+p)!}{M!p!}.$$

In the sequel, we give specific results for the case of Gaussian random variables and Hermite polynomials since this is most popular with practitioners (see [7], [16], [17], [20], [27]). However, this is not restrictive. Other distributions can be used in the same framework provided that the correct choice of orthogonal polynomial is made.

2.1. Karhunen-Loève expansion. A random field $K(\boldsymbol{x}, \omega)$ with continuous covariance function,

$$C(\boldsymbol{x}, \boldsymbol{y}) = \Big\langle \left(K(\boldsymbol{x}, \omega) - \langle K(\boldsymbol{x}) \rangle\right) \left(K(\boldsymbol{y}, \omega) - \langle K(\boldsymbol{y}) \rangle\right) \Big\rangle = \sigma^2 \varrho(\boldsymbol{x}, \boldsymbol{y}), \quad \boldsymbol{x}, \boldsymbol{y} \in D,$$

admits a proper orthogonal decomposition (see [24]), or Karhunen-Loève expansion,

$$K(\boldsymbol{x}, \omega) = \mu + \sigma \sum_{i=1}^{\infty} \sqrt{\lambda_i} c_i(\boldsymbol{x}) \xi_i, \tag{2.9}$$

where $\mu = \langle K(\boldsymbol{x}) \rangle$, the random variables $\{\xi_1, \xi_2, \dots\}$ are *uncorrelated* and $\{\lambda_i, c_i(\boldsymbol{x})\}$ are the set of eigenvalues and eigenfunctions of $\varrho(\boldsymbol{x}, \boldsymbol{y})$. $C(\cdot, \cdot)$ is non-negative definite, the eigenvalues are real and we label them in descending order $\lambda_1 > \lambda_2 > \dots$. Now, we can employ the truncated expansion,

$$K(\boldsymbol{x}, \boldsymbol{\xi}) \approx K_M(\boldsymbol{x}, \boldsymbol{\xi}) = \mu + \sigma \sum_{i=1}^{M} \sqrt{\lambda_i} c_i(\boldsymbol{x}) \xi_i, \tag{2.10}$$

for computational purposes in (2.5). This choice is motivated by the fact that (see [24]) quadratic mean square convergence of $K_M(\boldsymbol{x}, \boldsymbol{\xi})$ to $K(\boldsymbol{x}, \omega)$, is guaranteed as $M \to \infty$. Truncation criteria are usually based on the speed of decay of the eigenvalues since $|D|\text{Var}(\text{K}) = \sum \lambda_i$. However, care must be taken to ensure that for the chosen M, (2.5) is well-posed. For the conventional analysis, we require that the truncated coefficient is strictly positive and bounded, and thus satisfies,

$$0 < k_1 \leq K_M(\boldsymbol{x}, \omega) \leq k_2 < \infty \text{ a.e. in } D \times \Omega.$$

This is not the same as (2.1). In [1], it is shown that we require that $K_M(\boldsymbol{x}, \omega)$ converges to $K(\boldsymbol{x}, \omega)$ uniformly as $M \to \infty$. One sufficient condition for this (see [31]) is that the random variables $\xi_1, \dots, \xi_M$ have bounded images. Although Gaussian random variables do violate this condition, they are widely used in SFEMs, as we previously mentioned, because uncorrelated Gaussian random variables are independent. This simplifies (2.5). Uncorrelated non-Gaussian random variables are not necessarily independent and more complex expansions of $K(\boldsymbol{x}, \omega)$ are required (see [5].) However, using random variables with bounded images is not the whole story. At a discrete level, with an appropriate choice of discretisation parameters, Gaussian random variables can be employed. We will show that for a fixed variance, it is possible to choose the parameters M and p so that the system matrix to be defined in (3.4) is positive definite. However, other values of M and p can lead to an indefinite or singular system matrix.

2.1.1. Truncated KL expansion. To illustrate the positivity issue, consider the following example. The covariance function employed in [17], [7], [3] (and in the MATLAB-based code [10]) is,

$$C(\boldsymbol{x}, \boldsymbol{y}) = \sigma^2 \exp\left(-\frac{|x_1 - y_1|}{c_1} - \frac{|x_2 - y_2|}{c_2}\right), \tag{2.11}$$

where c_1 and c_2 are correlation lengths and $D = [-a, a] \times [-a, a]$. The attraction of working with (2.11) is that analytical expressions for the eigenfunctions and eigenvalues exist. To see this, note that the kernel is separable and so the eigenfunctions and eigenvalues can be expressed as the products of those of two corresponding one-dimensional problems. That is $c_i(\boldsymbol{x}) = c_k^1(x_1) c_j^2(x_2)$ and $\lambda_i = \lambda_k^1 \lambda_j^2$, where the eigenpairs $\left\{c_k^1(x_1), \lambda_k^1\right\}_{k=1}^{\infty}$ and $\left\{c_j^2(x_2), \lambda_j^2\right\}_{j=1}^{\infty}$ are solutions to,

$$\int_{-a}^{a} \exp\left(-b_1|x_1 - y_1|\right) c_k^1(y_1)\, dy_1 = \lambda_k^1 c_k^1(x_1) \qquad \int_{-a}^{a} \exp\left(-b_2|x_2 - y_2|\right) c_j^2(y_2)\, dy_2 = \lambda_j^2 c_j^2(x_2),$$

with $b_i = c_i^{-1}$, $i = 1, 2$. Solutions are given in [17]. As i increases, the eigenfunctions become more oscillatory. The more random variables we use to represent $K(\boldsymbol{x}, \boldsymbol{\xi})$, the more scales of fluctuation we incorporate. In Fig. 2.1 we plot a sample of the eigenfunctions for the case $c_1 = 1 = c_2$. In

5

Fig. 2.2, we plot three realisations of the corresponding truncated coefficient (2.10) with standard deviation $\sigma = 0.5$. Observe that in one of these realisations ($M = 50$) the truncated KL expansion is not strictly positive. This fits with theoretical arguments given in [1]. The truncated coefficient is not strictly-positive a.e in $D \times \Omega$. However, if the variance is 'not large' we can still choose M and p so that the discrete SFEM system has a positive definite system matrix. We investigate this issue further in sections 3 and 4.
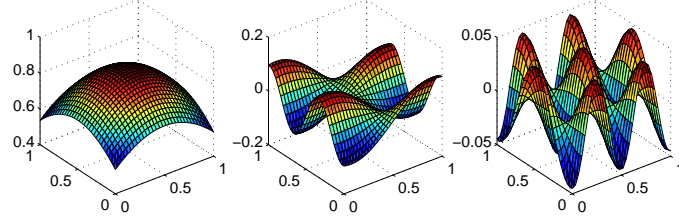

FIG. 2.1. *1st, 20th, 50th, eigenfunction of the covariance kernel in (2.11) with $c_1 = c_2 = 1$*
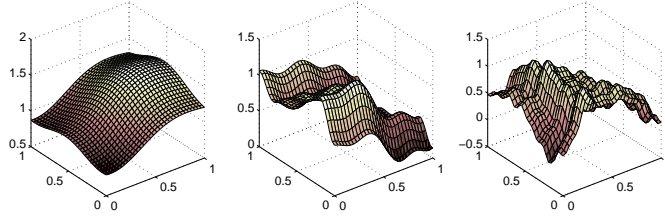

FIG. 2.2. *Realisations of $K_M(\boldsymbol{x}, \omega)$ with $M = 5, 20, 50$, $c_1 = 1 = c_2$, $\mu = 1$, $\sigma = 0.5$ and $\xi_i \sim N(0, 1), i = 1 : M$.*

3. Linear algebra aspects of spectral SFEM formulation. Given $K_M(\boldsymbol{x}, \boldsymbol{\xi})$ and bases for X_h and S, we now seek a finite-dimensional solution $u_{hp}(\boldsymbol{x}, \boldsymbol{\xi}) \in W^h = X_h \otimes S$ satisfying,

$$\int_\Gamma \rho(\boldsymbol{\xi}) \int_D K_M(\boldsymbol{x}, \boldsymbol{\xi}) \nabla u_{hp}(\boldsymbol{x}, \boldsymbol{\xi}) \cdot \nabla w(\boldsymbol{x}, \boldsymbol{\xi}) \, d\boldsymbol{x} \, d\boldsymbol{\xi} = \int_\Gamma \rho(\boldsymbol{\xi}) \int_D f(\boldsymbol{x}) w(\boldsymbol{x}, \boldsymbol{\xi}) \, d\boldsymbol{x} \, d\boldsymbol{\xi} \tag{3.1}$$

$\forall \, w(\boldsymbol{x}, \boldsymbol{\xi}) \in W^h$. Expanding the solution and the test functions in the chosen bases in (3.1),

$$u_{hp}(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{s=1}^{N_\xi} \sum_{r=1}^{N_x} u_{r,s} \phi_r(\boldsymbol{x}) \psi_s(\boldsymbol{\xi}) = \sum_{s=1}^{N_\xi} \underline{u}_s \psi_s(\boldsymbol{\xi}), \tag{3.2}$$

leads to a linear system $A\underline{u} = \underline{f}$ of dimension $N_x N_\xi \times N_x N_\xi$ with block-structure,

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,N_\xi} \\ A_{2,1} & A_{2,2} & \dots & A_{2,N_\xi} \\ \vdots & \ddots & & \vdots \\ A_{N_\xi,1} & A_{N_\xi,2} & \dots & A_{N_\xi,N_\xi} \end{pmatrix}, \qquad \underline{u} = \begin{pmatrix} \underline{u}_1 \\ \underline{u}_2 \\ \vdots \\ \underline{u}_{N_\xi} \end{pmatrix}, \qquad \underline{f} = \begin{pmatrix} \underline{f}_1 \\ \underline{f}_2 \\ \vdots \\ \underline{f}_{N_\xi} \end{pmatrix}. \tag{3.3}$$

The blocks of A are linear combinations of $M + 1$ weighted stiffness matrices of dimension N_x, each with a sparsity pattern equivalent to that of the corresponding deterministic problem. That is,

6

$$A_{r,s} = \langle \psi_r(\boldsymbol{\xi})\psi_s(\boldsymbol{\xi})\rangle K_0 + \sum_{k=1}^{M} \langle \xi_k \psi_r(\boldsymbol{\xi})\psi_s(\boldsymbol{\xi})\rangle K_k,$$

$$K_0(i,j) = \int_D \mu \nabla \phi_i(\boldsymbol{x})\nabla\phi_j(\boldsymbol{x})\,d\boldsymbol{x}, \quad K_k(i,j) = \sqrt{\lambda_k} \int_D c_k(\boldsymbol{x})\nabla\phi_i(\boldsymbol{x})\nabla\phi_j(\boldsymbol{x})\,d\boldsymbol{x}, \qquad (3.4)$$

where $k = 1 : M$ and $\mu = \langle K(\boldsymbol{x})\rangle$. K_0 contains the mean information of the permeability coefficient while the other K_k blocks represent fluctuations. In tensor product notation, we have

$$A = G_0 \otimes K_0 + \sum_{k=1}^{M} G_k \otimes K_k, \qquad f = \underline{g}_0 \otimes \underline{f}_0, \qquad (3.5)$$

where the stochastic matrices G_k are defined via,

$$G_0(r,s) = \langle \psi_r, \psi_s\rangle, \qquad G_k(r,s) = \langle \xi_k \psi_r \psi_s\rangle, \qquad k = 1 : M, \qquad (3.6)$$

and the vectors $\underline{g}_0$ and $\underline{f}_0$ are given by $\underline{g}_0(i) = \langle \psi_i\rangle$, $\underline{f}_0(i) = \int_D f(\boldsymbol{x})\phi(\boldsymbol{x})\,d\boldsymbol{x}$. Since the stochastic basis functions are orthogonal with respect to the probability measure of the distribution of the chosen random variables, G_0 is diagonal. If doubly-orthogonal polynomials are used (see [34]) then each G_k is diagonal, so that A is block-diagonal. This can be handled very easily by solving N_ξ decoupled systems of dimension N_x. We do not consider that case here.

The block-structure of A obtained from the spectral SFEM is illustrated in Fig. 3.1. Many of the coefficients in the summation in (3.4) are zero, due to the orthogonality properties of the stochastic basis functions (see section 3.1), and the matrix is highly sparse in a block sense. In particular, K_0 occurs only on the main diagonal blocks. It should also be noted that A is never fully assembled. As pointed out in [16], we store only $M + 1$ matrices of dimension $N_x \times N_x$ and the entries of each G_k in (3.6). If the discrete problem is wellposed then A is symmetric and positive definite but is ill-conditioned with respect to the discretisation parameters. We can solve the system iteratively, using the conjugate gradient method (performing matrix-vector products intelligently) but a preconditioner is required. We discuss this in the next section.
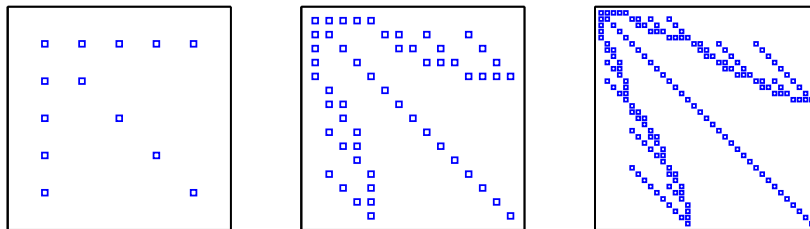

FIG. 3.1. *Matrix block-structure (each block has dimension $N_x \times N_x$), $M = 4$ with $p = 1, 2, 3$ (left to right)*

3.1. Matrix properties. We examine first the properties of the stochastic G matrices in (3.6). Assuming Gaussian random variables, each stochastic basis function $\psi_i(\boldsymbol{\xi})$ is the product of M one-dimensional Hermite polynomials[1]. That is $\psi_i(\boldsymbol{\xi}) = H_{i_1}(\xi_1)H_{i_2}(\xi_2)\cdots H_{i_M}(\xi_M)$, where the index i into the stochastic basis is identified with a multi-index $i = (i_1, \ldots, i_M)$, $\sum i_s \leq p$, where p is the total polynomial degree and M is the number of random variables retained in (2.10). The ordering of these multi-indices is not important for the calculations but some simple eigenvalue bounds for these matrices are obvious if a specific ordering is used.

---

[1]If uniform random variables are more appropriate, replace Hermite polynomials with Legendre polynomials.

Using only the orthogonality property of the one-dimensional Hermite polynomials and the independence of the random variables, a simple calculation yields,

$$G_0(i,j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} H_i(\boldsymbol{\xi}) H_j(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) \, d\boldsymbol{\xi} = \prod_{s=1}^{M} \int_{-\infty}^{\infty} H_{i_s}(\xi_s) H_{j_s}(\xi_s) \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi_s^2}{2}} \, d\xi_s$$

$$= \prod_{s=1}^{M} i_s! \, \delta_{i_s,j_s} = \left(\prod_{s=1}^{M} i_s! \right) \delta_{i,j} = \begin{cases} \prod_{s=1}^{M} i_s! & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

G_0 is a diagonal matrix whose ith entry is the product of the factorials of the components of the multi-index corresponding to the global index i (see tabulated values of $\langle \psi_i^2 \rangle$ in [17, ch.2]). If the stochastic basis functions are normalised, then G_0 is the identity matrix. Using, in addition, the three-term recurrence for the Hermite polynomials,

$$H_{k+1}(x) = x H_k(x) - k H_{k-1}(x), \tag{3.8}$$

for $k = 1 : M$, we obtain,

$$G_k(i,j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \xi_k H_{\underline{i}}(\boldsymbol{\xi}) H_{\underline{j}}(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) \, d\boldsymbol{\xi} = \left(\prod_{s=1,s\neq k}^{M} \langle H_{i_s}(\xi_s) H_{j_s}(\xi_s) \rangle \right) \langle \xi_k H_{i_k}(\xi_k) H_{j_k}(\xi_k) \rangle$$

$$= \begin{cases} \left(\prod_{s=1,s\neq k}^{M} i_s! \, \delta_{i_s,j_s} \right) (i_k + 1)! & \text{if } i_k = j_k - 1 \\ \left(\prod_{s=1,s\neq k}^{M} i_s! \, \delta_{i_s,j_s} \right) i_k! & \text{if } i_k = j_k + 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \left(\prod_{s=1,s\neq k}^{M} i_s! \right) (i_k + 1)! & \text{if } i_k = j_k - 1 \text{ and } i_s = j_s, \quad s = \{1 : M\} \setminus \{k\} \\ \left(\prod_{s=1}^{M} i_s! \right) & \text{if } i_k = j_k + 1 \text{ and } i_s = j_s, \quad s = \{1 : M\} \setminus \{k\} \\ 0 & \text{otherwise.} \end{cases}$$

Due to (3.8), G_k has at most two non-zero entries per row. $G_k(i,j)$ is non-zero only when the multi-indices for the global indices i and j agree in all components except the kth one, where the entries differ by one. (The same is true for bases built on alternative sets of orthogonal polynomials).

In section 3.2, it will be necessary to have a handle on the eigenvalues of $G_0^{-1} G_k$ or, equivalently, the symmetrically preconditioned matrices $\hat{G}_k = G_0^{-\frac{1}{2}} G_k G_0^{-\frac{1}{2}}$, $k = 1 : M$, for a fixed value of p. The next result relies on the well-known fact that roots of orthogonal polynomials are eigenvalues of certain tri-diagonal matrices (see [18] or a standard numerical analysis text e.g. [32]).

LEMMA 3.1. *If Hermite polynomials of total degree p in M Gaussian random variables are used for the stochastic basis, the eigenvalues of $\hat{G}_k = G_0^{-\frac{1}{2}} G_k G_0^{-\frac{1}{2}}$, for each $k = 1 : M$, lie in the interval $[-H_{p+1}^{max}, H_{p+1}^{max}]$ where H_{p+1}^{max} is the maximum positive root of the one-dimensional Hermite polynomial of degree $p + 1$.*

Proof. Using the definitions of G_0 and G_k observe that $\hat{G}_k$ has at most two non-zeroes per row:

$$\left[G_0^{-\frac{1}{2}} G_k G_0^{-\frac{1}{2}} \right](i,j) = \begin{cases} \dfrac{\left(\prod_{s=1,s\neq k}^{M} i_s! \right)(i_k+1)!}{\sqrt{\left(\prod_{s=1}^{M} i_s! \right)} \sqrt{\left(\prod_{s=1}^{M} j_s! \right)}} & \text{if } i_k = j_k - 1 \text{ and } i_s = j_s, \quad s = \{1 : M\} \setminus \{k\} \\[3ex] \dfrac{\left(\prod_{s=1}^{M} i_s! \right)}{\sqrt{\left(\prod_{s=1}^{M} i_s! \right)} \sqrt{\left(\prod_{s=1}^{M} j_s! \right)}} & \text{if } i_k = j_k + 1 \text{ and } i_s = j_s, \quad s = \{1 : M\} \setminus \{k\} \\[3ex] 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \sqrt{i_k + 1} & \text{if } i_k = j_k - 1 \text{ and } i_s = j_s, \quad s = \{1 : M\} \setminus \{k\} \\[2ex] \sqrt{i_k} & \text{if } i_k = j_k + 1 \text{ and } i_s = j_s, \quad s = \{1 : M\} \setminus \{k\} \\[2ex] 0 & \text{otherwise} \end{cases}$$

8

Let M and p be fixed but arbitrary and consider, first, the matrix $\hat{G}_1$. It is possible to choose an ordering of the stochastic basis functions that causes $\hat{G}_1$ to be block tri-diagonal. Recall that the sum of the multi-index components does not exceed p. First, list multi-indices with first component ranging from 0 to p with entries in the second to Mth components summing to zero: $(0, 0 \ldots 0)$, $(1, 0 \ldots 0)$ ... $(p, 0 \ldots 0)$. This accounts for $p+1$ basis functions. Given the definition of $\hat{G}_1$, the leading $(p+1) \times (p+1)$ block, T_{p+1}, is then necessarily tri-diagonal

$$T_{p+1} = \begin{pmatrix} 0 & 1 & & & \\ 1 & 0 & \sqrt{2} & & \\ & \ddots & \ddots & \ddots & \\ & & \sqrt{p-1} & 0 & \sqrt{p} \\ & & & \sqrt{p} & 0 \end{pmatrix}. \tag{3.9}$$

Next, we list multi-indices with first components ranging from 0 to $p-1$ and with entries in the second to Mth components that add up to one, but grouped to have the same entries in those components:

$$
\begin{matrix}
(0, 0 \ldots 0, 1) & (0, 0 \ldots 1, 0) & & (0, 1 \ldots 0, 0) \\
(1, 0 \ldots 0, 1) & (1, 0 \ldots 1, 0) & \ldots\ldots & (1, 1 \ldots 0, 0) \\
\vdots & \vdots & \ldots\ldots & \vdots \\
(p-1, 0 \ldots 0, 1) & (p-1, 0 \ldots 1, 0) & \ldots\ldots & (p-1, 1 \ldots 0, 0)
\end{matrix}
$$

This accounts for $(M-1) \times p$ basis functions. $\hat{G}_1$ then has $M-1$ copies of a tri-diagonal matrix T_p defined analogously to T_{p+1}. We continue to order the multi-indices in this way until, finally, we list multi-indices which are 0 in the first component and in the second to Mth components are the same and have entries that add up to p. Then, $\hat{G}_1$ is a symmetric block tri-diagonal matrix with multiple copies of the symmetric tri-diagonal matrices $T_{p+1}, T_p, \ldots, T_1 = 0$ as the diagonal blocks. The number of copies of T_{p+1} is one and the number of copies of T_j, $j = 1 : p$, that appear is:

$$\frac{1}{(p-j+1)!} \prod_{r=0}^{p-j} (M-1+r).$$

The eigenvalues of $\hat{G}_1$ are the eigenvalues of the $\{T_j\}$. The eigenvalues of each tri-diagonal block are just roots of a characteristic polynomial $p_j(\lambda)$ that satisfies the recursion (3.8). That is,

$$p_{j+1}(\lambda) = (\lambda - 0)\, p_j(\lambda) - (-\sqrt{j})^2 p_{j-1}(\lambda).$$

Hence $p_j(\lambda)$ is the Hermite polynomial of degree j (see [18] or [32, ch.3]). Since the roots of lower degree Hermite polynomials are bounded by the extremal eigenvalues of higher degree polynomials (see [32, ch 5.5]), the maximum eigenvalue of $\hat{G}_1$ is the maximum root of the $(p+1)^{\text{st}}$ degree polynomial H_{p+1} or equivalently, the maximum eigenvalue of T_{p+1}. The minimum eigenvalue is identical to the maximum eigenvalue but with a sign change.

Now, if the basis functions have not been chosen to give $\hat{G}_1$ explicitly as a block tri-diagonal matrix, there exists a permutation matrix P_1 (corresponding to a reordering of the stochastic basis functions) such that $\tilde{G}_1 = P_1 \hat{G}_1 P_1^T$ is the block tri-diagonal matrix described above. The eigenvalues of $\tilde{G}_1$ are the same as those of $\hat{G}_1$. The same argument applies for the other matrices $\hat{G}_k = G_0^{-\frac{1}{2}} G_k G_0^{-\frac{1}{2}}, k = 2 : M$. There exists a permutation matrix P_k so that $\tilde{G}_k = P_k \hat{G}_k P_k^T$ is block tri-diagonal and whose extremal eigenvalues are given by those of T_{p+1}. $\square$

We now examine the eigenvalues of the matrices K_k coming from the spatial discretisation.

LEMMA 3.2. Let K_0 and K_k be the stiffness matrices defined in (3.4). If $c_k(\boldsymbol{x}) \geq 0$ where $\{\lambda_k, c_k(\boldsymbol{x})\}$ is the kth eigenpair of $\varrho(\boldsymbol{x}, \boldsymbol{y})$ then

$$0 \leq \frac{\sigma}{\mu} \sqrt{\lambda_k} c_k^{min} \leq \frac{\underline{x}^T K_k \underline{x}}{\underline{x}^T K_0 \underline{x}} \leq \frac{\sigma}{\mu} \sqrt{\lambda_k} c_k^{max} \quad \forall\, \underline{x} \in \mathbb{R}^{N_x},$$

where $c_k^{min} = \inf_{\boldsymbol{x} \in D} c_k(\boldsymbol{x})$ and $c_k^{max} = \sup_{\boldsymbol{x} \in D} c_k(\boldsymbol{x}) = \| c_k(\boldsymbol{x}) \|_\infty$. *Alternatively, if $c_k(\boldsymbol{x})$ is not uniformly positive,*

$$-\frac{\sigma}{\mu}\sqrt{\lambda_k} \| c_k(\boldsymbol{x}) \|_\infty \leq \frac{\underline{x}^T K_k \underline{x}}{\underline{x}^T K_0 \underline{x}} \leq \frac{\sigma}{\mu}\sqrt{\lambda_k} \| c_k(\boldsymbol{x}) \|_\infty \quad \forall\, \underline{x} \in \mathbb{R}^{N_x}.$$

Proof. Given any $\underline{x} \in \mathbb{R}^{N_x}$, define a function $v \in X_h$ via $v = \sum x_i \phi_i(\boldsymbol{x})$. If $c_k(\boldsymbol{x}) \geq 0$ then,

$$\underline{x}^T K_k \underline{x} = \int_D \sigma\sqrt{\lambda_k} c_k(\boldsymbol{x})\nabla v \cdot \nabla v\, dD \leq \frac{\sigma}{\mu}\sqrt{\lambda_k} c_k^{max} \int_D \mu\nabla v \cdot \nabla v\, dD = \frac{\sigma}{\mu}\sqrt{\lambda_k} c_k^{max} \underline{x}^T K_0 \underline{x}$$

$$\underline{x}^T K_k \underline{x} = \int_D \sigma\sqrt{\lambda_k} c_k(\boldsymbol{x})\nabla v \cdot \nabla v\, dD \geq \frac{\sigma}{\mu}\sqrt{\lambda_k} c_k^{min} \int_D \mu\nabla v \cdot \nabla v\, dD = \frac{\sigma}{\mu}\sqrt{\lambda_k} c_k^{min} \underline{x}^T K_0 \underline{x}.$$

If μ is positive, dividing through by the quantity $\underline{x}^T K_0 \underline{x}$ gives the first result. Now, if $c_k(\boldsymbol{x})$ also takes on negative values, we have

$$\left|\underline{x}^T K_k \underline{x}\right| = \sigma\sqrt{\lambda_k} \left|\int_D c_k(\boldsymbol{x})\, \nabla v \cdot \nabla v\, dD\right| \leq \frac{\sigma}{\mu}\sqrt{\lambda_k} \| c_k(\boldsymbol{x}) \|_\infty \underline{x}^T K_0 \underline{x}.$$

Arguing as in the first case gives the second result. $\square$

3.2. Preconditioning. When the global matrix A is symmetric and positive definite, we can use the conjugate gradient (CG) method as a solver. However, the system is ill-conditioned and a preconditioner is required. In [27] and [16] it is noted that if the variance of $K(\boldsymbol{x}, \omega)$ is small then the preconditioner P composed of the diagonal blocks of A, i.e,

$$P = G_0 \otimes K_0, \tag{3.10}$$

is heuristically the simplest and most appropriate choice. Working under the assumption that the variance is 'sufficiently small' is not suitable for some applications but the user is, in fact, already limited to this if employing Hermite polynomials in Gaussian random variables. In this section, we obtain a theoretical handle on these observations. First, we explain why preconditioning is required.

LEMMA 3.3. *If G_0 is defined as in (3.7) using Hermite polynomials in Gaussian random variables and piecewise linear (or bilinear) approximation is used for the spatial discretisation, on quasi-uniform meshes, the eigenvalues of $(G_0 \otimes K_0)$ lie in the interval, $[\mu\,\alpha_1 h^2, \mu\,\alpha_2\, p\,!]$ where μ is the mean value of $K(\boldsymbol{x}, \omega)$, p is the degree of stochastic polynomials, h is the characteristic spatial mesh-size and α_1, and α_2 are constants independent of h, M and p.*

Proof. If $\underline{v}_\xi$ is an eigenvector of G_0 with corresponding eigenvalue λ_ξ and $\underline{v}_x$ is an eigenvector of K_0 with corresponding eigenvalue λ_x then $(G_0 \otimes K_0)\left(\underline{v}_\xi \otimes \underline{v}_x\right) = \lambda_\xi \lambda_x \left(\underline{v}_\xi \otimes \underline{v}_x\right)$. Using (3.7), we deduce that $1 \leq \lambda_\xi \leq p\,!$. A bound for the eigenvalues of K_0 can be obtained in the usual way (eg see [8, pp.57–59]) to give,

$$\mu\alpha_1 h^2 \leq \frac{\underline{v}_x^T K_0 \underline{v}_x}{\underline{v}_x^T \underline{v}_x} \leq \mu\alpha_2 \quad \forall\, \underline{v}_x \in \mathbb{R}^{N_x}.$$

The result immediately follows. $\square$

REMARK 1. *Note that if the polynomial chaos basis functions are normalised with $\langle \psi_i, \psi_j \rangle = \delta_{ij}$ then the stochastic mass matrix is the identity matrix and the above eigenvalue bound is simply $[\mu\alpha_1 h^2, \mu\alpha_2]$ and is independent of p. It is always worthwhile normalising the basis functions for this reason. We shall assume that this is the case in the sequel.*

Now, we can expect the eigenvalues of the global unpreconditioned system matrix (3.5) to be a perturbation of the eigenvalues of $G_0 \otimes K_0$.

LEMMA 3.4. *If G_k is defined as in (3.6) and the stochastic basis functions are normalised multi-dimensional Hermite polynomials in Gaussian random variables, and piecewise linear (or bilinear)*

10

approximation is used for the spatial discretisation, on quasi-uniform meshes, then the eigenvalues of the global stiffness matrix A in (3.5) are bounded and lie in the interval, $[\mu\alpha_1 h^2 - \delta, \mu\alpha_2 + \delta]$, where,

$$\delta = \alpha_2 \sigma H_{p+1}^{max} \sum_{k=1}^{M} \sqrt{\lambda_k} \parallel c_k(\boldsymbol{x}) \parallel_\infty$$

where H_{p+1}^{max} is the maximum root of the Hermite polynomial of degree $p+1$, h is the spatial discretisation parameter and α_1 and α_2 are constants independent of h, M and p.

Proof. First note that the maximum and minimum eigenvalues ν_{max} and ν_{min} of

$$\left((G_0 \otimes K_0) + \sum_{k=1}^{M} (G_k \otimes K_k) \right) \underline{v} = \nu \underline{v},$$

can be bounded in terms of the maximum and minimum eigenvalues of the matrices in the sum. Assuming normalised stochastic basis functions, the matrices $\hat{G}_k$ in Lemma 3.1 are the same as the matrices G_k in (3.6). Hence, the eigenvalues of G_k belong to the symmetric interval $\left[-H_{p+1}^{max}, H_{p+1}^{max} \right]$ and, using a similar argument to that presented in Lemma 3.2, the eigenvalues of K_k, $k = 1 : M$, lie in the bounded interval,

$$\left[\sigma \sqrt{\lambda_k} c_k^{min} \alpha_1 h^2, \quad \sigma \sqrt{\lambda_k} c_k^{max} \alpha_2 \right] \qquad \text{if} \qquad c_k(\boldsymbol{x}) \geq 0,$$

$$\left[-\sigma \sqrt{\lambda_k} \parallel c_k(\boldsymbol{x}) \parallel_\infty \alpha_2, \quad \sigma \sqrt{\lambda_k} \parallel c_k(\boldsymbol{x}) \parallel_\infty \alpha_2 \right] \qquad \text{otherwise.}$$

Denoting the minimum and maximum eigenvalues of $(G_k \otimes K_k)$ by γ_{min}^k and γ_{max}^k respectively, and applying the result of Lemma 3.3 we have,

$$\nu_{min} \geq \mu\,\alpha_1\,h^2 + \sum_{k=1}^{M} \gamma_{min}^k, \quad \nu_{max} \leq \mu\,\alpha_2 + \sum_{k=1}^{M} \gamma_{max}^k.$$

Now, noting that the eigenvalues of the Kronecker product of two matrices are the products of the eigenvalues of the individual matrices we have, for any k,

$$\nu_{min} \geq \mu\,\alpha_1\,h^2 - \sigma\,\alpha_2\,H_{p+1}^{max} \sum_{k=1}^{M} \sqrt{\lambda_k} \parallel c_k(\boldsymbol{x}) \parallel_\infty, \quad \nu_{max} \leq \mu\,\alpha_2 + \sigma\,\alpha_2\,H_{p+1}^{max} \sum_{k=1}^{M} \sqrt{\lambda_k} \parallel c_k(\boldsymbol{x}) \parallel_\infty.$$

□

Using the preceding arguments, we can now establish a result that determines the efficiency of the chosen preconditioner.

THEOREM 1. *The eigenvalues $\{\nu_i\}$ of the generalised eigenvalue problem, $A\underline{x} = \nu P\underline{x}$ where the stochastic matrices G_k are defined as in (3.6), with respect to the Gaussian probability measure, using normalised multi-dimensional Hermite polynomials in Gaussian random variables lie in the interval $[1 - \tau, 1 + \tau]$ where*

$$\tau = \frac{\sigma}{\mu} H_{p+1}^{max} \sum_{k=1}^{M} \sqrt{\lambda_k} \parallel c_k(\boldsymbol{x}) \parallel_\infty, \tag{3.11}$$

σ and μ are the mean and standard deviation of $K(\boldsymbol{x}, \omega)$, $\{\lambda_k, c_k(\boldsymbol{x})\}$ are the eigenpairs of $\rho(\boldsymbol{x}, \boldsymbol{y})$ and H_{p+1}^{max} is the maximum root of the one-dimensional Hermite polynomial of degree $p+1$.

Proof. First note that the eigenvalues we are seeking satisfy $\nu = \theta + 1$ where,

$$\sum_{k=1}^{M} (G_0 \otimes K_0)^{-1} (G_k \otimes K_k) \underline{v} = \theta \underline{v}.$$

11

Hence, using standard properties of the matrix Kronecker product, and assuming normalised stochastic basis functions, we have

$$\sum_{k=1}^{M}\Big(G_k \otimes K_0^{-1}K_k\Big)\underline{v} = \theta\underline{v}.$$

Now, let $\hat{K}_k = K_0^{-1}K_k$. Applying Lemmas 3.2 and 3.1, the eigenvalues of G_k belong to the symmetric interval $\big[-H_{p+1}^{max}, H_{p+1}^{max}\big]$ and the eigenvalues of $\hat{K}_k$ belong to the interval

$$\left[\frac{\sigma}{\mu}\sqrt{\lambda_k}c_k^{min}, \frac{\sigma}{\mu}\sqrt{\lambda_k}c_k^{max}\right] \quad \text{or} \quad \left[-\frac{\sigma}{\mu}\sqrt{\lambda_k}\parallel c_k(\boldsymbol{x})\parallel_\infty, \frac{\sigma}{\mu}\sqrt{\lambda_k}\parallel c_k(\boldsymbol{x})\parallel_\infty\right],$$

depending on the positivity of $c_k(\boldsymbol{x})$. Proceeding as in Lemma 3.4, and denoting the minimum and maximum eigenvalues of $G_k \otimes \hat{K}_k$ by γ_{min}^k and γ_{max}^k, we have, in both cases,

$$\theta_{min} \geq \sum_{k=1}^{M}\gamma_{min}^k \geq -\sum_{k=1}^{M}H_{p+1}^{max}\frac{\sigma}{\mu}\sqrt{\lambda_k}\parallel c_k(\boldsymbol{x})\parallel_\infty, \ \theta_{max} \leq \sum_{k=1}^{M}\gamma_{max}^k \leq \sum_{k=1}^{M}H_{p+1}^{max}\frac{\sigma}{\mu}\sqrt{\lambda_k}\parallel c_k(\boldsymbol{x})\parallel_\infty .$$

The eigenvalues we need are the values $\nu_i = 1 + \theta_i, i = 1 : N_x N_\xi$. $\square$

REMARK 2. *As $\sigma \to 0$, the bound collapses to a single cluster at one. This is intuitively correct, since the off-diagonal blocks, which are not represented in the preconditioner, become insignificant. For increasing σ, the upper and lower bounds move away from one. The lower bound may be negative. Note that for σ too large, the condition (2.1) is violated for even low values of M and the unpreconditioned matrix A is not positive definite.*

REMARK 3. *The bound depends on the value H_{p+1}^{max}. Note then that it is not strictly independent of p and in fact grows at the same rate as the quantity $\sqrt{p-1} + \sqrt{p}$ with increasing p. Since the Hermite polynomials are defined on the interval $(-\infty, \infty)$, H_{p+1}^{max} can become arbitrarily large for p arbitrarily large. Note that p influences the solution u_{hp} via the expansion (3.2). Increasing p in effect increases the support of the pdf of u_{hp}. Choosing p too large amounts to allowing u_{hp} to be truly normally distributed at any given spatial location, which is not desirable. Since the stochastic basis functions are normalised, P does not improve the conditioning of A with respect to p.*

REMARK 4. *The bounds are pessimistic in M, due to the fact that we have bounded the maximum eigenvalue of a sum of matrices by the sum of the maximum eigenvalues of the individual matrices (and similarly with the minimum eigenvalue). The bound is sharp when $M = 1$ with any p, μ and σ (since then there is no sum) and are tighter in M when the eigenvalues decay rapidly or when σ is very small.*

For the case $p = 1$, a tighter bound (with respect to M) can be established.

THEOREM 2. *When $p = 1$, for any M, the eigenvalues $\{\nu_i\}$ of the generalised eigenvalue problem, $A\underline{x} = \nu P\underline{x}$ where A and P are as defined in (3.5) and (3.10), and the matrices G_k are defined as in (3.6) using normalised Hermite polynomials in Gaussian random variables, lie in the interval $[1-\tau, 1+\tau]$ where*

$$\tau = \frac{\sigma}{\mu}\left(\sum_{k=1}^{M}\lambda_k\parallel c_k(\boldsymbol{x})\parallel_\infty^2\right)^{\frac{1}{2}}, \tag{3.12}$$

σ and μ are the mean and standard deviation of $K(\boldsymbol{x}, \omega)$ and $\{\lambda_k, c_k(\boldsymbol{x})\}$ are the eigenpairs of $\rho(\boldsymbol{x}, \boldsymbol{y})$.

Proof. When $p = 1$, each $\hat{G}_k$ is a permutation of an $(M + 1) \times (M + 1)$ block tri-diagonal matrix G_* with leading block

$$T_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

12

and all remaining rows and columns filled with zeros. Hence, following the proof of Theorem 1,

$$\sum_{k=1}^{M} \hat{G}_k \otimes \hat{K}_k = \begin{pmatrix} 0 & \hat{K}_M & \dots & \hat{K}_1 \\ \hat{K}_M & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \hat{K}_1 & 0 & \dots & 0 \end{pmatrix}$$

(or some block permutation thereof). It is then a trivial task to show that the eigenvalues of this sum are either 0, or $\pm\sqrt{\lambda(\sum_{k=1}^{M} \hat{K}_k^2)}$, or, in other words, the eigenvalues of the matrix:

$$\hat{G}_* \otimes \left(\sum_{k=1}^{M} \hat{K}_k^2 \right)^{\frac{1}{2}}$$

(since the eigenvalues of G_* are $-1, 0, 1$). Using the result of Lemma 3.2, noting that the eigenvalues of $\hat{K}_k^2$ are non-negative, and denoting the maximum eigenvalue of $\hat{K}_k$ by γ_{max}^k we have,

$$\theta_{max} \leq \left(\sum_{k=1}^{M} \left(\gamma_{max}^k \right)^2 \right)^{\frac{1}{2}} \leq \frac{\sigma}{\mu} \left(\sum_{k=1}^{M} \lambda_k \parallel c_k(\boldsymbol{x}) \parallel_\infty^2 \right)^{\frac{1}{2}},$$

$$\theta_{min} \geq - \left(\sum_{k=1}^{M} \left(\gamma_{max}^k \right)^2 \right)^{\frac{1}{2}} \geq -\frac{\sigma}{\mu} \left(\sum_{k=1}^{M} \lambda_k \parallel c_k(\boldsymbol{x}) \parallel_\infty^2 \right)^{\frac{1}{2}}.$$

The eigenvalues we need are the values $\nu_i = 1 + \theta_i, i = 1 : N_x N_\xi$. $\square$

REMARK 5. *The above bound is tighter with respect to M as the sequence $\lambda_1, \lambda_2, \dots$ decays more rapidly than the sequence $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots$. Unfortunately, for other values of p there is no nice representation for the sum of matrices in the form $G_* \otimes X$ for some matrix X that is easy to handle.*

We now explore the accuracy of the bounds. In each example below, we list the computed extremal eigenvalues of $P^{-1}A$ and bounds on those eigenvalues calculated using Theorems 1 and 2. For the stochastic basis, we employ Hermite polynomials in Gaussian random variables.

Example 1 We consider first the case where the covariance function is (2.11) with $\sigma = 0.1$, $\mu = 1$, and $c_1 = 1 = c_2$ and $h = \frac{1}{8}$.

M	p	$\nu_{min}(P^{-1}A)$	$\nu_{max}(P^{-1}A)$	**Bounds**	H_{p+1}^{max}
1	1	0.9155	1.0845	[0.9151, 1.0849]	1
-	2	0.8537	1.1463	[0.8529, 1.1471]	1.7321
-	3	0.8028	1.1972	[0.8017, 1.1983]	2.3344
-	4	0.7586	1.2414	[0.7573, 1.2427]	2.8570
2	1	0.9125	1.0875	[0.9037, 1.0963]	1
-	2	0.8485	1.1515	[0.7743, 1.2257]	1.7321
-	3	0.7959	1.2041	[0.6958, 1.3042]	2.3344
-	4	0.7502	1.2498	[0.6277, 1.3723]	2.8570
3	1	0.9107	1.0893	[0.8935, 1.1065]	1
-	2	0.8453	1.1547	[0.6957, 1.3043]	1.7321
-	3	0.7915	1.2085	[0.5899, 1.4101]	2.3344
-	4	0.7449	1.2551	[0.4981, 1.5019]	2.8570

TABLE 3.1
Example 1: Extremal eigenvalues of $P^{-1}A$ and bounds on extremal eigenvalues of $P^{-1}A$.

Example 2 Next, we consider the same example but with a very small standard deviation $\sigma = 0.01$.

M	p	$\nu_{min}(P^{-1}A)$	$\nu_{max}(P^{-1}A)$	**Bounds**	H_{p+1}^{max}
1	1	0.9916	1.0170	[0.9915, 1.0085]	1
-	2	0.9854	1.0146	[0.9853, 1.0147]	1.7321
-	3	0.9803	1.0197	[0.9802, 1.0198]	2.3344
-	4	0.9759	1.0241	[0.9757, 1.0243]	2.8570
2	1	0.9913	1.0176	[0.9904, 1.0096]	1
-	2	0.9849	1.0151	[0.9774, 1.0226]	1.7321
-	3	0.9796	1.0204	[0.9696, 1.0304]	2.3344
-	4	0.9750	1.0250	[0.9628, 1.0372]	2.8570
3	1	0.9911	1.0089	[0.9893, 1.0107]	1
-	2	0.9845	1.0155	[0.9696, 1.0304]	1.7321
-	3	0.9792	1.0208	[0.9590, 1.0410]	2.3344
-	4	0.9745	1.0255	[0.9498, 1.0502]	2.8570

TABLE 3.2

Example 2: Extremal eigenvalues of $P^{-1}A$ and bounds on extremal eigenvalues of $P^{-1}A$.

Example 3 Observe what happens when we use a large standard deviation $\sigma = \mathbf{0.3}$, and increase p, the stochastic polynomial degree. In this example, we also list the extremal eigenvalues of A.

M	p	$\nu_{min}(A)$	$\nu_{max}(A)$	$\nu_{min}(P^{-1}A)$	$\nu_{max}(P^{-1}A)$	**Bounds**	H_{p+1}^{max}
1	3	0.1431	5.9508	0.4083	1.5917	[0.4051, 1.5949]	2.3344
-	4	0.1080	6.4326	0.2758	1.7242	[0.2720, 1.7280]	2.8570
-	5	0.0756	6.8636	0.1574	1.8426	[0.1529, 1.8471]	3.3243
-	6	0.0427	7.2569	0.0493	1.9507	[0.0443,1.9557]	3.7504
-	**7**	**-0.1545**	**7.6206**	**-0.0506**	**2.0506**	**[-0.0561, 2.0561]**	4.1445
-	**8**	**-0.4640**	**7.9605**	**-0.1439**	**2.1439**	**[-0.1500, 2.1500]**	4.5127
2	3	0.1410	6.0175	0.3876	1.6124	[0.0874, 1.9126]	2.3344
-	4	0.1052	6.5085	0.2505	1.7495	[-0.1169, 2.1169]	2.8570
-	5	0.0717	6.9464	0.1279	1.8721	[-0.2996, 2.2996]	3.3243
-	6	0.0333	7.3450	0.0161	1.9839	[-0.4662, 2.4662]	3.7504
-	**7**	**-0.2725**	**7.7130**	**-0.0873**	**2.0873**	**[-0.6202,2.6202]**	4.1445
-	**8**	**-0.5972**	**8.0563**	**-0.1838**	**2.1838**	**[-0.7642, 2.7642]**	4.5127

TABLE 3.3

Example 3: Extremal eigenvalues of $P^{-1}A$ and bounds on extremal eigenvalues of $P^{-1}A$.

Thus, it can be seen that when Hermite polynomials (with infinite support) are employed, for fixed values of h, M and σ, we can always find a value of p, that causes the system matrix A, and the preconditioned system matrix $P^{-1}A$ to be indefinite. The eigenvalue bounds in Theorems 1 and 2 predict this. Figure 3.2 summaries this for the case $M = 1$.

Example 4 Finally, consider the case where the covariance function is (2.11) with $\sigma = 0.1$, $\mu = 1$, and $c_1 = 10 = c_2$ and $h = \frac{1}{8}$. Here, the eigenvalues decay more quickly than in the first two examples. Eigenvalues of the preconditioned system are listed in Table 3.4.

In all cases, the extremal eigenvalues of $P^{-1}A$ exhibit the behaviour anticipated by the bounds in Theorems 1 and 2. They are symmetric about one, increase very slightly with p and retract to one for small variance. For small values of σ, the dependence on p is not evident. These results,
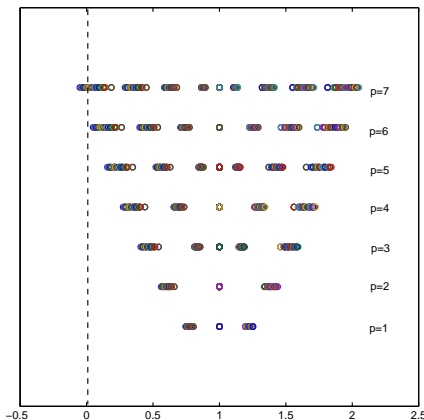
FIG. 3.2. $h = \frac{1}{8}$ Example 3: Eigenvalues of $P^{-1}A$, $\sigma = 0.3$, $M = 1$, **varying p**

M	p	$\nu_{min}(P^{-1}A)$	$\nu_{max}(P^{-1}A)$	**Bounds**	H^{max}_{p+1}
1	1	0.9017	1.0983	[0.9017, 1.0983]	1
-	2	0.8298	1.1702	[0.8297, 1.1703]	1.7321
-	3	0.7706	1.2294	[0.7704, 1.2296]	2.3344
-	4	0.7192	1.2808	[0.7190, 1.2810]	2.8570
2	1	0.9014	1.0986	[0.8998, 1.1002]	1
-	2	0.8291	1.1709	[0.7961, 1.2039]	1.7321
-	3	0.7697	1.2303	[0.7252, 1.2748]	2.3344
-	4	0.7182	1.2818	[0.6637, 1.3363]	2.8570
3	1	0.9010	1.0990	[0.8979, 1.1021]	1
-	2	0.8286	1.1714	[0.7626, 1.2374]	1.7321
-	3	0.7689	1.2311	[0.6800, 1.3200]	2.3344
-	4	0.7172	1.2828	[0.6084, 1.3916]	2.8570

TABLE 3.4
Example 4: Extremal eigenvalues of $P^{-1}A$ and bounds on extremal eigenvalues of $P^{-1}A$.

together with Theorem 1 give us confidence that the mean-based block diagonal preconditioner has sound theoretical justification provided that the variance and the polynomial degree are not too large. Now, we turn to the question of implementation and focus on cases where A is positive definite.

4. Numerical Results. In this section, we present iteration counts and timings for two test problems. We implement block-diagonal preconditioning with CG. The theoretical results above tell us that we can expect the iteration count to be independent of the spatial discretisation parameter, h, and almost independent of p (polynomial degree) and M (KL terms). It is required, however, in each CG iteration, to approximate the quantity $P^{-1}\underline{r}$, where $\underline{r}$ is a residual error vector. Applying the preconditioner therefore requires N_ξ approximate solutions of subsidiary systems with coefficient matrix K_0. The number of subproblems can be very large for increasing M and p. (See Table 4.1 for details.) Fortunately, approximately inverting each of the diagonal blocks of the preconditioner is equivalent to solving a standard diffusion problem. Exact solves are too costly for highly refined spatial meshes. However, we can benefit from our experience of solving deterministic problems by replacing the exact solves for K_0 with either an incomplete factorisation preconditioner (see [27] and [16]) or with a multigrid V-cycle. In fact, *any fast solver for a Poisson problem is a potential candidate*. Moreover, the N_ξ approximate solves required at each CG iteration are independent of one

15

another and can be performed in parallel. Crucially, set-up of the approximation to, or factorisation of K_0, only needs to be performed once.

p	$M = 2$	4	6	8	10	15	20	30
1	3	5	7	9	11	16	21	31
2	6	15	28	45	66	136	231	992
3	10	35	84	165	286	816	1,771	32,736

TABLE 4.1
Values of N_ξ, dimension of stochastic basis

Below, we implement the preconditioner using both incomplete Cholesky factorisation and one V-cycle of AMG with symmetric Gauss-Seidel (SGS) smoothing to approximately invert K_0. The latter method has the key advantage that the computational cost grows linearly in the problem size. Our particular AMG code (see [30]) is implemented in MATLAB and based on the traditional Ruge-Stüben algorithm (see [28]). No parameters are tuned. We apply the method as a black-box in each experiment. Other ways of incorporating geometric multigrid into fast solvers for these systems are discussed in [7] and [23]. All iterations are terminated when the relative residual error, measured in the Euclidean norm, is reduced to 10^{-10}. All computations are performed in serial using a MATLAB 7.3 on a standard laptop PC with 512MB of RAM.

4.1. Homogeneous Dirichlet boundary condition. First, we reproduce an experiment performed in [3]. The chosen covariance function is (2.11) with $c_1 = 1 = c_2$, standard deviation $\sigma = 0.1$ and mean $\mu = \langle K(\boldsymbol{x})\rangle = 1$. We solve (1.2) on $D = [-0.5, 0.5] \times [-0.5, 0.5]$ with homogenous Dirichlet boundary condition and $f = 2(0.5 - x^2 - y^2)$. Post-processing the coefficient blocks of the solution in the spectral expansion (3.2) to recover the mean and variance of the solution is trivial. Solutions obtained on a 32×32 uniform spatial grid are plotted in Fig. 4.1.
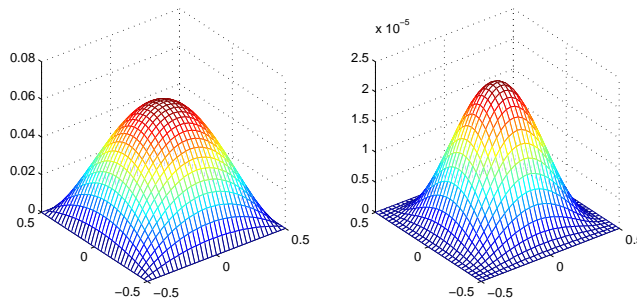

FIG. 4.1. *Mean (left) and variance(right) of pressure on a 32×32 mesh for the case $M = 4$ with $p = 2$*

The maximum values of the mean and variance obtained with $h^{-1} = 16$, $p = 4$ and $M = 6$ are 0.063113 and 2.3600e-05 respectively. Using the SFEM, a *single* system of dimension $15^2 \times 210$ is solved. In Table 4.2 we record the maximum values of the estimated mean and variance of the pressure solution obtained using a traditional Monte Carlo method, with N realisations of $K(\boldsymbol{x}, \omega)$. The random field inputs were generated using the circulant embedding method [4]. Each subproblem requires the solution of a small system of dimension 15^2, but more than $10,000$ systems need to be solved in order to obtain comparable accuracy with the SFEM results.

In Table 4.3 we record iteration counts and timings for preconditioned CG applied to the SFEM systems, with varying h, M and p. Now we can compare implementations based on incomplete Cholesky factorisation and on our suggested algebraic multigrid solver. Note that the performance of the former is sensitive to the choice of drop tolerance parameter and we have not sought to optimise this. The black-box AMG version of the preconditioning scheme proved to be optimal with respect to the spatial discretisation without tuning any parameters. Indeed, the matrix V corresponding to a single V-cycle of the AMG algorithm is a spectrally equivalent approximation to K_0 (see Table 4.4).

16

	$N = 100$	1,000	10,000	40,000
Max(sample mean)	0.063608	0.063299	0.063127	0.063134
Max(sample variance)	2.1611e-05	2.4065e-05	2.2584e-05	2.3160e-05

<div align="center">TABLE 4.2</div>

Maximum values of sample mean and standard deviation after N realisations

The maximum eigenvalue of $V^{-1}K_0$ is one, independently of h. The efficiency of this approximation is completely unaffected by the choice of p, M and standard deviation σ.

M=4				
Preconditioner	h	$p = 2$	3	4
None	$\frac{1}{4}$	19	30	55
	$\frac{1}{8}$	43	73	139
	$\frac{1}{16}$	92	161	314
	$\frac{1}{32}$	188	335	666
Block-diagonal	$\frac{1}{16}$	10 (0.00 + 0.33)	11 (0.00 + 1.05)	12 (0.00 + 2.88)
(`cholinc`, $1e^{-3}$)	$\frac{1}{32}$	11 (0.02 + 1.23)	13 (0.01 + 4.24)	15 (0.01 + 9.42)
	$\frac{1}{64}$	21 (0.10 + 9.68)	22 (0.10 + 30.33)	24 (0.09 + 67.74)
	$\frac{1}{128}$	38 (0.61 + 91.44)	42 (0.61 + 272.17)	45 (0.61 + 613.92)
Block-diagonal	$\frac{1}{16}$	10 (0.06 + 0.53)	12 (0.06 + 1.76)	13 (0.13 + 4.50)
(`amg`)	$\frac{1}{32}$	11 (0.20 + 1.60)	12 (0.20 + 5.71)	13 (0.29 + 13.41)
	$\frac{1}{64}$	11 (0.88 + 6.38)	12 (0.99 + 20.50)	13 (0.96 + 47.15)
	$\frac{1}{128}$	12 (6.72 + 36.40)	13 (6.84 + 104.64)	14 (5.18 + 233.44)
M=6				
Preconditioner	h	$p = 2$	3	4
None	$\frac{1}{4}$	19	29	55
	$\frac{1}{8}$	44	76	146
	$\frac{1}{16}$	93	169	332
	$\frac{1}{32}$	190	350	702
Block-diagonal	$\frac{1}{16}$	10 (0.07 + 0.90)	11 (0.00 + 4.11)	12 (0.00 + 17.21)
(`cholinc`, $1e^{-3}$)	$\frac{1}{32}$	13 (0.01 + 3.98)	13 (0.00 + 14.51)	15 (0.01 + 44.00)
	$\frac{1}{64}$	21 (0.09 + 24.17)	23 (0.10 + 91.21)	24 (0.60 + 242.06)
	$\frac{1}{128}$	38 (0.61 + 240.10)	42 (0.68 + 876.49)	46 (0.60 + 2610.48)
Block-diagonal	$\frac{1}{16}$	11 (0.06 + 1.39)	12 (0.06 + 6.06)	13 (0.06 + 24.02)
(`amg`)	$\frac{1}{32}$	11 (0.20 + 4.37)	12 (0.20 + 16.66)	13 (0.20 + 51.93)
	$\frac{1}{64}$	11 (0.88 + 15.67)	13 (0.98 + 58.32)	14 (6.75 + 180.48)
	$\frac{1}{128}$	12 (6.78 + 89.66)	13 (6.75 + 310.61)	14 (6.75 + 886.46)

<div align="center">TABLE 4.3</div>

Preconditioned CG iterations and timings in seconds (set-up + total iteration times)

The efficiency of both implementations of the block-diagonal preconditioner deteriorates with increasing variance. This is due to the fact that as σ increases, the off-diagonal blocks of A become more significant and they are not represented in the preconditioner. Choosing σ to be too large also causes A to become indefinite. In that case, CG breaks down. Iteration counts, for the exact version of the block-diagonal preconditioner, for fixed M and h and varying σ are listed in Table 4.5. Dimensions of the global systems for the problems considered are summarised in Table 4.6. Observe then that using our multigrid method, we can solve more than 3.5 million equations on a laptop PC in under 15 minutes. Furthermore, it should be noted that multigrid algorithms haver lower memory requirements than incomplete factorisation methods, even with optimised parameters.

h	$p = 2$	3	4
$\frac{1}{4}$	0.9996	0.9996	0.9996
$\frac{1}{8}$	0.9882	0.9882	0.9882
$\frac{1}{16}$	0.9707	0.9707	0.9707
$\frac{1}{32}$	0.9525	0.9525	0.9252

TABLE 4.4

$M = 4$ Minimum eigenvalue of $V^{-1}K_0$ where V is one V-cycle of AMG (with SGS smoothing)

σ	p=2	3	4
0.1	8	10	11
0.2	11	14	17
0.3	14	21	30
0.4	18	35	532

TABLE 4.5

CG iteration counts with exact block-diagonal preconditioning, $h = \frac{1}{16}$, $M = 4$

	h	$p = 2$	3	4		h	$p = 2$	3	4
M=4	$\frac{1}{16}$	4,335	10,115	20,230	M=6	$\frac{1}{16}$	8,092	24,276	60,690
	$\frac{1}{32}$	16,335	38,115	76,230		$\frac{1}{32}$	30,492	91,476	228,690
	$\frac{1}{64}$	63,375	147,875	295,750		$\frac{1}{64}$	121,968	365,904	887,250
	$\frac{1}{128}$	249,615	582,435	1,116,870		$\frac{1}{128}$	465,948	1,397,844	3,494,610

TABLE 4.6

Dimension of global stiffness matrix

4.2. Mixed boundary conditions. Next we consider steady flow from left to right on the domain $D = [0,1] \times [0,1]$ with $f = 0$, $\partial D_D = \{0,1\} \times [0,1]$ and $\partial D_N = \partial D \backslash \partial D_D$. We set $\vec{q} \cdot \vec{n} = 0$ at the two horizontal walls so that flow is tangent to those boundaries. The Dirichlet data is $u = 1$ on $\{0\} \times [0,1]$ and $u = 0$ on $\{1\} \times [0,1]$. Again, we employ the covariance function (2.11) with $c_1 = 1 = c_2$, $\sigma = 0.1$, and $\mu = 1$. The mean and variance of the primal variable, obtained on a 32×32 uniform grid using four terms in the KL expansion of $K(\boldsymbol{x}, \omega)$ and quadratic Hermite polynomial chaos functions for the stochastic discretisation, are plotted in Fig. 4.2. Preconditioned CG iteration counts and timings are recorded in Table 4.7. Our multigrid-based implementation of the preconditioner performs as in the previous example. That is, convergence is identical to that of the exact implementation, insensitive to M and h and only very slightly dependent on p. The efficiency deteriorates for increasing standard deviation, σ.

5. Conclusions. The focus of this work was the design of a fast and robust solver for the model elliptic stochastic boundary value problem (1.2). Our goals were to provide a theoretical basis for a simple, popular preconditioning scheme employed by other authors and suggest a practical, efficient implementation based on multigrid. We described the classical spectral SFEM discretisation, outlined the structure of the resulting symmetric linear systems and established a lower bound for the minimum eigenvalue of the system matrix in the case that Gaussian random variables are employed to represent the diffusion coefficient. We analysed the exact block-diagonal preconditioner proposed in [16], based on the mean component of the system matrix, and established an eigenvalue bound for the preconditioned system. Those eigenvalues are independent of h but depend on p and σ. The bound is slightly pessimistic in M, the number of terms retained in the truncated Karhunen-Loève expansion of $K(\boldsymbol{x}, \omega)$, but the dependence on all other SFEM parameters is sharp. We tested the robustness of the preconditioner, with approximate solves for the mean stiffness matrix computed via incomplete Cholesky factorisation and using a V-cycle of black-box algebraic multigrid. The black-box multigrid scheme was robust with respect to the spatial discretisation parameter without
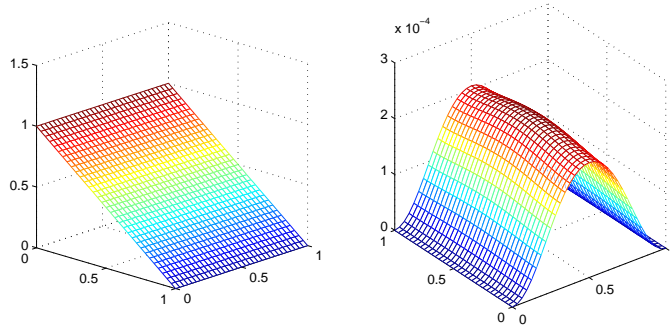
FIG. 4.2. *Mean (left) and variance (right) of pressure on a 32×32 mesh for the case $M = 4$ with $p = 2$*

M=4 Preconditioner	h	$p = 2$	3	4
None	$\frac{1}{4}$	39	61	118
-	$\frac{1}{8}$	73	129	247
-	$\frac{1}{16}$	139	246	498
-	$\frac{1}{32}$	266	485	984
Block-diagonal	$\frac{1}{16}$	10 (0.00 + 0.45)	11 (0.00 + 1.11)	11 (0.00 + 2.75)
(`cholinc`, $1e^{-3}$)	$\frac{1}{32}$	13 (0.02 + 1.55)	14 (0.18 + 4.71)	15 (0.02 + 10.97)
	$\frac{1}{64}$	23 (0.10 + 11.63)	24 (0.12 + 32.78)	26 (0.11 + 76.11)
	$\frac{1}{128}$	42 (0.66 + 106.59)	45 (0.66 + 284.94)	49 (0.66 + 650.37)
Block-diagonal	$\frac{1}{16}$	10 (0.07 + 0.75)	11 (0.07 + 1.89)	12 (0.07 + 4.70)
(`amg`)	$\frac{1}{32}$	10 (0.27 + 1.95)	11 (0.21 + 5.63)	12 (0.21 + 12.25)
	$\frac{1}{64}$	10 (0.91 + 6.36)	11 (1.00 + 18.63)	12 (0.91 + 40.64)
	$\frac{1}{128}$	10 (5.25 + 32.62)	11 (6.89 + 84.65)	12 (6.84 + 193.24)
M=6 Preconditioner	h	$p = 2$	3	4
None	$\frac{1}{4}$	40	68	128
-	$\frac{1}{8}$	75	138	270
-	$\frac{1}{16}$	142	264	533
-	$\frac{1}{32}$	273	511	1029
Block-diagonal	$\frac{1}{16}$	10 (0.00 + 0.88)	11 (0.00 + 4.42)	11 (0.00 + 16.51)
(`cholinc`, $1e^{-3}$)	$\frac{1}{32}$	13 (0.02 + 3.44)	14 (0.18 + 16.69)	15 (0.02 + 53.79)
	$\frac{1}{64}$	22 (0.11 + 26.93)	24 (0.11 + 103.18)	26 (0.11 + 307.86)
	$\frac{1}{128}$	42 (0.66 + 260.22)	45 (0.66 + 877.54)	48 (0.66 + 2566.47)
Block-diagonal	$\frac{1}{16}$	10 (0.07 + 1.54)	11 (0.07 + 6.40)	12 (0.07 + 23.21)
(`amg`)	$\frac{1}{32}$	10 (0.22 + 4.53)	11 (0.22 + 16.91)	12 (0.22 + 52.48)
	$\frac{1}{64}$	10 (1.00 + 15.34)	11 (0.92 + 54.60)	12 (1.02 + 166.29)
	$\frac{1}{128}$	10 (6.90 + 733.30)	11 (7.05 + 270.38)	12 (6.74 + 767.19)

TABLE 4.7

Preconditioned CG *iterations and timings in seconds (set-up + total iteration times)*

tuning any parameters. It also has lower memory requirements than factorisation methods for fine spatial meshes.

REFERENCES

[1] I. Babuška and P. Chatzipantelidis, On solving elliptic stochastic partial differential equations, *Comput. Methods Appl. Mech. Engrg.,* 191 (2002), pp.4093–4122.

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

[3] M.K. Deb, I.M. Babuška and J.T. Oden, Solution of stochastic partial differential equations using Galerkin finite element techniques, *Comput. Methods Appl. Mech. Engrg.,* 90, 48(2001), pp.6359-6372.

[4] C.R. Dietrich and G. N. Newsam, Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix, *SIAM J. Sci. Comput.,* 18, 4(1997), pp.1088–1107.

[5] M. Eiermann, O.G. Ernst and E. Ullmann, Computational aspects of the stochastic finite element method, *Proceedings of ALGORITHMY 2005*, pp.1–10, 2005.

[6] H. Elman, O.G. Ernst, D.P. O'Leary and M. Stewart, Efficient iterative algorithms for the stochastic finite element method with application to acoustic scattering, *Comput. Methods Appl. Mech. and Engrg,* 18(2005), pp.1037–1055.

[7] H. Elman and D. Furnival, Solving the stochastic steady-state diffusion problem using multigrid, *IMA J. Numer. Anal.,* To appear, 2007.

[8] H. Elman, D. Silvester and A.Wathen, Finite Elements and Fast Iterative Solvers, Oxford University Press, 2005.

[9] Ewing R.E., Wheeler M.F., Computational aspects of mixed finite element methods. In: R. STEPLEMAN, ed. *Numerical methods for Scientific Computing,* North-Holland, 163–172, 1983.

[10] Ferum user's guide, `http://www.ce.berkeley.edu/ ~ haukaas/FERUM/User_s_Guide/user_s_guide.html`

[11] P. Frauenfelder, C. Schwab and R.A. Todor, Finite elements for elliptic problems with stochastic coefficients, *Comput. Methods Appl. Mech. and Engrg,* 194(2005), pp.205–228.

[12] R.G. Ghanem, Probabilistic characterization of transport in heterogeneous media, *Comput. Methods Appl. Mech. Engrg.,* 158(1998), pp.199-220.

[13] R.G. Ghanem, Hybrid stochastic finite elements and generalized monte carlo simulation, *Journal of Applied Mechanics - Transactions of the ASME,* 65, 4(1998), pp.1004-1009.

[14] R.G. Ghanem, Ingredients for a general purpose stochastic finite elements implementation, *Comput. Methods Appl. Mech. Engrg.,* 168(1999), pp.19-34.

[15] R.G. Ghanem and S. Dham, Stochastic finite element analysis for multiphase flow in heterogeneous porous media, *Transport in Porous Media,* 32(1998), pp.239-262.

[16] R.G. Ghanem and R.M. Kruger, Numerical solution of spectral stochastic finite element systems, *Comput. Methods Appl. Mech. Engrg.,* 129(1996), pp.289-303.

[17] R.G. Ghanem and P.D. Spanos, Stochastic Finite Elements A Spectral Approach, Dover Publications, 2003.

[18] G.H. Golub and J.H. Welsch, Calculation of Gauss Quadrature Rules, *Math. Comp.,* 23, 106(1969), pp.221–230.

[19] A. Keese, A review of recent developments in the numerical solution of stochastic partial differential equations (stochastic finite elements), *Technical Report 2003-6,* Institute of scientific computing, Technical University Braunschweig, Germany, 2003.

[20] A. Keese, Numerical solution of systems with stochastic uncertainties: a general purpose framework for stochastic finite elements, PhD. thesis, TU Braunschweig, Germany, Fachbereich Mathematik und Informatik, 2004.

[21] A. Keese and H.G. Matthies, Efficient solvers for nonlinear stochastic problems, *Proceedings of the NIC Symposium, Julich,Germany,* 2004.

[22] A. Keese and H.G. Matthies, Parallel computation of stochastic groundwater flow, *Fifth world congress on computational mechanics, Vienna, Austria,* 2002.

[23] O.P. Le Maitre, O.M. Knio, B.J. Debusschere, H.N. Najm, and R.G. Ghanem, A multigrid solver for two-dimensional stochastic diffusion equations, *Comput. Methods Appl. Mech. Engrg.,* 192 (2003), pp.4723–4744.

[24] M. Loève, Probability Theory, New York, 1960.

[25] H.G. Matthies and C. Bucher, Finite elements for stochastic media problems, *Comput. Methods Appl. Mech. Engrg.,* 168(1999), pp.3–17.

[26] H.G. Matthies, C.E. Brenner, C. Bucher and C.G. Soares, Uncertainties in probabilistic numerical analysis of structures and solids-stochastic finite elements, *Structural Safety,* 18, 3(1997), pp.283–336.

[27] M.F. Pellissetti and R.G. Ghanem, Iterative solution of systems of linear equations arising in the context of stochastic finite elements, *Advances in Engineering Software.* 313, (2000), pp.607–616.

[28] J.W. Ruge and K. Stüben, *Efficient solution of finite difference and finite element equations by algebraic multigrid (AMG).* In: Multigrid Methods for Integral and Differential Equations, The Institute of Mathematics and its Applications Conference Series, D.J. Paddon, H. Holstein (Eds.), New Series 3, Clarendon Press, Oxford, 1985, pp.169–212.

[29] T.F. Russell and M.F. Wheeler, Finite element and finite difference methods for continuous flows in porous media. In: R.E. EWING, ed. *The Mathematics of Reservoir Simulation.* Philadelphia: SIAM, 35–106, 1983.

[30] D.J. Silvester and C.E. Powell, PIFISS Potential (Incompressible) Flow & Iterative Solution Software guide. *MIMS technical report 2007.14,* University of Manchester, UK, (2007).

[31] P. Frauenfelder, C. Schwab and R.A. Todor, Finite elements for elliptic problems with stochastic coefficients, *Comput. Methods Appl. Mech. Engrg.,* 194, (2005), pp.205–228.

[32] J. Stoer and R. Bulirsch, Introduction to Numerical Anlaysis, Springer-Verlag, New York, 1980.

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

[34] D. Xiu and G.E. Karniadakis, Modeling uncertainty in steady state diffusion problems via generalised polynomial chaos, *Comput. Methods Appl. Mech. Engrg.,* 191(2003), pp.4927–4948.