KRYLOV SUBSPACE RECYCLING METHODS IN STOCHASTIC FINITE ELEMENT COMPUTATIONS *

E. ULLMANN †

Abstract. Certain discretizations of second-order elliptic stochastic partial differential equations in primal or mixed formulation, where the diffusion coefficient is modelled via a truncated Karhunen-Loève expansion, result in a sequence of linear systems. We employ Krylov subspace recycling methods that utilize information generated during the solution of a previous system in order to improve the convergence behaviour of the next system. In particular we discuss the ordering of the (independent) systems. In addition we report on our experiences when using Krylov subspace recycling in combination with different preconditioners.

Keywords. stochastic finite element method, Krylov subspace recycling

1. Introduction. The solution of partial differential equations (PDEs) with random input data by Stochastic Finite Element methods (SFEMs) for uncertainty quantification has attracted much interest in the last few years. However, the size of the resulting Galerkin systems is equal to the number of deterministic degrees of freedom (DOFs) times the number of stochastic DOFs.

For an important special case the large Galerkin matrix can be decoupled. Then, the task is to solve a sequence of linear systems where each system matrix is a linear combination of a small number of finite element stiffness matrices. Therefore we employ preconditioned Krylov subspace recycling iterative solvers for each linear system in the sequence. These methods have already been studied in [8] and [15] for the primal SFEM discretization of second-order elliptic PDEs. In [15] the authors partition the sequences into several groups and apply the GCRO-DR method [22] together with additive Schwarz domain decomposition preconditioners within each group. In [8] we use GCROT with recycling [22] and GCRO-DR together with *one* preconditioner based on an incomplete Cholesky decomposition of the mean stiffness matrix. No particular ordering of the linear systems has been applied. The question of how a more powerful preconditioner works in combination with recycling and the application of recycling to mixed SFEM discretizations are the motivations of this work.

Consider the second-order elliptic boundary value problem

$$\begin{aligned} -\nabla\cdot(T\,\nabla p) &= F && \text{in } D\subset\mathbb{R}^2,\\ p &= g && \text{on } \partial D_D\neq\emptyset,\\ \boldsymbol{n}\cdot(T\,\nabla p) &= 0 && \text{on } \partial D_N. \end{aligned} \tag{1.1}$$

This steady-state diffusion problem can be reformulated by introducing the variable $\boldsymbol{u}=-T\,\nabla p$ as

$$\begin{aligned} T^{-1}\boldsymbol{u}+\nabla p &= 0,\\ \nabla\cdot\boldsymbol{u} &= F && \text{in } D,\\ p &= g && \text{on } \partial D_D,\\ \boldsymbol{n}\cdot\boldsymbol{u} &= 0 && \text{on } \partial D_N. \end{aligned} \tag{1.2}$$

In the context of groundwater flow modelling the variable p is the hydraulic head and $\boldsymbol{u}$ is the volumetric flux, respectively. We refer to (1.1) as the *primal formulation* and to (1.2) as the *mixed formulation* of a second-order elliptic boundary value problem.

Since in many applications only limited information about the diffusion coefficient T is available, we assume $T=T(\boldsymbol{x},\omega)$ to be a random field, i.e., a family of random variables $T(\boldsymbol{x},\cdot)$ with index variable $\boldsymbol{x}\in\overline{D}$. Each random variable takes on values in $\mathbb{R}$ and is defined on a probability space $(\Omega,\mathfrak{A},P)$, where Ω denotes the set of elementary events, $\mathfrak{A}$ is a σ-algebra on Ω and P is a probability measure. In addition we assume that the random field T is bounded and strictly positive, that is

$$0<T_\ell\leq T(\boldsymbol{x},\omega)\leq T_u<\infty \quad \text{a.e. in } D\times\Omega. \tag{1.3}$$

*This research was supported by grant 1279 from the British Council and grant D/06/12721 from the DAAD, as part of the British-German Academic Research Collaboration Programme.

†Institut für Numerische Mathematik und Optimierung, Technische Universtität Bergakademie Freiberg, D-09596 Freiberg, Germany

Note that this assumption also implies the reciprocal field T^{-1} to be bounded and strictly positive. A consequence of the diffusion coefficient's randomness is that the output variables p and eventually $\boldsymbol{u}$ are random fields as well. The primal formulation (1.1) transforms to the problem of finding a random field $p = p(\boldsymbol{x}, \omega)$ such that, P-almost surely,

$$\begin{aligned} -\nabla\cdot(T(\boldsymbol{x},\omega)\nabla p(\boldsymbol{x},\omega)) &= F(\boldsymbol{x}) \quad \text{in } D\times\Omega, \\ p(\boldsymbol{x},\omega) &= g(\boldsymbol{x}) \quad \text{on } \partial D_D\times\Omega, \\ \boldsymbol{n}\cdot(T(\boldsymbol{x},\omega)\nabla p(\boldsymbol{x},\omega)) &= 0 \quad \text{on } \partial D_N\times\Omega. \end{aligned} \tag{1.4}$$

Analogously in the mixed formulation (1.5) we now look for random fields $u = u(\boldsymbol{x}, \omega)$ and $p = p(\boldsymbol{x}, \omega)$ such that, P-a.s.,

$$\begin{aligned} T^{-1}(\boldsymbol{x},\omega)\boldsymbol{u}(\boldsymbol{x},\omega) + \nabla p(\boldsymbol{x},\omega) &= 0, \\ \nabla\cdot\boldsymbol{u}(\boldsymbol{x},\omega) &= F(\boldsymbol{x}) \quad \text{in } D\times\Omega, \\ p(\boldsymbol{x},\omega) &= g(\boldsymbol{x}) \quad \text{on } \partial D_D\times\Omega, \\ \boldsymbol{n}\cdot\boldsymbol{u}(\boldsymbol{x},\omega) &= 0 \quad \text{on } \partial D_N\times\Omega. \end{aligned} \tag{1.5}$$

In this paper we discretize problems (1.4) and (1.5) using the Stochastic Finite Element Method (SFEM) [12]. In Section 2 we review its basic discretization steps and give the resulting sequence of independent linear systems having decoupled the global Galerkin matrix with respect to the stochastic DOFs. This is achieved via a so called doubly orthogonal basis of the stochastic variational space but for diffusion random fields that are *linear* in a finite number of independent variables only, see Section 2.2. This approach has been introduced in [2]. In Section 3 we summarize the main ideas of Krylov subspace recycling methods that we utilize to solve the decoupled Galerkin system, introduce the preconditioners employed for the primal and mixed discretization and discuss ordering issues for the linear systems in the presence of recycling. In Section 4 we give some numerical results.

2. Stochastic Galerkin formulation. In this Section we review essential discretization steps of the SFEM for the primal problem (1.4). Since the discretization of the mixed problem (1.5) proceeds analoguously we give only the corresponding Galerkin equations at the end of this Section and refer to [10] for further details.

The variational formulation of the deterministic problem (1.1), that can be found for example in [4], serves as starting point for the stochastic Galerkin formulation of (1.4): Find $p \in H^1_0(D)$, such that

$$\int_D T\,\nabla p\cdot\nabla q\,d\boldsymbol{x} = \int_D Fq\,d\boldsymbol{x} \quad \forall q \in H^1_0(D). \tag{2.1}$$

The variational space $H^1_0(D) = \{q \in H^1(D) : q|_{\partial D_N} = 0\}$ is a subspace of the Sobolev space $H^1(D)$. Since in (1.4) the we look for random fields, one chooses the tensor product stochastic variational space $H^1_0(D)\otimes L^2(\Omega)$. Applying the expectation operator $\langle\cdot\rangle$ on both sides of (2.1) we arrive at the stochastic variational formulation of (1.4): Find $p \in H^1_0(D)\otimes L^2(\Omega)$ such that

$$\left\langle \int_D T\,\nabla p\cdot\nabla q\,d\boldsymbol{x}\right\rangle = \left\langle \int_D Fq\,d\boldsymbol{x}\right\rangle \quad \forall q \in H^1_0(D)\otimes L^2(\Omega). \tag{2.2}$$

2.1. Finite dimensional noise. We follow a well established approach in the SFEM [2, 3, 18, 11] by assuming that all input and output random fields can be represented in terms of a finite number $M \in \mathbb{N}$ of independent random variables $\{\xi_m\}_{m=1}^M$ with given probability density functions $\rho_m : \Gamma_m \to \mathbb{R}^+_0$, $m = 1,\dots,M$. Hence their joint probability density function $\rho : \Gamma_1\times\cdots\times\Gamma_M \to \mathbb{R}^+_0$ is given by $\rho(\boldsymbol{\xi}) = \rho_1(\xi_1)\cdots\rho_M(\xi_M)$ and the space $L^2(\Omega)$ can be identified with the weighted L^2-space

$$L^2_\rho(\Gamma) = \left\{ f : \Gamma\to\mathbb{R} : \int_\Gamma f^2(\boldsymbol{\xi})\rho(\boldsymbol{\xi})\,d\boldsymbol{\xi} < \infty\right\}.$$

In addition the expectation operator reads $\langle\boldsymbol{\xi}\rangle = \int_\Gamma \boldsymbol{\xi}\rho(\boldsymbol{\xi})\,d\boldsymbol{\xi}$. Reformulating the variational problem (2.2) we arrive at: Find $p \in H^1_0(D)\otimes L^2_\rho(\Gamma)$, such that

$$\int_\Gamma\int_D T(\boldsymbol{x},\boldsymbol{\xi})\nabla p(\boldsymbol{x},\boldsymbol{\xi})\cdot\nabla q(\boldsymbol{x},\boldsymbol{\xi})\rho(\boldsymbol{\xi})\,d\boldsymbol{x}\,d\boldsymbol{\xi} = \int_\Gamma\int_D F(\boldsymbol{x},\boldsymbol{\xi})q(\boldsymbol{x},\boldsymbol{\xi})\rho(\boldsymbol{\xi})\,d\boldsymbol{x}\,d\boldsymbol{\xi} \quad \forall q \in H^1_0(D)\otimes L^2_\rho(\Gamma). \tag{2.3}$$

This problem (and its mixed case analogue (2.12)) is to be discretized.

2.2. Input random field model. For the representation of the diffusion coefficient T (or its inverse T^{-1}) we require only second-order information, that is the mean value $t_0(\boldsymbol{x}) = \langle T(\boldsymbol{x}, \cdot)\rangle$ and the covariance function $\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y}) = \langle (T(\boldsymbol{x}, \cdot) - t_0(\boldsymbol{x}))(T(\boldsymbol{y}, \cdot) - t_0(\boldsymbol{y}))\rangle$ of a random field. Common covariance models are

$$\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y}) = \sigma^2 \exp\left(-\frac{r^2}{c^2}\right), \tag{2.4a}$$

$$\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y}) = \sigma^2 \left(\frac{r}{c}\right) K_1\left(\frac{r}{c}\right), \tag{2.4b}$$

where r denotes the Eucledian distance between $\boldsymbol{x}$ and $\boldsymbol{y}$; c is a correlation length and K_1 denotes the modified Bessel function of second kind and order one. Note that in all models the variance is constant: $\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{x}) = \sigma^2$. We represent T by means of its truncated Karhunen-Loève (KL) expansion [17]:

$$T(\boldsymbol{x}, \boldsymbol{\xi}) = t_0(\boldsymbol{x}) + \sigma \sum_{m=1}^{M} \sqrt{\lambda_m} t_m(\boldsymbol{x}) \xi_m, \tag{2.5}$$

where (λ_m, t_m), $m = 1, \ldots, M$, are eigenpairs of the integral operator

$$C : L^2(D) \to L^2(D), \quad (Cu)(\boldsymbol{x}) = \int_D c(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{y})\, d\boldsymbol{y}, \tag{2.6}$$

whose kernel function is the correlation function of the random field under consideration: $c(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y})/\sigma^2$. If the kernel function is continuous, then the integral operator is compact having a countable set of eigenpairs, where the eigenvalues decay to zero. We assume that in expansion (2.5) the eigenvalues are arranged in decreasing order. In addition we use normalized eigenfunctions $||t_m||_{L^2(D)} = 1$, hence $\sum_{m=1}^{\infty} \lambda_m = |D| = \int_D c(\boldsymbol{x}, \boldsymbol{x})\, d\boldsymbol{x}$, meaning that the truncation can be performed in a way such that a given amount of the random field's total variance is captured.

2.3. Galerkin approximation of primal problem. In the SFEM one selects two finite-dimensional subspaces $X_h = \mathrm{span}\{\phi_1, \ldots, \phi_{N_{\boldsymbol{x}}}\} \subset H_0^1(D)$ and $S_d = \{\psi_1, \ldots, \psi_{N_{\boldsymbol{\xi}}}\} \subset L^2_\rho(\Gamma)$ and uses the variational space $X_h \otimes S_d$ for the discretization of problem (2.3). Inserting the Ansatz $p(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{i=1}^{N_{\boldsymbol{x}}} \sum_{j=1}^{N_{\boldsymbol{\xi}}} p_{i,j} \phi_i(\boldsymbol{x}) \psi_j(\boldsymbol{\xi})$ together with the KL expansion (2.5) of T into (2.3) we arrive at a linear system of equations in $N_{\boldsymbol{x}} \cdot N_{\boldsymbol{\xi}}$ unknowns:

$$\begin{aligned} &\left[\int_D t_0 \nabla \phi_i \cdot \nabla \phi_k\, d\boldsymbol{x}\, \langle \psi_j \psi_\ell \rangle + \sigma \sum_{m=1}^{M} \int_D \sqrt{\lambda_m} t_m \nabla \phi_i \cdot \nabla \phi_k\, d\boldsymbol{x}\, \langle \xi_m \psi_j \psi_\ell \rangle \right] p_{i,j} \\ &= \int_D F \phi_k\, d\boldsymbol{x}\, \langle \psi_\ell \rangle, \quad k = 1, \ldots, N_{\boldsymbol{x}}, \ell = 1, \ldots, N_{\boldsymbol{\xi}}. \end{aligned} \tag{2.7}$$

The deterministic variational space X_h is free of choice. For the stochastic subspace S_d we employ global M-variate polynomials on Γ. Regarding the degree of these polynomials we shall use tensor product polynomials in this paper, i.e.,

$$S_d = \left\{\boldsymbol{\xi}^{\boldsymbol{\alpha}} : \alpha_m \le d_m, \alpha \in \mathbb{N}_0^M\right\} \tag{2.8}$$

is the space of all M-variate polynomials of bounded separate degree d_m, $m = 1, \ldots, M$, which is also discussed in [2, 3, 8, 15, 10]. Clearly we then have $N_{\boldsymbol{\xi}} = (d_1 + 1) \cdots (d_M + 1)$ stochastic DOFs. It is well known [2] that this space possesses a doubly orthogonal basis $\left\{\widehat{\psi}_1, \ldots, \widehat{\psi}_{N_{\boldsymbol{\xi}}}\right\}$ with

$$\left\langle \widehat{\psi}_j \widehat{\psi}_\ell \right\rangle = \delta_{j,\ell}, \quad \left\langle \xi_m \widehat{\psi}_j \widehat{\psi}_\ell \right\rangle = c_{m,\ell} \delta_{j,\ell}, \quad j, \ell = 1, \ldots, N_{\boldsymbol{\xi}}, m = 1, \ldots, M. \tag{2.9}$$

Using this doubly orthogonal basis the linear system (2.7) decouples into a sequence of $N_{\boldsymbol{\xi}}$ linear systems:

$$A^{(\ell)}\boldsymbol{p}^{(\ell)} = \boldsymbol{b}^{(\ell)}, \quad \ell = 1, \ldots, N_{\boldsymbol{\xi}}, \tag{2.10}$$

where each system matrix $A^{(\ell)} \in \mathbb{R}^{N_{\boldsymbol{x}} \times N_{\boldsymbol{x}}}$ is a linear combination of $M+1$ finite element stiffness matrices:

$$A^{(\ell)} = A_0 + \sigma \sum_{m=1}^{M} \sqrt{\lambda_m} c_{m,\ell} A_m, \quad \ell = 1, \ldots, N_{\boldsymbol{\xi}}, \tag{2.11}$$

$$[A_0]_{i,k} = \int_D t_0 \, \nabla \phi_i \cdot \nabla \phi_k \, d\boldsymbol{x},$$

$$[A_m]_{i,k} = \int_D t_m \, \nabla \phi_i \cdot \nabla \phi_k \, d\boldsymbol{x}, \quad i,k = 1, \ldots, N_{\boldsymbol{x}},\ m = 1, \ldots, M.$$

The vectors $\boldsymbol{p}^{(\ell)} \in \mathbb{R}^{N_{\boldsymbol{x}}}$ and right hand side vectors $\boldsymbol{b}^{(\ell)} \in \mathbb{R}^{N_{\boldsymbol{x}}}$ are defined as

$$[\boldsymbol{p}^{(\ell)}]_k = p_{k,\ell} \quad [\boldsymbol{b}^{(\ell)}]_k = \int_D F\phi_k \, d\boldsymbol{x} \left\langle \widehat{\psi}_\ell \right\rangle, \quad k = 1, \ldots, N_{\boldsymbol{x}}, \ell = 1, \ldots, N_{\boldsymbol{\xi}}.$$

2.4. Galerkin approximation of mixed problem. Having reviewed the Galerkin approximation of the primal stochastic variational problem (2.3) we give the system of linear equations arising from the discretization of the mixed stochastic formulation (1.5). We discretize the following problem: Find a pair $(\boldsymbol{u}, p) \in (H_0(div; D) \times L^2(D)) \otimes L^2_\rho(\Gamma)$, such that

$$\begin{aligned} \int_\Gamma \int_D T^{-1}\boldsymbol{u} \cdot \boldsymbol{v} \rho \, d\boldsymbol{x} \, d\boldsymbol{\xi} - \int_\Gamma \int_D p \, \nabla\cdot \boldsymbol{v} \rho \, d\boldsymbol{x} \, d\boldsymbol{\xi} &= -\int_\Gamma \int_{\partial D_D} g\boldsymbol{v} \cdot \boldsymbol{n} \rho \, ds \, d\boldsymbol{\xi} \ \forall \boldsymbol{v} \in H_0(div; D) \otimes L^2_\rho(\Gamma), \\ -\int_\Gamma \int_D q \, \nabla\cdot \boldsymbol{u} \rho \, d\boldsymbol{x} \, d\boldsymbol{\xi} &= -\int_\Gamma \int_D Fq\rho \, d\boldsymbol{x} \, d\boldsymbol{\xi} \qquad \forall q \in L^2(D) \otimes L^2_\rho(\Gamma). \end{aligned} \tag{2.12}$$

The deterministic part $H_0(div; D) \times L^2(D)$ of the variational space is composed using the space $H_0(div; D) = \{\boldsymbol{v} \in H(div; D) : \boldsymbol{v} \cdot \boldsymbol{n}|_{\partial D_N} = 0\}$ which is a subspace of the Sobolev space $H(div; D)$, see [5, 4]. Again we use a truncated KL expansion for the representation of the inverse T^{-1} of the random diffusion coefficient:

$$T^{-1}(\boldsymbol{x}, \boldsymbol{\xi}) = \widetilde{t}_0(\boldsymbol{x}) + \sigma \sum_{m=1}^{M} \sqrt{\widetilde{\lambda}_m} \widetilde{t}_m(\boldsymbol{x}) \xi_m. \tag{2.13}$$

The stochastic discretization remains unchanged, i.e., we choose S_d as in (2.8) in conjunction with the doubly orthogonal basis (2.9), hence the global Galerkin system decouples exactly in the same fashion as for the primal formulation. Having chosen deterministic subspaces $V_h = \operatorname{span}\left\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{N_{\boldsymbol{x}}^{\boldsymbol{u}}}\right\} \subset H_0(div; D)$ and $W_h = \left\{\phi_1, \ldots, \phi_{N_{\boldsymbol{x}}^p}\right\} \subset L^2(D)$ we arrive at the sequence of linear systems

$$\begin{bmatrix} \widetilde{A}^{(\ell)} & B^T \\ B & O \end{bmatrix} \begin{bmatrix} \boldsymbol{u}^{(\ell)} \\ \boldsymbol{p}^{(\ell)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}^{(\ell)} \\ \boldsymbol{f}^{(\ell)} \end{bmatrix} \quad \ell = 1, \ldots, N_{\boldsymbol{\xi}}, \tag{2.14}$$

where each matrix $\widetilde{A}^{(\ell)} \in \mathbb{R}^{N_{\boldsymbol{x}}^{\boldsymbol{u}}} \times \mathbb{R}^{N_{\boldsymbol{x}}^{\boldsymbol{u}}}$ is a linear combination of $M+1$ weighted $(L^2(D))^2$ mass matrices:

$$\widetilde{A}^{(\ell)} = \widetilde{A}_0 + \sigma \sum_{m=1}^{M} \sqrt{\widetilde{\lambda}_m} c_{m,\ell} \widetilde{A}_m, \quad \ell = 1, \ldots, N_{\boldsymbol{\xi}}, \tag{2.15}$$

$$\left[\widetilde{A}_0\right]_{i,k} = \int_D \widetilde{t}_0 \boldsymbol{\phi}_i \cdot \boldsymbol{\phi}_k \, d\boldsymbol{x}, \quad \left[\widetilde{A}_m\right]_{i,k} = \int_D \widetilde{t}_m \boldsymbol{\phi}_i \cdot \boldsymbol{\phi}_k \, d\boldsymbol{x}, \quad i,k = 1, \ldots, N_{\boldsymbol{x}}^{\boldsymbol{u}},\ m = 1, \ldots, M,$$

$$[B]_{i,k} = -\int_D \phi_i \, \nabla\cdot \boldsymbol{\phi}_k \, d\boldsymbol{x}, \quad i = 1, \ldots, N_{\boldsymbol{x}}^p,\ k = 1, \ldots, N_{\boldsymbol{x}}^{\boldsymbol{u}}.$$

The vectors $\boldsymbol{u}^{(\ell)} \in \mathbb{R}^{N_x^{\boldsymbol{u}}}$ and $\boldsymbol{p}^{(\ell)} \in \mathbb{R}^{N_x^p}$ contain the coefficients of $\boldsymbol{u}$ and p with respect to the finite element Ansatz functions. The right hand side vectors $\boldsymbol{g}^{(\ell)} \in \mathbb{R}^{N_x^{\boldsymbol{u}}}$ and $\boldsymbol{f}^{(\ell)} \in \mathbb{R}^{N_x^p}$ are defined as

$$\left[\boldsymbol{g}^{(\ell)}\right]_i = -\int_{\partial D_D} g\boldsymbol{\phi}_i \cdot \boldsymbol{n}\, ds \left\langle \widehat{\psi}_\ell \right\rangle$$
$$\left[\boldsymbol{f}^{(\ell)}\right]_j = -\int_D F\phi_j\, d\boldsymbol{x} \left\langle \widehat{\psi}_\ell \right\rangle, \quad i = 1, \ldots, N_{\boldsymbol{x}}^{\boldsymbol{u}},\ j = 1, \ldots, N_{\boldsymbol{x}}^p,\ \ell = 1, \ldots, N_{\boldsymbol{\xi}}.$$

REMARK 2.1. *The stochastic collocation method is recently becoming popular for the discretization of stochastic PDEs of form* (1.4) *or* (1.5)*. In [1] it was shown that a particular choice of the collocation points leads to the same sequences of linear systems to be solved as in* (2.10) *and* (2.14)*, respectively. All recycling methods and preconditioners discussed in Section 3 can be applied in this case.*

3. Iterative solution. In this Section we address solving the sequence of linear systems (2.10) and (2.14), respectively, by preconditioned Krylov subspace iterations. One can apply the CG method [14] (as the system matrices in (2.10) are symmetric positive definite by assumption (1.3)) or the MINRES method [21] (as the system matrices in (2.14) are symmetric and indefinite by (1.3)) to every system in the sequence independently. Since the system matrices are linear combinations of a small number of fixed finite element matrices in both cases, cf. Section 2, we expect the corresponding Krylov subspaces to be similar. Therefore we employ Krylov subspace recycling methods such as GCROT [22], GCRO-DR [22] and R-MINRES [27] that reuse information generated during the solve of one linear system for the solution of the next system in the sequence. After a short review of these methods we consider mean-based preconditioners for the primal and mixed problem. Finally we discuss ordering strategies for the sequences of linear systems among which is a grouping algorithm proposed in [15].

3.1. Krylov subspace recycling methods. The GCR algorithm [9] serves as the common basis of all iterative methods under consideration. Given an initial approximation $\boldsymbol{x}_0$ to the solution of $A\boldsymbol{x} = \boldsymbol{b}$ it selects corrections $\boldsymbol{c}_k \in \mathcal{C}_k$, such that $||\boldsymbol{r}_k||_2 = ||\boldsymbol{r}_0 - A\boldsymbol{c}_k||_2$ is minimized among all possible corrections $\boldsymbol{c}_k \in \mathcal{C}_k$. Therefore the method maintains two bases of the correction space $\mathcal{C}_k = \text{span}\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\}$ and approximation space $\mathcal{W}_k = \text{span}\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\}$ that satisfy

$$AC_k = W_k \qquad W_k^T W_k = I_k, \tag{3.1}$$

where we have introduced the notation $C_k := \left[\boldsymbol{c}_1 \cdots \boldsymbol{c}_k\right]$, $W_k := \left[\boldsymbol{w}_1 \cdots \boldsymbol{w}_k\right]$. The properties (3.1) imply that $\boldsymbol{r}_k = (I_k - W_k W_k^T)\boldsymbol{r}_0$ solves the minimization problem. Choosing Krylov correction spaces $\mathcal{C}_k = \mathcal{K}_k(A, \boldsymbol{r}_0) = \text{span}\{\boldsymbol{r}_0, A\boldsymbol{r}_0, \cdots, A^{k-1}\boldsymbol{r}_0\}$ results in the GMRES method [26] as a special case of GCR.

The GCRO algorithm [6] combines an outer GCR iteration (with Krylov correction spaces) and an inner GMRES iteration where the projected residual equation

$$(I - W_k W_k^T)A\boldsymbol{c} = (I - W_k W_k^T)\boldsymbol{r}_k \Leftrightarrow (I - W_k W_k^T)A\boldsymbol{c} = \boldsymbol{r}_k$$

is solved approximately. Then the outer approximation space is extended by the residual correction that comes from the inner iteration. The basic idea is to keep the correction space of the inner iteration orthogonal to the approximation space of the outer iteration in order to utilize the outer approximation vectors for the inner iteration.

The GCROT algorithm [7] provides in addition a strategy for selecting more that one vector after an inner iteration to extend the outer approximation space, as well as a rule for the truncation of the outer approximation space.

Since for GCR one can choose any pair of approximation and correction spaces whose k dimensional bases satisfy (3.1), the idea of recycling methods for solving a sequence of linear systems is obvious now. One solves the first system in the sequence by GCR or one of its refinements and reuses the corresponding outer correction space to construct a new pair of (outer) correction and approximation spaces for the next system in turn. For details on this construction we refer to [22].

The Krylov subspace recycling methods GCROT-rec. and GCRO-DR introduced in [22] are based on the GCRO algorithm. Both methods recycle the outer GCR approximation and correction spaces, but the construction of these spaces follows different approaches. For GCROT-rec. the recycled spaces $\mathcal{C}_k$ and $\mathcal{W}_k$ are simply the outer correction and approximation spaces created by the GCROT algorithm. For GCRO-DR the recycled correction space contains approximate eigenvectors of the current system matrix.

All aforementioned methods do not exploit the symmetry of the system matrix neither for outer nor inner iterations. In [27] a Krylov subspace recycling method named R-MINRES was introduced. Recall that in GCRO a fixed number n of (inner) GMRES iterations is performed with the linear operator $(I - W_k W_k^T)A$. This results in the modified Arnoldi decomposition $(I - W_k W_k^T)AV_n = V_{n+1}\widetilde{H}_n$. For symmetric matrices $A = A^T$ the inner iteration simplifies since the matrix $\widetilde{H}_n$ is tridiagonal. Indeed for $\boldsymbol{x} \in \mathcal{W}_k^\perp$ we have $(I - W_k W_k^T)A\boldsymbol{x} = (I - W_k W_k^T)A(I - W_k W_k^T)\boldsymbol{x}$, meaning that $(I - W_k W_k^T)A$ is symmetric on $\mathcal{W}_k^\perp$. Consequently there is no need to store all vectors created during an inner GMRES iteration in case of symmetric matrices. On the other hand it is impossible to generate a useful recycling space while discarding most of the Krylov subspace basis vectors at the same time. Therefore in [27] a fixed number m of outer and inner MINRES iterations is performed, respectively. The stored Lanczos vectors are utilized together with the outer (recycled) vectors to compute approximate eigenvectors of the current system matrix. However, in contrast to GCRO-DR, the outer approximation space is never extended or even updated since this would destroy the short term recurrence relation in the Lanczos decomposition for the matrix $(I - W_k W_k^T)A$.

In addition to the R-MINRES method we will also use a modified version R-MINRES-mod. that does not recycle approximate eigenvectors but uses a cheaper approach. At the end of R-MINRES the current correction to $\boldsymbol{x}_0$ is added to the outer correction space that is stored for the next linear system in turn. If the dimension of the outer correction space exceeds m only the latest k correction vectors are retained.

3.2. Mean-based preconditioning. Having decoupled the global Galerkin matrix with respect to the stochastic DOFs, the system matrices in (2.10) and (2.14) always have the first term of the sum in common, namely the deterministic finite element stiffness matrix corresponding to the mean problem where the input random fields T and T^{-1} have been replaced by their mean values. We study preconditioners based on these mean stiffness matrices. This is an efficient approach provided the input random field's standard deviation σ is not too large.

The idea of mean-based preconditioning in the context of SFEM has already been studied in [13, 23, 16, 24], though in conjunction with a stochastic space different from (2.8). There, complete global multivariate polynomials are employed, i.e., polynomials with bounded total degree, and the fully coupled global Galerkin system is solved.

For the primal problems (2.10) we use two preconditioners: $P_{chol} = \texttt{cholinc}(A_0,\texttt{0})$, where $\texttt{cholinc}(A_0,\texttt{0})$ denotes an incomplete Cholesky decomposition of A_0 with no fill-in, and P_{amg}. The inverse of P_{amg} is formed by applying one V-cycle of an algebraic multigrid method (AMG) to the matrix A_0.

For the mixed problems (2.14) we apply

$$\widetilde{P}_{chol} = \begin{bmatrix} D_0 & O \\ O & \texttt{cholinc}(S,\texttt{0}) \end{bmatrix} \quad \text{and} \quad \widetilde{P}_{amg} = \begin{bmatrix} D_0 & O \\ O & V \end{bmatrix}$$

where $D_0 = \texttt{diag}(\widetilde{A}_0)$ and $S = BD_0^{-1}B^T$ is an approximation to the Schur complement matrix. Finally the action of V^{-1} is one single AMG V-cycle applied to S. The preconditioner $\widetilde{P}_{amg}$ is analyzed in [10].

REMARK 3.1. *Under assumption* (1.3) *the matrices* A_0 *and* $\widetilde{A}_0$ *are positive definite.*

3.3. Ordering the linear systems. The sequences of linear systems in (2.10) and (2.14) are independent from each other meaning that we can choose a heuristic ordering that is likely to be useful for the Krylov subspace recycling methods. Consider the matrices (2.11) or the 1-1 block (2.15). For both the primal and the mixed formulation the system matrices are linear combinations of $M + 1$ finite element matrices and the only thing that changes from system to system are the

coefficients $c_{m,\ell} = \left\langle \xi_m \widehat{\psi}_\ell^2 \right\rangle$, $m = 1, \ldots, M$. The idea is to divide the linear systems into groups such that a specific subset of coefficients remains unchanged within one group. All the recycling techniques are applied only within a group. We consider two grouping strategies.

(a) In [15] a grouping algorithm was proposed based on the observation that the coefficients $c_{m,\ell}$ in (2.11) and (2.15) are weighted by the square roots of λ_m and $\widetilde{\lambda}_m$, $m = 1, \ldots, M$, respectively, which form a decreasing sequence, cf. Section 2.2. Hence in general the mth term in the sum of finite element matrices carries more weight than the $(m+1)$st term. Recall that the number of linear systems in a sequence is $N_{\boldsymbol{\xi}} = (d_1 + 1) \cdots (d_M + 1)$. Having decided that the first p eigenvalues are relevant, one divides the sequence into $(d_1 + 1) \cdots (d_p + 1)$ groups having $(d_{p+1} + 1) \cdots (d_M + 1)$ members each such that within a single group the first p coefficients $c_{m,\ell}$ in front of the first p eigenvalues remain constant. However, in [15] there is no specification on the ordering of the systems within one single group, i.e., the ordering of the coefficients $c_{m,\ell}$, $m = p+1, \ldots, M$. In our examples we sort the tuples $(c_1, \ldots, c_M)$ by radix sort (decreasing order) before applying a grouping algorithm.

(b) Another way to group the linear systems is based on the same observation as before. Having selected p relevant eigenvalues resulting in $(d_1 + 1) \cdots (d_p + 1)$ groups we know that within a group only the $M - p + 1$ last terms in the sums (2.11) or (2.15) change. Now consider all first systems in each group. The only thing that changes from the first system in one group to the first system in the next group is the coefficient $c_{1,\ell}$ whereas all other coefficients remain fixed. We apply a kind of *reverse grouping* now and decide to collect all first systems of each group into a new group, then collect all second systems of each group into another new group and so on. This reverse grouping strategy results in $(d_{p+1} + 1) \cdots (d_M + 1)$ groups having $(d_1 + 1) \cdots (d_p + 1)$ members each.

4. Numerical results. In this Section we present numerical results for a primal and a mixed test problem. We use preconditioned Krylov subspace recyling methods discussed in Section 3.1. The stopping criterion for all methods was a reduction of the initial residual norm by a certain factor *tol*, though the norm used depends on the kind of preconditioning. For GMRES, GCROT-rec. and GCRO-DR we apply right preconditioning and measure the Euclidean norm of the residual vector whereas for R-MINRES we use the P^{-1}-norm. For GCRO-DR and R-MINRES we always recycle harmonic Ritz vectors corresponding to harmonic Ritz values of smallest magnitude.

We compare the average, minimal and maximal iteration count of all iterative methods applied together with the AMG or incomplete Cholesky decomposition version of the mean-based preconditioners introduced in Section 3.2. The reported results were performed using MATLAB 7.4 together with a MATLAB version of the AMG code `HSL_MI20` [19].

EXAMPLE 4.1. *We consider the primal problem* (1.4) *with* $D = (0,1)^2$, $F = 1$, $\partial D_D = \partial D$ *and* $g \equiv 0$. *The spatial discretization uses a mesh of 10* x *10 hp square elements with 7 Gauss-Legendre-Lobatto (GLL) nodes per element in each spatial direction yielding* $N_{\boldsymbol{x}} = 3,481$. *The diffusion coefficient* T *is given by a truncated Karhunen-Loève expansion in* $M = 11$ *uniform random variables with mean* $\langle T \rangle = 3 + \sin(\pi x)$ *and covariance function* (2.4a), *where* $c = 1$ *and* $\sigma = 1$. *The maximal degree of the tensor polynomials in each variable is* $\boldsymbol{d} = (3,2,2,1,1,1,1,1,1,1,1)$ *yielding* $N_{\boldsymbol{\xi}} = 9,216$. *We will also use* $M = 4$ *random variables together with* $\boldsymbol{d} = (3,2,2,1)$ *yielding* $N_{\boldsymbol{\xi}} = 72$. *With* $M = 11$ *random variables we capture 99 % of the random field's variance and with* $M = 4$ *we have 98 % captured. The integral eigenproblem associated with the operator* (2.6) *is discretized using the Nystrom method [20] on the same mesh used for the spatial discretization. This example is taken from [15].*

In Tables 4.1 and 4.2 we give the results for Example 4.1 with $M = 4$ and $M = 11$ random variables, respectively. The stopping tolerance is $tol = 10^{-6}$. The parameters of all recycling methods are chosen such that at most 40 vectors are stored and at most 20 vectors are recycled. We compare three grouping strategies for the linear systems: no grouping, strategy (a) above with $p = 1$ and strategy (b) with $p = 1$, cf. Section 3.3. Unlike in [15], where in addition to the grouping strategy (a) all nearly singular linear systems are singled out into a separate group, this is not necessary here, because the minimal diffusion coefficient at the random field's GLL nodes taken over all linear systems in a sequence is equal to 0.6748 ($M = 4$) or 0.0972 ($M = 11$).

We observe that recycling methods in combination with the P_{amg} preconditioner do not decrease the average iteration count significantly independent of the number of random variables used. How-

	GMRES	GCRO-DR			GCROT-rec.			R-MINRES		
grouping	-	-	(a)	(b)	-	(a)	(b)	-	(a)	(b)
P_{amg} preconditioner										
av. iters	10.5	11.0	10.9	9.9	9.7	9.6	8.3	7.9	7.8	5.6
max iters	17	18	18	15	16	15	12	14	14	8
min iters	7	7	7	6	6	7	4	5	4	2
P_{chol} preconditioner										
av. iters	68.6	38.1	40.8	47.7	56.6	56.3	52.7	28.3	31.7	41.1
max iters	83	74	83	71	84	84	79	67	81	71
min iters	48	30	28	28	34	36	16	22	22	27

TABLE 4.1
Example 4.1: iteration counts obtained for $M = 4$.

	GMRES	GCRO-DR			GCROT-rec.			R-MINRES		
grouping	-	-	(a)	(b)	-	(a)	(b)	-	(a)	(b)
P_{amg} preconditioner										
av. iters	11.0	11.3	11.3	10.2	8.4	8.4	8.6	8.6	8.6	5.9
max iters	25	26	26	23	25	25	20	19	19	14
min iters	7	7	7	6	4	4	4	3	3	2
P_{chol} preconditioner										
av. iters	69.9	33.5	33.4	45.7	50.7	50.7	53.9	25.2	25.3	42.3
max iters	86	76	84	74	88	88	85	69	83	73
min iters	53	23	23	26	24	24	15	20	20	24

TABLE 4.2
Example 4.1: iteration counts obtained for $M = 11$.

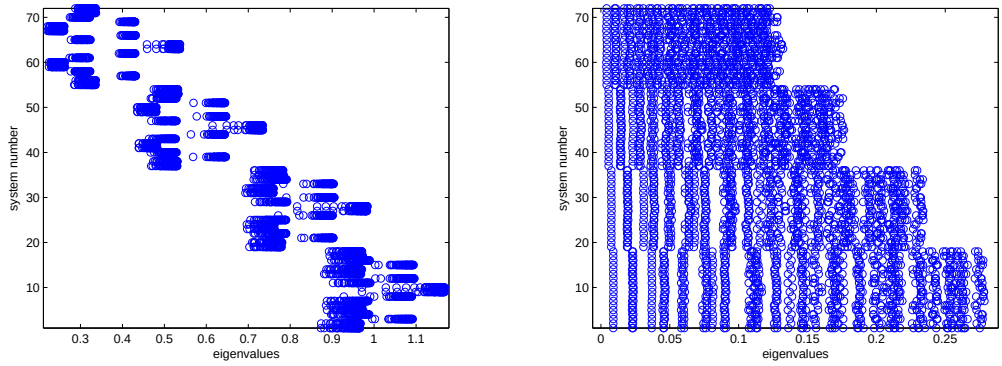
ever, the proposed reverse grouping strategy (b) performs best in almost every case. This may be due to the fact that this strategy leads to a large number of small groups hence recycling is applied rarely.

In contrast, when using the P_{chol} preconditioner together with GCRO-DR or R-MINRES we can save half of the iterations in the best case. The grouping strategy (a) is more successful than reverse grouping (b). For GCROT-rec. the reduction is not as large as for the other recycling methods, whereas GCROT-rec. in combination with P_{amg} performs better than GCRO-DR.

If we compare the grouping strategy (a) and no grouping there is no significant difference in the average iteration count in all cases. This may be due to the fact that not enough vectors are recycled.

However, the results suggest that for the P_{amg} preconditioner, GCRO-DR and R-MINRES seem not to be as helpful as in combination with the P_{chol} preconditioner. This may be due to the fact, that when using the P_{chol} preconditioner, the preconditioned system matrices still have eigenvalues that are small in magnitude compared to the P_{amg}-preconditioned system matrices, see for example Figure 4.1. Hence recycling methods are not likely to improve the convergence behaviour of the iterative solvers since the eigenvalues of the preconditioned system matrices for P_{amg} are already well separated from zero.

EXAMPLE 4.2. *We consider the mixed problem* (1.5) *with* $D = (0,1)^2$, $F = 0$, $\partial D_D = \{0,1\} \times [0,1]$ *and* $\partial D_N = [0,1] \times \{0,1\}$. *We fix* $\boldsymbol{n} \cdot \boldsymbol{u} = 0$ *on the horizontal boundaries and prescribe* $p = 1$ *on* $\{0\} \times [0,1]$ *and* $p = 0$ *on* $\{1\} \times [0,1]$. *The spatial discretization uses the lowest order Raviart-Thomas approximation [25] on a 50* x *50 mesh of square elements for the approximation of* $\boldsymbol{u}$ *and* p *respectively, yielding* $N_{\boldsymbol{x}} = 7,500$. *The reciprocal field* T^{-1} *is modelled as Gaussian random field with constant mean* $\langle T^{-1} \rangle = 1$, *standard deviation* $\sigma = 0.1$ *and covariance function* (2.4b) *with* $c = 1$. *We use* $M = 5$ *random variables which captures 97 % of the field's total variance. The maximal degree of the tensor polynomials in each variable is* $\boldsymbol{d} = (3,3,3,3,3)$ *yielding* $N_{\boldsymbol{\xi}} = 1,024$. *The integral eigenproblem associated with the operator* (2.6) *is discretized using the Nystrom method [20] on a 16*



(a) P_{amg} preconditioner (b) P_{chol} preconditioner

FIG. 4.1. *Example 4.1: The 50 smallest eigenvalues of the preconditioned system matrices estimated using MATLAB's* `eigs` *for* $M = 4$.

x *16 mesh of square elements with 4 GLL nodes per element in each spatial direction.*

	MINRES	R-MINRES			R-MINRES-mod.		
grouping	-	-	(a)	(b)	-	(a)	(b)
$\widetilde{P}_{amg}$ preconditioner							
av. iters	40.5	40.5	40.4	36.1	89.1	76.2	32.5
max iters	46	46	46	46	119	112	46
min iters	35	31	24	21	4	4	6
$\widetilde{P}_{chol}$ preconditioner							
av. iters	247.5	76.4	77.5	175.6	241.3	190.5	194.3
max iters	281	271	271	281	315	315	281
min iters	209	66	65	111	8	8	25

TABLE 4.3
Example 4.2: iteration counts.

In Table 4.3 we give the results for Example 4.2. The stopping tolerance is $tol = 10^{-8}$. The parameters of all recycling methods are chosen such that at most 80 vectors are stored and at most 40 vectors are recycled. We compare three grouping strategies for the linear systems: no grouping, strategy (a) with $p = 1$, and reverse grouping (b) with $p = 1$, cf. Section 3.3.

For the $\widetilde{P}_{amg}$ preconditioner we observe no significant reduction of the average iteration count. However, R-MINRES-mod. applied with reverse grouping (b) performs best. For the $\widetilde{P}_{chol}$ preconditioner together with R-MINRES and grouping stategy (a) there is a reduction of the average iteration count. R-MINRES-mod. or reverse grouping (b) perform worse in that case, although the smallest minimal iteration count is achieved for R-MINRES-mod.

5. Conclusions. We have studied the performance of Krylov subspace recycling methods for the solution of sequences of linear systems arising from primal or mixed SFEM discretizations of second-order elliptic stochastic PDEs. In both cases recycling does not decrease the average iteration count significantly when applied with a preconditioner that separates the eigenvalues of the preconditioned system matrices from zero, whereas in combination with a less efficient preconditioner, recycling can save more than half of the iteration count in the best case. On the other hand, for the AMG version of the preconditioner the modified R-MINRES method or the reverse grouping strategy can outperform other recycling methods. In some cases recycling techniques can improve the convergence behaviour of the applied iterative solvers, and sometimes recycling is not helpful.

Acknowledgements. The support of several people who contributed to this research by supplying code as well as helpful discussion is gratefully acknowledged: my supervisor Oliver Ernst for providing the MATLAB code for GMRES and GCROT, Catherine Powell for the Raviart-Thomas mixed approximation code and the AMG code and finally, Eric de Sturler for the R-MINRES code, that has been rewritten for left preconditioning by the author.

REFERENCES

[1] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034.
[2] I. Babuška, R. Tempone, and G. E. Zouraris, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Math., 42 (2004), pp. 800–825.
[3] ———, *Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 1251–1294.
[4] D. Braess, *Finite elements. Theory, fast solvers, and applications in elasticity theory*, Cambridge University Press, Cambridge, 1997.
[5] F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, Springer-Verlag, New York, 1991.
[6] E. de Sturler, *Nested Krylov methods based on GCR*, J. Comput. Appl. Math., 67 (1994), pp. 15–41.
[7] ———, *Truncation strategies for optimal Krylov subspace methods*, SIAM J. Numer. Math., 36 (1999), pp. 864–889.
[8] M. Eiermann, O. G. Ernst, and E. Ullmann, *Computational aspects of the stochastic finite element method*, Comput. Visual. Sci., 10 (2007), pp. 3–15.
[9] S. C. Eisenstat, H. C. Elman, and M. H. Schultz, *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 20 (1983), pp. 345–357.
[10] O. G. Ernst, C. E. Powell, D. J. Silvester, and E. Ullmann, *Efficient solvers for a linear stochastic Galerkin mixed formulation of diffusion problems with random data*, Tech. Report MIMS EPrint 2007.126, Manchester Institute for Mathematical Sciences, School of Mathematics, University of Manchester, October 2007.
[11] 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.
[12] R.G. Ghanem and P.D. Spanos, *Stochastic finite elements: a spectral approach*, Springer-Verlag, New York, 1991.
[13] 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.
[14] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand., 49 (1952), pp. 409–436.
[15] C. Jin, X.-C. Cai, and C. Li, *Parallel domain decomposition methods for stochastic elliptic equations*, SIAM J. Sci. Comput., 29 (2007), pp. 2096–2114.
[16] A. Keese, *Numerical Solution of Systems with Stochastic Uncertainties – A General Purpose Framework for Stochastic Finite Elements*, PhD thesis, Fachbereich für Mathematik und Informatik, Technische Universität Braunschweig, 2004.
[17] M. Loève, *Probability Theory*, vol. II, Springer-Verlag, New York - Heidelberg - Berlin, 4th ed., 1978.
[18] H. G. Matthies and A. Keese, *Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 1295–1331.
[19] http://www.maths.manchester.ac.uk/∼djs/hsl_mi20.pdf.
[20] E. J. Nyström, *Über die praktische Auflösung von Integralgleichungen mit Anwendungen auf Randwertaufgaben*, Acta Math., 54 (1930), pp. 185–204.
[21] C. C. Paige and M. A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
[22] M.L. Parks, E. de Sturler, G. Mackey D.D. Johnson, and S. Maiti, *Recycling Krylov subspaces for sequences of linear systems*, SIAM J. Sci. Comput., 28 (2006), pp. 1651–1674.
[23] M. F. Pellissetti and R. G. Ghanem, *Iterative solution of linear systems arising in the context of stochastic finite elements*, Adv. Eng. Softw., 31 (2000), pp. 607–616.
[24] C. E. Powell and H. C. Elman, *Block-diagonal preconditioning for spectral stochastic finite element systems*, Tech. Report TR-4879, Department of Computer Science, University of Maryland, May 2007.
[25] P. A. Raviart and J. M. Thomas, *A mixed finite element method for 2nd order elliptic problems*, in Mathematical Aspects of Finite Element Methods, no. 606 in Lect. Notes Math., Berlin - Heidelberg - New York, 1977, Springer-Verlag, pp. 292–315.
[26] Y. Saad and M. H. Schultz, *Conjugate gradient-like algorithms for solving nonsymmetric linear systems*, Math. Comput., 44 (1985), pp. 417–424.
[27] S. Wang, E. de Sturler, and G. H. Paulino, *Large-scale topology optimization using preconditioned Krylov subspace methods with recycling*, Int. J. Numer. Meth. Engng., 69 (2007), pp. 2441–2468.