Solving log-transformed random diffusion problems by stochastic Galerkin mixed finite element methods

Ullmann, Elisabeth and Powell, Catherine

2014

MIMS EPrint: **2014.76**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. xx, pp. x

Solving log-transformed random diffusion problems by stochastic Galerkin mixed finite element methods

Elisabeth Ullmann* and Catherine E. Powell†

Abstract. Stochastic Galerkin finite element discretisations of PDEs with stochastically nonlinear coefficients lead to linear systems of equations with block dense matrices. In contrast, stochastic Galerkin finite element discretisations of PDEs with stochastically linear coefficients lead to linear systems of equations with block sparse matrices which are cheaper to manipulate and precondition in the framework of Krylov subspace iteration. In this paper we focus on mixed formulations of second-order elliptic problems, where the diffusion coefficient is the exponential of a random field, and the priority is to approximate the flux. We build on the previous work [Efficient iterative solvers for stochastic Galerkin discretizations of log-transformed random diffusion problems, SIAM J. Sci. Comput., 34(2012), pp.A659–A682] and reformulate the PDE model as a first-order system in which the logarithm of the diffusion coefficient appears on the left-hand side. We apply a stochastic Galerkin mixed finite element method and discuss block triangular and block diagonal preconditioners for use with GMRES iteration. In particular, we analyse a practical approximation to the Schur complement of the Galerkin matrix and provide spectral inclusion bounds. Numerical experiments reveal that the preconditioners are completely insensitive to the spatial mesh size, and are only slightly sensitive to the statistical parameters of the diffusion coefficient. As a result, the computational cost required to approximate the flux when the diffusion coefficient is stochastically nonlinear grows only linearly with respect to the total problem size.

Key words. generalised saddle point problems, PDEs with random data, convection-diffusion, stochastic finite elements, mixed finite elements, preconditioning, Schur complement approximation

AMS subject classifications. 35R60, 60H15, 60H35, 65N30, 65F10, 65F08

1. Introduction.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with sample space Ω and let $D \subset \mathbb{R}^2$ be the chosen computational domain. We begin by considering the stationary diffusion problem: find $u : D \times \Omega \to \mathbb{R}$ such that $\mathbb{P}$-a.s.,

$$-\nabla \cdot (e^{a_M} \nabla u) = f \quad \text{in } D, \tag{1.1}$$

$$u = 0 \quad \text{on } \partial D, \tag{1.2}$$

where the diffusion coefficient is the exponential of a random field $a_M : D \times \Omega \to \mathbb{R}$ of the form

$$a_M(\boldsymbol{x}, \omega) = a_0(\boldsymbol{x}) + \sigma \sum_{k=1}^{M} \sqrt{\lambda_k} a_k(\boldsymbol{x}) \xi_k(\omega). \tag{1.3}$$

We assume that a_M represents a truncated Karhunen-Loève expansion of an underlying mean-square continuous random field $a : D \times \Omega \to \mathbb{R}$ with mean a_0 and standard deviation σ whose covariance operator $\mathscr{C} : D \times D \to \mathbb{R}$ has eigenfunctions a_k and eigenvalues λ_k. We further

*Fachbereich Mathematik, Universität Hamburg, Bundesstr. 55, 20146 Hamburg, Germany (elisabeth.ullmann@uni-hamburg.de).

†School of Mathematics, University of Manchester, Oxford Road, Manchester M20 2WL, United Kingdom (c.powell@manchester.ac.uk).

assume that the random variables $\xi_k : \Omega \to \Gamma_k \subset \mathbb{R}$ are bounded and independent. For simplicity, f is assumed to be a deterministic function of $\boldsymbol{x} \in D$.

It is well known (e.g., see [14]) that applying stochastic Galerkin finite element methods (SGFEMs) to PDEs with coefficients of the form e^{a_M} results in very large linear systems $A\boldsymbol{x} = \boldsymbol{b}$ with coefficient matrices of the form

$$A = I \otimes A_0 + \sum_{\ell=1}^{L} G_\ell \otimes A_\ell.$$

These matrices are both block dense and ill-conditioned and the number of terms L depends nonlinearly on M. In the case of (1.1)–(1.2), if the finite-dimensional variational problem associated with the chosen SGFEM is well-posed, then A is symmetric and positive definite. Preconditioners have been suggested for these matrices in [16, 19, 20, 23] for use with the conjugate gradient (CG) method. However, in addition to the application of the chosen preconditioner, one CG iteration requires a matrix-vector product with the SGFEM matrix. Since A is block dense, and the number of blocks depends nonlinearly on M, the cost of this operation becomes unmanageable as $M \to \infty$. Hence, even with a robust preconditioner, the solution time can be far too slow. A commonly cited application of (1.1)–(1.2) is groundwater flow modelling. Here, one frequently encounters cases where a_M, the logarithm of the diffusion coefficient, is a random field with a small correlation length. This exacerbates the problem since M then has to be large to control the error between a and the approximation a_M. The more random variables are needed to parameterise the uncertainty in the logarithm of the diffusion coefficient, the higher the cost of a matrix-vector product with the SGFEM matrix and hence, the higher the cost of one CG iteration (with or without a preconditioner).

The lack of efficient solvers for SGFEM systems associated with PDEs with so-called stochastically nonlinear coefficients (coefficients that are nonlinear functions of the random variables ξ_k) has spawned significant interest in stochastic collocation and Monte Carlo FEMs. These methods can be implemented for (1.1)–(1.2) by solving sequences of discretised deterministic PDEs which all have sparse symmetric and positive definite system matrices. Designing good solvers for these is straight-forward if one is already available for the corresponding deterministic problem. However, since Galerkin approximation leads to a best approximation and yields a favourable framework for error estimation [2], it is worthwhile pursuing more computationally efficient ways to solve stochastically nonlinear problems using SGFEMs.

1.1. Convection-diffusion reformulation. To increase the number of random variables that can be handled efficiently in (1.1)–(1.2) when standard SGFEMs and Krylov iteration are applied, a remedy is suggested in [24]. The idea is to reformulate the boundary-value problem before discretisation so that only the logarithm of the diffusion coefficient appears on the left-hand side. Multiplying both sides of (1.1) by e^{-a_M} and differentiating yields the *convection-diffusion* problem: find $u : D \times \Omega \to \mathbb{R}$ such that $\mathbb{P}$-a.s.,

$$-\Delta u + \boldsymbol{w} \cdot \nabla u = f e^{-a_M} \quad \text{in } D, \tag{1.4}$$

$$u = 0 \quad \text{on } \partial D, \tag{1.5}$$

with convective velocity $\boldsymbol{w} := -\nabla a_M$. The PDE operator in (1.1) has a stochastically nonlinear coefficient. However, (1.4)–(1.5) has only the gradient of the logarithm of the diffusion

coefficient on the left-hand side. Crucially,

$$\boldsymbol{w}(\boldsymbol{x},\omega) = -\nabla a_M(\boldsymbol{x},\omega) = -\nabla a_0(\boldsymbol{x}) - \sigma \sum_{k=1}^{M} \sqrt{\lambda_k}\nabla a_k(\boldsymbol{x})\xi_k(\omega),$$

is a *linear* function of ξ_k. Now, when SGFEMs are applied to PDEs like (1.4) with stochastically linear coefficients, we obtain linear systems of equations $A\boldsymbol{x} = \boldsymbol{b}$ with block *sparse* matrices A. Moreover, these matrices tend to be better conditioned than their stochastically nonlinear counterparts with respect to the standard deviation σ of the random field a and the polynomial degree d chosen for the stochastic part of the Galerkin approximation space.

The fact that SGFEM matrices associated with discretisations of (1.4)–(1.5) are sparser and more well-conditioned than those associated with (1.1)–(1.2) gives us hope that we can approximate u more efficiently by solving a convection-diffusion problem. Of course, swapping a diffusion problem for a convection-diffusion problem gives rise to some new issues. For instance, the SGFEM matrices are non-symmetric. Establishing the well-posedness of the variational problem associated with (1.4)–(1.5) requires careful consideration of $\boldsymbol{w}$. It is well known that solutions of deterministic convection-diffusion problems can exhibit steep layers (depending on the size of $\boldsymbol{w}$ and the chosen boundary conditions). Standard finite element methods may be unstable if the problem is convection-dominated and the mesh size is not compatible with $\boldsymbol{w}$, see [7]. The authors of [24] argue that the well-posedness of the weak formulation of the stochastic problem (1.4)–(1.5) follows from that of the diffusion problem (1.1)–(1.2). The latter is readily established via the Lax-Milgram lemma by assuming that e^{a_M} is positive and bounded almost everywhere in $D \times \Omega$. An SGFEM comprising bilinear finite elements on D and global polynomial approximation on the parameter domain $\Gamma = \Gamma_1 \times \cdots \times \Gamma_M$ is also discussed in [24]. Numerical investigation of the mesh Peclet numbers reveals that stabilisation is not required for representative test problems.

1.2. Darcy flux reformulation. In flow modelling, it is usually of interest to approximate the Darcy flux $\boldsymbol{q} := -e^{a_M}\nabla u$. By introducing the variable $\boldsymbol{q}$, (1.1)–(1.2) can also be reformulated as the first-order system: find $\boldsymbol{q} : D \times \Omega \to \mathbb{R}^2$, $u : D \times \Omega \to \mathbb{R}$ such that $\mathbb{P}$-a.s.,

$$e^{-a_M}\boldsymbol{q} + \nabla u = 0 \quad \text{in } D, \tag{1.6}$$

$$\nabla \cdot \boldsymbol{q} = f \quad \text{in } D, \tag{1.7}$$

$$u = 0 \quad \text{on } \partial D. \tag{1.8}$$

We shall refer to (1.6)–(1.8) as the *standard mixed formulation* of (1.1)–(1.2). The weak formulation is a saddle point problem whose well-posedness can be established using classical analysis [3] which involves an inf-sup condition. This has been done in [6] and [1], under the assumption that the diffusion coefficient is both positive and bounded on $D \times \Omega$. That is,

$$0 < a_{min} \leq \exp(a_M(\boldsymbol{x},\omega)) \leq a_{max}, \qquad \text{a.e. in } D \times \Omega, \tag{1.9}$$

where a_{min} and a_{max} are constants. To achieve this, the random variables ξ_k in (1.3) must be uniformly bounded. We will assume that the ξ_k are independent and each has a truncated

Gaussian density of the form

$$\rho_k(\xi_k) = (2\Phi(c/s) - 1)^{-1} \times \frac{1}{\sqrt{2\pi}s} e^{-\frac{\xi_k^2}{2s^2}} \times \mathbb{1}_{[-c,c]}(\xi_k), \qquad k = 1, \ldots, M. \tag{1.10}$$

Here, $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution function, $c > 0$ is a cut-off parameter and $s > 0$ is chosen so that $\mathrm{Var}(\xi_k) = 1$. For this choice of ρ_k, we have $\xi_k \in \Gamma_k := [-c, c]$ for $k = 1, \ldots, M$. To discretise (1.6)–(1.8), one possibility is to apply a stochastic Galerkin *mixed* finite element method (SGMFEM) comprising lowest-order Raviart-Thomas mixed finite elements on D and global polynomial approximation on Γ. If (1.9) holds then the stability of this scheme can be established straight-forwardly (see [6] and [1]). As explained in [14], however, applying SGMFEMs to (1.6)–(1.8) leads to linear systems with block dense and ill-conditioned indefinite matrices that are highly expensive to solve.

Our goal is to approximate $\boldsymbol{q}$ efficiently when the diffusion coefficient is the stochastically nonlinear function e^{a_M}. Inspired by [24], we begin by reformulating (1.1)–(1.2). First, we define a rescaled pressure $\widetilde{u} := e^{a_M} u$. Hence, $u = e^{-a_M}\widetilde{u}$ and differentiating gives

$$e^{a_M}\nabla u = e^{a_M}\left(e^{-a_M}\nabla\widetilde{u} - e^{-a_M}\nabla a_M \widetilde{u}\right) = \nabla\widetilde{u} + \boldsymbol{w}\,\widetilde{u}.$$

Substituting this expression into (1.1)–(1.2) gives

$$-\nabla\cdot(\nabla\widetilde{u} + \boldsymbol{w}\,\widetilde{u}) = f \qquad \text{in } D, \tag{1.11}$$

$$\widetilde{u} = 0 \qquad \text{on } \partial D. \tag{1.12}$$

Introducing the flux $\boldsymbol{q} = -e^{a_M}\nabla u = -\nabla\widetilde{u} - \boldsymbol{w}\,\widetilde{u}$ yields the *alternative mixed formulation*: find $\boldsymbol{q} : D \times \Omega \to \mathbb{R}^2$, $\widetilde{u} : D \times \Omega \to \mathbb{R}$ such that $\mathbb{P}$-a.s.,

$$\boldsymbol{q} + \nabla\widetilde{u} + \boldsymbol{w}\,\widetilde{u} = 0 \qquad \text{in } D, \tag{1.13}$$

$$\nabla\cdot\boldsymbol{q} = f \qquad \text{in } D, \tag{1.14}$$

$$\widetilde{u} = 0 \qquad \text{on } \partial D. \tag{1.15}$$

Notice that we can also obtain (1.13)–(1.15) from the standard mixed formulation by multiplying both sides of (1.6) by e^{a_M} and changing variable in (1.6) and (1.8) from u to $\widetilde{u}$. Notice that now, only $\boldsymbol{w}$ (a stochastically linear function) appears on the left hand-side of the system of PDEs. Applying standard SGMFEMs to (1.13)–(1.15) yields block *sparse* coefficient matrices that are better conditioned than those obtained for (1.6)–(1.8), and for which the cost of a matrix-vector product is more manageable as $M \to \infty$.

1.3. Outline. In Section 2 we review some well known results about the deterministic analogue of the boundary-value problem (1.13)–(1.15) and approximation using lowest-order Raviart-Thomas elements. In Section 3 we present the mixed variational formulation of (1.13)–(1.15), which takes the form of a generalised (non-symmetric) saddle point problem, and discuss well-posedness. In Section 4 we apply an SGMFEM that builds on the deterministic finite element method introduced in Section 2 and derive the associated linear systems of equations. In Section 5 we introduce and analyse efficient block triangular and block diagonal preconditioners for use with generalised minimal residual (GMRES) iteration. A key point is that both preconditioners can be implemented by solving only deterministic problems. In Section 6 we present numerical results. Finally, in Section 7 we present our conclusions.

2. The deterministic problem. The deterministic analogue of (1.11)–(1.12), where $\boldsymbol{w}$ is simply a function of $\boldsymbol{x} \in D$, and the associated deterministic first-order system (1.13)–(1.15) have been studied by many authors (see, [18], [5], [21], [17]). Note that when $\nabla \cdot \boldsymbol{w} = 0$, (1.11)–(1.12) is a convection-diffusion problem. Otherwise, it is a convection-diffusion-reaction problem. For the random fields of interest here, $\nabla \cdot \boldsymbol{w} \neq 0$. The well-posedness of the standard weak form of the deterministic version of (1.11)–(1.12) can only be established using the Lax-Milgram lemma if one can prove that $\|\boldsymbol{w}\|_\infty < \infty$ and, for coercivity,

$$1 - \frac{1}{2} K_p \nabla \cdot \boldsymbol{w} > 0, \qquad \text{a.e. in } D, \tag{2.1}$$

where K_p is the Poincaré constant. Often, one may not be able to establish (2.1). Nevertheless, for sufficiently regular and bounded $\boldsymbol{w}$ and $f \in L^2(D)$ one may still show that there exists a unique $\widetilde{u} \in H^2(D) \cap H_0^1(D)$ that satisfies the weak form of (1.11)–(1.12), and that (1.13)–(1.15) is also well-posed. The existence and uniqueness of finite element approximations to (1.11)–(1.12) and (1.13)–(1.15) can usually only be established, however, under the assumption that the mesh size h is sufficiently small (see [18], [5]). Such analysis does not make it clear what happens for large mesh Peclet numbers, that is when $\|\boldsymbol{w}\|_\infty h/2$ is large.

The finite-dimensional variational formulation of the deterministic analogue of (1.13)–(1.15) is : find $(\boldsymbol{q}_h, \widetilde{u}_h) \in \boldsymbol{V}_h \times W_h$ such that

$$(\boldsymbol{q}_h, \boldsymbol{v}) - (\widetilde{u}_h, \nabla \cdot \boldsymbol{v}) + (\widetilde{u}, \boldsymbol{w} \cdot \boldsymbol{v}) = 0, \qquad \forall \boldsymbol{v} \in \boldsymbol{V}_h, \tag{2.2}$$

$$-(\nabla \cdot \boldsymbol{q}_h, z) = -(f, z), \qquad \forall z \in W_h, \tag{2.3}$$

where $(\cdot, \cdot)$ denotes the $L^2(D)$ inner-product, $\boldsymbol{V}_h \subset \boldsymbol{V} := H(div; D)$ and $W_h \subset W := L^2(D)$. In particular, we will consider the lowest-order Raviart-Thomas mixed finite element method. Assume that D can be partitioned into non-overlapping rectangular elements $\square_k$ and let $\mathscr{T}_h$ denote the finite element mesh. The lowest-order Raviart-Thomas space is

$$\boldsymbol{V}_h = \boldsymbol{RT}_0(D) := \left\{ \boldsymbol{v} \in H(div; D) : \boldsymbol{v}\,|_{\square_k} = \begin{bmatrix} a + bx \\ c + dy \end{bmatrix} \text{ for all } \square_k \in \mathscr{T}_h \right\}$$

and we choose $W_h = P_0(D)$, the set of piecewise constant functions on D. In [5], approximations obtained with this scheme are shown to converge, and a priori error estimates are established, but only under the assumption that h is small enough. The following simple numerical experiment illustrates this point well.

Let $D = [0,1] \times [0,1]$ and set $f = 1$. To mimic the structure of $\boldsymbol{w}$ for the stochastic problem of interest here, we set $\boldsymbol{w} = -\nabla a_0$ where $a_0 = 1 + \alpha x^2$. The approximations to the scalar variable $\widetilde{u}$ in (2.2)–(2.3) obtained for $\alpha = 10$ and $\alpha = 100$ using a uniform mesh with $h = 1/32$ are shown in Figure 2.1. Note that the Peclet numbers are 10/32 and 100/32 when $\alpha = 10$ and 100, respectively. As α increases, $\boldsymbol{w}$ exerts a strong convective force in the x-direction and the scalar solution develops a layer at the boundary $x = 1$. When the mesh does not resolve this layer, and the Peclet number is greater than one, the numerical solution has non-physical oscillations. Stabilisation strategies for mixed finite element approximations of the deterministic version of (1.13)–(1.15) are discussed in [21] and [17]. When α, and hence $\|\boldsymbol{w}\|_\infty$ is small, there are no numerical difficulties and stabilisation is not needed.

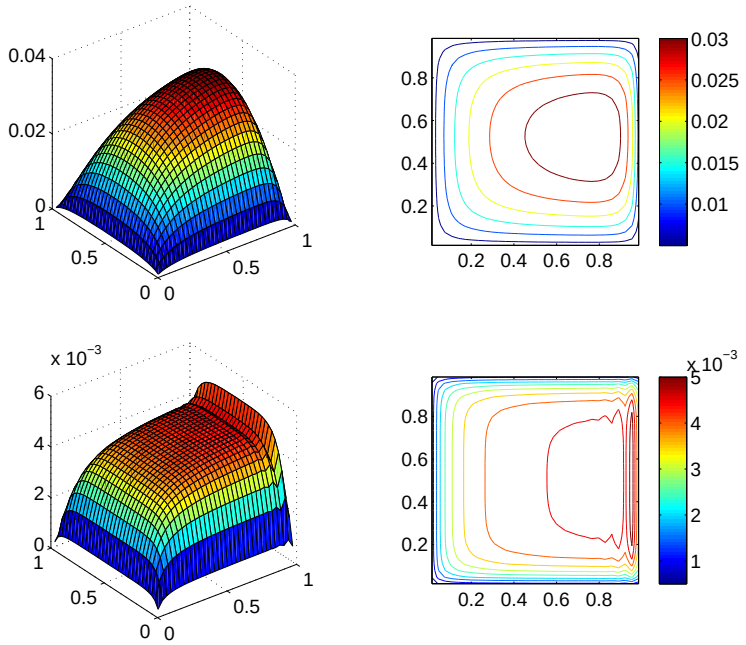

Figure 2.1. *Mixed finite element approximation to $\widetilde{u}$ satisfying (2.2)–(2.3) with $\boldsymbol{w} = -\nabla a_0$ and $a_0 = 1+\alpha x^2$ for $\alpha = 10$ (top), and $\alpha = 100$ (bottom).*

3. Mixed variational problem.

Mixed variational formulations of the stochastic problem (1.13)–(1.15) lead to generalised saddle point problems (gSPs), which have been studied in an abstract setting in [13] and [4]. Let $\boldsymbol{V}$ and W be Hilbert spaces, equipped with norms $\|\cdot\|_{\boldsymbol{V}}$ and $\|\cdot\|_W$, respectively. We use boldface to indicate that $\boldsymbol{V}$ contains vector-valued functions. A gSP is a variational problem of the form: find $(\boldsymbol{q}, u) \in \boldsymbol{V} \times W$ satisfying

$$a(\boldsymbol{q}, \boldsymbol{v}) + b_1(u, \boldsymbol{v}) = g(\boldsymbol{v}), \qquad \forall \boldsymbol{v} \in \boldsymbol{V}, \tag{3.1}$$

$$b_2(z, \boldsymbol{q}) \qquad\quad = f(z), \qquad \forall z \in W, \tag{3.2}$$

where $a : \boldsymbol{V} \times \boldsymbol{V} \to \mathbb{R}$, $b_1, b_2 : W \times \boldsymbol{V} \to \mathbb{R}$, $f : W \to \mathbb{R}$ and $g : \boldsymbol{V} \to \mathbb{R}$. Such problems arise in the study of many physical processes governed by systems of PDEs with two coupled solution variables (e.g., groundwater flow, Stokes flow and Navier–Stokes flow).

Sufficient (but not necessary) conditions for the well-posedness of gSPs are supplied in [13]. Specifically, it can be shown that (3.1)–(3.2) is well-posed if f and g are bounded, if there exist positive constants α, β_1, β_2 such that:

$$|a(\boldsymbol{q}, \boldsymbol{q})| \leq \alpha \,\|\, \boldsymbol{q} \,\|_{\boldsymbol{V}} \,\|\, \boldsymbol{q} \,\|_{\boldsymbol{V}} \qquad \forall \boldsymbol{q} \in \boldsymbol{V}, \tag{3.3}$$

$$|b_1(z, \boldsymbol{q})| \leq \beta_1 \,\|\, z \,\|_W \,\|\, \boldsymbol{q} \,\|_{\boldsymbol{V}} \qquad \forall \boldsymbol{q} \in \boldsymbol{V}, \forall z \in W, \tag{3.4}$$

$$|b_2(z, \boldsymbol{q})| \leq \beta_2 \,\|\, z \,\|_W \,\|\, \boldsymbol{q} \,\|_{\boldsymbol{V}} \qquad \forall \boldsymbol{q} \in \boldsymbol{V}, \forall z \in W, \tag{3.5}$$

if the following coercivity conditions hold:

$$\sup_{\boldsymbol{v} \in \boldsymbol{Z}_1} \frac{a(\boldsymbol{q}, \boldsymbol{v})}{\|\, \boldsymbol{v} \,\|_{\boldsymbol{V}}} \geq c_1 \,\|\, \boldsymbol{q} \,\|_{\boldsymbol{V}} \quad \forall \boldsymbol{q} \in \boldsymbol{Z}_2, \qquad \sup_{\boldsymbol{q} \in \boldsymbol{Z}_2} a(\boldsymbol{q}, \boldsymbol{v}) > 0 \quad \forall \boldsymbol{v} \in \boldsymbol{Z}_1 \setminus \{\boldsymbol{0}\}, \tag{3.6}$$

for some $c_1 > 0$, where

$$\boldsymbol{Z}_1 := \{\boldsymbol{q} \in \boldsymbol{V} \,|\, b_1(z, \boldsymbol{q}) = 0 \quad \forall\, z \in W\}, \quad \boldsymbol{Z}_2 := \{\boldsymbol{q} \in \boldsymbol{V} \,|\, b_2(z, \boldsymbol{q}) = 0 \quad \forall\, z \in W\}, \tag{3.7}$$

and finally, if the following inf-sup conditions hold:

$$\sup_{\boldsymbol{q} \in \boldsymbol{V}} \frac{b_1(z, \boldsymbol{q})}{\| \boldsymbol{q} \|_{\boldsymbol{V}}} \geq \gamma_1 \| z \|_W \qquad \forall z \in W, \tag{3.8}$$

$$\sup_{\boldsymbol{q} \in \boldsymbol{V}} \frac{b_2(z, \boldsymbol{q})}{\| \boldsymbol{q} \|_{\boldsymbol{V}}} \geq \gamma_2 \| z \|_W \qquad \forall z \in W, \tag{3.9}$$

for $\gamma_1, \gamma_2 > 0$. These conditions coincide with the classical ones [3] for symmetric saddle point problems when $b_1(\cdot, \cdot) = b_2(\cdot, \cdot)$ and $\boldsymbol{Z}_1 = \boldsymbol{Z}_2$. A more general class of gSPs is discussed in [4].

After changing co-ordinates from $\boldsymbol{\xi}(\omega)$ to $\boldsymbol{y} \in \Gamma$ in the usual way, the mixed variational formulation of (1.13)–(1.15) is : find $(\boldsymbol{q}, \widetilde{u}) \in \boldsymbol{V} \times W$ such that

$$\mathbb{E}\left[(\boldsymbol{q}, \boldsymbol{v})\right] - \mathbb{E}\left[(\widetilde{u}, \nabla \cdot \boldsymbol{v})\right] + \mathbb{E}\left[(\widetilde{u}, \boldsymbol{w} \cdot \boldsymbol{v})\right] = 0, \qquad \forall \boldsymbol{v} \in \boldsymbol{V}, \tag{3.10}$$

$$-\mathbb{E}\left[(\nabla \cdot \boldsymbol{q}, z)\right] = -\mathbb{E}\left[(f, z)\right], \qquad \forall z \in W, \tag{3.11}$$

where $(\cdot, \cdot)$ denotes the $L^2(D)$ inner-product and $\mathbb{E}\left[\cdot\right] = \int_\Gamma \rho(\boldsymbol{y}) \cdot d\boldsymbol{y}$. The appropriate spaces are $\boldsymbol{V} := L^2_\rho(\Gamma, H(div; D))$ and $W := L^2_\rho(\Gamma, L^2(D))$. The natural norm on $\boldsymbol{V}$ is defined by

$$\| \boldsymbol{q} \|^2_{\boldsymbol{V}} = \mathbb{E}\left[\| \boldsymbol{q} \|^2_{L^2(D)}\right] + \mathbb{E}\left[\| \nabla \cdot \boldsymbol{q} \|^2_{L^2(D)}\right] = \mathbb{E}\left[\| \boldsymbol{q} \|^2_{H(div;D)}\right],$$

and the norm on W is defined by

$$\| z \|^2_W = \mathbb{E}\left[\| z \|^2_{L^2(D)}\right].$$

Clearly, (3.10)–(3.11) is a gSP with $a\left(\boldsymbol{q}, \boldsymbol{v}\right) := \mathbb{E}\left[(\boldsymbol{q}, \boldsymbol{v})\right]$, $b_2\left(z, \boldsymbol{v}\right) := -\mathbb{E}\left[(z, \nabla \cdot \boldsymbol{v})\right]$ and

$$b_1\left(z, \boldsymbol{v}\right) : = b_2\left(z, \boldsymbol{v}\right) + \mathbb{E}\left[(z, \boldsymbol{w} \cdot \boldsymbol{v})\right], \tag{3.12}$$

together with $g\left(\boldsymbol{v}\right) := 0$ and $f\left(z\right) := -\mathbb{E}\left[(f, z)\right]$.

Now, since

$$|f(z)| \leq \|f\|_{L^2_\rho(\Gamma, L^2(D))} \|z\|_{L^2_\rho(\Gamma, L^2(D))} = \|f\|_{L^2(D)} \|z\|_W, \tag{3.13}$$

$f : W \to \mathbb{R}$ is bounded if $f \in L^2(D)$. It is straight-forward to show that (3.3) holds with $\alpha = 1$ and for any $z \in W$ and $\boldsymbol{q} \in \boldsymbol{V}$, we have

$$|b_2\left(z, \boldsymbol{q}\right)| \leq \mathbb{E}\left[\| z \|^2_{L^2(D)}\right]^{1/2} \mathbb{E}\left[\| \nabla \cdot \boldsymbol{q} \|^2_{L^2(D)}\right]^{1/2} \leq \| z \|_W \| \boldsymbol{q} \|_{\boldsymbol{V}}.$$

Hence (3.4) holds with $\beta_1 = 1$. We also have

$$\begin{aligned} |b_1\left(z, \boldsymbol{q}\right)| &\leq \| z \|_W \| \boldsymbol{q} \|_{\boldsymbol{V}} + \| \boldsymbol{w} \|_\infty \| z \|_W \ \mathbb{E}\left[\| \boldsymbol{q} \|^2_{L^2(D)}\right]^{1/2} \\ &\leq (1 + \| \boldsymbol{w} \|_\infty) \| z \|_W \| \boldsymbol{q} \|_{\boldsymbol{V}}, \end{aligned}$$

where we define,

$$\| \boldsymbol{w} \|_\infty := \max_{i=1,2} \left(\operatorname*{ess\,sup}_{(\boldsymbol{x},\boldsymbol{y})\in D\times\Gamma} |w_i(\boldsymbol{x},\boldsymbol{y})| \right).$$

Hence (3.5) holds with $\beta_2 = 1 + \| \boldsymbol{w} \|_\infty$, provided $\|\boldsymbol{w}\|_\infty = \|\nabla a_M\|_\infty < \infty$.

Now, the null spaces of $b_1(\cdot,\cdot)$ and $b_2(\cdot,\cdot)$ are

$$\boldsymbol{Z}_1 := \{\boldsymbol{q} \in \boldsymbol{V} \,|\, \mathbb{E}\left[(z, \nabla\cdot\boldsymbol{q})\right] = \mathbb{E}\left[(z, \boldsymbol{w}\cdot\boldsymbol{q})\right] \quad \forall\, z \in W\}, \tag{3.14}$$

$$\boldsymbol{Z}_2 := \{\boldsymbol{q} \in \boldsymbol{V} \,|\, \mathbb{E}\left[(z, \nabla\cdot\boldsymbol{q})\right] = 0 \quad \forall\, z \in W\}. \tag{3.15}$$

Since $\nabla\cdot\boldsymbol{V} \subset W$, we have

$$\|\nabla\cdot\boldsymbol{q}\|^2_{L^2_\rho(\Gamma,L^2(D))} = \mathbb{E}\left[\| \nabla\cdot\boldsymbol{q} \|^2_{L^2(D)}\right] = 0, \qquad \forall \boldsymbol{q} \in \boldsymbol{Z}_2, \tag{3.16}$$

so $\boldsymbol{Z}_2$ contains divergence-free fluxes. However, there is no straight-forward interpretation of $\boldsymbol{Z}_1$ and no clear way to use the abstract analysis from [13] to establish the well-posedness of (3.10)–(3.11). It is worth noting however that the condition (3.9) associated with $b_2(\cdot,\cdot)$ is the usual inf-sup condition for the standard mixed formulation (1.6)–(1.8). It is known (see [1] and references therein) that this holds with γ_2 depending only on the domain D.

Fortunately, we may exploit the fact that the gSP (3.10)–(3.11) is a reformulation of the *standard* problem (1.6)–(1.8) which is well-posed, provided e^{a_M} is bounded on $D\times\Gamma$, with

$$0 < a_{min} \le \exp(a_M(\boldsymbol{x},\boldsymbol{y})) \le a_{max}, \qquad \text{a.e. in } D\times\Gamma. \tag{3.17}$$

That is, there exists a unique $(\boldsymbol{q}, u) \in \boldsymbol{V}\times W$ such that

$$\mathbb{E}\left[\left(e^{-a_M}\boldsymbol{q}, \boldsymbol{v}\right)\right] - \mathbb{E}\left[(u, \nabla\cdot\boldsymbol{v})\right] = 0, \qquad \forall \boldsymbol{v} \in \boldsymbol{V}, \tag{3.18}$$

$$-\mathbb{E}\left[(\nabla\cdot\boldsymbol{q}, z)\right] = -\mathbb{E}\left[(f, z)\right], \qquad \forall z \in W. \tag{3.19}$$

If (3.17) holds then $\widetilde{u} = e^{a_M}u \in W$. Hence, there exists a unique $(\boldsymbol{q}, \widetilde{u}) \in \boldsymbol{V}\times W$ such that

$$\mathbb{E}\left[\left(e^{-a_M}\boldsymbol{q}, \boldsymbol{v}\right)\right] - \mathbb{E}\left[\left(e^{-a_M}\widetilde{u}, \nabla\cdot\boldsymbol{v}\right)\right] = 0, \qquad \forall \boldsymbol{v} \in \boldsymbol{V}, \tag{3.20}$$

$$-\mathbb{E}\left[(\nabla\cdot\boldsymbol{q}, z)\right] = -\mathbb{E}\left[(f, z)\right], \qquad \forall z \in W. \tag{3.21}$$

If (3.17) holds then for any $\boldsymbol{v} \in \boldsymbol{V}$, we also have $\widehat{\boldsymbol{v}} = e^{a_M}\boldsymbol{v} \in \boldsymbol{V}$. Hence, there exists a unique $(\boldsymbol{q}, \widetilde{u}) \in \boldsymbol{V}\times W$ such that

$$\mathbb{E}\left[\left(e^{-a_M}\boldsymbol{q}, e^{a_M}\boldsymbol{v}\right)\right] - \mathbb{E}\left[\left(e^{-a_M}\widetilde{u}, \nabla\cdot(e^{a_M}\boldsymbol{v})\right)\right] = 0, \qquad \forall \boldsymbol{v} \in \boldsymbol{V}, \tag{3.22}$$

$$-\mathbb{E}\left[(\nabla\cdot\boldsymbol{q}, z)\right] = -\mathbb{E}\left[(f, z)\right], \qquad \forall z \in W. \tag{3.23}$$

Hence, we conclude that there exists a unique $(\boldsymbol{q}, \widetilde{u}) \in \boldsymbol{V}\times W$ satisfying (3.10)–(3.11). It is essential here that $e^{a_M}\boldsymbol{v} \in \boldsymbol{V}$ and $e^{a_M}w \in W$ for all $\boldsymbol{v} \in \boldsymbol{V}$ and $w \in W$. Unfortunately the same argument does not hold when we replace $\boldsymbol{V}$ and W with finite-dimensional spaces.

3.1. Numerical investigation of convective velocity. In this section, we investigate numerically the mesh Peclet numbers associated with two stochastic test problems. Again, we choose $D = [0,1] \times [0,1]$ and $f = 1$. We consider the isotropic covariance function

$$C(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{\sigma^2}{2} \left(\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2}{\ell} \right)^2 K_2 \left(\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2}{\ell} \right),$$

where K_2 is the usual Bessel function, σ is the standard deviation and ℓ is the correlation length. This is the so-called Whittle-Matérn covariance function with parameter $\nu = 2$. Mean-zero random fields $a : D \times \Omega \to \mathbb{R}$ with this covariance are n times mean-square differentiable with $n < 2$. We select two values for ℓ and choose M in (1.3) such that 97% of the variance of a is incorporated in the truncated expansion a_M in (1.3).

Let $\boldsymbol{x}_j$ denote the centroid of the jth finite element and define $\boldsymbol{w}_j(\omega) := \boldsymbol{w}(\boldsymbol{x}_j, \omega)$. We investigate the element Peclet numbers

$$P_j(\omega) := \frac{\|\boldsymbol{w}_j(\omega)\|_\infty h}{2},$$

(which are random variables) where $\|\boldsymbol{w}_j(\omega)\|_\infty = \max\{\alpha_{j,1}, \alpha_{j,2}\}$ and

$$\alpha_{j,i} = \operatorname*{ess\,sup}_{\omega \in \Omega} \left| \frac{\partial a_0(\boldsymbol{x}_j)}{\partial x_i} + \sigma \sum_{k=1}^{M} \sqrt{\lambda_k} \frac{\partial a_k(\boldsymbol{x}_j)}{\partial x_i} \xi_k(\omega) \right|, \qquad i = 1, 2.$$

For each element in the finite element mesh, we estimate the probability

$$Pr_j := \mathbb{P}\left(\omega \in \Omega : P_j(\omega) \leq 1\right)$$

using the standard Monte Carlo method (with 10^5 samples of $\boldsymbol{w}_j(\omega)$) as well as the expected element Peclet number $\mathbb{E}\left[P_j\right]$.

Test problem 1. In the first test problem, we fix the mean to be $a_0 = 0$. In this case a is an isotropic random field and there is no directional dependence in its realisations. The vector field $\nabla a_0 = 0$ also obviously has no strong directional component. In Table 3.1 we record the minimum value of Pr_j over all elements in the mesh and in Table 3.2 we record the maximum value of $\mathbb{E}\left[P_j\right]$. The results indicate that in this test problem the mesh Peclet number is less than or equal to one on all elements, with probability one. In addition, the highest expected mesh Peclet number is observed to be $O(\sigma h)$. This is intuitive since

$$\alpha_{j,i} \leq c\sigma \sum_{k=1}^{M} \sqrt{\lambda_k} \left| \frac{\partial a_k(\boldsymbol{x}_j)}{\partial x_i} \right|, \qquad i = 1, 2.$$

One might therefore query whether for a very large value of σ and a coarse enough mesh, we would obtain $\mathbb{E}\left[P_j\right] > 1$. However, since σ is the standard deviation of a, we can be assured that $\sigma \ll h^{-1}$ in physical applications.

Test problem 2. In the second test problem, we set $a_0 = 1 + 10x^2$ so that $\nabla a_0 = (20x, 0)^\top$. In Table 3.3 we record the minimum value of Pr_j over all elements in the mesh and in Table 3.4

Table 3.1
Test problem 1: Minimum value of Pr_j over all elements in an $n \times n$ mesh.

	$M = 6$, $\ell = 1$			$M = 10$, $\ell = 0.7$		
$n = h^{-1}$	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$
16	1	1	1	1	1	1
32	1	1	1	1	1	1
64	1	1	1	1	1	1

Table 3.2
Test problem 1: Maximum value of $\mathbb{E}[P_j]$ over all elements in an $n \times n$ mesh.

	$M = 6$, $\ell = 1$			$M = 10$, $\ell = 0.7$		
$n = h^{-1}$	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$
16	0.0062	0.0623	0.1246	0.0089	0.0887	0.1777
32	0.0031	0.0312	0.0625	0.0044	0.0445	0.0889
64	0.0016	0.0156	0.0312	0.0022	0.0223	0.0446

we record the maxium value of $\mathbb{E}[P_j]$. The results indicate that for all but one combination of σ and h, the mesh Peclet number is less than or equal to one on all elements, with probability one. In addition, the highest expected mesh Peclet number is $O(h)$. For a fixed h, it is $O(1)$ with respect to σ. In this case,

$$\alpha_{j,1} \leq 20x_{j,1} + c\sigma \sum_{k=1}^{M} \sqrt{\lambda_k} \left| \frac{\partial a_k(\boldsymbol{x}_j)}{\partial x_i} \right|, \qquad \alpha_{j,2} \leq c\sigma \sum_{k=1}^{M} \sqrt{\lambda_k} \left| \frac{\partial a_k(\boldsymbol{x}_j)}{\partial x_i} \right|,$$

and so close to the boundary $x_1 = 1$ the contribution from ∇a_0 dominates. When $h^{-1} = 16$ and $\sigma = 2$, there are elements where $Pr_j < 1$. Here, h is too small compared to ∇a_0 and σ.

Table 3.3
Test problem 2: Minimum value of Pr_j over all elements in an $n \times n$ mesh.

	$M = 6$, $\ell = 1$			$M = 10$, $\ell = 0.7$		
$n = h^{-1}$	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$
16	1	1	1	1	1	0.9995
32	1	1	1	1	1	1
64	1	1	1	1	1	1

In summary, the above experiments tell us that when $\nabla a_0 = 0$ (when the mean of the diffusion coefficient is a constant and hence a is isotropic) it is unlikely that a physically relevant value of σ will lead to a convection dominated problem. However, when ∇a_0 is large in some region of D, then care should be taken.

Table 3.4
Test Problem 2: Maximum value of $\mathbb{E}[P_j]$ over all elements in an $n \times n$ mesh.

	$M = 6$, $\ell = 1$			$M = 10$, $\ell = 0.7$		
$n = h^{-1}$	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$
16	0.6055	0.6056	0.6052	0.6055	0.6054	0.6055
32	0.3076	0.3077	0.3076	0.3076	0.3076	0.3077
64	0.1550	0.1550	0.1551	0.1550	0.1550	0.1551

4. Finite-dimensional mixed variational formulation. We now apply a SGMFEM and solve (1.13)–(1.15) using finite-dimensional spaces. Specifically, we combine the lowest-order Raviart-Thomas scheme from Section 2 (square $\boldsymbol{RT}_0$–P_0 elements with edge length h) with global polynomial approximation on Γ. Hence, the spatial components of the flux are approximated by piecewise linear functions, such that the normal component is continuous across the edges of the finite element mesh, and the spatial component of the rescaled pressure is approximated by a piecewise constant function. Tensorising these finite element spaces,

$$\boldsymbol{V}_h = \operatorname{span}\{\boldsymbol{\varphi}_i(\boldsymbol{x})\}_{i=1}^{n_q} \subset H(div; D), \qquad W_h = \operatorname{span}\{\phi_j(\boldsymbol{x})\}_{j=1}^{n_u} \subset L^2(D),$$

with the set of polynomials of total degree d or less on Γ, denoted by

$$S_d = \operatorname{span}\{\psi_i(\boldsymbol{\xi})\}_{j=1}^{n_\xi} \subset L^2_\rho(\Gamma),$$

we look for approximations $\boldsymbol{q}_{hd} \in \boldsymbol{V}_h \otimes S_d$ and $\widetilde{u}_{hd} \in W_h \otimes S_d$ of the form

$$\boldsymbol{q}_{hd}(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{i=1}^{n_q} \sum_{j=1}^{n_\xi} q_{i,j} \boldsymbol{\varphi}_i(\boldsymbol{x}) \psi_j(\boldsymbol{\xi}), \qquad \widetilde{u}_{hd}(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{i=1}^{n_u} \sum_{j=1}^{n_\xi} \widetilde{u}_{i,j} \phi_i(\boldsymbol{x}) \psi_j(\boldsymbol{\xi}). \tag{4.1}$$

Note that $n_\xi = \dim(S_d) = (M+d)!/d!M!$ The polynomials ψ_k are chosen to be orthonormal with respect to the joint density ρ of the vector of independent random variables $(\xi_1, \ldots, \xi_M)$ and are collectively known as a polynomial chaos. For the truncated Gaussian density (1.10) these polynomials are known as Rys polynomials (see [10] and the references therein).

Now, the finite-dimensional variational formulation of (1.13)–(1.15) is: find $(\boldsymbol{q}_{hd}, \widetilde{u}_{hd}) \in (\boldsymbol{V}_h \otimes S_d) \times (W_h \otimes S_d)$ such that

$$\mathbb{E}\left[(\boldsymbol{q}_{hd}, \boldsymbol{v})\right] - \mathbb{E}\left[(\widetilde{u}_{hd}, \nabla \cdot \boldsymbol{v})\right] + \mathbb{E}\left[(\widetilde{u}_{hd}, \boldsymbol{w} \cdot \boldsymbol{v})\right] = 0, \tag{4.2}$$

$$-\mathbb{E}\left[(\nabla \cdot \boldsymbol{q}_{hd}, z)\right] = -\mathbb{E}\left[(f, z)\right], \tag{4.3}$$

$\forall \boldsymbol{v} \in \boldsymbol{V}_h \otimes S_d$ and $\forall z \in W_h \otimes S_d$. This is the gSP (3.10)–(3.11) with $\boldsymbol{V}$ now replaced by $\boldsymbol{V}_h \otimes S_d$ and W by $W_h \otimes S_d$. Note that the pairing $\boldsymbol{V}_h \otimes S_d$ and $W_h \otimes S_d$ is known to be inf-sup stable for the *standard* mixed formulation (1.6)–(1.8) and hence the inf-sup condition (3.9) is satisfied for these finite-dimensional spaces (again, see [1]). However, as we know, the finite element pairing $\boldsymbol{V}_h \times W_h$ is unstable for the deterministic analogue of (4.2)–(4.3) when the problem is convection-dominated. The pairing $(\boldsymbol{V}_h \otimes S_d) \times (W_h \otimes S_d)$ will therefore

not always provide a stable approximation for (4.2)–(4.3) for arbitrary choices of $\boldsymbol{w}$ and h. However, the results in Section 3.1 do indicate that for the stochastic problems of interest, with $\boldsymbol{w} = -\nabla a_M$ where a_M is a truncated random field with $\nabla a_0 = 0$ or ∇a_0 not too large and $\sigma \ll h^{-1}$, the mesh Peclet number is not likely to be a concern. Hence, stabilisation is not likely to be necessary. Note that if the assumptions on ∇a_M and f from Section 3 hold then (3.3)–(3.5) are satisfied on the finite-dimensional spaces and the linear functionals f and g are bounded, because $W_h \otimes S_d \subset W$ and $\boldsymbol{V}_h \otimes S_d \subset \boldsymbol{V}$.

To set up the linear system for (4.2)–(4.3), we define the mass matrix $A \in \mathbb{R}^{n_q \times n_q}$ by

$$[A]_{i,k} := \int_D \boldsymbol{\varphi}_i \cdot \boldsymbol{\varphi}_k \, d\boldsymbol{x}, \qquad i,k = 1,\dots,n_q, \tag{4.4}$$

which is symmetric and positive definite, and the rectangular matrix $B \in \mathbb{R}^{n_u \times n_q}$ by

$$[B]_{\ell,k} := -\int_D \phi_\ell \, \nabla \cdot \boldsymbol{\varphi}_k \, d\boldsymbol{x} = -\int_{\square_\ell} \nabla \cdot \boldsymbol{\varphi}_k \, d\boldsymbol{x}, \qquad k = 1,\dots,n_q,\ \ell = 1,\dots,n_u,$$

where $\square_\ell$ denotes the ℓth finite element. B is a discrete divergence operator with $\operatorname{rank}(B) = n_u$ and $B^\top$ is a discrete gradient operator. We also define N_0 and $N_1,\dots,N_M \in \mathbb{R}^{n_u \times n_q}$, by

$$[N_0]_{\ell,k} := -\int_D \phi_\ell \nabla a_0 \cdot \boldsymbol{\varphi}_k \, d\boldsymbol{x} = -\int_{\square_\ell} \nabla a_0 \cdot \boldsymbol{\varphi}_k \, d\boldsymbol{x}, \tag{4.5}$$

$$[N_m]_{\ell,k} := -\sigma\sqrt{\lambda_m} \int_D \phi_\ell \nabla a_m \cdot \boldsymbol{\varphi}_k \, d\boldsymbol{x} = -\sigma\sqrt{\lambda_m} \int_{\square_\ell} \nabla a_m \cdot \boldsymbol{\varphi}_k \, d\boldsymbol{x}, \tag{4.6}$$

for $k = 1,\dots,n_q$ and $\ell = 1,\dots,n_u$. Notice that if a_0 in (1.3) (the mean of the logarithm of the diffusion coefficient) is constant, then $\nabla a_0 = 0$ and $N_0 = 0$.

The linear system associated with (4.2)–(4.3) can now be written as

$$\begin{bmatrix} \widehat{A} & \widehat{B}^\top + \widehat{N}^\top \\ \widehat{B} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{q} \\ \widetilde{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{f} \end{bmatrix} \tag{4.7}$$

where $\widehat{A} := I \otimes A$, $\widehat{B} := I \otimes B$, and

$$\widehat{N} := I \otimes N_0 + \sum_{m=1}^{M} G_m \otimes N_m.$$

Here, I denotes the $n_\xi \times n_\xi$ identity matrix and hence $\widehat{A}$ and $\widehat{B}$ are block diagonal. The matrices $G_1,\dots,G_M \in \mathbb{R}^{n_\xi \times n_\xi}$ are defined by

$$[G_m]_{r,s} = \mathbb{E}[y_m \psi_r(\boldsymbol{y}) \psi_s(\boldsymbol{y})], \qquad m = 1,\dots,M, \quad r,s = 1,\dots,n_\xi.$$

These are the so-called stochastically linear G-matrices [9], that always arise in stochastic finite element discretisations of linear PDEs with stochastically linear coefficients. Each G_m is symmetric and indefinite and crucially, has at most two non-zero entries per row. As a result, $\widehat{N}$ is sparse. That is, block sparse with sparse blocks.

The right-hand side vector $\boldsymbol{f}$ in (4.7) has the form

$$\boldsymbol{f} = \begin{bmatrix} \boldsymbol{f}^1 \\ \vdots \\ \boldsymbol{f}^{n_\xi} \end{bmatrix} \tag{4.8}$$

and when f is deterministic, the blocks of this vector are defined by

$$\left[\boldsymbol{f}^k\right]_j = -\mathbb{E}\left[\psi_k\right] \int_{\square_j} f \; d\boldsymbol{x}, \qquad j = 1, \ldots, n_u.$$

Since the polynomials ψ_k satisfy $\mathbb{E}\left[\psi_k\right] = 0$ for $k > 1$ and we assume ψ_1 corresponds to the polynomial of degree zero in each of $y_1, \ldots, y_M$, only the block $\boldsymbol{f}^1$ is non-zero.

The matrix $\widehat{A}$ in (4.7) is symmetric and positive definite (because A is) and it gives a discrete representation of the $L^2_\rho(\Gamma, L^2(D))$ norm on $\boldsymbol{V}_h \otimes S_d$. That is,

$$\mathbb{E}\left[(\boldsymbol{v}_{hd}, \boldsymbol{v}_{hd})\right] = \mathbb{E}\left[\| \boldsymbol{v}_{hd} \|^2_{L^2(D)}\right] = \boldsymbol{v}^\top \widehat{A} \boldsymbol{v}, \tag{4.9}$$

for $\boldsymbol{v}_{hd} \in \boldsymbol{V}_h \otimes S_d$, where $\boldsymbol{v} \in \mathbb{R}^{n_q n_\xi}$ is the vector of coefficients associated with $\boldsymbol{v}_{hd}$ when it is expanded in the chosen basis. We also have

$$b_1(z_{hd}, \boldsymbol{v}_{hd}) = \boldsymbol{z}^\top \widehat{B} \boldsymbol{v}, \qquad b_2(z_{hd}, \boldsymbol{v}_{hd}) = \boldsymbol{z}^\top \left(\widehat{B} + \widehat{N}\right) \boldsymbol{v}.$$

Since B has full row rank, so does $\widehat{B}$ and $\widehat{B}^\top$ has full column rank.

5. Linear algebra and preconditioning. We now discuss how to solve the non-symmetric saddle point system (4.7) with coefficient matrix

$$\widehat{C} := \begin{bmatrix} \widehat{A} & \widehat{B}^\top + \widehat{N}^\top \\ \widehat{B} & 0 \end{bmatrix}. \tag{5.1}$$

We will assume that the diffusion coefficient a_M has been chosen so that the reformulated problem is not convection-dominated and the discretisation parameters h and d have been chosen so that the linear system is uniquely solvable. Since $\widehat{C}$ is non-symmetric and indefinite, we will use GMRES iteration. A preconditioner is required, however, as $\widehat{C}$ is ill-conditioned. The finite element matrices A, B, and N_m are all ill-conditioned with respect to the mesh parameter h. The matrices N_m are also ill-conditioned with respect to the statistical parameters associated with a_M, in particular, the variance σ^2. However, the symmetric matrices G_m each have eigenvalues that lie in the bounded interval $\Gamma_m = [-\pi_{d+1}, \pi_{d+1}]$, where $\pi_{d+1} > 0$ denotes the largest root of the univariate, orthonormal Rys polynomial of exact degree $d+1$ (see e.g. [9]). Note that $\pi_{d+1} \leq c$, where c is the cut-off parameter associated with the truncated Gaussian density in (1.10). Thus spectral bounds for G_m are independent of the number of random variables M and the polynomial degree d.

In the experiments in Section 6 we use right-preconditioned GMRES. Each iteration entails a matrix-vector product with $\widehat{C}$ and the application of the chosen preconditioner. In our log-transformed mixed approximation framework, $\widehat{C}$ is sparse and to perform multiplications with

it we only need to perform multiplications with A, B, N_m and G_m. That is, we do not need to assemble $\widehat{C}$. To compute $\left(G_m \otimes N_m^\top\right) \boldsymbol{v}$ where $\boldsymbol{v} \in \mathbb{R}^{n_\xi n_u}$, we use the standard identity

$$\left(G_m \otimes N_m^\top\right) \boldsymbol{v} = \operatorname{vec}\left(N_m^\top V G_m^\top\right) = \operatorname{vec}\left(N_m^\top \left(G_m V^\top\right)^\top\right),$$

where $V \in \mathbb{R}^{n_u \times n_\xi}$ is the matrix whose columns, when stacked on top of one another, yield $\boldsymbol{v}$. When $M \ll n_\xi$, the cost of performing a matrix vector product with $\widehat{C}$ is $O\left(n_\xi \left(n_q + n_u\right)\right)$, and this scales linearly with respect to the problem size. See Table 5.1 for details.

Table 5.1
The costs of matrix-vector products with the blocks of the saddle point matrix $\widehat{C}$.

$\widehat{A} = I \otimes A$	$O(n_\xi n_q)$
$\widehat{B} = I \otimes B$	$O(n_\xi n_q)$
$\widehat{B}^\top = I \otimes B^\top$	$O(n_\xi n_u)$
$\widehat{N}^\top = I \otimes N_0^\top + \sum_{m=1}^{M} G_m \otimes N_m^\top$	$O((M+1) n_\xi n_u)$

We now investigate two preconditioning strategies. First, we note that the Schur complement associated with $\widehat{C}$ in (5.1) is

$$\widehat{S} := \widehat{B}\widehat{A}^{-1}(\widehat{B} + \widehat{N})^\top = \widehat{B}\widehat{A}^{-1}\widehat{B}^\top + \widehat{B}\widehat{A}^{-1}\widehat{N}^\top = I \otimes (BA^{-1}B^\top) + \widehat{B}\widehat{A}^{-1}\widehat{N}^\top.$$

This is non-symmetric, although the first term $\widehat{B}\widehat{A}^{-1}\widehat{B}^\top$ (which is the Schur complement associated with the standard mixed approximation with unit diffusion coefficient) is both symmetric and positive definite. An *ideal* block triangular preconditioner for $\widehat{C}$ is

$$\widehat{P}_U := \begin{bmatrix} \widehat{A} & \widehat{B}^\top + \widehat{N}^\top \\ 0 & -\widehat{S} \end{bmatrix},$$

whose inverse is given by

$$\widehat{P}_U^{-1} := \begin{bmatrix} \widehat{A}^{-1} & \widehat{A}^{-1}\left(\widehat{B}^\top + \widehat{N}^\top\right)\widehat{S}^{-1} \\ 0 & -\widehat{S}^{-1} \end{bmatrix}.$$

The corresponding right-preconditioned saddle point matrix is

$$\widehat{C}\widehat{P}_U^{-1} = \begin{bmatrix} \widehat{I} & 0 \\ \widehat{B}\widehat{A}^{-1} & \widehat{I} \end{bmatrix}$$

and this has a single eigenvalue $\lambda = 1$. Since the matrix is not diagonalisable, the eigenvalues do not give us information about how GMRES converges, see [11, Chapter 3]. However, since $\widehat{C}\widehat{P}_U^{-1}$ has a minimum polynomial of degree two (see [12]), GMRES with this preconditioner would converge in just two iterations, independently of the problem size and the chosen discretisation and statistical parameters. In this sense, the preconditioner is optimal.

We also consider an ideal block-diagonal preconditioner

$$\widehat{P}_D := \begin{bmatrix} \widehat{A} & 0 \\ 0 & -\widehat{S} \end{bmatrix},$$

which is slightly cheaper to apply than $\widehat{P}_U$. The right-preconditioned matrix in this case is

$$\widehat{C}\widehat{P}_D^{-1} = \begin{bmatrix} \widehat{I} & -(\widehat{B}^\top + \widehat{N}^\top)\widehat{S}^{-1} \\ \widehat{B}\widehat{A}^{-1} & 0 \end{bmatrix}$$

and since $\widehat{B}\widehat{A}^{-1}(\widehat{B}^\top + \widehat{N}^\top)\widehat{S}^{-1} = \widehat{I}$ it is easy to show that $\widehat{C}\widehat{P}_D^{-1}$ has three distinct eigenvalues. These are $\lambda = 1$ and $\lambda = \frac{1}{2}\left(1 \pm \sqrt{3}i\right)$. The preconditioned matrix is diagonalisable in this case (see [12]) and right-preconditioned GMRES would converge in three iterations, again independently of the problem size and the chosen discretisation and statistical parameters.

Notice that applying the action of $\widehat{P}_U^{-1}$ requires solves with $\widehat{A}$ and $\widehat{S}$ and a multiplication with $\widehat{B}^\top + \widehat{N}^\top$, whereas applying the action of $\widehat{P}_D^{-1}$ requires only solves with $\widehat{A}$ and $\widehat{S}$. Recall that $\widehat{A} = I \otimes A$ is block-diagonal, so a solve with $\widehat{A}$ simply requires n_ξ decoupled solves with the sparse mass matrix A. We also know $\widehat{B}^\top + \widehat{N}^\top$ is sparse and that the cost of a multiplication with it scales linearly with respect to the problem size (see Table 5.1). However, performing exact solves with the dense matrix $\widehat{S}$ is infeasible and neither of the ideal preconditioners can be used in practice. Suppose then that we can find a sparse symmetric and positive definite approximation $\widehat{S}_{approx}$ to $\widehat{S}$ for which the action of $\widehat{S}_{approx}^{-1}$ is cheap to apply and consider

$$\widehat{P}_{U,approx} := \begin{bmatrix} \widehat{A} & \widehat{B}^\top + \widehat{N}^\top \\ 0 & -\widehat{S}_{approx} \end{bmatrix}, \qquad \widehat{P}_{D,approx} := \begin{bmatrix} \widehat{A} & 0 \\ 0 & -\widehat{S}_{approx} \end{bmatrix}.$$

The following results characterise the eigenvalues of the preconditioned saddle point systems in terms of the eigenvalues of $\widehat{S}_{approx}^{-1}\widehat{S}$. The proofs of Theorems 5.1 and 5.2 follow standard arguments (e.g., see [7, Chapter 8]) that are commonly applied to non-symmetric saddle point systems associated with deterministic Navier–Stokes problems.

Theorem 5.1. *The eigenvalues λ of the generalised eigenvalue problem*

$$\begin{bmatrix} \widehat{A} & \widehat{B}^\top + \widehat{N}^\top \\ \widehat{B} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{q} \\ \boldsymbol{u} \end{bmatrix} = \lambda \begin{bmatrix} \widehat{A} & \widehat{B}^\top + \widehat{N}^\top \\ 0 & -\widehat{S}_{approx} \end{bmatrix} \begin{bmatrix} \boldsymbol{q} \\ \boldsymbol{u} \end{bmatrix}$$

are $\lambda = 1$ and $\lambda = \nu$ where $\widehat{S}\boldsymbol{u} = \nu\widehat{S}_{approx}\boldsymbol{u}$.

Proof. The two block matrix equations are

$$(1-\lambda)\widehat{A}\boldsymbol{q} = (\lambda - 1)\left(\widehat{B}^\top + \widehat{N}^\top\right)\boldsymbol{u}, \qquad \widehat{B}\boldsymbol{q} = -\lambda\widehat{S}_{approx}\boldsymbol{u}.$$

Either $\lambda = 1$ or $\lambda \neq 1$. If $\lambda \neq 1$ then, since $\widehat{A}$ is symmetric and positive definite, and hence invertible, we can combine the two equations to give

$$\widehat{B}\widehat{A}^{-1}\left(\widehat{B}^\top + \widehat{N}^\top\right)\boldsymbol{u} = \lambda\widehat{S}_{approx}\boldsymbol{u}.$$

$\lambda = 1$ is an eigenvalue of multiplicity $n_\xi\,(n_q - n_u)$ corresponding to eigenvectors with $\boldsymbol{q} \in \text{null}(\widehat{B})$ and $\boldsymbol{u} = \boldsymbol{0}$. If $\widehat{B} = I \otimes B$ has full (row) rank (which is true here), then $\text{rank}(\widehat{B}) = n_\xi n_u$ and by the rank theorem, the dimension of $\text{null}(\widehat{B})$ is $n_\xi\,(n_q - n_u)$. ■

Theorem 5.2. *The eigenvalues λ of the generalised eigenvalue problem*

$$\begin{bmatrix} \widehat{A} & \widehat{B}^\top + \widehat{N}^\top \\ \widehat{B} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{q} \\ \boldsymbol{u} \end{bmatrix} = \lambda \begin{bmatrix} \widehat{A} & 0 \\ 0 & -\widehat{S}_{approx} \end{bmatrix} \begin{bmatrix} \boldsymbol{q} \\ \boldsymbol{u} \end{bmatrix}$$

are $\lambda = 1$ and $\lambda = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - 4\nu}$ where $\widehat{S}\boldsymbol{u} = \nu \widehat{S}_{approx}\boldsymbol{u}$.

Proof. As in Theorem 5.1, there are $n_\xi\,(n_q - n_u)$ eigenvalues $\lambda = 1$. If $\lambda \neq 1$ then,

$$(1 - \lambda)\widehat{A}\boldsymbol{q} = -\left(\widehat{B}^\top + \widehat{N}^\top\right)\boldsymbol{u}, \qquad \widehat{B}\boldsymbol{q} = -\lambda \widehat{S}_{approx}\boldsymbol{u},$$

and we can combine these equations to obtain $\widehat{B}\widehat{A}^{-1}\left(\widehat{B}^\top + \widehat{N}^\top\right)\boldsymbol{u} = \lambda\,(1 - \lambda)\,\widehat{S}_{approx}\boldsymbol{u}$. The result follows by setting $\nu = \lambda\,(1 - \lambda)$ and solving for λ. Note that each of the $n_\xi n_u$ eigenvalues ν yields $2n_\xi n_u$ eigenvalues λ of the preconditioned saddle point system. ■

Theorems 5.1 and 5.2 tell us that if we can choose $\widehat{S}_{approx}$ so that the eigenvalues of $\widehat{S}_{approx}^{-1}\widehat{S}$ are tightly clustered in the complex plane, then we can expect the eigenvalues of both preconditioned systems to also be tightly clustered in the complex plane. $\widehat{P}_{U,approx}$ will produce two clusters and $\widehat{P}_{D,approx}$ will produce three. A natural starting point is to consider

$$\widehat{S}_{approx} = \widehat{B}\widehat{A}^{-1}\widehat{B}^\top = I \otimes S, \qquad S := BA^{-1}B^\top. \tag{5.2}$$

This yields $\nu = 1 + \tau$ where $\widehat{B}\widehat{A}^{-1}\widehat{N}^\top\boldsymbol{u} = \tau\widehat{B}\widehat{A}^{-1}\widehat{B}^\top\boldsymbol{u}$, and it is clear that τ will depend on the statistical information about a_M that is encoded in $\widehat{N}$. We now investigate this. The main result is Theorem 5.5 but first, we establish two preliminary results. Lemma 5.3 makes use of the mass matrix Q_u associated with the finite element space W_h defined by

$$[Q_u]_{i,k} := \int_D \phi_k \phi_i \, d\boldsymbol{x}, \qquad i, k = 1, \ldots, n_u \ .$$

For a vector-valued function $\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{x})$ we also define

$$\|\boldsymbol{z}\|_{2,\infty} := \sup_{\boldsymbol{x} \in D} \|\boldsymbol{z}(\boldsymbol{x})\|_2 \ . \tag{5.3}$$

Lemma 5.3. *Let $\boldsymbol{u} \in \mathbb{R}^{n_u}$, $\boldsymbol{q} \in \mathbb{R}^{n_q}$, then for $m = 1, \ldots, M$,*

$$|\boldsymbol{q}^\top N_m^\top \boldsymbol{u}| \leq \sigma\sqrt{\lambda_m}\|\nabla a_m\|_{2,\infty}(\boldsymbol{u}^\top Q_u \boldsymbol{u})^{1/2}(\boldsymbol{q}^\top A \boldsymbol{q})^{1/2}. \tag{5.4}$$

Moreover,

$$|\boldsymbol{q}^\top N_0^\top \boldsymbol{u}| \leq \|\nabla a_0\|_{2,\infty}(\boldsymbol{u}^\top Q_u \boldsymbol{u})^{1/2}(\boldsymbol{q}^\top A \boldsymbol{q})^{1/2}. \tag{5.5}$$

Proof. For a given vector $\boldsymbol{u} \in \mathbb{R}^{n_u}$ we define the function $u \in W_h$ by $u(\boldsymbol{x}) = \sum_i u_i \phi_i(\boldsymbol{x})$. Likewise, for $\boldsymbol{q} \in \mathbb{R}^{n_q}$ we define $\boldsymbol{s} \in \boldsymbol{V}_h$ by $\boldsymbol{s}(\boldsymbol{x}) = \sum_j q_j \boldsymbol{\varphi}_j(\boldsymbol{x})$. Now, let $\boldsymbol{z}_0 := -\nabla a_0$ and

$\boldsymbol{z}_m := -\sigma\sqrt{\lambda_m}\nabla a_m$, $m = 1, \ldots, M$ and consider the matrix $N_m^\top$ for m fixed. We use the Cauchy Schwarz inequality to obtain

$$|\boldsymbol{q}^\top N_m^\top \boldsymbol{u}| = \left| \int_D u \boldsymbol{z}_m \cdot \boldsymbol{s}\, d\boldsymbol{x} \right| \leq \int_D |u|(z_{m,1}^2 + z_{m,2}^2)^{1/2}(s_1^2 + s_2^2)^{1/2}\, d\boldsymbol{x}$$
$$\leq \|\boldsymbol{z}_m\|_{2,\infty} \left(\int_D u^2\, d\boldsymbol{x} \right)^{1/2} \left(\int_D \boldsymbol{s} \cdot \boldsymbol{s}\, d\boldsymbol{x} \right)^{1/2} = \|\boldsymbol{z}_m\|_{2,\infty} (\boldsymbol{u}^\top Q_u \boldsymbol{u})^{1/2} (\boldsymbol{q}^\top A \boldsymbol{q})^{1/2} .$$

Using the definition of $\boldsymbol{z}_m$ in (5.3) the bounds in (5.4) and (5.5) follow. ■

Our second preliminary result exploits the matrix representation of the inf-sup condition associated with the weak form of the deterministic version of the standard mixed formulation (1.6)–(1.8). With our specific choices for $\boldsymbol{V}_h$ and W_h ($\boldsymbol{RT}_0$–P_0 elements), it is known that there exists a constant $\gamma > 0$ independent of the mesh parameter h satisfying

$$\gamma \|u\|_{L^2(D)} \leq \sup_{\boldsymbol{s} \in \boldsymbol{V}_h} \frac{(u, -\nabla \cdot \boldsymbol{s})}{\|\boldsymbol{s}\|_{H(div;D)}} \qquad \forall u \in W_h . \tag{5.6}$$

For a discussion of this result, see [3] and [15].

Lemma 5.4. *Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{n_u}$, then for $m = 1, \ldots, M$,*

$$|\boldsymbol{v}^\top B A^{-1} N_m^\top \boldsymbol{u}| \leq \gamma^{-1} \sigma \sqrt{\lambda_m} \|\nabla a_m\|_{2,\infty} (\boldsymbol{v}^\top S \boldsymbol{v})^{1/2} (\boldsymbol{u}^\top S \boldsymbol{u})^{1/2}, \tag{5.7}$$

where S is defined in (5.2) and $\gamma > 0$ is the discrete inf-sup constant in (5.6). Moreover,

$$|\boldsymbol{v}^\top B A^{-1} N_0^\top \boldsymbol{u}| \leq \gamma^{-1} \|\nabla a_0\|_{2,\infty} (\boldsymbol{v}^\top S \boldsymbol{v})^{1/2} (\boldsymbol{u}^\top S \boldsymbol{u})^{1/2}. \tag{5.8}$$

Proof. Consider the matrix $BA^{-1}N_m^\top$ for m fixed. Given $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{n_u}$ we may choose $\boldsymbol{q} = A^{-1}B^\top \boldsymbol{v}$ in (5.4) and (5.5). This yields

$$|\boldsymbol{v}^\top B A^{-1} N_m^\top \boldsymbol{u}| \leq \|\boldsymbol{z}_m\|_{2,\infty} (\boldsymbol{u}^\top Q_u \boldsymbol{u})^{1/2} (\boldsymbol{v}^\top S \boldsymbol{v})^{1/2}, \tag{5.9}$$

where $\boldsymbol{z}_m$ is defined as in the proof of Lemma 5.3. It remains to bound $\boldsymbol{u}^\top Q_u \boldsymbol{u}$ in terms of $\boldsymbol{u}^\top S \boldsymbol{u}$. For the chosen pairing $\boldsymbol{V}_h \times W_h$, (5.6) holds. Hence, $\exists \gamma > 0$ such that

$$\gamma \|u\|_{L^2(D)} \leq \sup_{\boldsymbol{s} \in \boldsymbol{V}_h} \frac{(u, -\nabla \cdot \boldsymbol{s})}{\|\boldsymbol{s}\|_{L^2(D)}} \qquad \forall u \in W_h . \tag{5.10}$$

For a vector $\boldsymbol{u} \in \mathbb{R}^{n_u}$ we define $u \in W_h$ by $u(\boldsymbol{x}) = \sum_i u_i \phi_i(\boldsymbol{x})$ and for $\boldsymbol{q} \in \mathbb{R}^{n_q}$ we define $\boldsymbol{s} \in \boldsymbol{V}_h$ by $\boldsymbol{s}(\boldsymbol{x}) = \sum_j q_j \boldsymbol{\varphi}_j(\boldsymbol{x})$. Then, the estimate in (5.10) reads

$$\gamma\, (\boldsymbol{u}^\top Q_u \boldsymbol{u})^{1/2} \leq \max_{\boldsymbol{q} \in \mathbb{R}^{n_q}} \frac{\boldsymbol{u}^\top B \boldsymbol{q}}{(\boldsymbol{q}^\top A \boldsymbol{q})^{1/2}} = \max_{\boldsymbol{y} = A^{1/2}\boldsymbol{q}} \frac{\boldsymbol{u}^\top B A^{-1/2} \boldsymbol{y}}{(\boldsymbol{y}^\top \boldsymbol{y})^{1/2}} = \frac{\boldsymbol{u}^\top S \boldsymbol{u}}{(\boldsymbol{u}^\top S \boldsymbol{u})^{1/2}}$$
$$= (\boldsymbol{u}^\top S \boldsymbol{u})^{1/2} . \tag{5.11}$$

Combining (5.11) and (5.9) gives the desired results, for $m = 0, 1, \ldots, M$. ■

We can now provide a bound for the complex eigenvalues of $\widehat{S}_{approx}^{-1}\widehat{S}$.

Theorem 5.5. *For $\widehat{S}_{approx} = \widehat{B}\widehat{A}^{-1}\widehat{B}^\top$ the eigenvalues ν of the generalised eigenvalue problem $\widehat{S}\boldsymbol{u} = \nu\widehat{S}_{approx}\boldsymbol{u}$ are contained in the circle*

$$\left\{z \in \mathbb{C}\colon |z-1| \leq 2\,\gamma^{-1}\,\delta\right\}, \qquad \delta := \|\nabla a_0\|_{2,\infty} + c\,\sigma\sum_{m=1}^{M}\sqrt{\lambda_m}\|\nabla a_m\|_{2,\infty}, \tag{5.12}$$

where c is the cut-off parameter in (1.10) and $\gamma > 0$ is the discrete inf-sup constant in (5.6).

Proof. Let $\nu \in \mathbb{C}$ and $\boldsymbol{u} \in \mathbb{C}^{n_u n_\xi}\backslash\{\boldsymbol{0}\}$ satisfy $\widehat{S}\boldsymbol{u} = \nu\widehat{S}_{approx}\boldsymbol{u}$. Noting that $\widehat{S}_{approx}$ is symmetric and positive definite, we have the generalised Raleigh quotient

$$\begin{aligned}\nu &= \frac{\boldsymbol{u}^H\widehat{S}\boldsymbol{u}}{\boldsymbol{u}^H\widehat{S}_{approx}\boldsymbol{u}} = \frac{\boldsymbol{u}^H\widehat{S}_{approx}\boldsymbol{u} + \boldsymbol{u}^H\widehat{B}\widehat{A}^{-1}\widehat{N}^\top\boldsymbol{u}}{\boldsymbol{u}^H\widehat{S}_{approx}\boldsymbol{u}}\\ &= 1 + \frac{\boldsymbol{u}^H I\otimes BA^{-1}N_0^\top\boldsymbol{u} + \sum_{m=1}^{M}\boldsymbol{u}^H G_m\otimes BA^{-1}N_m^\top\boldsymbol{u}}{\boldsymbol{u}^H I\otimes S\boldsymbol{u}}\end{aligned}$$

where S is defined in (5.2). Hence we obtain the bound

$$|\nu-1| \leq \left|\frac{\boldsymbol{u}^H I\otimes BA^{-1}N_0^\top\boldsymbol{u}}{\boldsymbol{u}^H I\otimes S\boldsymbol{u}}\right| + \sum_{m=1}^{M}\left|\frac{\boldsymbol{u}^H G_m\otimes BA^{-1}N_m^\top\boldsymbol{u}}{\boldsymbol{u}^H I\otimes S\boldsymbol{u}}\right|. \tag{5.13}$$

We aim to bound $|\boldsymbol{u}^H I\otimes BA^{-1}N_0^\top\boldsymbol{u}|$ and $|\boldsymbol{u}^H G_m\otimes BA^{-1}N_m^\top\boldsymbol{u}|$, $m = 1,\dots,M$, in terms of $|\boldsymbol{u}^H I\otimes S\boldsymbol{u}|$. By linearity and the properties of the Kronecker product it suffices to establish a bound for vectors of the form $\boldsymbol{u} = \boldsymbol{u}_\ell\otimes\boldsymbol{u}_r$, where $\boldsymbol{u}_\ell \in \mathbb{C}^{n_\xi}$ and $\boldsymbol{u}_r \in \mathbb{C}^{n_u}$. Observe that

$$\boldsymbol{u}^H I\otimes BA^{-1}N_0^\top\boldsymbol{u} = (\boldsymbol{u}_\ell^H\boldsymbol{u}_\ell)(\boldsymbol{u}_r^H BA^{-1}N_0^\top\boldsymbol{u}_r)$$

and $\boldsymbol{u}^H I\otimes S\boldsymbol{u} = (\boldsymbol{u}_\ell^H\boldsymbol{u}_\ell)(\boldsymbol{u}_r^H S\boldsymbol{u}_r)$. Hence it suffices to bound $|\boldsymbol{u}_r^H BA^{-1}N_0^\top\boldsymbol{u}_r|$ in terms of $\boldsymbol{u}_r^H S\boldsymbol{u}_r$. Analogously, $\boldsymbol{u}^H G_m\otimes BA^{-1}N_m^\top\boldsymbol{w} = (\boldsymbol{u}_\ell^H G_m\boldsymbol{u}_\ell)(\boldsymbol{u}_r^H BA^{-1}N_m^\top\boldsymbol{u}_r)$. Thus, it is sufficient to bound $|\boldsymbol{u}_r^H BA^{-1}N_0^\top\boldsymbol{u}_r|$ in terms of $\boldsymbol{u}_r^H S\boldsymbol{u}_r$. In addition, we require a bound for $|\boldsymbol{u}_\ell^H G_m\boldsymbol{u}_\ell|$. We proceed by collecting the required results.

In Lemma 5.4 we have proved bounds of the form

$$|\boldsymbol{v}^\top BA^{-1}N_m^\top\boldsymbol{u}| \leq c_m(\boldsymbol{v}^\top S\boldsymbol{v})^{1/2}(\boldsymbol{u}^\top S\boldsymbol{u})^{1/2}, \qquad m = 0,1,\dots,M, \tag{5.14}$$

with $c_0 := \gamma^{-1}\|\nabla a_0\|_{2,\infty}$, and $c_m := \gamma^{-1}\sigma\sqrt{\lambda_m}\|\nabla a_m\|_{2,\infty}$, for $m = 1,\dots,M$. Now, decomposing $\boldsymbol{u}_r = \boldsymbol{a} + i\boldsymbol{b}$, $\boldsymbol{a},\boldsymbol{b} \in \mathbb{R}^{n_u}$, and utilising (5.14) it follows

$$\begin{aligned}|\boldsymbol{u}_r^H BA^{-1}N_m^\top\boldsymbol{u}_r| &\leq |\boldsymbol{a}^\top BA^{-1}N_m^\top\boldsymbol{a}| + |\boldsymbol{b}^\top BA^{-1}N_m^\top\boldsymbol{b}| + |\boldsymbol{a}^\top BA^{-1}N_m^\top\boldsymbol{b}| + |\boldsymbol{b}^\top BA^{-1}N_m^\top\boldsymbol{a}|\\ &\leq c_m\left(\boldsymbol{a}^\top S\boldsymbol{a} + \boldsymbol{b}^\top S\boldsymbol{b} + 2(\boldsymbol{a}^\top S\boldsymbol{a})^{1/2}(\boldsymbol{b}^\top S\boldsymbol{b})^{1/2}\right)\\ &\leq c_m\left(\boldsymbol{a}^\top S\boldsymbol{a} + \boldsymbol{b}^\top S\boldsymbol{b} + \boldsymbol{a}^\top S\boldsymbol{a} + \boldsymbol{b}^\top S\boldsymbol{b}\right)\\ &= 2c_m(\boldsymbol{a}^\top S\boldsymbol{a} + \boldsymbol{b}^\top S\boldsymbol{b}) = 2c_m\,\boldsymbol{u}_r^H S\boldsymbol{u}_r\ .\end{aligned}$$

In summary, we have the estimates

$$|\boldsymbol{u}_r^H B A^{-1} N_m^\top \boldsymbol{u}_r| \leq 2c_m \, \boldsymbol{u}_r^H S \boldsymbol{u}_r, \qquad m = 0, 1, \ldots, M. \tag{5.15}$$

Finally, the spectrum of G_m is contained in the interval $[-\pi_{d+1}, \pi_{d+1}]$, where π_{d+1} denotes the largest root of the univariate Rys polynomial of degree $d+1$ (see, e.g., [9]). Since $\pi_{d+1} \leq c$ we have $|\boldsymbol{u}_\ell^H G_m \boldsymbol{u}_\ell| \leq c \boldsymbol{u}_\ell^H \boldsymbol{u}_\ell$. Combining this with (5.15) and (5.13) gives

$$|\nu - 1| \leq 2\gamma^{-1} \left(\|\nabla a_0\|_{2,\infty} + c\,\sigma \sum_{m=1}^{M} \sqrt{\lambda_m} \|\nabla a_m\|_{2,\infty} \right),$$

which completes the proof. ■

Theorem 5.5 tells us that when we choose $\widehat{S}_{approx} = I \otimes S$, the eigenvalues of $\widehat{S}_{approx}^{-1} \widehat{S}$ are bounded in a circle in the complex plane whose radius depends, in particular on σ but not on the discretisation parameters h and d of the chosen SGMFEM.

Now, to reduce the cost of implementing our preconditioners $\widehat{P}_{U,approx}$ and $\widehat{P}_{D,approx}$, we will also choose an approximation $\widehat{A}_{approx}$ to $\widehat{A}$. In particular, we choose $\widehat{A}_{approx} = I \otimes A_{diag}$, where A_{diag} is the diagonal of A. Hence,

$$\widehat{A}_{approx}^{-1} \widehat{A} = I \otimes A_{diag}^{-1} A.$$

Since A is a mass matrix, this approximation is optimal with respect to h, see [8]. If we also use this approximation to $\widehat{A}$ in the Schur complement approximation we obtain

$$\widehat{S}_{approx} := \widehat{B} \widehat{A}_{diag}^{-1} \widehat{B}^\top = I \otimes S_0, \qquad S_0 := B A_{diag}^{-1} B^\top.$$

$\widehat{S}_{approx}$ is now block-diagonal and S_0 is sparse. In the next result, we obtain a bound for the eigenvalues of the preconditioned Schur complement with this new approximation.

Theorem 5.6. *If uniform meshes of square $\boldsymbol{RT}_0$–P_0 elements are used for the spatial discretisation then, for $\widehat{S}_{approx} = I \otimes S_0$ the eigenvalues ν of the generalised eigenvalue problem $\widehat{S}\boldsymbol{u} = \nu \widehat{S}_{approx} \boldsymbol{u}$ are contained in the rectangle*

$$\left\{ z \in \mathbb{C} \colon |\mathrm{Re}(z) - 4/3| \leq 2/3 + 4\gamma^{-1}\delta, \, |\mathrm{Im}(z)| \leq 4\gamma^{-1}\delta \right\}, \tag{5.16}$$

where γ is the inf-sup constant of the pairing $\boldsymbol{V}_h \times W_h$ in (5.6), and δ is defined in (5.12).

Proof. Let $\nu \in \mathbb{C}$ and $\boldsymbol{u} \in \mathbb{C}^{n_u n_\xi} \backslash \{\boldsymbol{0}\}$ satisfy $\widehat{S}\boldsymbol{u} = \nu \widehat{S}_{approx} \boldsymbol{u}$ and recall $S := B A^{-1} B^\top$. The generalised Raleigh quotient associated with ν reads

$$\nu = \frac{\boldsymbol{u}^H \widehat{S} \boldsymbol{u}}{\boldsymbol{u}^H \widehat{S}_{approx} \boldsymbol{u}} = \frac{\boldsymbol{u}^H I \otimes S \boldsymbol{u}}{\boldsymbol{u}^H I \otimes S_0 \boldsymbol{u}} + \frac{\boldsymbol{u}^H \widehat{B} \widehat{A}^{-1} \widehat{N}^\top \boldsymbol{u}}{\boldsymbol{u}^H I \otimes S_0 \boldsymbol{u}}.$$

When uniform meshes of square $\boldsymbol{RT}_0$–P_0 elements are used, the eigenvalues of $A_{diag}^{-1} A$ are contained in the bounded interval $[1/2, 3/2]$, see [8]. Hence, the first term in the expression on the right-hand side is contained in the interval $[2/3, 2]$ on the real line, because S and S_0 are

symmetric positive definite, and because the eigenvalues of $S_0^{-1}S$ are contained in the interval $[2/3, 2]$. The second term can be bounded as follows

$$\left|\frac{\boldsymbol{u}^H\widehat{B}\widehat{A}^{-1}\widehat{N}^\top\boldsymbol{u}}{\boldsymbol{u}^H I\otimes S_0\boldsymbol{u}}\right| = \frac{|\boldsymbol{u}^H\widehat{B}\widehat{A}^{-1}\widehat{N}^\top\boldsymbol{u}|}{\boldsymbol{u}^H I\otimes S\boldsymbol{u}}\times\frac{\boldsymbol{u}^H I\otimes S\boldsymbol{u}}{\boldsymbol{u}^H I\otimes S_0\boldsymbol{u}} \leq 2\gamma^{-1}\delta\times 2\ .$$

Here, we have used again the eigenvalue bound for $S_0^{-1}S$, and, following the lines of proof of Theorem 5.5, obtained a bound for $|\boldsymbol{u}^H\widehat{B}\widehat{A}^{-1}\widehat{N}^\top\boldsymbol{u}|/(\boldsymbol{u}^H I\otimes S\boldsymbol{u})$. The result follows. ■

Theorem 5.6 tells us that when we choose the cheaper approximation $\widehat{S}_{approx} = I\otimes S_0$, the eigenvalues of $\widehat{S}_{approx}^{-1}\widehat{S}$ are again bounded in a region in the complex plane whose size depends, in particular on the standard deviation σ of the diffusion coefficient, but not on the discretisation parameters h and d of the chosen SGMFEM.

We now return to the consideration of the preconditioned saddle point matrices. For diagonalisable preconditioned matrices, it is known that if the eigenvalues are bounded in an ellipse in the complex plane which does not contain the origin, then the asymptotic convergence factor for GMRES (the rate by which the residual error is reduced at each iteration) is bounded by a constant that depends on the size of the ellipse, see [11, Chapter 3]. However, this argument does not help us much here and so we do not pursue a full eigenvalue analysis of the preconditioned systems. Nevertheless, based on the above analysis of the chosen Schur complement approximation, one would expect GMRES convergence to be sensitive to the statistical parameter σ but not to the discretisation parameters.

Using $\widehat{A}_{approx} = I\otimes A_{diag}$ and $\widehat{S}_{approx} = I\otimes S_0$, our practical preconditioners are

$$\widehat{P}_{U,approx} := \begin{bmatrix} I\otimes A_{diag} & I\otimes B^\top + \widehat{N}^\top \\ 0 & -I\otimes S_0 \end{bmatrix},\quad \widehat{P}_{D,approx} := \begin{bmatrix} I\otimes A_{diag} & 0 \\ 0 & -I\otimes S_0 \end{bmatrix}. \quad (5.17)$$

We now consider the costs of implementing them. Computing the action of $\widehat{P}_{U,approx}^{-1}$ involves:

(i) n_ξ solves with the $n_q\times n_q$ diagonal matrix A_{diag},
(ii) n_ξ solves with the $n_u\times n_u$ sparse matrix S_0,
(iii) n_ξ multiplications with the $n_q\times n_u$ sparse matrix $B^\top$,
(iv) a multiplication with the sparse matrix $\widehat{N}^\top$,

whereas implementing $\widehat{P}_{D,approx}$ in each GMRES iteration requires only the operations in (i) and (ii). Note that since S_0 is a discrete representation of the Laplace operator, the solves with S_0 can be done inexactly using any number of off-the-shelf optimal solvers for elliptic problems. In particular, we will approximate the action of S_0^{-1} by applying a single V-cycle of algebraic multigrid (AMG, [22]). The resulting costs associated with $\widehat{P}_{U,approx}$ and $\widehat{P}_{D,approx}$ are shown in Table 5.2. In summary, the theoretical cost of applying both preconditioners is $O(n_\xi\,(n_q+n_u))$ and hence scales linearly with respect to the problem size. Note that all the operations associated with the application of the preconditioners can be performed in a completely decoupled way, by manipulating the component matrices A, $B^\top$, N_m and G_m, $m=1,\ldots,M$. Moreover, there are several possibilities to paralellise the computations.

Table 5.2 suggests that $\widehat{P}_{D,approx}$ is cheaper to apply than $\widehat{P}_{U,approx}$ per iteration. However, from existing preconditioning studies (see [7, Theorem 8.2]) for non-symmetric saddle point

Table 5.2
Costs associated with applying the preconditioners $\widehat{P}_{U,approx}$ and $\widehat{P}_{D,approx}$.

Solve with $I \otimes A_{diag}$	$O(n_\xi n_q)$
Solve with $I \otimes S_0$ using AMG for solves with S_0	$O(n_\xi n_u)$
Multiplication with $\widehat{B}^\top = I \otimes B^\top$	$O(n_\xi n_u)$
Multiplication with $\widehat{N}^\top = I \otimes N_0^\top + \sum_{m=1}^{M} G_m \otimes N_m^\top$	$O((M+1)n_\xi n_u)$

systems associated with deterministic systems of PDEs (such as Navier–Stokes equations) we anticipate that the upper-triangular preconditioner will yield lower GMRES iteration counts. We now investigate this.

6. Numerical results. All experiments reported below were performed on a single processor of a four processor quad-core Linux machine with 8 GB RAM using MATLAB 8.3.

6.1. Preconditioned GMRES. We solve the equations (4.2)–(4.3) corresponding to test problems 1 and 2 defined in Section 3.1. Recall, in test problem 1 we have $\nabla a_0 = 0$ and the diffusion coefficient a is isotropic and in test problem 2, we have $\nabla a_0 \neq 0$. We explore the performance of right-preconditioned GMRES in conjunction with the block triangular preconditioner $\widehat{P}_{U,approx}$ and the block diagonal preconditioner $\widehat{P}_{D,approx}$ in (5.17). The stopping criterion is $\|\boldsymbol{r}_k\|_2 < 10^{-8}\|\boldsymbol{b}\|_2$, where $\boldsymbol{r}_k$ denotes the residual error at the kth step and $\boldsymbol{b}$ is the right-hand side vector of the linear system. The initial guess is always $\boldsymbol{x}_0 = \boldsymbol{0}$.

Table 6.1
Test problem 1: GMRES iteration counts with the preconditioners $\widehat{P}_{U,approx}$ and $\widehat{P}_{D,approx}$.

		$M=6$, $\ell=1$						$M=10$, $\ell=0.7$					
		$\widehat{P}_{U,approx}$			$\widehat{P}_{D,approx}$			$\widehat{P}_{U,approx}$			$\widehat{P}_{D,approx}$		
$n=h^{-1}$	d	σ=0.1	1.0	2.0	0.1	1.0	2.0	0.1	1.0	2.0	0.1	1.0	2.0
32	1	25	27	28	36	38	41	25	28	30	36	39	43
64	-	25	27	29	36	37	40	25	28	30	36	39	42
128	-	25	27	29	35	37	39	25	28	30	35	38	41
32	2	25	28	31	36	39	44	25	30	34	36	42	48
64	-	25	28	31	36	39	43	25	30	34	36	41	48
128	-	25	29	31	35	39	43	25	30	35	35	41	48
32	3	25	30	33	36	41	48	26	32	37	36	44	55
64	-	25	30	34	36	41	47	25	32	38	36	44	55
128	-	25	30	34	35	41	48	25	32	39	35	44	56
32	4	25	31	35	36	42	51	26	33	41	36	46	62
64	-	25	31	36	36	42	51	25	33	41	36	46	62
128	-	25	31	36	35	42	53	25	33	42	36	46	64

First, we investigate the robustness of the preconditioners with respect to h, d, σ, and

M. GMRES iteration counts are recorded in Tables 6.1 and 6.2. In all cases, the iteration counts are completely insensitive to the mesh width h. For $\sigma \leq 1$ the iteration counts are also almost insensitive to M and d. For $\sigma = 2$, however, they are a little sensitive to those parameters. This is consistent with the discussion in Section 5, since the spectral inclusion bounds for the Schur complement approximation depend on σ and will contain the origin for some σ large enough. Note that the solves with S_0 were performed with AMG. When these solves are performed exactly there is essentially no difference in the iteration counts.

Table 6.2

Test problem 2: GMRES iteration counts with the preconditioners $\widehat{P}_{U,approx}$ and $\widehat{P}_{D,approx}$.

		$M=6,\ \ell=1$						$M=10,\ \ell=0.7$					
		$\widehat{P}_{U,approx}$			$\widehat{P}_{D,approx}$			$\widehat{P}_{U,approx}$			$\widehat{P}_{D,approx}$		
$n=h^{-1}$	d	σ=0.1	1.0	2.0	0.1	1.0	2.0	0.1	1.0	2.0	0.1	1.0	2.0
32	1	32	33	35	50	54	57	32	34	36	50	56	60
64	-	34	35	36	52	56	59	34	36	38	52	57	62
128	-	35	36	37	53	57	60	35	37	39	53	59	63
32	2	33	36	39	52	60	67	33	37	40	52	62	71
64	-	35	38	40	54	62	69	35	38	42	54	64	73
128	-	36	39	41	55	63	70	36	39	43	56	65	75
32	3	33	37	41	53	63	73	33	38	45	54	66	80
64	-	35	39	43	54	65	76	35	40	46	55	69	81
128	-	36	40	44	56	67	77	36	41	47	56	70	83
32	4	33	38	44	54	65	79	33	40	48	54	69	88
64	-	35	40	45	55	67	81	35	41	50	56	71	90
128	-	36	41	46	56	69	83	36	42	51	57	73	93

Now, we compare the performances of the two preconditioners. In Tables 6.1 and 6.2 we see that the number of iterations for $\widehat{P}_{U,approx}$ is less than for $\widehat{P}_{D,approx}$ in all tests. The difference in iteration counts is also more pronounced for test problem 2 where $\nabla a_0 \neq 0$. Observe that in this case our chosen Schur complement approximation is not as good as for test problem 1 since δ in (5.12) involves the additional term $\|\nabla a_0\|_{2,\infty} > 0$.

Table 6.3

Average preconditioner time (in seconds) per GMRES iteration divided by n_ξ. The mesh width is $h = 1/64$.

	$M=6,\ \ell=1$				$M=10,\ \ell=0.7$			
	$d=1$	$d=2$	$d=3$	$d=4$	$d=1$	$d=2$	$d=3$	$d=4$
$\widehat{P}_{U,approx}$	0.0039	0.0028	0.0031	0.0031	0.0039	0.0037	0.0039	0.0040
$\widehat{P}_{D,approx}$	0.0028	0.0018	0.0017	0.0016	0.0023	0.0016	0.0016	0.0016
n_ξ	7	28	84	210	11	66	286	1,001

In Table 6.3 we record the average time required to apply the preconditioners, divided by

n_ξ. We see that the timings scale linearly with n_ξ. This is consistent with the discussion in Section 5. As expected, the timings for $\widehat{P}_{D,approx}$ are lower than those for $\widehat{P}_{U,approx}$. However, this does not mean that the total GMRES iteration time is lower for $\widehat{P}_{D,approx}$. Looking at Table 6.4, we see that the total solve time with $\widehat{P}_{U,approx}$ is consistently lower across the range of values chosen for d, σ, and M. This is because the GMRES iteration count is significantly lower with $\widehat{P}_{U,approx}$. Since the cost of an iteration increases as the iteration number increases in (unrestarted) GMRES, it is cheaper overall to perform a small number of more expensive preconditioned iterations then a larger number of less expensive ones.

Table 6.4

Total GMRES iteration time (in seconds). In this example, $M = 6$, $\ell = 1$, and $h = 1/64$.

		$\widehat{P}_{U,approx}$				$\widehat{P}_{D,approx}$			
	σ	d=1	d=2	d=3	d=4	d=1	d=2	d=3	d=4
$a_0 = 0$	0.1	1.5	4.1	14.2	37.9	1.8	5.3	18.6	46.0
	1.0	1.4	4.7	17.8	46.7	1.6	5.7	21.0	53.8
	2.0	1.4	5.4	19.8	55.0	1.8	6.5	25.0	70.4
$a_0 = 1 + 10x^2$	0.1	1.8	6.2	21.1	54.0	2.5	8.8	30.6	78.2
	1.0	1.9	6.8	23.9	64.8	2.7	10.6	37.6	99.2
	2.0	1.9	7.2	28.6	76.4	2.9	11.9	46.7	131.8

6.2. Comparison of formulations. Finally, we compare our SGMFEM approximation $\boldsymbol{q}_{hd}$ to the solution of (1.13)–(1.15) to numerical solutions of (1.6)–(1.8). One approximation to (1.6)–(1.8) is obtained with a Monte Carlo mixed finite element (MCMFEM) with 10^5 samples, and a second is obtained with the same SGMFEM used for (1.13)–(1.15). In Table 6.5 we present the relative errors for the expected value and variance of the Darcy flux $\boldsymbol{q}$. We consider test problem 2 where $\nabla a_0 \neq 0$. We see that our proposed alternative mixed formulation delivers an approximation of essentially the same accuracy as both a MCMFEM approximation and a SGMFEM approximation of the solution to the standard formulation. The costs associated with the alternative mixed formulation are significantly lower, however. Recall that the coefficient matrix $\widehat{C}$ is block sparse (with sparse blocks), whereas the coefficient matrix associated with the SGMFEM for the standard mixed formulation is block dense. The matrix-vector multiplications are much cheaper in the log-transformed setting. To illustrate this key point, we present in Table 6.6 the average time (divided by n_ξ) for a matrix-vector product with the system matrices in both formulations. The cost for the SGMFEM on the log-transformed problem is optimal; it grows linearly with n_ξ. In contrast, the cost for the SGMFEM in the standard formulation increases rapidly as d and M increase.

7. Conclusions. To avoid the computational bottleneck encountered when solving the diffusion problem with stochastically nonlinear coefficients by SGMFEMs, we proposed a novel strategy for reformulating the problem as a stochastically linear one. The weak form of the so-called log-transformed problem is a generalised saddle point problem. We applied a SGMFEM and introduced a block triangular and a block diagonal preconditioner for the associated linear

Table 6.5

Test problem 2: Comparison of the SGMFEM solution $\boldsymbol{q}_{hd}$ to a MCMFEM reference solution $\boldsymbol{q}_h^{(std)}$ and the SGMFEM solution $\boldsymbol{q}_{hd}^{(std)}$ of the standard mixed formulation. $M = 6$, $\ell = 1$, and $h = 1/64$.

d	σ	$\frac{\|\mathbb{E}[\boldsymbol{q}_{hd}]-\bar{\boldsymbol{q}}_h^{(std)}\|_\infty}{\|\bar{\boldsymbol{q}}_h^{(std)}\|_\infty}$	$\frac{\|\mathbb{V}[\boldsymbol{q}_{hd}]-s^2(\boldsymbol{q}_h^{(std)})\|_\infty}{\|s^2(\boldsymbol{q}_h^{(std)})\|_\infty}$	$\frac{\|\mathbb{E}[\boldsymbol{q}_{hd}]-\mathbb{E}[\boldsymbol{q}_{hd}^{(std)}]\|_\infty}{\|\mathbb{E}[\boldsymbol{q}_{hd}^{(std)}]\|_\infty}$	$\frac{\|\mathbb{V}[\boldsymbol{q}_{hd}]-\mathbb{V}[\boldsymbol{q}_{hd}^{(std)}]\|_\infty}{\|\mathbb{V}[\boldsymbol{q}_{hd}^{(std)}]\|_\infty}$
1	0.1	7.0206×10^{-3}	6.8489×10^{-3}	7.0208×10^{-3}	3.7159×10^{-3}
2	-	7.0206×10^{-3}	6.9741×10^{-3}	7.0206×10^{-3}	3.9410×10^{-3}
3	-	7.0206×10^{-3}	6.9742×10^{-3}	7.0209×10^{-3}	3.9099×10^{-3}
4	-	7.0206×10^{-3}	6.9742×10^{-3}	7.0209×10^{-3}	3.9632×10^{-3}
1	1.0	7.0971×10^{-3}	2.6212×10^{-2}	7.0870×10^{-3}	1.5237×10^{-1}
2	-	7.1013×10^{-3}	5.8551×10^{-3}	7.0752×10^{-3}	4.1331×10^{-3}
3	-	7.1011×10^{-3}	6.3210×10^{-3}	7.0752×10^{-3}	4.2027×10^{-3}
4	-	7.1011×10^{-3}	6.3078×10^{-3}	7.0774×10^{-3}	3.7696×10^{-3}
1	2.0	7.2308×10^{-3}	9.8295×10^{-2}	1.0378×10^{-2}	6.2904×10^{-1}
2	-	7.2937×10^{-3}	1.0801×10^{-2}	7.2594×10^{-3}	6.3715×10^{-2}
3	-	7.2878×10^{-3}	3.8782×10^{-3}	7.2156×10^{-3}	9.2148×10^{-3}
4	-	7.2882×10^{-3}	3.7711×10^{-3}	7.2237×10^{-3}	4.2949×10^{-3}

Table 6.6

Average matrix-vector product time (in seconds) divided by n_ξ for the standard and alternative mixed formulations. The mesh width is $h = 1/64$.

	$M = 6$, $\ell = 1$				$M = 10$, $\ell = 0.7$			
	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 1$	$d = 2$	$d = 3$	$d = 4$
standard	0.0037	0.0257	0.1211	0.4271	0.0072	0.1374	1.2938	-
alternative	0.0020	0.0020	0.0025	0.0028	0.0029	0.0036	0.0042	0.0044
n_ξ	7	28	84	210	11	66	286	1,001

systems. We partially analysed the preconditioners by obtaining spectral inclusion bounds for an efficient Schur complement approximation. The bounds are insensitive to the discretisation parameters h and d and only slightly sensitive to the number of random variables M and the standard deviation σ of the log-transformed diffusion coefficient. Numerical tests showed that the block triangular preconditioner outperforms the block diagonal one in terms of both GMRES iteration counts, and total solve time. The availability of a robust and cheap iterative solver for the log-transformed problem means that it is possible to solve some stochastically nonlinear problems as efficiently as stochastically linear ones with stochastic Galerkin methods.

REFERENCES

[1] A. Bespalov, C. E. Powell, and D. Silvester, *A priori error analysis of stochastic Galerkin mixed approximations of elliptic PDEs with random data*, SIAM J. Numer. Anal., 50 (2012), pp. 2039–2063.

[2] ———, *Energy norm a posteriori error estimation for parametric operator equations*, SIAM J. Sci. Com-

put., 36 (2013), pp. A339–A363.
[3] F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, vol. 15 of Springer Series in Computational Mathematics, Springer-Verlag, New York, 1991.
[4] P. Ciarlet, Jr., J. Huang, and J. Zou, *Some observations on generalized saddle-point problems*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 224–236.
[5] J. Douglas, Jr. and J. E. Roberts, *Mixed finite element methods for second order elliptic problems*, Mat. Apl. Comput., 1 (1982), pp. 91–103.
[6] H. C. Elman, D. G. Furnival, and C. E. Powell, *H(div) preconditioning for a mixed finite element formulation of the diffusion problem with random data*, Math. Comp., 79 (2010), pp. 733–760.
[7] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2005.
[8] 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*, SIAM J. Sci. Comput., 31 (2008/09), pp. 1424–1447.
[9] O. G. Ernst and E. Ullmann, *Stochastic Galerkin matrices*, SIAM J. Matrix Anal. Appl., 31 (2009/10), pp. 1848–1872.
[10] W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*, Oxford University Press, Oxford, 2004.
[11] A. Greenbaum, *Iterative methods for solving linear systems*, vol. 17 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
[12] M. F. Murphy, G. H. Golub, and A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput., 21 (2000), pp. 1969–1972 (electronic).
[13] R. A. Nicolaides, *Existence, uniqueness and approximation for generalized saddle point problems*, SIAM J. Numer. Anal., 19 (1982), pp. 349–357.
[14] C. E. Powell and E. Ullmann, *Preconditioning stochastic Galerkin saddle point systems*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2813–2840.
[15] P.-A. Raviart and J. M. Thomas, *A mixed finite element method for 2nd order elliptic problems*, in Mathematical aspects of finite element methods (Proc. Conf., Consiglio Naz. delle Ricerche (C.N.R.), Rome, 1975), Springer, Berlin, 1977, pp. 292–315. Lecture Notes in Math., Vol. 606.
[16] E. Rosseel and S. Vandewalle, *Iterative Solvers for the stochastic finite element method*, SIAM J. Sci. Comput., 32 (2010), pp. 372–397.
[17] R. Sacco and F. Saleri, *Stabilization of mixed finite elements for convection-diffusion problems*, CWI Quarterly, 10 (1997), pp. 301–315. International Workshop on the Numerical Solution of Thin-layer Phenomena (Amsterdam, 1997).
[18] A. H. Schatz, *An observation concerning Ritz-Galerkin methods with indefinite bilinear forms*, Math. Comp., 28 (1974), pp. 959–962.
[19] B. Sousedik and R.G. Ghanem, *Truncated hierarchical preconditioning for the stochastic Galerkin FEM*, Int. J. Uncertain. Quantif., 4 (2014), pp. 333–348.
[20] B. Sousedik, R.G. Ghanem, and E.T. Phipps, *Hierarchical Schur complement preconditioner for the stochastic Galerkin finite element methods*, Numer. Linear Algebra Appl., 21 (2014), pp. 136–151.
[21] J.-M. Thomas, *Mixed finite elements methods for convection-diffusion problems*, in Numerical approximation of partial differential equations (Madrid, 1985), vol. 133 of North-Holland Math. Stud., North-Holland, Amsterdam, 1987, pp. 241–250.
[22] U. Trottenberg, C. W. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, Inc., San Diego, CA, 2001. With contributions by A. Brandt, P. Oswald and K. Stüben.
[23] E. Ullmann, *A Kronecker product preconditioner for stochastic Galerkin finite element discretizations*, SIAM J. Sci. Comput., 32 (2010), pp. 923–946.
[24] E. Ullmann, H. C. Elman, and O. G. Ernst, *Efficient iterative solvers for stochastic Galerkin discretizations of log-transformed random diffusion problems*, SIAM J. Sci. Comput., 34 (2012), pp. A659–A682.