Efficient Solvers for Stochastic Finite Element Saddle Point Problems

Catherine E. Powell

c.powell@manchester.ac.uk

School of Mathematics

University of Manchester, UK

Collaborators:

- Elisabeth Ullmann,
  University of Freiberg, Germany.

- Andrew Gordon,
  University of Manchester, UK.

MANCHESTER 1824
The University of Manchester

Deterministic PDEs

Galerkin mixed finite element approximation for two field PDE problems on $D \subset \mathbb{R}^{2,3}$, leads to variational problems of the form

find $u_h(\boldsymbol{x}) \in V_h$ and $p_h(\boldsymbol{x}) \in W_h$ such that:

$$\begin{aligned} a\left(u_h, v\right) + b\left(p_h, v\right) &= g\left(v\right), \qquad \forall v \in V_h \\ b\left(w, u_h\right) &= f\left(w\right), \qquad \forall w \in W_h \end{aligned}$$

and saddle point systems

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{p} \end{pmatrix} = \begin{pmatrix} \underline{g} \\ \underline{f} \end{pmatrix}$$

for which the study of efficient solvers is mature.

PDEs with Random Data

Stochastic Galerkin mixed finite element approximation for two field PDE problems with random data, on $D \times \Gamma$, with $\Gamma \subset \mathbb{R}^M$, leads to

find $u_{hd}(\boldsymbol{x}, \boldsymbol{\xi}) \in V_h \otimes S_d$ and $p_{hd}(\boldsymbol{x}, \boldsymbol{\xi}) \in W_h \otimes S_d$ such that:

$$\begin{aligned} \hat{a}\left(u_{hd}, v\right) + \hat{b}\left(p_{hd}, v\right) &= \hat{g}\left(v\right), \qquad \forall v \in V_h \otimes S_d \\ \hat{b}\left(w, u_{hd}\right) &= \hat{f}\left(w\right), \qquad \forall w \in W_h \otimes S_d \end{aligned}$$

and saddle point systems

$$\begin{pmatrix} \hat{A} & \hat{B}^T \\ \hat{B} & 0 \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{p} \end{pmatrix} = \begin{pmatrix} \underline{g} \\ \underline{f} \end{pmatrix}$$

which are much larger and more complicated to solve.

Model Problem: Darcy Flow

$$
\begin{aligned}
-T\nabla p &= \boldsymbol{u}, && \text{in } D\\
\nabla\cdot\boldsymbol{u} &= 0, && \text{in } D\\
p &= g, && \text{on } \partial D_D\\
T\nabla p\cdot\boldsymbol{n} &= 0, && \text{on } \partial D_N = \partial D \backslash \partial D_D
\end{aligned}
$$

Solution variables:

p	$=$	$p(\boldsymbol{x})$	hydraulic head (pressure)
$\boldsymbol{u}$	$=$	$\boldsymbol{u}(\boldsymbol{x})$	velocity field

Data:

T	$=$	$T(\boldsymbol{x})$	transmissivity coefficients
g	$=$	$g(\boldsymbol{x})$	boundary data
D	$\subset$	$\mathbb{R}^d$	spatial domain

Outline

- Stochastic Galerkin methods for:

$$T(\boldsymbol{x},\omega)^{-1}\,\boldsymbol{u}(\boldsymbol{x},\omega) - \nabla p(\boldsymbol{x},\omega) = 0,\ \nabla\cdot\boldsymbol{u}(\boldsymbol{x},\omega) = 0$$

  ▷ Solving single very large saddle point system

- Stochastic collocation methods for:

$$T(\boldsymbol{x},\omega)^{-1}\,\boldsymbol{u}(\boldsymbol{x},\omega) - \nabla p(\boldsymbol{x},\omega) = 0,\ \nabla\cdot\boldsymbol{u}(\boldsymbol{x},\omega) = 0$$

  ▷ Solving many small deterministic saddle point systems (see A. Gordon's poster)

Stochastic PDE Problem

Let $T(\boldsymbol{x}, \omega) : D \times \Omega \rightarrow \mathbb{R}$ be a correlated random field.

Approximate $T^{-1}(\boldsymbol{x}, \omega)$ by a function $T_M^{-1}(\boldsymbol{x}, \boldsymbol{\xi})$ involving a finite number of random variables $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_M)$ taking values in $\Gamma = \prod_{i=1}^{M} \Gamma_i \subset \mathbb{R}^M$.

Find $p(\boldsymbol{x}, \boldsymbol{\xi}),\ \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{\xi})$ such that P-a.s.

$$
\begin{aligned}
T_M^{-1}(\boldsymbol{x}, \boldsymbol{\xi})\, \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{\xi}) - \nabla p(\boldsymbol{x}, \boldsymbol{\xi}) &= 0, \\
\nabla \cdot \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{\xi}) &= 0 && \text{in } D \times \Gamma, \\
p(\boldsymbol{x}, \boldsymbol{\xi}) &= g(\boldsymbol{x}) && \text{on } \partial D_D \times \Gamma, \\
\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{\xi}) \cdot \boldsymbol{n} &= 0 && \text{on } \partial D_N \times \Gamma.
\end{aligned}
$$

Finite-dimensional Noise Assumption

Popular choices are:

- Truncated Karhunen-Loève expansion (linear)

$$T_M^{-1}(\boldsymbol{x}, \boldsymbol{\xi}) = \mu(\boldsymbol{x}) + \sum_{k=1}^{M} t_k(\boldsymbol{x})\, \xi_k, \qquad t_k(\boldsymbol{x}) = \sqrt{\lambda_k} c_k(\boldsymbol{x})$$

$\{\lambda_k, c_k(\boldsymbol{x})\}, k = 1, 2, \ldots$ are the eigenvalues and eigenfunctions of covariance

$\{\xi_1, \xi_2, \ldots\}$ are uncorrelated, with mean zero and unit variance

- Truncated Polynomial Chaos expansion (nonlinear)

$$T_M^{-1}(\boldsymbol{x}, \boldsymbol{\xi}(\omega)) = \mu(\boldsymbol{x}) + \sum_{k=1}^{N} t_k(\boldsymbol{x})\, \psi_k(\boldsymbol{\xi}), \qquad t_k(\boldsymbol{x}) = \left\langle T^{-1} \psi_k(\boldsymbol{\xi}) \right\rangle$$

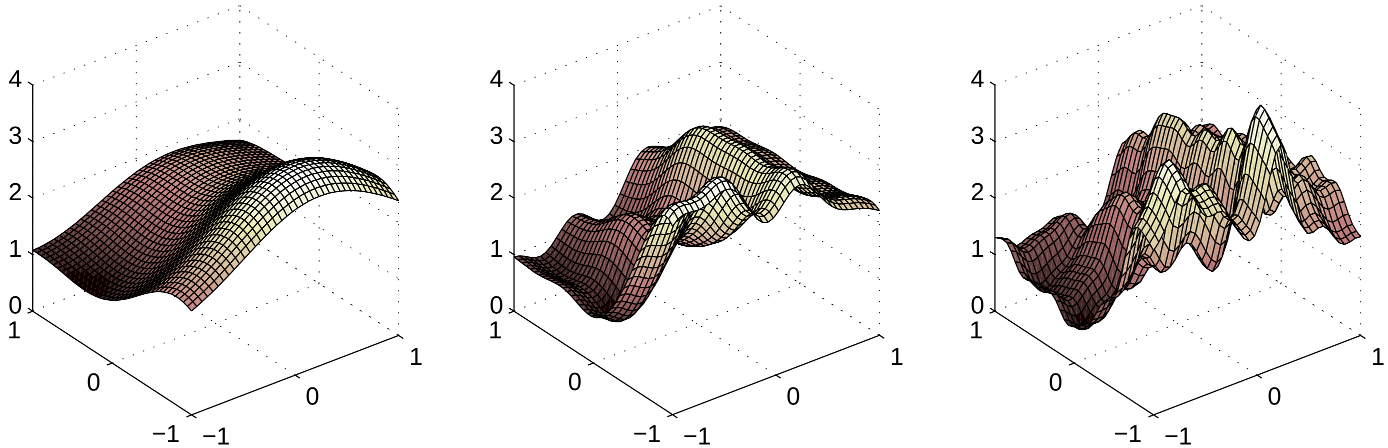
Example - KL expansion

Consider the covariance function

$$C_{T^{-1}}(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(- \parallel \boldsymbol{x} - \boldsymbol{y} \parallel_1\right)$$

$D = [0,1] \times [0,1]$ and Gaussian random variables.

Realisations of T_M^{-1} with $M = 5, 20, 50$, with $\mu(\boldsymbol{x}) = 0$



Polynomial Chaos

If the random variables are independent then the joint density function has the form:

$$\rho(\boldsymbol{\xi}) = \prod_i \rho_i(\xi_i)$$

and

$$E\left[g(\boldsymbol{\xi})\right] = \langle g(\boldsymbol{\xi})\rangle = \int_\Gamma \rho(\boldsymbol{y})g(\boldsymbol{y})\,d\boldsymbol{y}.$$

Orthonormal polynomials in $\boldsymbol{\xi}$ are constructed via

$$\psi_i(\boldsymbol{\xi}) = \prod_{s=1}^{M} \psi_{i_s}(\xi_s)$$

where $\psi_{i_s}(\xi_s)$ is a univariate polynomial of degree i_s, orthonormal w.r.t to $\langle \cdot, \cdot \rangle = E\left[\cdot\right]$.

Weak Formulation

$$V = L^2_\rho(\Gamma, H(div; D)) \qquad W = L^2_\rho(\Gamma, L^2(D))$$

We seek $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{\xi}) \in V$ and $p(\boldsymbol{x}, \boldsymbol{\xi}) \in W$ such that:

$$\begin{aligned} \left\langle \left(T_M^{-1} \boldsymbol{u}, \boldsymbol{v} \right) \right\rangle + \langle (p, \nabla \cdot \boldsymbol{v}) \rangle &= \left\langle (g, \boldsymbol{v} \cdot \boldsymbol{n})_{\partial \Gamma_D} \right\rangle, \\ \langle (w, \nabla \cdot \boldsymbol{u}) \rangle &= 0, \end{aligned}$$

$\forall\ \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\xi}) \in V$ and $w(\boldsymbol{x}, \boldsymbol{\xi}) \in W$.

MANCHESTER 1824

The University of Manchester

Stochastic Galerkin Equations

Find $\boldsymbol{u}_{hd}(\boldsymbol{x},\boldsymbol{\xi}) \in V_h \otimes S_d$ and $p_{hd}(\boldsymbol{x},\boldsymbol{\xi}) \in W_h \otimes S_d$ satisfying:

$$\begin{aligned} \left\langle \left(T_M^{-1} \boldsymbol{u}_{hd}, \boldsymbol{v} \right) \right\rangle + \langle (p_{hd}, \nabla \cdot \boldsymbol{v}) \rangle &= \left\langle (g, \boldsymbol{v} \cdot \boldsymbol{n})_{\partial \Gamma_D} \right\rangle, \\ \langle (w, \nabla \cdot \boldsymbol{u}_{hd}) \rangle &= 0, \end{aligned}$$

$\forall\ \boldsymbol{v} \in V_h \otimes S_d$ and $w \in W_h \otimes S_d$.

MANCHESTER 1824

The University of Manchester

Stochastic Galerkin Equations

Find $\boldsymbol{u}_{hd}(\boldsymbol{x}, \boldsymbol{\xi}) \in V_h \otimes S_d$ and $p_{hd}(\boldsymbol{x}, \boldsymbol{\xi}) \in W_h \otimes S_d$ satisfying:

$$\begin{aligned} \left\langle \left(T_M^{-1} \boldsymbol{u}_{hd}, \boldsymbol{v} \right) \right\rangle + \langle (p_{hd}, \nabla \cdot \boldsymbol{v}) \rangle &= \left\langle (g, \boldsymbol{v} \cdot \boldsymbol{n})_{\partial \Gamma_D} \right\rangle, \\ \langle (w, \nabla \cdot \boldsymbol{u}_{hd}) \rangle &= 0, \end{aligned}$$

$\forall\ \boldsymbol{v} \in V_h \otimes S_d$ and $w \in W_h \otimes S_d$.

- $V_h \subset H(div; D)$, $W_h \subset L^2(D)$ are a deterministic inf-sup stable pairing
- $S_d \subset L^2(\Gamma)$

Total degree polynomials

For S_d we choose M-variate polynomials of total degree d.

Use orthonormal polynomial chaos basis

$$S_d = \text{span}\left\{\psi_1(\boldsymbol{\xi}), \ldots, \psi_{N_\xi}(\boldsymbol{\xi})\right\}, \qquad \boxed{N_\xi = \frac{(M+d)!}{M!d!}}$$

	$M=1$			$M=5$			$M=10$		
	$d=1$	$d=2$	$d=3$	$d=1$	$d=2$	$d=3$	$d=1$	$d=2$	$d=3$
	2	3	4	6	21	56	11	66	286

Saddle point systems

$$V_h = \text{span}\{\boldsymbol{\varphi}_i(\boldsymbol{x})\}_{i=1}^{N_u}, \qquad W_h = \text{span}\{\phi_j(\boldsymbol{x})\}_{j=1}^{N_p}, \qquad S_d = \text{span}\{\psi_k(\boldsymbol{\xi})\}_{k=1}^{N_\xi}$$

$$\begin{pmatrix} \hat{A} & \hat{B}^T \\ \hat{B} & 0 \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{p} \end{pmatrix} = \begin{pmatrix} \underline{g} \\ \underline{f} \end{pmatrix}$$

(never assembled) of dimension $N_x N_\xi \times N_x N_\xi$ where $N_x = N_u + N_p$ and

MANCHESTER 1824
The University of Manchester

Saddle point systems

$$V_h = \text{span}\{\boldsymbol{\varphi}_i(\boldsymbol{x})\}_{i=1}^{N_u}, \qquad W_h = \text{span}\{\phi_j(\boldsymbol{x})\}_{j=1}^{N_p}, \qquad S_d = \text{span}\{\psi_k(\boldsymbol{\xi})\}_{k=1}^{N_\xi}$$

$$\begin{pmatrix} \hat{A} & \hat{B}^T \\ \hat{B} & 0 \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{p} \end{pmatrix} = \begin{pmatrix} \underline{g} \\ \underline{f} \end{pmatrix}$$

(never assembled) of dimension $N_x N_\xi \times N_x N_\xi$ where $N_x = N_u + N_p$ and

$$\begin{pmatrix} \hat{A} & \hat{B}^T \\ \hat{B} & 0 \end{pmatrix} = \begin{pmatrix} I \otimes A_0 + \sum_{k=1}^{N} G_k \otimes A_k & I \otimes B^T \\ I \otimes B & 0 \end{pmatrix}$$

$$[A_0]_{ij} = \int_D \mu(\boldsymbol{x})\, \boldsymbol{\varphi}_i \cdot \boldsymbol{\varphi}_j \, d\boldsymbol{x}, \qquad [A_k]_{ij} = \int_D t_k(\boldsymbol{x})\, \boldsymbol{\varphi}_i \cdot \boldsymbol{\varphi}_j \, d\boldsymbol{x}.$$

Linear (KL) Case

$$T_M^{-1}(\boldsymbol{x},\boldsymbol{\xi}) = \mu(\boldsymbol{x}) + \sum_{k=1}^{M} t_k(\boldsymbol{x})\xi_k$$

- $N = M$ (the number of random variables)
- G_k matrices have at most two non-zeros entries per row.

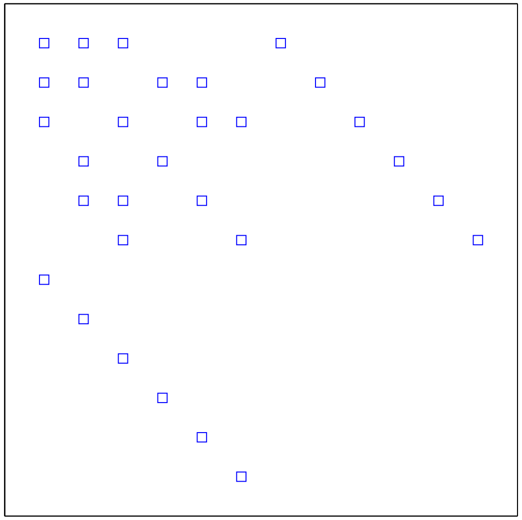
$$[G_k]_{rs} = \langle\, \xi_k \psi_r(\boldsymbol{\xi})\psi_s(\boldsymbol{\xi})\,\rangle\,, \qquad k = 1 : M$$

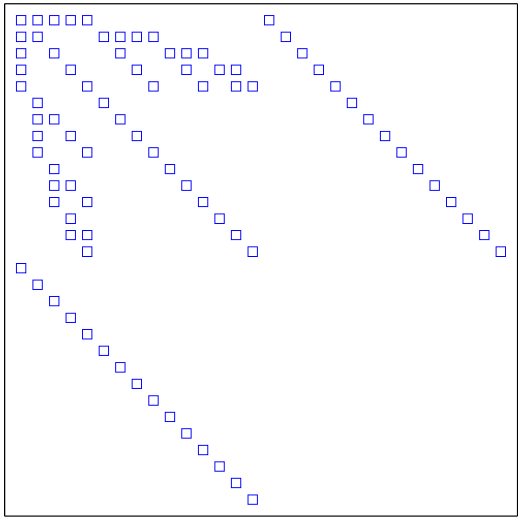
- $\sum_{k=1}^{N} G_k$ is sparse so $\hat{A}$ is block sparse.
- Matrix-vector products are cheap

$\hat{A}$	$O(M(N_u N_\xi))$
$\hat{B},\, \hat{B}^T$	$O(N_p N_\xi)$

Matrix structure

$M = 2, d = 2$ (left) and $M = 4, d = 2$ (right)





Nonlinear (PC) Case

$$T_M^{-1}(\boldsymbol{x}, \boldsymbol{\xi}) = \mu(\boldsymbol{x}) + \sum_{k=1}^{N} t_k(\boldsymbol{x})\psi_k$$

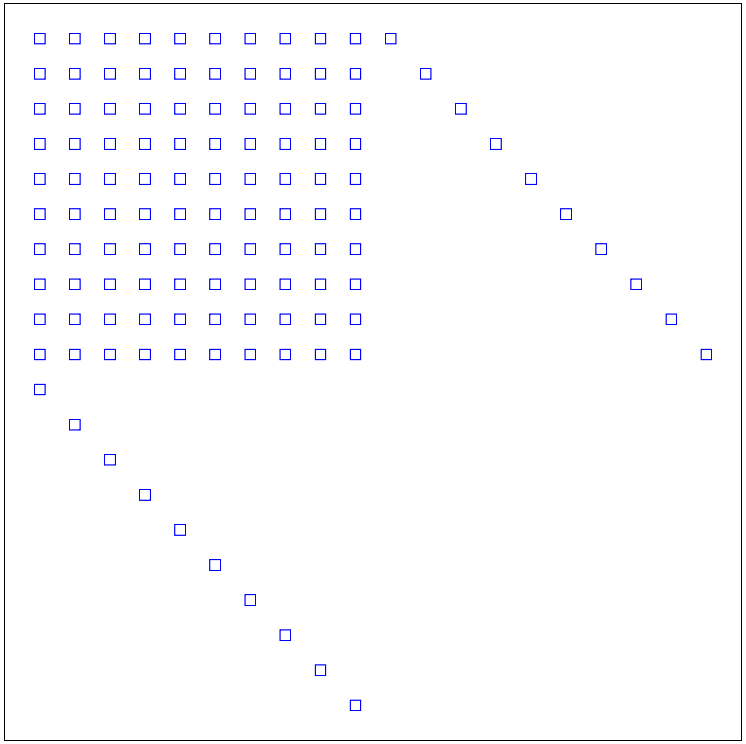
- $N = \frac{(M+2d)!}{M!(2d)!} >> N_\xi$
- G_k matrices have non-trivial sparsity pattern

$$[G_k]_{rs} = \langle\, \psi_k(\boldsymbol{\xi})\psi_r(\boldsymbol{\xi})\psi_s(\boldsymbol{\xi})\,\rangle\,, \qquad k = 1 : N$$

- $\sum_{k=1}^{M} G_k$ is dense so $\hat{A}$ is block dense.
- Matrix-vector products are expensive

$\hat{A}$	$O(N(N_u N_\xi))$—$O(N(N_u N_\xi^2))$
$\hat{B}$, $\hat{B}^T$	$O(N_p N_\xi)$

Matrix structure



Want to perform as few matrix vector products with this as possible!

(Minres) preconditioning strategies

There are two well-known classes of block-diagonal preconditioners.

Schur-Complement Preconditioning

$$\begin{pmatrix} \hat{A} & 0 \\ 0 & \hat{B}\hat{A}^{-1}\hat{B}^T \end{pmatrix}$$

Augmented Preconditioning

$$\begin{pmatrix} \hat{A} + \gamma^{-1}\hat{B}^T\hat{W}^{-1}\hat{B} & 0 \\ 0 & \gamma\hat{W} \end{pmatrix}$$

(Minres) preconditioning strategies

There are two well-known classes of block-diagonal preconditioners.

Schur-Complement Preconditioning

$$\begin{pmatrix} \hat{A} & 0 \\ 0 & \hat{B}\hat{A}^{-1}\hat{B}^T \end{pmatrix}$$

Augmented Preconditioning

$$\begin{pmatrix} \hat{A} + \gamma^{-1}\hat{B}^T\hat{W}^{-1}\hat{B} & 0 \\ 0 & \gamma\hat{W} \end{pmatrix}$$

Both require cheap, robust approximations to $\hat{A}$

The University of Manchester

Schur-Complement Preconditioning

Approximate $\hat{A} \approx L \otimes \texttt{diag}(A_0)$ where L is s.p.d. A practical preconditioner is:

$$
\begin{aligned}
P &= \begin{pmatrix} L \otimes \texttt{diag}(A_0) & 0 \\ 0 & \hat{B}\,(L \otimes \texttt{diag}(A_0))^{-1}\,\hat{B}^T \end{pmatrix} \\
&= \begin{pmatrix} L \otimes \texttt{diag}(A_0) & 0 \\ 0 & L \otimes \left(B\texttt{diag}(A_0)^{-1}B^T\right) \end{pmatrix}.
\end{aligned}
$$

MANCHESTER 1824

The University of Manchester

Schur-Complement Preconditioning

Approximate $\hat{A} \approx L \otimes \texttt{diag}(A_0)$ where L is s.p.d. A practical preconditioner is:

$$P = \begin{pmatrix} L \otimes \texttt{diag}(A_0) & 0 \\ 0 & \hat{B}\left(L \otimes \texttt{diag}(A_0)\right)^{-1} \hat{B}^T \end{pmatrix}$$

$$= \begin{pmatrix} L \otimes \texttt{diag}(A_0) & 0 \\ 0 & L \otimes \left(B\texttt{diag}(A_0)^{-1}B^T\right) \end{pmatrix}.$$

For the model problem, $B\texttt{diag}(A_0)^{-1}B^T \approx \nabla \cdot \mu \nabla$ and optimal deterministic elliptic PDE solvers (eg. AMG) can be applied.

Schur-Complement Preconditioning

Approximate $\hat{A} \approx L \otimes \texttt{diag}(A_0)$ where L is s.p.d. A practical preconditioner is:

$$P = \begin{pmatrix} L \otimes \texttt{diag}(A_0) & 0 \\ 0 & \hat{B}\left(L \otimes \texttt{diag}(A_0)\right)^{-1} \hat{B}^T \end{pmatrix}$$

$$= \begin{pmatrix} L \otimes \texttt{diag}(A_0) & 0 \\ 0 & L \otimes \left(B\texttt{diag}(A_0)^{-1} B^T\right) \end{pmatrix}.$$

For the model problem, $B\texttt{diag}(A_0)^{-1}B^T \approx \nabla \cdot \mu \nabla$ and optimal deterministic elliptic PDE solvers (eg. AMG) can be applied.

Simplest choice: $L = I$. Approximating $P^{-1}\underline{r}$ requires:

- One solve with a diagonal matrix
- N_ξ V-cycles of multigrid (e.g. AMG) on a deterministic matrix of dimension N_p.

Cost is $O(N_\xi(N_u + N_p))$.

Convergence analysis

$$P = \begin{pmatrix} I \otimes \mathtt{diag}(A_0) & 0 \\ 0 & I \otimes \mathtt{amg}(B\mathtt{diag}(A_0)^{-1}B^T) \end{pmatrix}$$

Theorem

Let $0 < \hat{\nu}_1 \ldots \leq \hat{\nu}_N$ be the eigenvalues of $\mathtt{diag}(\hat{A})^{-1}\hat{A}$. The eigenvalues of the preconditioned saddle-point matrix lie in the union of the bounded intervals,

$$\left[\frac{1}{2}\left(\hat{\nu}_1 - \sqrt{\hat{\nu}_1^2 + 4\Theta^2}\right), \frac{1}{2}\left(\hat{\nu}_N - \sqrt{\hat{\nu}_N^2 + 4\theta^2}\right)\right]$$
$$\cup \left[\hat{\nu}_1, \frac{1}{2}\left(\hat{\nu}_N + \sqrt{\hat{\nu}_N^2 + 4\Theta^2}\right)\right].$$

When $L = I$, this preconditioner is ‘mean-based’.

MANCHESTER 1824

The University of Manchester

Linear (KL) case

Theorem

E.g. if lowest-order Raviart-Thomas elements are used, the eigenvalues of $\texttt{diag}(\hat{A})^{-1}\hat{A}$ lie in the bounded interval $\left[\frac{1}{2} - c_d\tau, \frac{3}{2} + c_d\tau\right]$ where

$$\tau = \frac{3\sigma}{2\mu} \sum_{k=1}^{M} \sqrt{\lambda_k} \parallel t_k(\boldsymbol{x}) \parallel_\infty,$$

σ and μ are the standard deviation and mean of $T^{-1}(\boldsymbol{x}, \omega)$, and c_d is a constant (possibly) depending on d.

Linear (KL) case

Theorem

E.g. if lowest-order Raviart-Thomas elements are used, the eigenvalues of $\texttt{diag}(\hat{A})^{-1}\hat{A}$ lie in the bounded interval $\left[\frac{1}{2} - c_d\tau, \frac{3}{2} + c_d\tau\right]$ where

$$\tau = \frac{3\sigma}{2\mu} \sum_{k=1}^{M} \sqrt{\lambda_k} \parallel t_k(\boldsymbol{x}) \parallel_\infty,$$

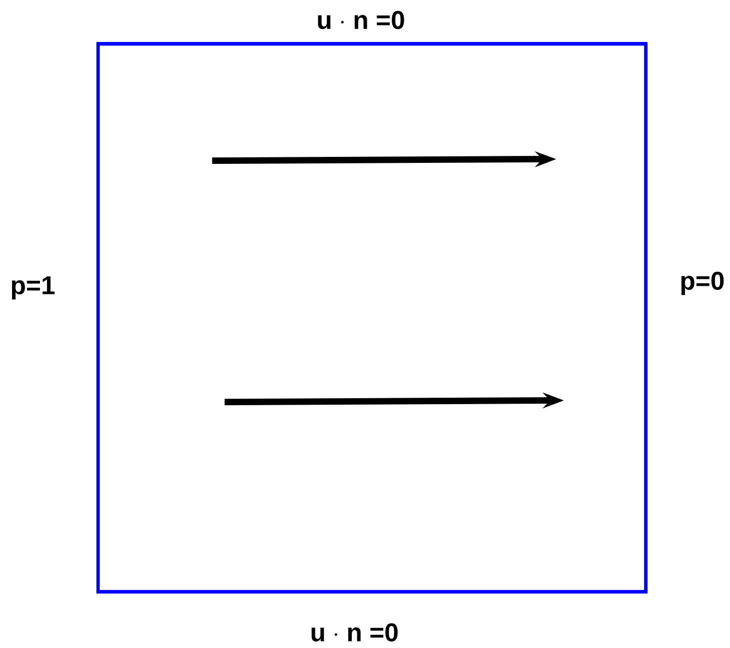
σ and μ are the standard deviation and mean of $T^{-1}(\boldsymbol{x}, \omega)$, and c_d is a constant (possibly) depending on d.

When T_M^{-1} is a KL expansion this is OK because

- $\frac{\sigma}{\mu}$ is small for well-posed problem.
- $c_d = O(1)$ (bounded rvs), $c_d = O(\sqrt{d})$ (Gaussian rvs)

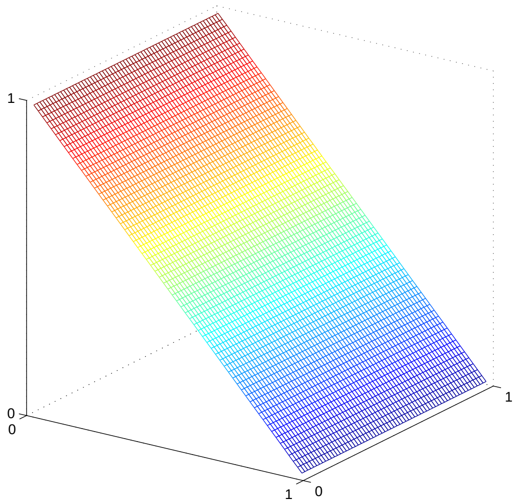
Example - KL case

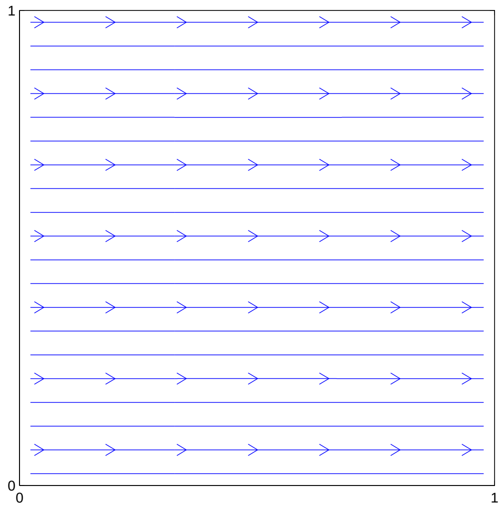
Let $D = [0,1] \times [0,1]$, $f = 0$ with mixed bcs. We choose a Bessel covariance function for the random input with $\mu(\boldsymbol{x}) = 1$ and $\lambda = 1 \Rightarrow M = 6$ random variables capture 98% of the variance.



Mean of numerical solution

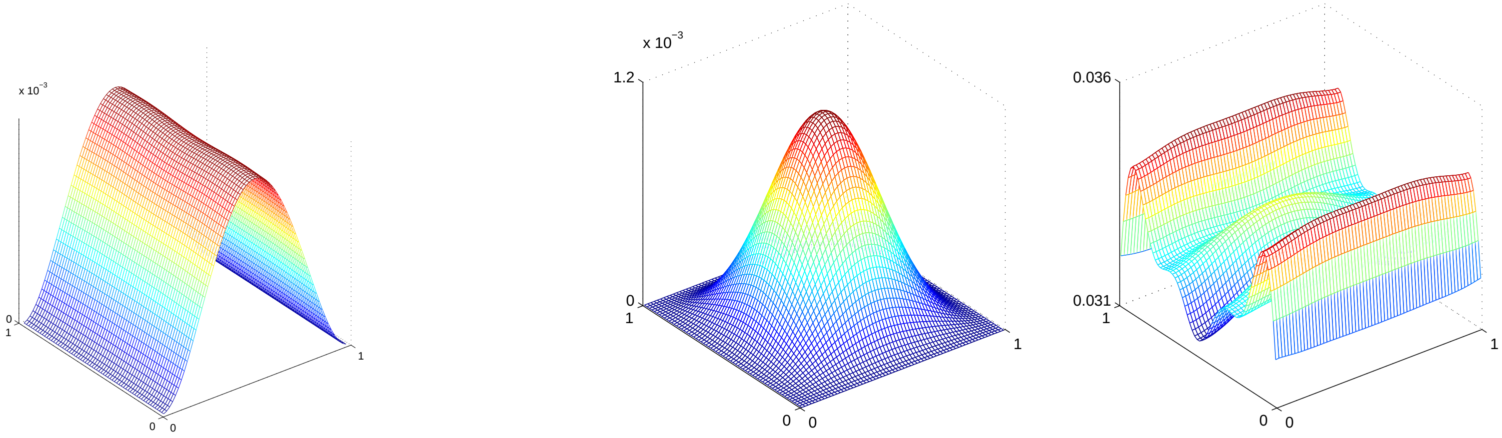
Pressure (left), Flux (right)





Variance of numerical solution

Pressure (left), y component (middle) and x component (right) of the Flux



Preconditioned minres iterations

	d	2	3	4	5
	$N_\xi(N_u + N_p)$	344,064	1,032,192	2,580,480	5,677,056
$\frac{\sigma}{\mu} = 0.1$	total iters	45	46	48	48
	N_V	1,260	3,864	10,080	22,176
	total solve time	14.0s	45.35s	119.01s	262.04s
$\frac{\sigma}{\mu} = 0.2$	total iters	55	59	62	63
	N_V	1,540	4,956	13,020	29,106
	total solve time	17.18s	58.51s	154.82s	379.01
$\frac{\sigma}{\mu} = 0.3$	total iters	66	74	80	86
	N_V	1,848	6,216	16,800	39,732
	total solve time	20.66s	72.97s	199.75s	486.74

Nonlinear (PC case)

Theorem

E.g. if lowest-order Raviart-Thomas elements are used, the eigenvalues of $\texttt{diag}(\hat{A})^{-1}\hat{A}$ lie in the bounded interval $\left[\frac{1}{2} - c_d\tau, \frac{3}{2} + c_d\tau\right]$ where, for lognormal diffusion coefficients in particular,

- τ depends nonlinearly on σ_G
- $c_d = O\left(\exp(dM)\right)$

The Schur-complement preconditioner is too weak for large d and σ_G requiring an excessive number of matrix-vector products.

Preconditioned minres iterations

Previous test problem. $T_M(\boldsymbol{x},\boldsymbol{\xi}) = \exp(G_M(\boldsymbol{x},\boldsymbol{\xi}))$ where underlying Gaussian field $G(\boldsymbol{x},\omega)$ has Bessel covariance function. $M = 5$.

	d	1	2	3	4
	$N_\xi(N_u + N_p)$	31,104	108,864	290,304	653,184
	N	21	126	462	1287
$\sigma_G = 0.2$	total iters	47	57	65	74
	total solve time	2s	22s	245s	2979s
$\sigma_G = 0.4$	total iters	61	89	121	148
	total solve time	3s	35s	534s	6321s
$\sigma_G = 0.6$	total iters	77	139	225	345
	total solve time	3s	53s	1369s	13794s
$\sigma_G = 0.8$	total iters	99	219	425	747
	total solve time	4s	84s	2604s	29808s

Augmented Preconditioning

Consider the ‘ideal’ preconditioner

$$P = \begin{pmatrix} \hat{A} + \gamma^{-1}\hat{B}^T\hat{W}^{-1}\hat{B} & 0 \\ 0 & \gamma\hat{W} \end{pmatrix}$$

Theorem

Let $\hat{W} = I \otimes M$, where $\underline{w}^T\,(I \otimes M)\,\underline{w} = \parallel w \parallel^2_{L^2(\Gamma, L^2(D))}$. The eigenvalues of the preconditioned saddle-point matrix are bounded and lie in

$$\left(-1, -\frac{\hat{\beta}^2 t_{min}}{\gamma + \hat{\beta}^2 t_{min}}\right] \cup \{1\}$$

where $\hat{\beta}$ is the inf-sup constant and $T_M \geq t_{min}$ a.e.

Augmented Preconditioning

Suppose we approximate $\hat{A}$ by $L \otimes A_0$ where L is any s.p.d $N_\xi \times N_\xi$ matrix and replace $\hat{W}$ with $L^{-1} \otimes M$ then,

$$P = \begin{pmatrix} L \otimes \left(A_0 + \gamma^{-1} B^T M^{-1} B\right) & 0 \\ 0 & L^{-1} \otimes \gamma M \end{pmatrix}$$

Augmented Preconditioning

Suppose we approximate $\hat{A}$ by $L \otimes A_0$ where L is any s.p.d $N_\xi \times N_\xi$ matrix and replace $\hat{W}$ with $L^{-1} \otimes M$ then,

$$P = \begin{pmatrix} L \otimes (A_0 + \gamma^{-1} B^T M^{-1} B) & 0 \\ 0 & L^{-1} \otimes \gamma M \end{pmatrix}$$

For this model problem,

$$\underline{v}^T \left(A_0 + \gamma^{-1} B^T M^{-1} B \right) \underline{v} = \left(T_M^{-1} \boldsymbol{v}, \boldsymbol{v} \right) + \gamma^{-1} \left(\nabla \cdot \boldsymbol{v}, \nabla \cdot \boldsymbol{v} \right).$$

The matrix in red is a discretisation of a weighted $H(div)$ operator on the velocity space.

Optimal deterministic solvers (geometric multigrid) can be exploited.

Augmented Preconditioning

Simplest choice: $L = I$ ('mean-based' preconditioner)

$$P = \begin{pmatrix} L \otimes (A_0 + \gamma^{-1} B^T M^{-1} B) & 0 \\ 0 & L \otimes \gamma M \end{pmatrix}$$

Approximating $P^{-1}\underline{r}$ requires:

- N_ξ solves with M and N_p multiplications with L.
- N_ξ V-cycles of multigrid (e.g. Arnold-Falk-Winther) on a deterministic matrix of dimension N_u and N_u solves with L.

Even if L is dense, still cheaper than a matrix-vector product with saddle point matrix!

Convergence analysis

$$P = \begin{pmatrix} L \otimes \left(A_0 + \gamma^{-1} B^T M^{-1} B\right) & 0 \\ 0 & L \otimes \gamma M \end{pmatrix}$$

Spectral bounds for preconditioned system matrix depend on choice of L and the efficiency of the approximation $\hat{A} \approx L \otimes A_0$ and γ.

MANCHESTER 1824
The University of Manchester

Convergence analysis

$$P = \begin{pmatrix} L \otimes \left(A_0 + \gamma^{-1} B^T M^{-1} B\right) & 0 \\ 0 & L \otimes \gamma M \end{pmatrix}$$

Spectral bounds for preconditioned system matrix depend on choice of L and the efficiency of the approximation $\hat{A} \approx L \otimes A_0$ and γ.

However, asymptotically, as $\gamma \to 0$, we can show

- all $N_\xi N_p$ negative eigenvalues cluster at -1.
- $N_\xi N_p$ positive eigenvalues cluster at $+1$
- $N_\xi (N_u - N_p)$ positive eigenvalues lie in $[\hat{\nu}_1, \hat{\nu}_n]$ where

$$\hat{\nu}_1 \leq \frac{\underline{v}^T \hat{A} \underline{v}}{\underline{v}^T \left(L \otimes A_0\right) \underline{v}} \leq \hat{\nu}_n$$

Preconditioned minres iterations

Previous test problem. Now $T_M(\boldsymbol{x}, \boldsymbol{\xi}) = \exp(G_M(\boldsymbol{x}, \boldsymbol{\xi}))$ where underlying Gaussian field $G(\boldsymbol{x}, \omega)$ has Bessel covariance function. $M = 5$. Choose $L = I$ and $\gamma = 10^{-3}$.

	d	1	2	3	4
	$N_\xi(N_u + N_p)$	31,104	108,864	290,304	653,184
	N	21	126	462	1287
$\sigma_G = 0.2$	total iters	24	19	16	16
	total solve time	3s	9s	65s	399s
$\sigma_G = 0.4$	total iters	28	24	24	27
	total solve time	3s	12s	97s	668s
$\sigma_G = 0.6$	total iters	33	32	35	43
	total solve time	4s	17s	140s	1063s
$\sigma_G = 0.8$	total iters	39	42	51	67
	total solve time	5s	19s	203s	1658s

Improved choice of L

Could use Pitsianis-Van Loan so-called ‘best Kronecker product approximation’:

$$L = \text{argmin}\{H \in \mathbb{R}^{N_\xi \times N_\xi} : ||\hat{A} - H \otimes A_0||_F\}$$

L is the symmetric and positive definite matrix

$$L = I + \sum_{k=1}^{N} \frac{\mathtt{tr}(A_k^T A_0)}{\mathtt{tr}(A_0^T A_0)} G_k$$

Preconditioned minres iterations

Previous test problem. Now $T_M(\boldsymbol{x}, \boldsymbol{\xi}) = \exp(G_M(\boldsymbol{x}, \boldsymbol{\xi}))$ where underlying Gaussian field $G(\boldsymbol{x}, \omega)$ has Bessel covariance function. $M = 5$.

	d	1	2	3	4
	$N_\xi(N_u + N_p)$	31,104	108,864	290,304	653,184
	N	21	126	462	1287
$\sigma_G = 0.2$	total iters	21	15	13	13
	total solve time	3s	8s	54s	328s
$\sigma_G = 0.4$	total iters	24	19	18	17
	total solve time	3s	11s	73s	429s
$\sigma_G = 0.6$	total iters	26	21	21	23
	total solve time	4s	12s	86s	576s
$\sigma_G = 0.8$	total iters	28	24	26	29
	total solve time	4s	14s	106s	723s

Average time in seconds per minres iteration and (av. mat-vec time + av. preconditioner time)

	d=2	d=3	d=4
$L = I$	0.58 (0.33 + 0.22)	4.00 (3.40 + 0.50)	24.80 (23.30 + 1.09)
Improved L	0.59 (0.33 + 0.23)	4.08 (3.40 + 0.53)	25.02 (23.30 + 1.23)

Summary

If random inputs are KL expansions:

- Mean-based preconditioners are cheap and robust within the range of statistical parameters where the problem is well-posed.
- Schur-complement preconditioners are OK as matrix-vector products are cheap.

Summary

If random inputs are KL expansions:

- Mean-based preconditioners are cheap and robust within the range of statistical parameters where the problem is well-posed.
- Schur-complement preconditioners are OK as matrix-vector products are cheap.

If the random inputs are PC expansions

- Mean-based preconditioners are still cheap but are inadequate as σ and d increase.
- Augmented preconditioners are promising because fewer matrix-vector products are required

Summary

If random inputs are KL expansions:

- Mean-based preconditioners are cheap and robust within the range of statistical parameters where the problem is well-posed.
- Schur-complement preconditioners are OK as matrix-vector products are cheap.

If the random inputs are PC expansions

- Mean-based preconditioners are still cheap but are inadequate as σ and d increase.
- Augmented preconditioners are promising because fewer matrix-vector products are required

But:

- Also need to look at stopping criteria for iteration.

MANCHESTER 1824

The University of Manchester

References

- C.E.Powell, E. Ullmann, Preconditioning stochastic Galerkin saddle point systems, Submitted.

- H. Elman, D. Furnival, C.E. Powell, $H(div)$ preconditioning for a mixed finite element formulation of the stochastic diffusion equation. Math. Comp. 2010.

- O.G. Ernst, C.E. Powell, D. Silvester and E. Ullmann, Efficient solvers for a Linear Stochastic Galerkin Mixed Formulation of the Steady-State Diffusion Equation, SIAM J. Sci. Comput. 2009.

- C.E. Powell, H. Elman, Block-diagonal preconditioning for spectral stochastic finite element systems, IMA J. Numer Anal. 2008/9.