Modeling Techniques and Numerical Solution Algorithms for Discrete Partial Differential Equations with Random Data

II. Solution Algorithms

Howard Elman
University of Maryland

Diffusion Equation with Random Data

$$-\nabla \cdot (a \nabla u) = f \quad \text{in } \mathcal{D} \subset R^d$$

$$u = g_D \text{ on } \partial\mathcal{D}_D, \quad (a\nabla u)\cdot n = 0 \text{ on } \partial\mathcal{D}_N = \partial\mathcal{D} \setminus \partial\mathcal{D}_D$$

Diffusion coefficient: truncated *Karhunen-Loève* expansion:

$$a(x,\omega) = a_0(x) + \sigma \sum_{r=1}^{m} \sqrt{\lambda_r}\, a_r(x) \xi_r(\omega)$$

$$a_0(x) = \mu(x) = E(a(x,\cdot)) \quad \text{mean}$$

$$\sigma(x) = E(a(x,\cdot)^2) - \mu^2 \quad \text{standard deviation}$$

$$a_r(x), \lambda_r = \text{eigenfunctions/eigenvalues of covariance operator}$$

$$(Ca)(x) = \lambda a(x), \quad (Ca)(x) = \int_D c(x,y) a(y)\, dy$$

Assumption: coercivity and boundedness $0 < \alpha_1 \le a \le \alpha_2 < \infty$

Stochastic Galerkin Method

Weak formulation

$$\int_\Gamma \int_{\mathcal{D}} a(x,\underline{\xi})\nabla u \cdot \nabla v \, dx\, \rho(\underline{\xi}) d\underline{\xi} = \int_\Gamma \int_{\mathcal{D}} f\, v\, dx\, \rho(\underline{\xi}) d\underline{\xi} \qquad (*)$$

Finite dimensional spaces:

- spatial discretization: $S_h \subset H_0^1(\mathcal{D})$, spanned by $\{\varphi_j\}_{j=1}^{N_x}$

  for example: piecewise linear on triangles

- stochastic discretization: $T_p \subset L^2(\Gamma)$, spanned by $\{\psi_l\}_{l=1}^{N_\xi}$

  determined by m-variate tensor product polynomials whose components are orthogonal wrt density measure

Discrete weak formulation: Find $u_{hp} \in S_h \otimes T_p$ such that (*) holds for all $v_{hp} \in S_h \otimes T_p$

$$u_{hp} = \sum_{j=1}^{N_x} \sum_{l=1}^{N_\xi} u_{jl} \varphi_j(x) \psi_l(\xi)$$

Matrix Equation *Au=f*

$$a(x,\xi(\omega)) = a_0(x) + \sigma \sum_{r=1}^{m} \sqrt{\lambda_r}\, a_r(x)\xi_r(\omega)$$

$$A = G_0 \otimes A_0 + \sum_{r=1}^{m} G_r \otimes A_r$$

$$[A_0]_{jk} = \int_{\mathcal{D}} a_0(x) \nabla\varphi_j(x) \cdot \nabla\varphi_k(x)\, dx$$

$$[A_r]_{jk} = \sqrt{\lambda_r} \int_{\mathcal{D}} \sigma(x) a_r(x) \nabla\varphi_j(x) \cdot \nabla\varphi_k(x)\, dx$$

$$[G_0]_{lq} = \langle \psi_l, \psi_q \rangle, \quad [G_r]_{lq} = \langle \xi_r \psi_l, \psi_q \rangle$$

$$[f]_{kq} = \int_{\Gamma}\int_{\mathcal{D}} f(x,\xi)\, \varphi_k(x) \psi_q(\xi)\, dx\, \rho(\xi) d\xi$$

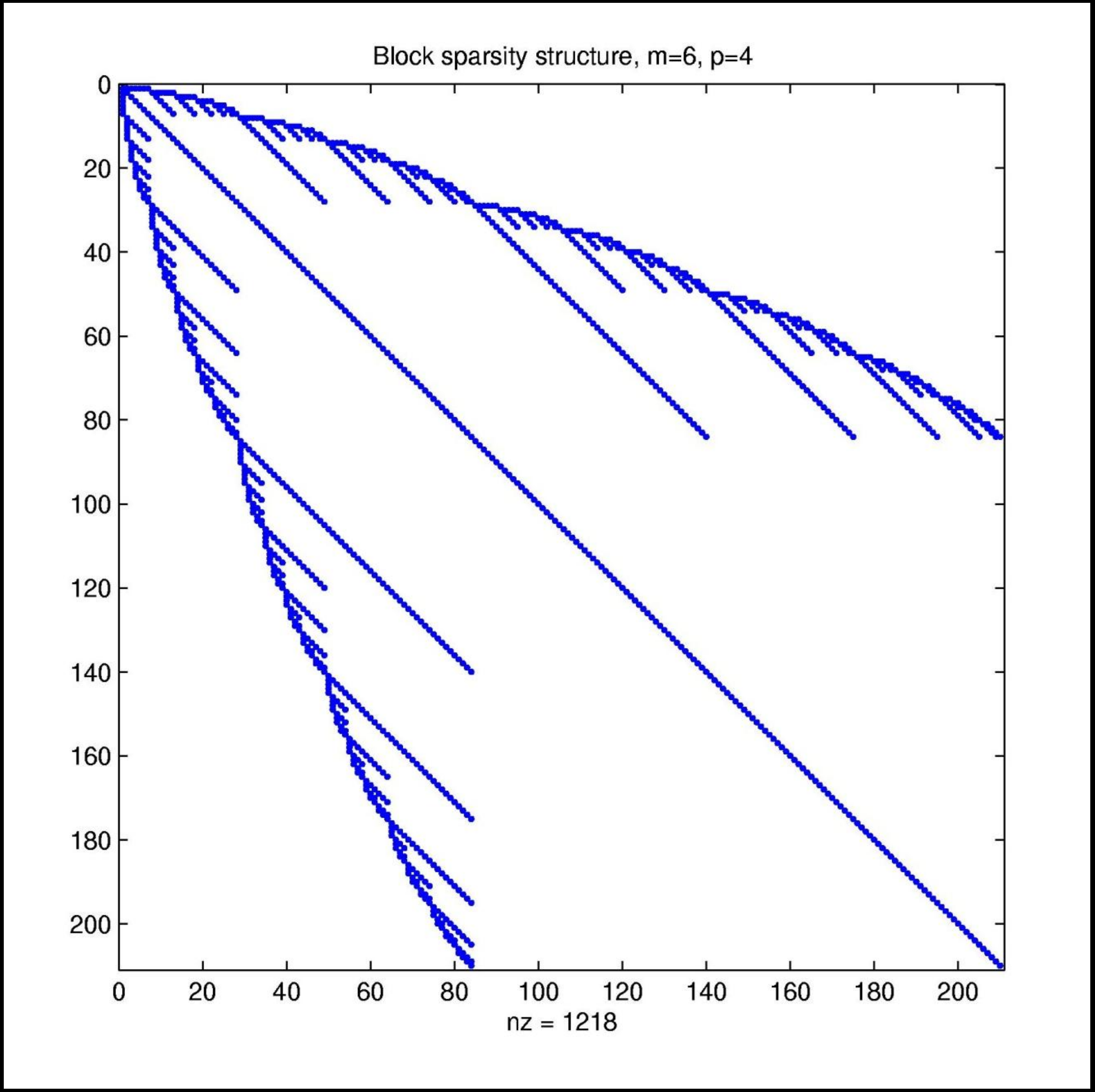
Properties of *A:*

- order = $N_x \times N_\xi$ = (size of spatial basis) *x* (size of stochastic basis)
- sparsity: inherited from that of $\{G_r\}$ and $\{A_r\}$

Example of Sparsity Pattern

For m-variate polynomials of total degree p:

$$N_\xi = \frac{(m+p)!}{m!p!} = \frac{10!}{6!4!} = 210$$



Multigrid Solution of Matrix Equation I (E. & Furnival)

Solving ***Au=f***

$$A = G_0 \otimes A_0 + \sum_{r=1}^{m} G_r \otimes A_r$$

$$[A_r]_{jk} = \sqrt{\lambda_r}\,\sigma \int_D a_r(x)\nabla\varphi_j(x)\cdot\nabla\varphi_k(x)\,dx,$$

$$[G_r]_{lq} = \int_\Omega \psi_l(\xi)\psi_q(\xi)\xi_r\rho(\xi)\,d\xi$$

$A_r = A_r^{(h)}, \quad A = A^{(h)},$ spatial discretization parameter h

$A_r = A_r^{(2h)}, \quad A = A^{(2h)},$ spatial discretization parameter $2h$

Develop MG algorithm for spatial component of the problem

Multigrid algorithm (two-grid)

Let $A^{(h)} = Q - N,$ Q = smoothing operator

for $i=0,1,\ldots$

for $j=1:k$ k smoothing steps

$$u^{(h)} \leftarrow (I - Q^{-1}A^{(h)})u^{(h)} + Q^{-1}f^{(h)}$$

end

$r^{(2h)} = \mathcal{R}(f^{(h)} - A^{(h)}u^{(h)})$ Restriction

Solve $A^{(2h)}c^{(2h)} = r^{(2h)}$ Coarse grid correction

$u^{(h)} \leftarrow u^{(h)} + \mathcal{P}c^{(2h)}$ Prolongation

end

Prolongation and restriction:

$\mathcal{P} = I \otimes P,$ induced by natural inclusion in spatial domain

$\mathcal{R} = \mathcal{P}^T = I \otimes R, \quad R = P^T$

Convergence analysis: use "standard" approach

Error propagation matrix:

$$e^{(i+1)} = [(A^{(h)})^{-1} - \mathcal{P}(A^{(2h)})^{-1}\mathcal{R})]\,[A^{(h)}(I - Q^{-1}A^{(h)})^k]\,e^{(i)}$$

Establish ***approximation property***

$$\left\| [(A^{(h)})^{-1} - \mathcal{P}(A^{(2h)})^{-1}\mathcal{R}]\,y \right\|_{A^{(h)}} \le C \|y\|_2 \quad \forall y$$

and ***smoothing property***

$$\left\| [A^{(h)}(I - Q^{-1}A^{(h)})^k]\,y \right\|_2 \le \eta(k) \|y\|_{A^{(h)}} \quad \forall y, \quad \eta(k) \xrightarrow{k \text{ increases}} 0$$

Analysis is:

$$\begin{aligned} \| e^{(i+1)} \|_{A^{(h)}} &\le \| (A^{(h)})^{-1} - \mathcal{P}(A^{(2h)})^{-1}\mathcal{R})][A^{(h)}(I - Q^{-1}A^{(h)})^k]\,e^{(i)} \|_{A^{(h)}} \\ &\le C \| [A^{(h)}(I - Q^{-1}A^{(h)})^k]\,e^{(i)} \|_2 \\ &\le C\eta(k) \| e^{(i)} \|_{A^{(h)}} \end{aligned}$$

Convergence analysis: smoothing property

Assumption of boundedness $0 < \alpha_1 \leq a \leq \alpha_2 < \infty \Longrightarrow$

$$\frac{(A^{(h)}\underline{v},\underline{v})}{(\underline{v},\underline{v})} = \frac{\iint_{\Gamma,\mathcal{D}} a(x,\underline{\xi}) \nabla v_{hp} \cdot \nabla v_{hp}\, dx\, \rho(\underline{\xi}) d\underline{\xi}}{\iint_{\Gamma,\mathcal{D}} v_{hp}^2\, dx\, \rho(\underline{\xi}) d\underline{\xi}}$$

$$\leq \alpha_2 \frac{\iint_{\Gamma,\mathcal{D}} \nabla v_{hp} \cdot \nabla v_{hp}\, dx\, \rho(\underline{\xi}) d\underline{\xi}}{\iint_{\Gamma,\mathcal{D}} v_{hp}^2\, dx\, \rho(\underline{\xi}) d\underline{\xi}} \leq \alpha_2 c$$

Thus, maximum eigenvalue of $A^{(h)}$ is bounded

For smoothing property (Braess):

$Q = \theta I$ (Richardson iteration) works with $\theta \geq \max(\lambda(A^{(h)}))$

In experiments described below:

we use *damped Jacobi*, $Q = diag(A)/\omega$

Approximation property

"Standard" MG analysis for deterministic problem:

$$\left\| [(A^{(h)})^{-1} - \mathcal{P}(A^{(2h)})^{-1}\mathcal{R}]\, y \right\|_{A^{(h)}} = \left\| u^{(h)} - u^{(2h)} \right\|_{A^{(h)}}$$

$$= \left\| u_h - u_{2h} \right\|_a \quad (= a(u_h - u_{2h}, u_h - u_{2h})^{1/2})$$

$$\leq \left\| u_h - u \right\|_a + \left\| u - u_{2h} \right\|_a$$

Approximability $\leq \sqrt{\alpha_2}\left(Ch \left\| D^2 u \right\|_{L^2(\mathcal{D})} + C2h \left\| D^2 u \right\|_{L^2(\mathcal{D})} \right)$

Regularity $\leq Ch \left\| f \right\|_{L^2(\mathcal{D})}$

Property of mass matrix $\leq C \left\| y \right\|_2$

For approximation property in stochastic case

Introduce *semi-discrete* space $H_0^1(\mathcal{D}) \otimes \underbrace{T_p}$ → Discrete stochastic space

Weak formulation

$$a(u_p, v_p) = \ell(v_p) \quad \text{for all } v_p \in H_0^1(\mathcal{D}) \otimes T_p$$

Solution u_p

$$\left\| [(A^{(h)})^{-1} - \mathcal{P}(A^{(2h)})^{-1}\mathcal{R}]\, y \right\|_{A^{(h)}} = \left\| u_{hp} - u_{2h,p} \right\|_a$$
$$\leq \left\| u_h - u_p \right\|_a + \left\| u_p - u_{2h} \right\|_a$$

Approximation (in 2D):

$$\left\| u_p - u_{hp} \right\|_a \leq Ch \left\| D^2 u_p \right\|_{L^2(\mathcal{D}) \otimes L^2(\Gamma)}$$

Established using best approximation property of u_{hp}
and interpolant $\tilde{u}_p(x_j, \xi) = u_p(x_j, \xi) \ \forall \xi$

Similarly for other steps used for deterministic analysis

Comments

- Establishes convergence of multigrid with rate independent of spatial discretization size h

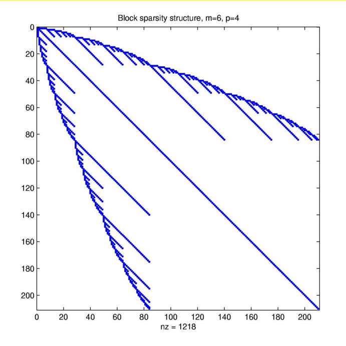
- No dependence on stochastic parameters m, p

- Applies to any basis of stochastic space

  For polynomial chaos $N_\xi = \dfrac{(m+p)!}{m!\,p!}$ degrees of freedom

- N.B. Coarse grid solve is with matrix (at least) this large

  This is a topic of ongoing study
  (E. & Ullmann)



Experimental results

Problem:

$$-\nabla \cdot (a \nabla u) = f \quad \text{in } D \times \Omega, \qquad D = (-1,1)^2$$
$$u = 0 \qquad \text{on } \partial D \times \Omega$$

a derived from an exponential covariance function

Tested both uniform and normal distributions with orthogonal basis (multidimensional Legendre / Hermite polynomials)

Multigrid scheme:

V-cycle
Damped Jacobi smoother, damping parameter 2/3
2 times 2 coarsest grid
3 presmoothing, 3 postsmoothing steps
Stop when relative residual ≤ 1.d-6

Iteration counts / normal distribution

terms (m) in KL-expansion

h=1/16

Polynomial degree

	m=1	m=2	m=3	m=4
p=1	8	8	8	8
p=2	8	8	8	8
p=3	9	9	9	9
p=4	9	10	10	10

h=1/32

Polynomial degree

	m=1	m=2	m=3	m=4
p=1	7	7	8	8
p=2	8	8	8	8
p=3	8	8	9	9
p=4	9	9	9	9

Multigrid Solution of Matrix Equation II

Solving ***Au=f***

$$A = G_0 \otimes A_0 + \sum_{r=1}^{m} G_r \otimes A_r$$

$$[A_r]_{jk} = \sqrt{\lambda_r}\,\sigma \int_{\mathcal{D}} a_r(x) \nabla \varphi_j(x) \cdot \nabla \varphi_k(x)\, dx,$$

$$[G_r]_{lq} = \int_{\Omega} \psi_l(\xi) \psi_q(\xi) \xi_r \rho(\xi)\, d\xi$$

Preconditioner for use with CG: $Q = G_0 \otimes A_0$ (Kruger, Pellissetti, Ghanem)

$$A_0 \sim \int_{\mathcal{D}} a_0(x) \nabla \varphi_j(x) \cdot \nabla \varphi_k(x)\, dx$$

Deterministic diffusion, from mean

$$G_0 = I$$

Analysis (Powell & E.)

Recall $a(x,\omega) = a_0(x) + \sigma \sum_{r=1}^{m} \sqrt{\lambda_r}\, a_r(x)\xi_r(\omega)$

$$\longrightarrow \qquad A = G_0 \otimes A_0 + \sum_{r=1}^{m} G_r \otimes A_r$$

$$Q = G_0 \otimes A_0$$

Theorem : For μ constant, the Rayleigh quotient satisfies

$$1 - \tau \le \frac{(w, Aw)}{(w, Qw)} \le 1 + \tau$$

$$\tau = (\sigma/\mu)\, c(p) \sum_{r=1}^{m} \sqrt{\lambda_r}\, \| a_r \|_\infty$$

Consequence: $\kappa \le \dfrac{1+\tau}{1-\tau}$ dictates convergence of PCG

Sketch of proof

$$\tau = (\sigma/\mu)\, c(p) \sum_{r=1}^{m} \sqrt{\lambda_r}\, \| a_r \|_\infty$$

$$A = G_0 \otimes A_0 + \sum_{r=1}^{m} G_r \otimes A_r$$

In spatial domain:

$$(\varphi, A_r\varphi) \sim \sigma\sqrt{\lambda_r} \int_D a_r(x) \nabla\varphi(x) \cdot \nabla\varphi(x)\, dx$$

$$\leq \sigma\sqrt{\lambda_r}\, \| a_r \|_\infty \int_D \nabla\varphi(x) \cdot \nabla\varphi(x)\, dx$$

$$= (\sigma/\mu)\sqrt{\lambda_r}\, \| a_r \|_\infty\, (\varphi, A_0\varphi)$$

Sketch of proof

$$\tau = (\sigma/\mu)\underline{c(p)}\sum_{r=1}^{m}\sqrt{\lambda_r}\,\| a_r \|_\infty$$

$$A = G_0 \otimes A_0 + \sum_{r=1}^{m} G_r \otimes A_r$$

In stochastic space:

$$(\psi, G_r\psi) \sim \int_\Gamma \xi_r \psi(\xi)\psi(\xi)\,\rho(\xi)d\xi$$

G_r derives from basis of multivariate polynomials of total degree p, orthogonal wrt probability measure $\rho(\xi)d\xi$

Consequence:

- three-term recurrence handles $\xi_r\psi(\xi)$
- maximum eigenvalue is largest root of scalar orthogonal polynomial
- $\sim \boxed{\sqrt{p}}$ for Gaussian measure (Hermite polynomials)
- $\sim \boxed{1}$ for uniform measure (Legendre polynomials)

Example of Eigenvalues &Bounds of Preconditioned Operator

h=1/8
σ=.01

m (#KL)	p	min λ	min bound	max λ	max bound
1	1	.92	.92	1.08	1.08
	2	.85	.85	1.15	1.15
	3	.80	.80	1.20	1.20
	4	.76	.76	1.24	1.24
2	1	.91	.90	1.09	1.10
	2	.85	.77	1.15	1.23
	3	.80	.70	1.20	1.30
	4	.75	.63	1.25	1.37
3	1	.91	.89	1.09	1.11
	2	.85	.70	1.15	1.30
	3	.79	.59	1.21	1.41
	4	.74	.50	1.26	1.50

Multigrid Variant of this Idea

Replace action of A_0^{-1} with multigrid ⟶ preconditioner

$$Q_{MG} = G_0 \otimes A_{0,MG}$$ (Le Maitre, et al.)

Analysis: $$\frac{(w, Aw)}{(w, Q_{MG} w)} = \frac{(w, Aw)}{(w, Qw)} \underbrace{\frac{(w, Qw)}{(w, Q_{MG} w)}}_{\in [\beta_1, \beta_2]}$$

Spectral equivalence of MG approximation to diffusion operator

$$\Longrightarrow \quad \kappa \le \frac{(1+\tau)}{(1-\tau)} \frac{\beta_2}{\beta_1}$$

Crimes and misdemeanors

For coercivity in $-\nabla \cdot (a \nabla u) = f$ require $0 < a_1 \le a \le a_2 < \infty$

For either problem definition:
if distribution of a is Gaussian, then a is unbounded

Example: eigenvalues of A and $Q^{-1}A$ (method II)

h=1/8
m=2
σ=.3

p	min λ(A)	max λ(A)	min λ(A/Q)	max λ(A/Q)
4	.11	6.51	.25	1.75
5	.07	6.95	.13	1.87
6	.03	7.35	.02	1.98
7	-.27	7.71	-.09	2.09
8	-.60	8.06	-.18	2.18

Experiment

Starting with *a* with specified covariance and small σ (=*.01*):

Compare Monte-Carlo simulation with SFEM, for

$$-\nabla \cdot (a\nabla u) = f$$

N.B.: No negative samples of diffusion obtained in MC

		# Samples *s*			
Max	SFEM	100	1000	10,000	40,000
Mean	.06311	.06361	.06330	.06313	.06313
Variance	2.360(-5)	2.161(-5)	2.407(-5)	2.258(-5)	2.316(-5)

Solve one system of order 210x225

Solve *s* systems of size 225

Return to crimes and misdemeanors

For coercivity in $-\nabla\cdot(a\nabla u)=f$ require $0<\alpha_1\le a\le\alpha_2<\infty$

Fix: use a different distribution, e.g. uniform

Require joint density function associated with $\underline{\xi}$

Difficulty: recall

$$a(x,\omega)=a_0(x)+\sigma\sum_{r=1}^{m}\sqrt{\lambda_r}\,a_r(x)\xi_r(\omega)$$

$$\langle a(u,v)\rangle=\int_\Omega\int_{\mathcal{D}} a\nabla u\cdot\nabla v\,dx\,dP(\omega)=\int_\Gamma\int_{\mathcal{D}} a(x,\underline{\xi})\nabla u\cdot\nabla v\,dx\,\rho(\underline{\xi})d\underline{\xi}$$

Except for Gaussian distribution $\{\xi_r\}$ are not independent

$$\rho(\underline{\xi})\ne\underbrace{\rho_1(\xi_1)\cdots\rho_m(\xi_m)}$$

In common use

Alternative Fixes

Enforce coercivity by requiring $a(x,\omega)$ to come from a *truncated Gaussian* distribution: omit tails so that

$$0 < \alpha_1 \le a \le \alpha_2 < \infty$$

$$\langle a(u,v) \rangle = \int_\Omega \int_\mathcal{D} a \nabla u \cdot \nabla v \, dx \, dP(\omega) = \int_\Gamma \int_\mathcal{D} a(x, \underline{\xi}) \nabla u \cdot \nabla v \, dx \, \rho(\underline{\xi}) d \underline{\xi}$$

Still have $\rho(\underline{\xi}) \neq \rho_1(\xi_1) \cdots \rho_m(\xi_m)$

but (perhaps) less of a crime

Note: can work with such a distribution even though orthogonal polynomials are not known (Gautschi code); Rys polynomials

Comparison of Galerkin and Collocation

Recall from last talk:

Stochastic Collocation, alternative to method discussed so far

Discrete solution $u_{hp}(x,\underline{\xi}) = \sum_{k=1}^{N_\xi} u_h^{(k)}(x) L_k(\underline{\xi})$

Obtained by solving

$$\int_{\mathcal{D}} (a_0(x) + \sigma \sum_{r=1}^{m} \sqrt{\lambda_r}\, a_r(x)\xi_r) \nabla u \cdot \nabla v \, dx = \int_{\mathcal{D}} f\, v\, dx$$

For set of samples $\{\underline{\xi}^{(k)}\}$ situated in a sparse grid

Advantage of this approach: simpler (decoupled) systems

Disadvantage: larger stochastic space for comparable accuracy

larger by factor approximately 2^p

Dimensions of Stochastic Space

m (#KL)	p	Galerkin	Collocation Sparse	Collocation Tensor
4	1	5	9	16
	2	15	41	81
	3	35	137	256
	4	70	401	625
10	1	11	21	1024
	2	66	221	59,049
	3	286	1582	1,048,576
	4	1001	8,801	9,765,625
30	1	31	61	1.07(9)
	2	496	1861	2.06(14)
	3	5456	37,941	1.15(18)
	4	46,376	582,801	9.31(20)

~ size of coarse grid space for MG / Version 1

systems for collocation MG / Version II

Experiment

(E., Miller, Phipps, Tuminaro)

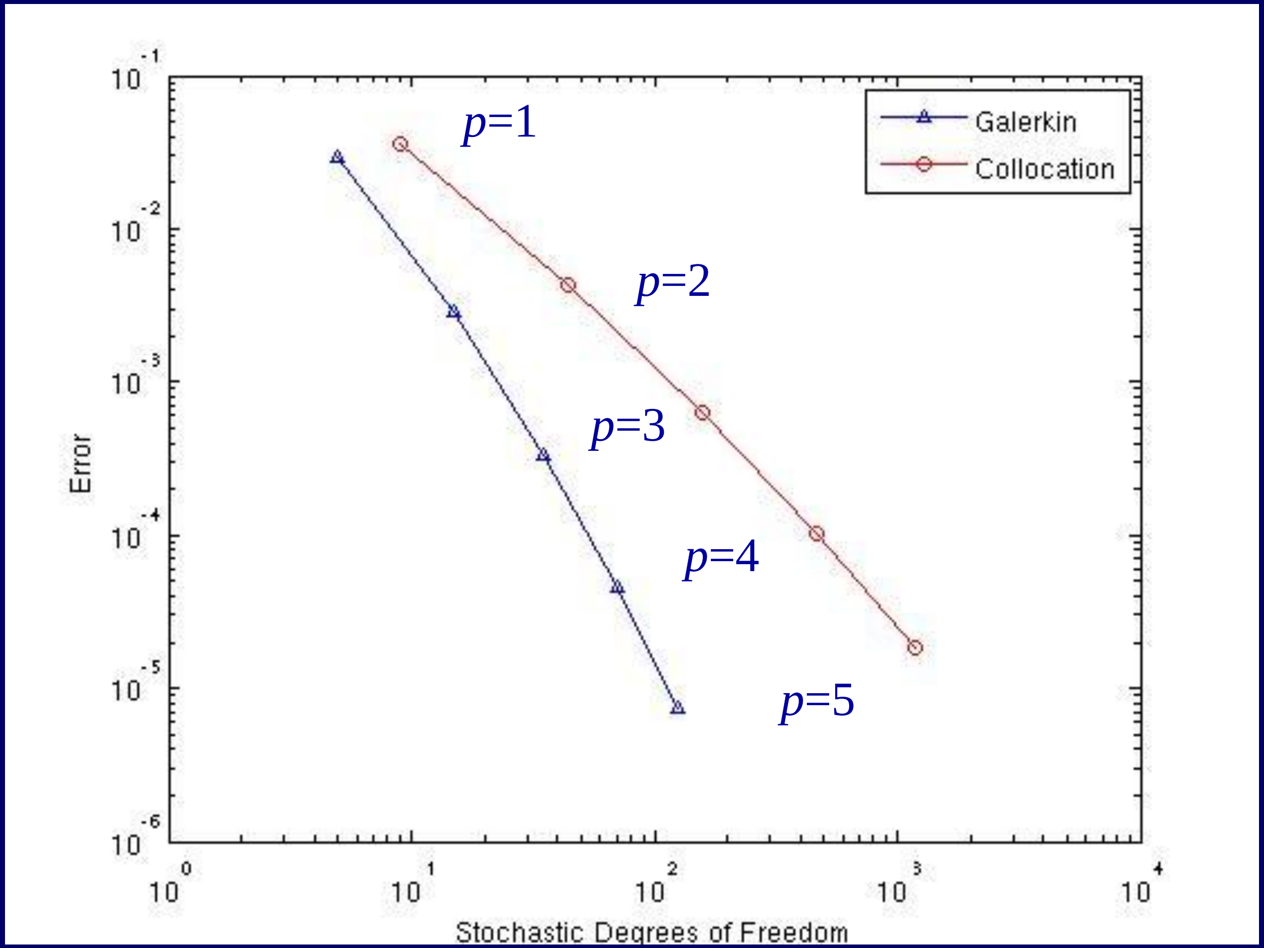
- Solve the stochastic diffusion equation by both methods
- Compare the accuracy achieved for different parameter sets[1]
- For parameter choices giving comparable accuracy, compare solution costs
- Spatial discretization fixed (32x32 finite difference grid)

Solution algorithm for both discretizations:

Preconditioned conjugate gradient with mean-based preconditioning, using AMG for the approximate diffusion solve

[1]Estimated using a high-degree (p=10) Galerkin solution.

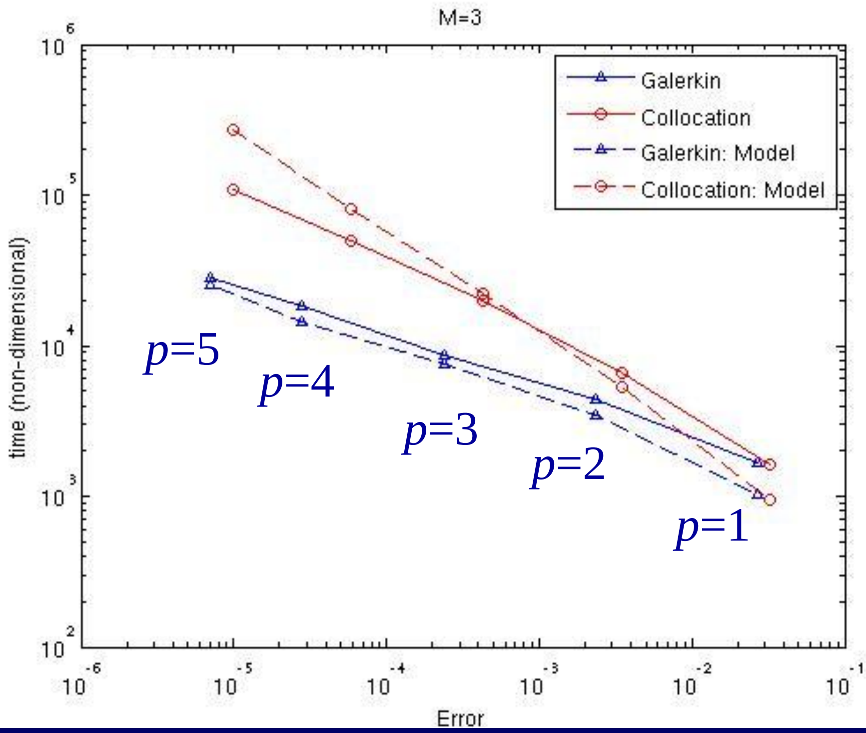
Experimental results: accuracy

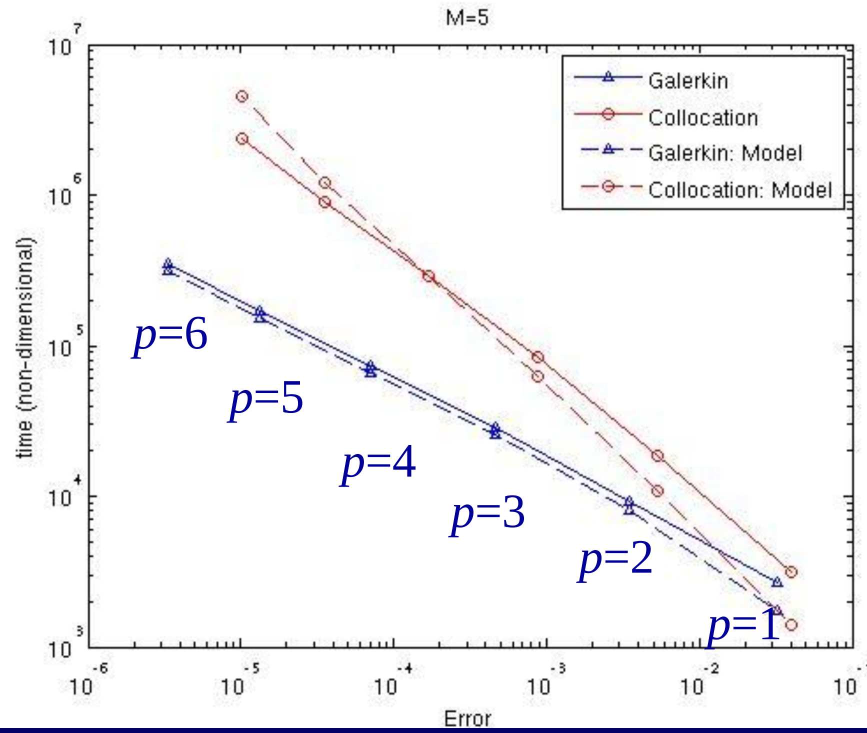


For fixed m=4:

Similar p = { polynomial degree for SG / "level" for collocation } produces comparable errors

Experimental results: performance





Performed on a serial machine with C-code and CG/AMG code from Trilinos

Observation: Galerkin faster, more so as number of stochastic variables (KL terms) grows

More General Problems

For the problem discussed, based on a KL expansion, has a *linear* dependence on the stochastic variable $\underline{\xi}$

Other models have *nonlinear* dependence. For example

$$a(x,\xi) = a_{\min} + \underbrace{e^{c(x,\xi)}}_{\text{Nonlinear}}$$

$$c(x,\xi) = a_0(x) + \sigma \sum_{r=1}^{m} \sqrt{\lambda_r}\, a_r(x)\xi_r$$

For Gaussian *c*, called a *log-normal* distribution

In particular: coercivity is guaranteed with this choice

More General Problems

For stochastic Galerkin, need a finite term *expansion* for *a*

$$a(x,\underline{\xi}) = a_0(x) + \sigma \sum_{r=1}^{M} \sqrt{\lambda_r}\, a_r(x) \psi_r(\underline{\xi})$$

Note: not ξ_r

→ matrix

$$A = G_0 \otimes A_0 + \sum_{\mathrm{r=1}}^{\mathrm{M}} G_r \otimes A_r$$

$$[G_r]_{ij} = \langle \psi_r \psi_i \psi_j \rangle$$

Less sparse

More importantly: # terms M will be larger
perhaps as large as $2N_\xi$

⟹ mvp will be more expensive

In contrast

Collocation is less dependent on this expansion

$A^{(k)}$ comes from $\int_{\mathcal{D}} a(x, \underline{\xi}^{(k)}) \nabla u \cdot \nabla v \, dx$ for each sparse grid point $\underline{\xi}^{(k)}$

Many matrices to assemble, but mvp is not a difficulty

Concluding remarks

- Exciting new developments models of PDEs with uncertain coefficients
- Replace pure simulation (Monte Carlo) with finite-dimensional models that simulate sampling at potentially lower cost
- Two techniques, the *stochastic Galerkin* method and the *stochastic collacation* method, were presented, each with some advantages
- Solution algorithms are available for both methods, and work continues in this direction