Improving Robustness in Multiscale Methods

Scott MacLachlan

maclachl@colorado.edu

Department of Applied Mathematics, University of Colorado at Boulder

Support and Collaboration

- This work has been supported by the University of Colorado Chancellor's Fellowship program, the DOE SciDAC TOPS program, the Center for Applied Scientific Computing at Lawrence Livermore National Lab, and Los Alamos National Laboratory.
- This work has been performed in collaboration with Steve McCormick, Tom Manteuffel, John Ruge, Marian Brezina at CU-Boulder, J. David Moulton at LANL, and Rob Falgout at CASC.

Outline

- Modern Scientific Computing
- Multigrid Methods
- Upscaling and Homogenization
- Adaptive Multigrid Methods
- Future Work

Why Compute?

- Interested in modeling physical processes
  - Diffusion (Heat, Energy, Chemical)
  - Fluid Flow
  - Particle Transport
  - Elastic Materials
- Can describe these processes through differential equations (both ODEs and PDEs)
- Cannot write down closed form solutions
- Need to find (approximate) solutions in other ways

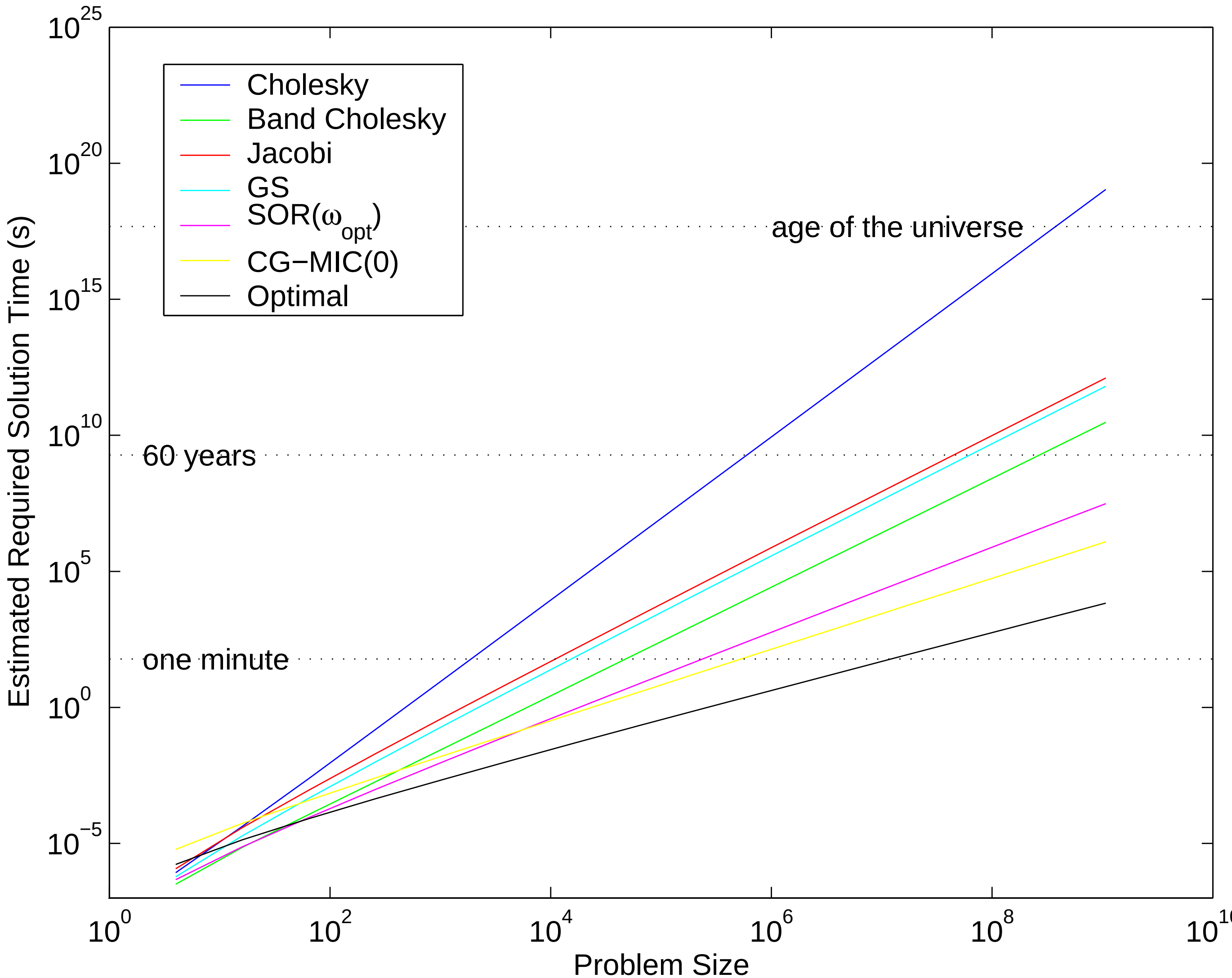
Scientific Computation

- Significant interest in simulating complex physical systems with features, and hence solutions, that vary on multiple scales
- Accuracy constraints lead to discretizations with tens of millions, or even billions, of degrees of freedom (DOFs)
  - 3D Tsunami Model: 200 million cells
  - Transport: 500 million to 1 billion DOFs

Properties of Discretizations

- We consider discretizations of the underlying continuum models (differential equations) via finite elements (or finite differences)
- The matrices from these discretizations tend to be sparse and ill-conditioned
- The matrices inherit properties of the continuum model (e.g. symmetry, definiteness)

Classical Methods do not Suffice



Porous Media Flow

- Saturated flow in a reservoir can be modeled via Darcy's Law:

$$\mathbf{u}(\mathbf{x}) = -\mathcal{K}(\mathbf{x})\nabla p(\mathbf{x})$$
$$\nabla \cdot \mathbf{u}(\mathbf{x}) = Q(\mathbf{x})$$

- Simulation domain: $\backsim 10^3$ meters in each dimension
- Material properties ($\mathcal{K}(\mathbf{x})$) vary on millimeter scales

Porous Media Flow

- Saturated flow in a reservoir can be modeled via Darcy's Law:

$$\mathbf{u}(\mathbf{x}) = -\mathcal{K}(\mathbf{x})\nabla p(\mathbf{x})$$
$$\nabla \cdot \mathbf{u}(\mathbf{x}) = Q(\mathbf{x})$$

- Simulation domain: $\backsim 10^3$ meters in each dimension
- Material properties ($\mathcal{K}(\mathbf{x})$) vary on millimeter scales
- For 3-D, fully-resolved flow calculation, need $\backsim 10^{18}$ DOFs
- Such a simulation is at the limits of the capability of modern supercomputers (the fastest of which performs 3.5×10^{13} floating point operations per second)

Scientific Computation

- Significant interest in simulating complex physical systems with features, and hence solutions, that vary on multiple scales
- Accuracy constraints lead to discretizations with tens of millions, or even billions, of degrees of freedom (DOFs)
  - 3D Tsunami Model: 200 million cells
  - Transport: 500 million to 1 billion DOFs
  - Fully-resolved Porous Media Flow: 10^{18} DOFs
- Without optimal methods, solving such problems can be prohibitively expensive

Scientific Computation

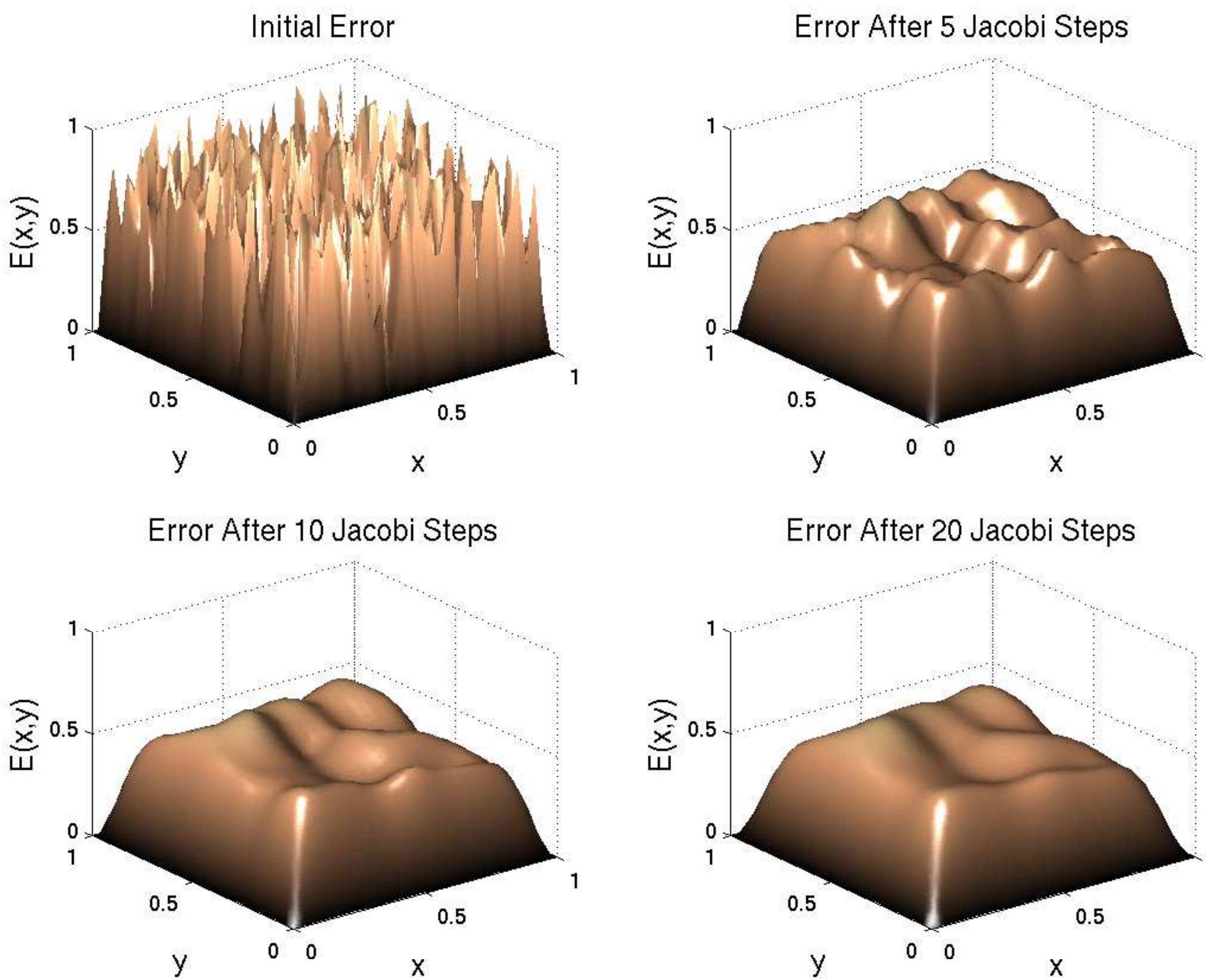
- Significant interest in simulating complex physical systems with features, and hence solutions, that vary on multiple scales
- Accuracy constraints lead to discretizations with tens of millions, or even billions, of degrees of freedom (DOFs)
  - 3D Tsunami Model: 200 million cells
  - Transport: 500 million to 1 billion DOFs
  - Fully-resolved Porous Media Flow: 10^{18} DOFs
- Without optimal methods, solving such problems can be prohibitively expensive
- Bottom Line:

Accuracy per computational cost

Stationary Iterative Methods

- The Jacobi and Gauss-Seidel iterations do converge for FE discretizations of elliptic operators, but require $O(N^2)$ operations for 2-D problems and $O(N^{\frac{5}{3}})$ operations for 3-D
- These methods do, however, resolve some components much faster than others
- For the Laplacian operator, it is the geometrically smoothest components of the solution that are the slowest to be resolved
- For this reason, Jacobi and Gauss-Seidel are often called smoothers - they smooth the error in the approximation

Smoother Performance



Multigrid

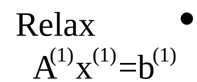
Multigrid Methods achieve optimality through complementarity

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

- Relaxation



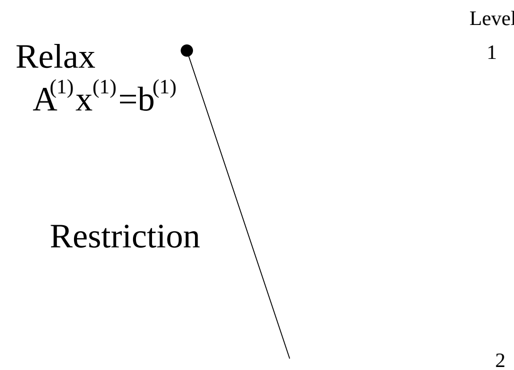
- Use a smoothing process (such as Gauss-Seidel) to eliminate oscillatory errors
- Remaining error satisfies $Ae = r \equiv b - Ax$

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

- Relaxation
- Restriction



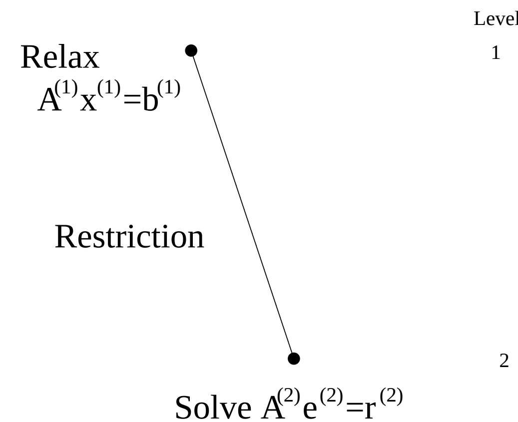
- Transfer residual to coarse grid

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction



- Use coarse-grid correction to eliminate smooth errors
- To solve for error on coarse grid, use residual equation

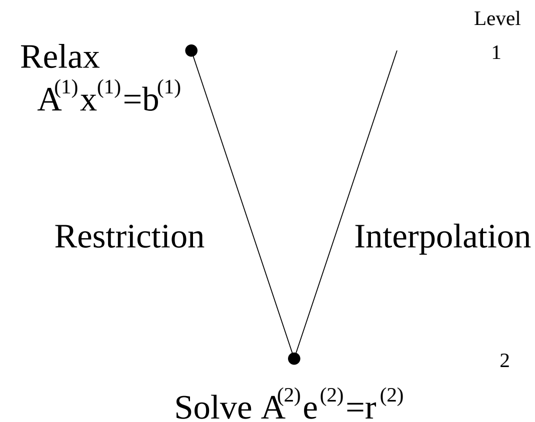
$$A^{(2)}e^{(2)} = r^{(2)}$$

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction
- Interpolation

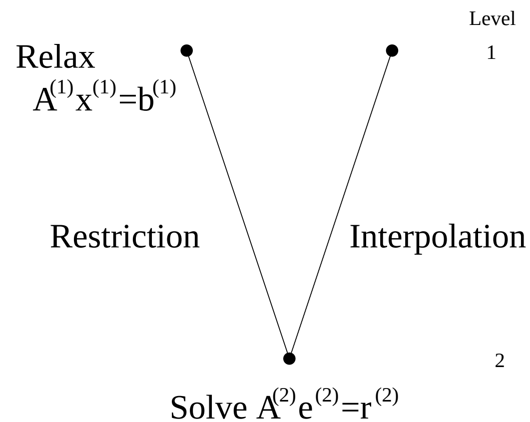


- Transfer correction to fine grid

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components



- Relaxation
- Restriction
- Coarse-Grid Correction
- Interpolation
- Relaxation

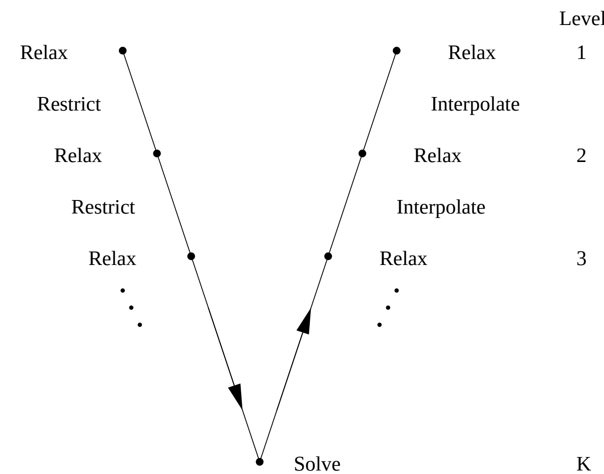
- Relax once again to remove oscillatory error introduced in coarse-grid correction

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction
- Interpolation
- Relaxation



Obtain optimal efficiency through recursion

Geometric Multigrid

- When $Ax = b$ comes from a geometrically regular discretization of a DE, that information can be used in the coarse-grid problems
- Coarse grids chosen by removing points from the fine grid in a geometrically regular fashion
- Restriction (R) and interpolation (P) operators computed using geometric locations
- Coarse-grid operators determined by rediscretization on the reduced space

BoxMG

- The Black Box Multigrid Algorithm (BoxMG) was developed by Dendy for discontinuous coefficient operators, such as $-\nabla \cdot \mathcal{K}(\mathbf{x})\nabla$
- Coarsening is geometric
- Interpolation is chosen to approximately preserve continuity of normal flux
- Variational formulation

Variational Multigrid

- Multigrid with $R = P^T$ and $A_c = RAP$ is called a *variational formulation*
- Terminology comes from minimization form of $Ax = b$:

$$F(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle$$
$$x = \arg\min_{v \in \mathcal{H}} F(v)$$

- Given an approximation, v, to the solution on the fine level, it can be shown that the optimal coarse grid correction, Pw, solves

$$(P^T AP)w = P^T(b - Av)$$

Coarse-Scale Models

- Quality of coarse-grid corrections depends on representation of fine-scale details in coarse-scale operator
- Rediscretization requires explicit averaging of fine-scale structures
- Variational coarsening allows multiscale information to be encoded in the coarse-grid operators
- Coarse-scale operator needs to reflect information about the fine-scale operator relating to the low-energy modes for which a coarse-scale correction is being computed.
- Practical interest is in coarse-scale properties, such as total net flux

The Need for Coarse-Scale Models

- Fine-scale discretization requires too many DOFs
- Coarse-scale properties are influenced by fine-scale variations
- Need physically-meaningful solutions
- Cannot directly (explicitly) capture effect of fine-scale variations in a coarse-scale discretization
- Goal: derive effective, coarse-scale models

Interpretation of Multigrid CGOs

- Fine-scale, finite-element discretization of porous-media problem:

$$A_{ij} = \mathbf{e}_j^T A \mathbf{e}_i = \int_\Omega \langle \mathcal{K}(\mathbf{x}) \nabla \phi_i, \nabla \phi_j \rangle d\Omega$$

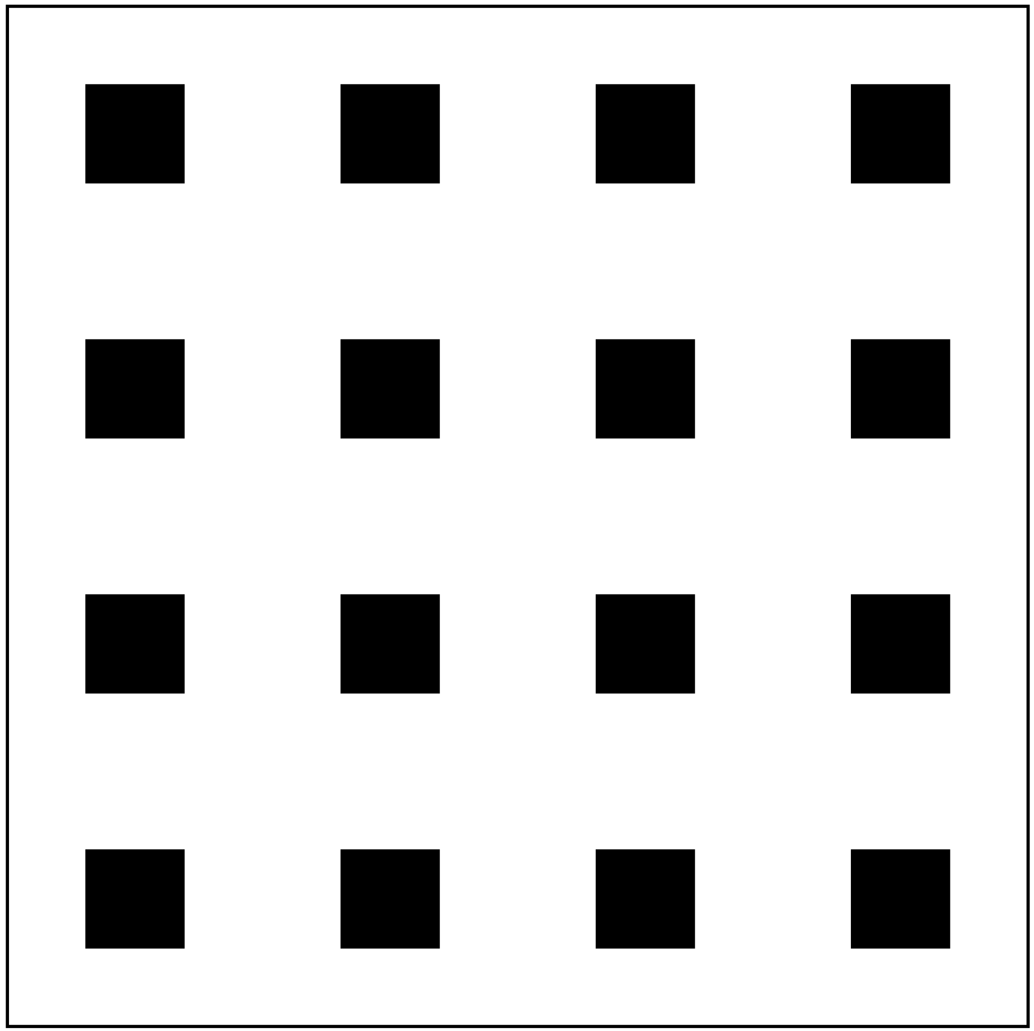
- Variational coarsening gives finite-element discretizations on coarse grids:

$$\begin{aligned}
(A_c)_{ij} = (P^T A P)_{ij} &= (P\hat{\mathbf{e}}_j)^T A (P\hat{\mathbf{e}}_i) \\
&= \int_\Omega \left\langle \mathcal{K}(\mathbf{x}) \nabla \left(\sum_l p_{li} \phi_l \right), \nabla \left(\sum_k p_{kj} \phi_k \right) \right\rangle d\Omega \\
&= \int_\Omega \langle \mathcal{K}(\mathbf{x}) \nabla \hat{\phi}_i, \nabla \hat{\phi}_j \rangle d\Omega
\end{aligned}$$

- Coarse-grid basis functions are linear combinations of fine-grid basis functions (weighted by the interpolation operators)

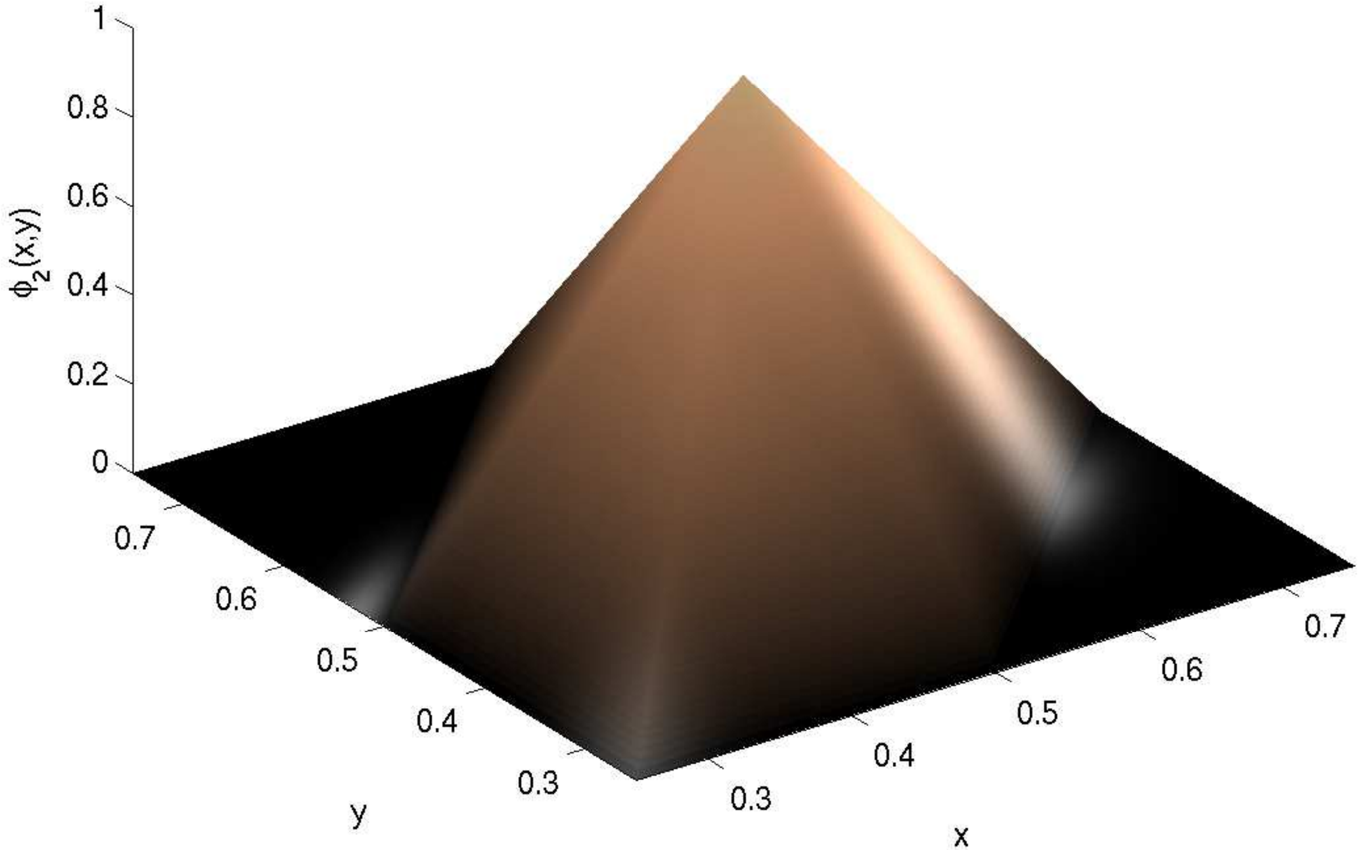
Sample Basis Functions

Periodic permeability field, $\mathcal{K}(\mathbf{x})$, with jump of 10^3



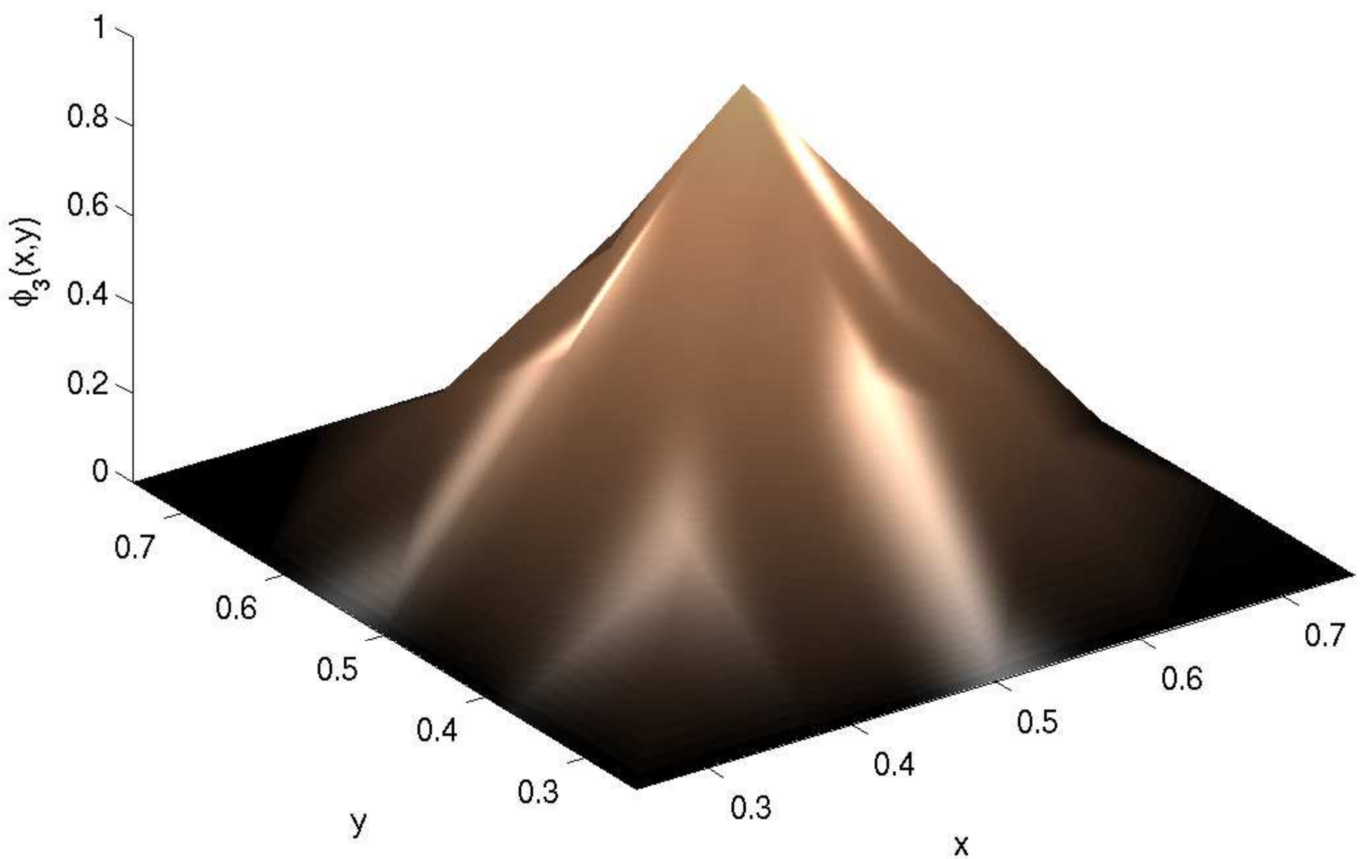
Sample Basis Functions

Bilinear basis function on 4×4 grid



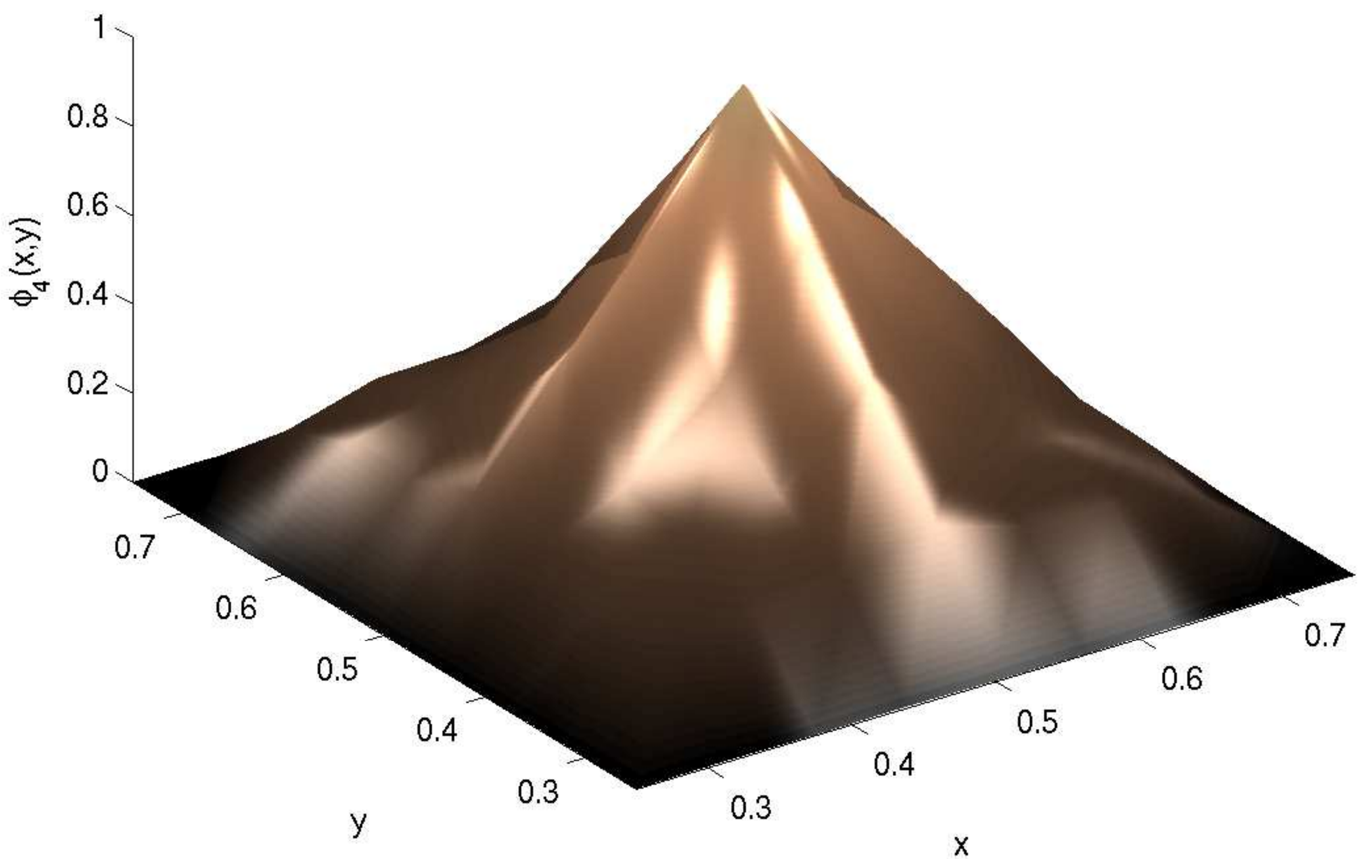
Sample Basis Functions

8×8 grid multiscale basis function



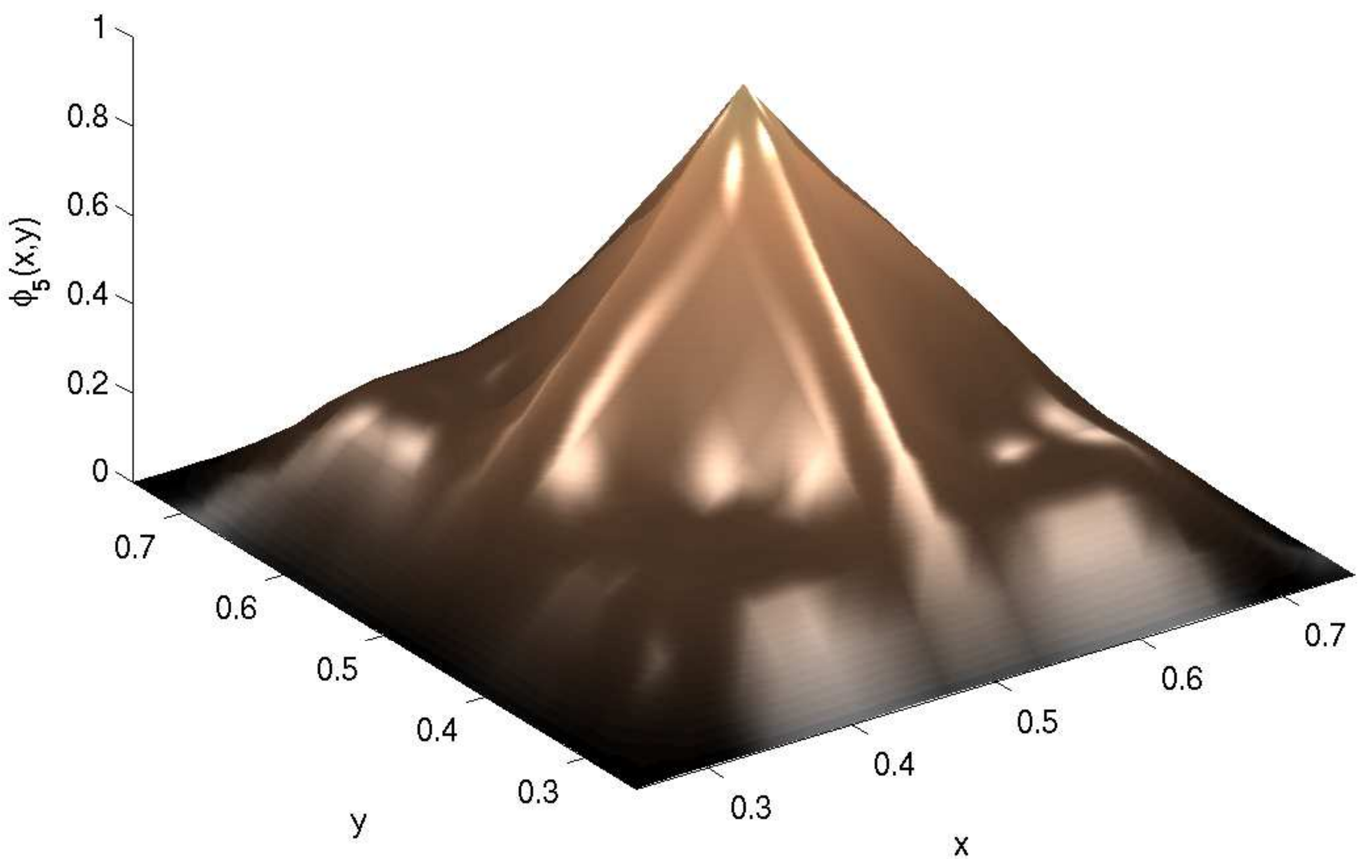
Sample Basis Functions

16×16 grid multiscale basis function



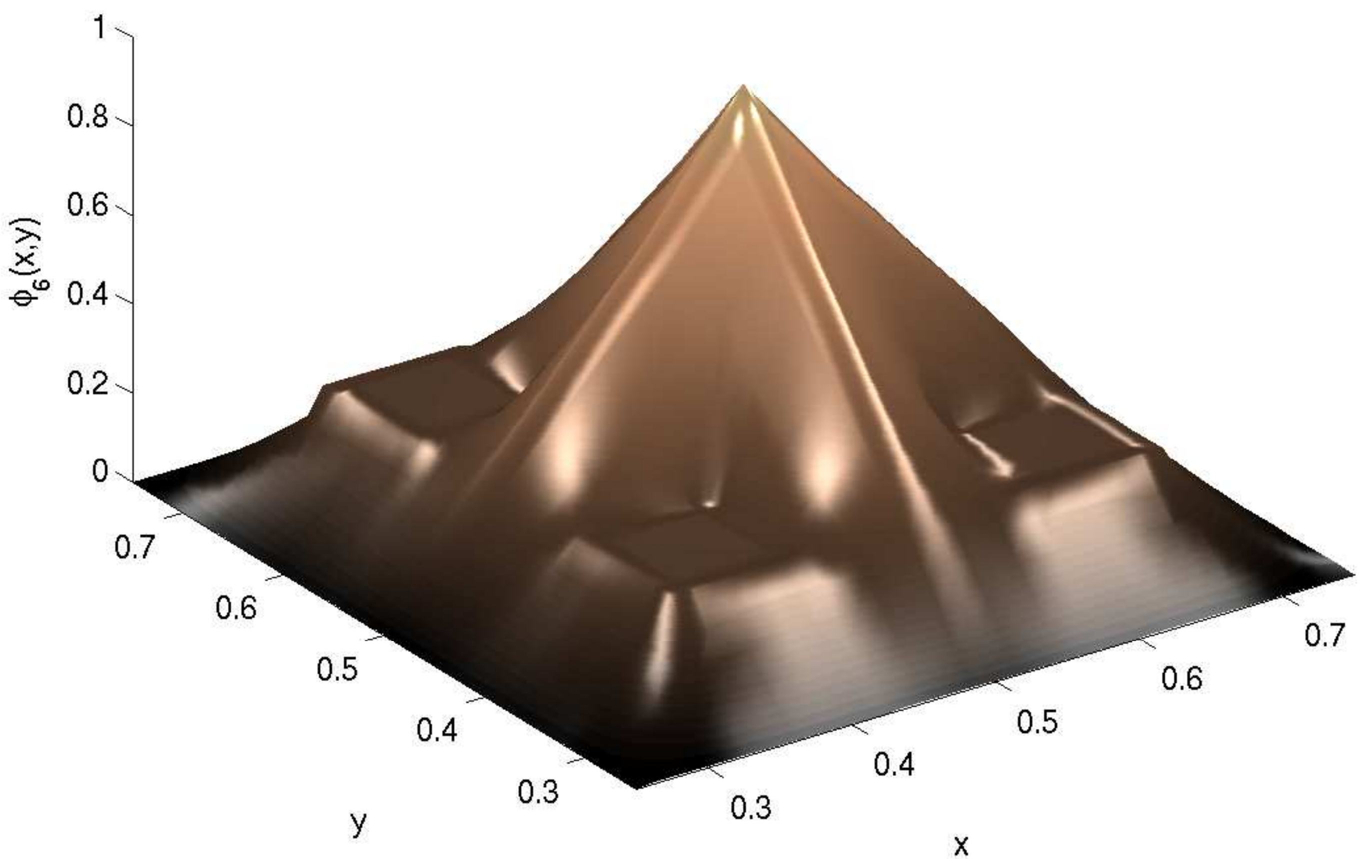
Sample Basis Functions

32×32 grid multiscale basis function



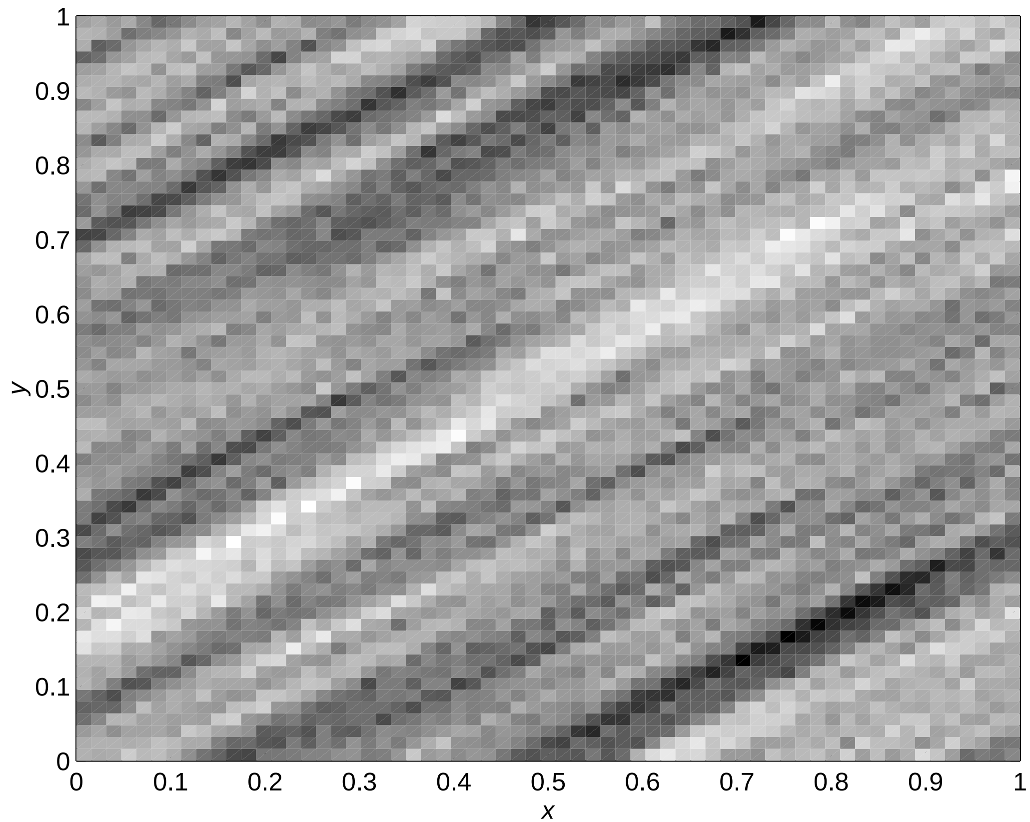
Sample Basis Functions

64×64 grid multiscale basis function



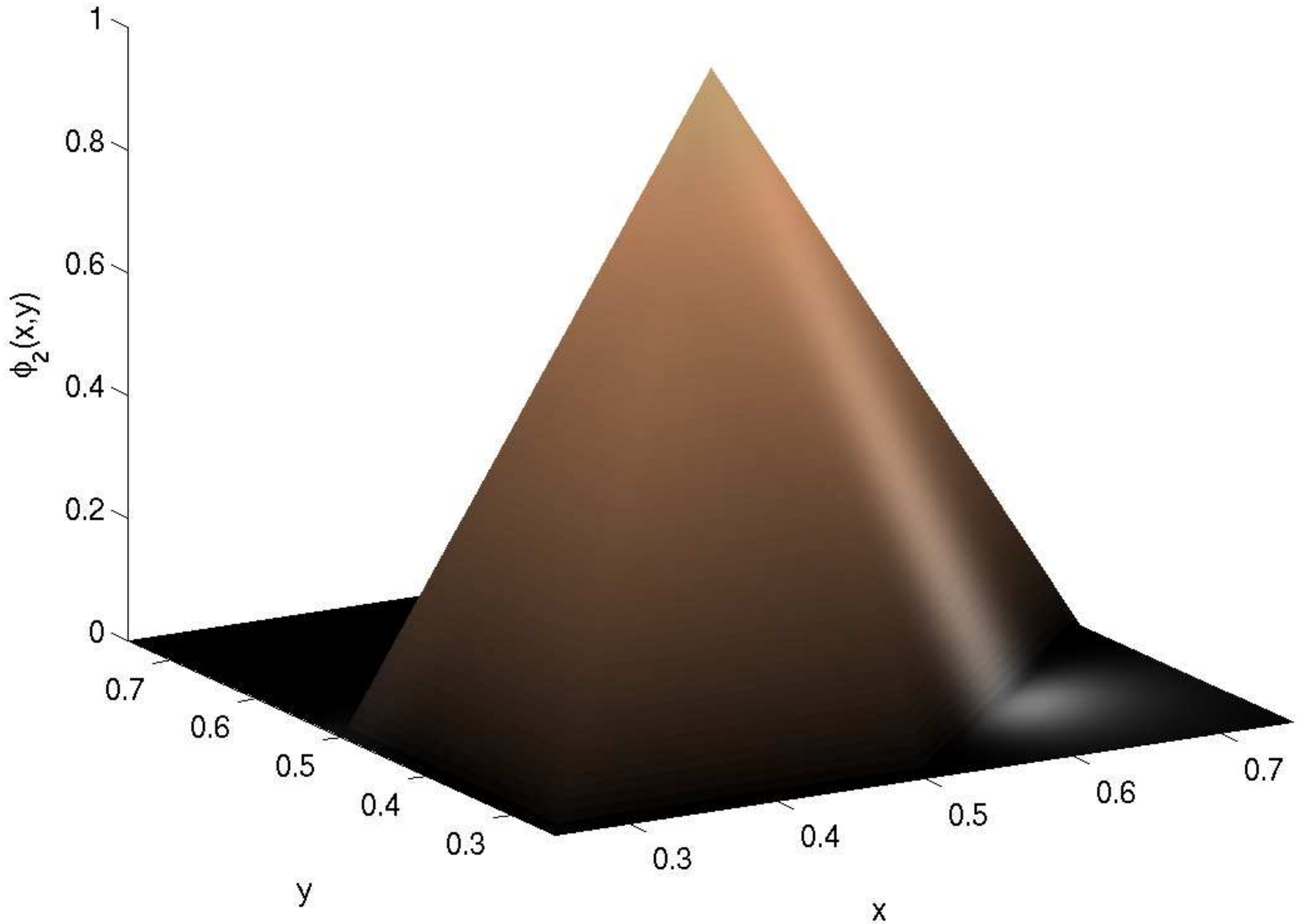
Sample Basis Functions

Geostatistical permeability field, $\mathcal{K}(\mathbf{x})$, with range of $[10^{-2}, 10^{2}]$
(Black pixels correspond to $\mathcal{K} = 10^{-2}$)



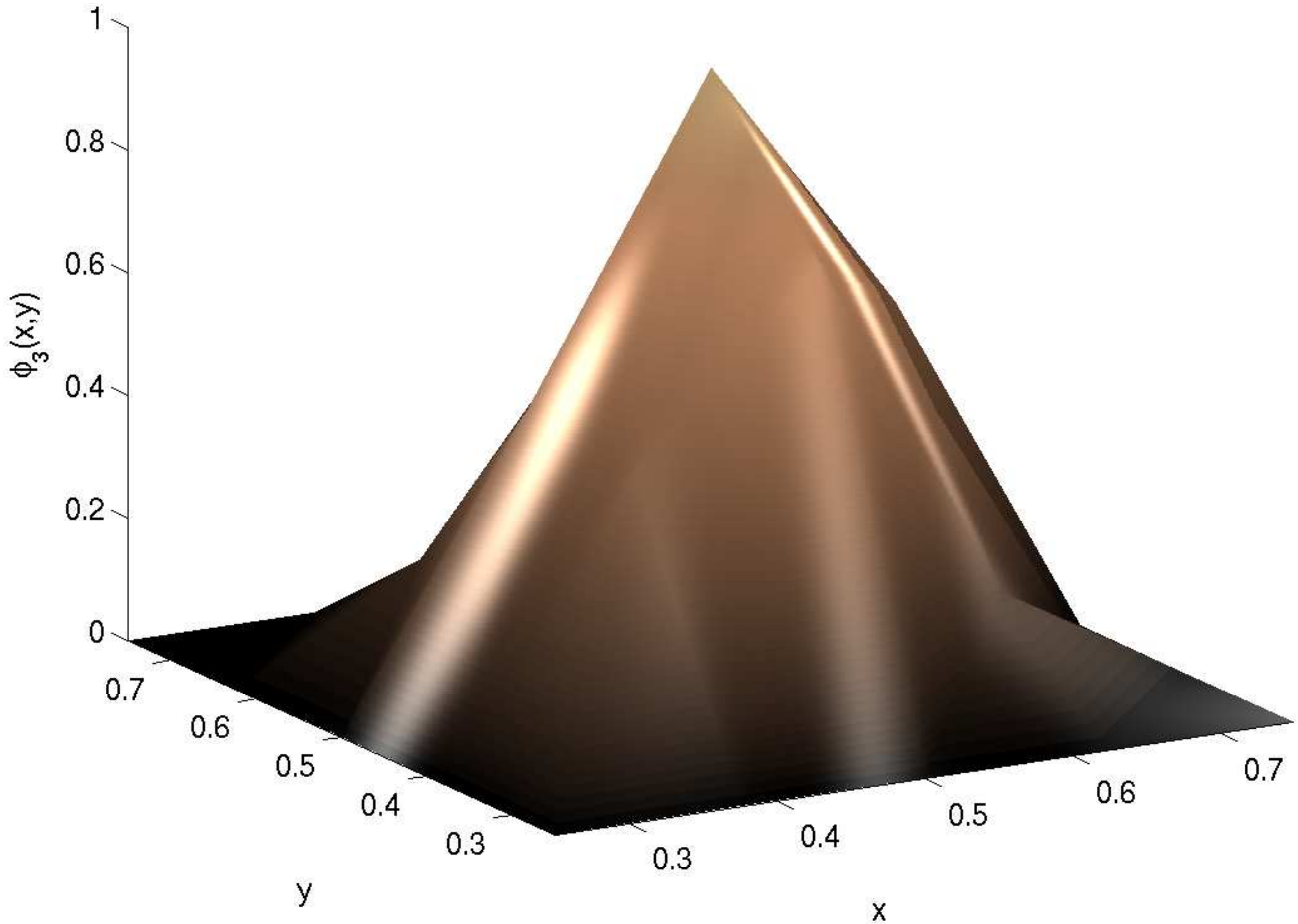
Sample Basis Functions

Bilinear basis function on 4×4 grid



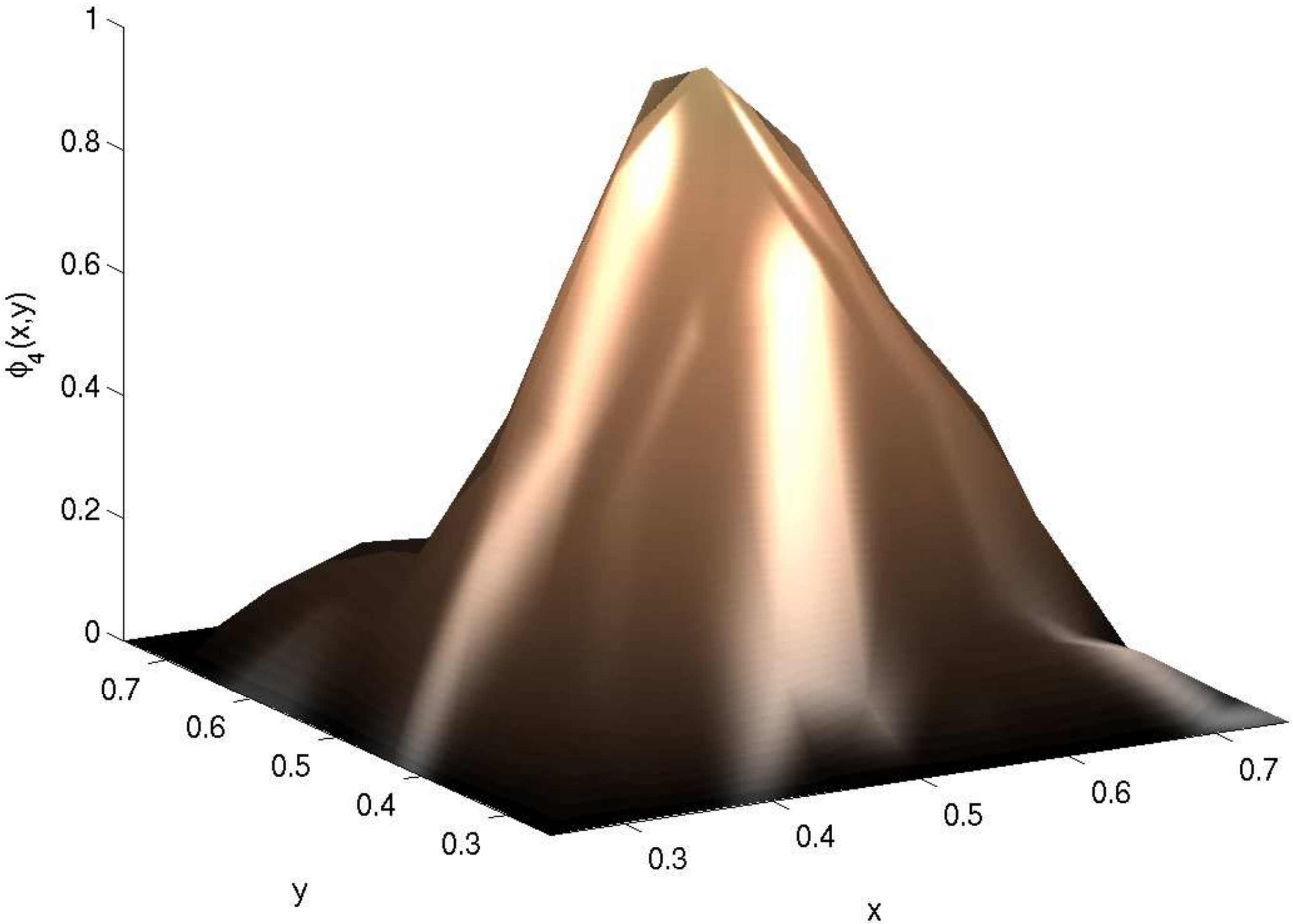
Sample Basis Functions

8×8 grid multiscale basis function



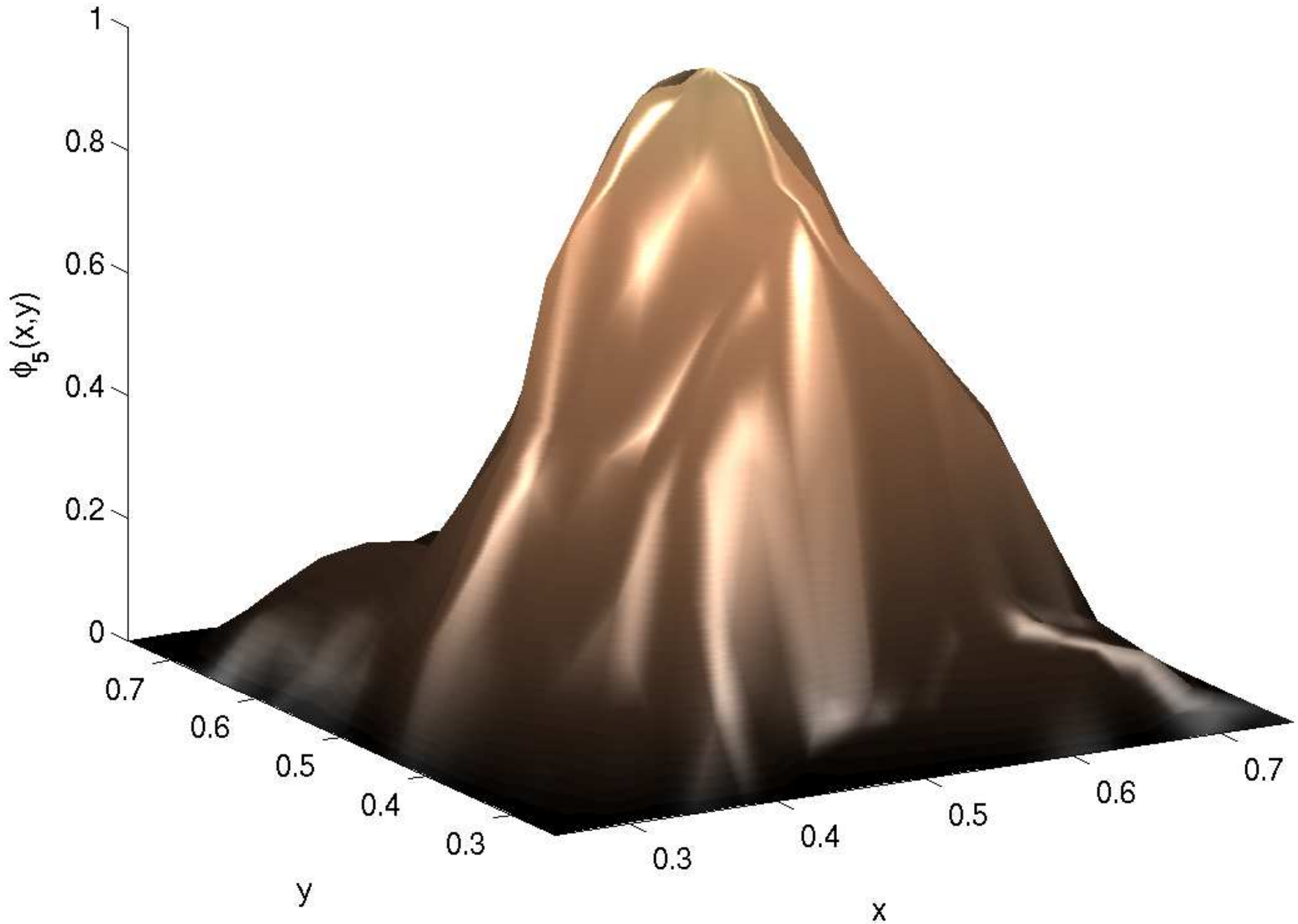
Sample Basis Functions

16×16 grid multiscale basis function



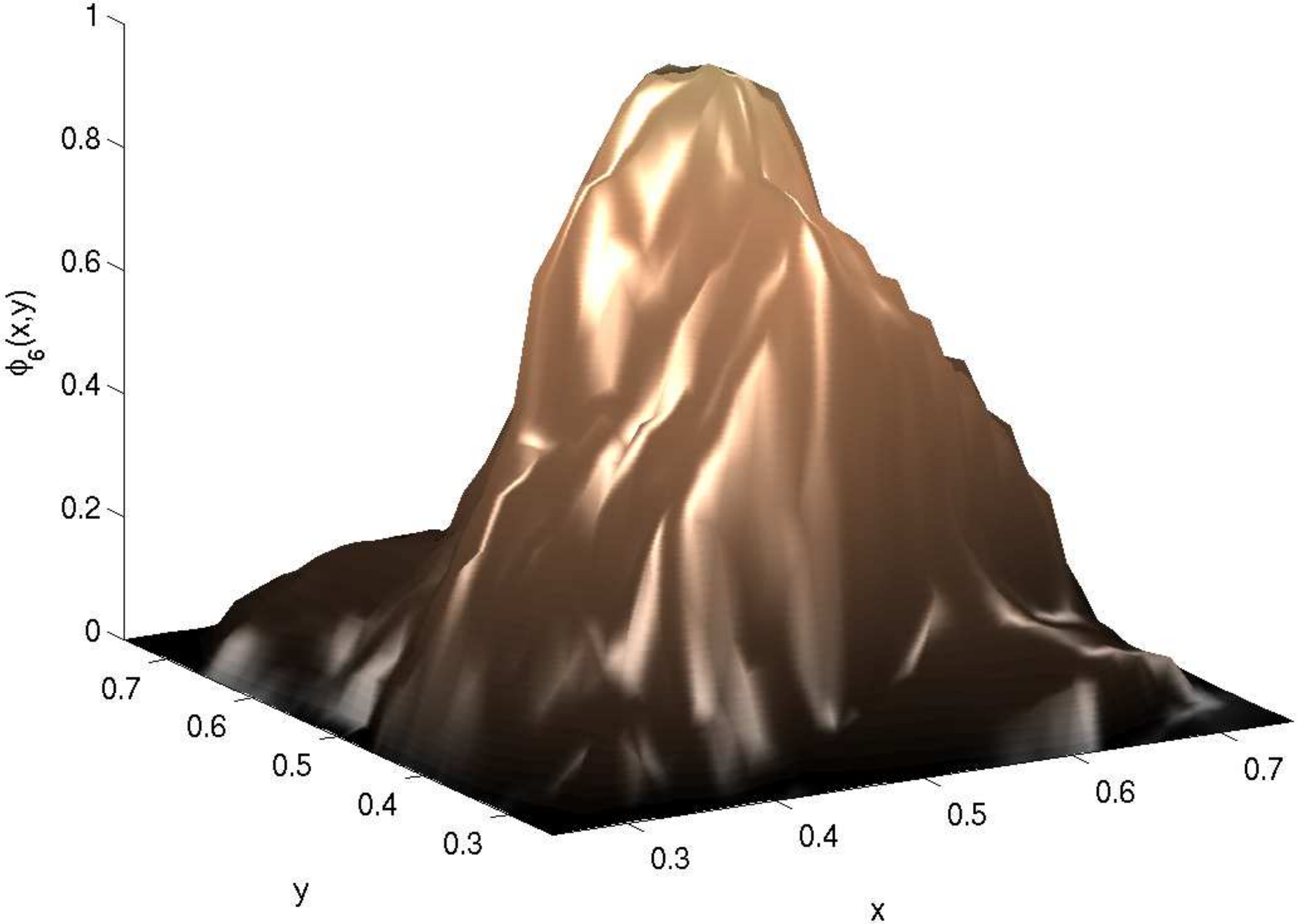
Sample Basis Functions

32×32 grid multiscale basis function



Sample Basis Functions

64×64 grid multiscale basis function



Multiscale Goals

- Capture macro-scale properties of fine-scale operator
  - Modeling flow, so compute net flux through domain of interest
  - Boundary conditions: $p = 1$ on left, $p = 0$ on right, no-flow (Neumann) on top and bottom
  - Integrate $(\mathcal{K}\nabla p) \cdot \mathbf{n}$ along outflow
- Linear model is a simplification of unsaturated/multi-phase flow where $\mathcal{K} = \mathcal{K}(p, \mathbf{x})$
  - Local pressure fluctuations important
  - Seek to match local maxima and minima of p

Periodic Theory Approach

- Two-scale asymptotic analysis to compute homogenized permeability
- Upscaled equation, $-\nabla \cdot \hat{\mathcal{K}} \nabla p_0(\mathbf{x}) = Q(\mathbf{x})$, arises from solvability conditions with periodic BCs
- Upscaled permeability, $\hat{\mathcal{K}}$, given in terms of particular solutions of fine-scale problem over averaging subdomain
- Coarse-scale model created by rediscretization on that scale
- We’ve shown equivalence between this approach and that of Durlofsky

Flux Calculations

- Multigrid upscaling procedure
  - Discretize on fine scale
  - Use BoxMG to coarsen to given coarse scale
  - Solve coarse-scale problem
  - Interpolate solution to fine scale
  - Compute Outflow Flux
- Periodic theory approach
  - Discretize on fine scale
  - Solve two fine-scale problems per coarse-scale element to get $\hat{\mathcal{K}}$
  - Rediscretize on coarse scale
  - Solve coarse-scale problem
  - Compute flux with coarse-scale permeability and pressure

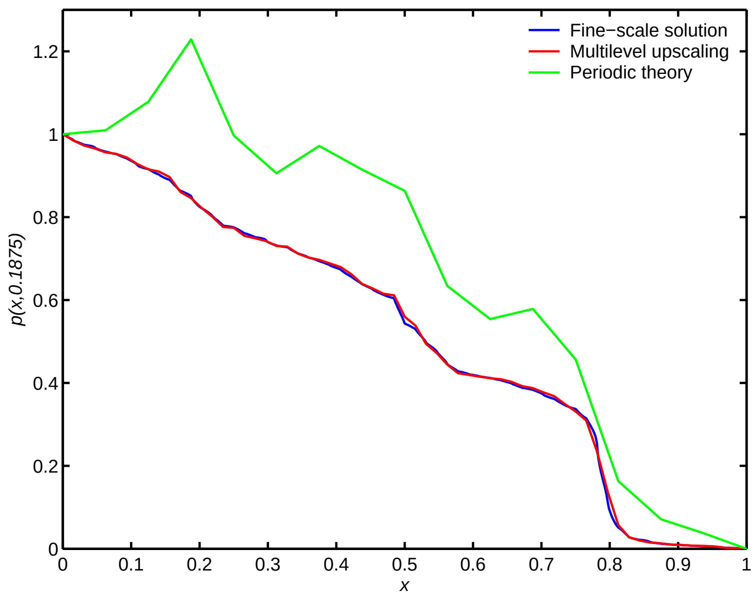
Flux Calculations

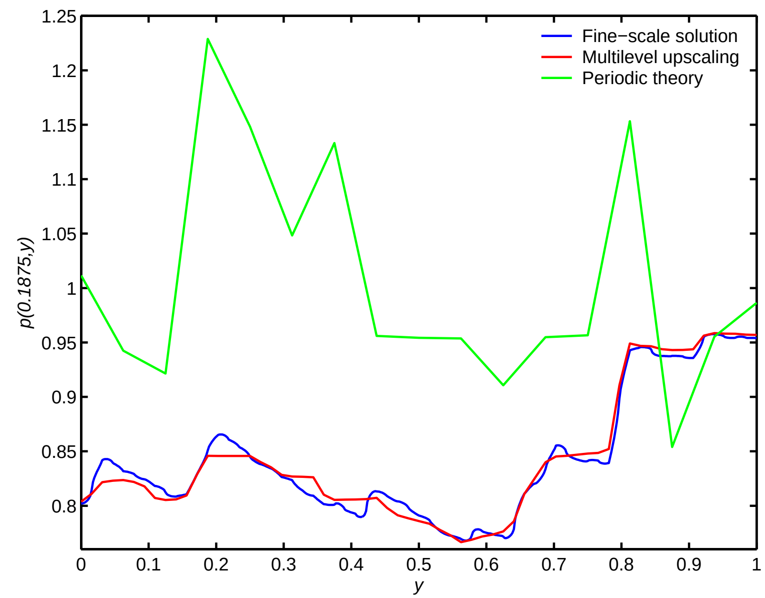
- Computed Flux for 512×512 element discretization is 2.229

Coarse Grid	Multilevel Upscaling		Periodic Theory	
	Flux	% Error	Flux	% Error
32×32	2.430	9.0%	2.319	4.0%
16×16	2.558	14.8%	3.482	56.2%
8×8	2.599	16.6%	4.923	120.8%
4×4	2.493	11.8%	3.124	40.1%

Fine-Scale Structure

- Accurate reconstruction of fine-scale structure needed to address nonlinearities in unsaturated and multi-phase flows





Accuracy of Coarse-Scale Models

- Performance of multilevel upscaling technique comes from accuracy of coarse-scale model
- In variational multigrid setting, good coarse-scale models come from good choices in interpolation
- Good interpolation must complement relaxation

Algebraically-Smooth Error

- Multigrid methods reduce error through
  - Relaxation (Jacobi, Gauss-Seidel)
  - Coarse-grid correction (variational)
- Error which is not efficiently reduced by relaxation is called *algebraically smooth* and must be reduced by coarse-grid correction
- Pointwise relaxation implies that algebraically-smooth error, $\mathbf{e}$, satisfies $A\mathbf{e} \approx \mathbf{0}$, relative to $\mathbf{e}$
- If the origins of the matrix are known, so is character of algebraically-smooth error

Algebraic Multigrid

- Assume no knowledge of grid geometry
- Interpolation and coarse grids chosen based only on the entries of the matrix
- Primary goal is to interpolate suitable corrections from the coarse grids
- Assume algebraically-smooth error is locally constant
- Equivalently, assume global near null space is the constant vector

Adaptive Multigrid

- If we don’t know what algebraically-smooth error looks like, can we still develop an effective multigrid method?

Adaptive Multigrid

- If we don't know what algebraically-smooth error looks like, can we still develop an effective multigrid method? Yes!
- Use relaxation on $Ax = 0$ to expose algebraic smoothness
- Fine-grid relaxation quickly exposes local character of algebraic smoothness
- Use this representation to determine interpolation

Adapting Interpolation

- AMG and BoxMG choose interpolation by a fixed formula
- Now need interpolation that depends on both the matrix, A, and the prototypical algebraically-smooth error, $\mathbf{v}$
- Algebraic smoothness still means that $A\mathbf{e} \approx \mathbf{0}$, or

$$a_{ii}e_i \approx -\sum_{k \in C_i} a_{ik}e_k - \sum_{j \in F_i} a_{ij}e_j$$

- Approximate e_j by values in $C_i \cap C_j$, weighted by a_{jk} and v_k

Scaling Property

- If we scale $A \to DAD$ for diagonal matrix D, adaptive AMG performance need not suffer
- If $A\mathbf{v} = \mathbf{0}$, then $DAD(D^{-1}\mathbf{v}) = 0$
- Replacing $\mathbf{v}$ with $D^{-1}\mathbf{v}$, the coarse-grid matrices for DAD are diagonally-rescaled versions of those for A
- Pointwise relaxation is also invariant to such scaling
- If we could generate $D^{-1}\mathbf{v}$ as easily $\mathbf{v}$, overall performance wouldn't degrade

Test Problems

- $-\nabla \cdot \mathcal{K}(\mathbf{x})\nabla p(\mathbf{x}) = 0$ on $[0,1]^2$
- Problem 1:
  - $\mathcal{K}(\mathbf{x}) = 1$ (Laplace), Full Dirichlet BCs
- Problem 2:
  - $\mathcal{K}(\mathbf{x}) = 10^{-8}$ on 20% of elements chosen randomly, $\mathcal{K}(\mathbf{x}) = 1$ otherwise
  - Dirichlet BCs on left and right, Neumann on top and bottom
- Setup Phase: Single V-cycle, # pre-relaxations chosen for optimal performance
- Solution Phase: V(1,1) cycles until residual reduced by 10^{10} or 200 iterations
- Geometric choice of coarse grids

Numerical Results - Solution Time

Total time to setup MG method and reduce residual by 10^{10}

	Problem 1		Problem 2	
h	Standard AMG	Adapted AMG	Standard AMG	Adapted AMG
$1/64$	0.04s	0.04s	0.05s	0.03s
$1/128$	0.22s	0.25s	0.28s	0.31s
$1/256$	0.91s	0.89s	1.04s	1.09s
$1/512$	3.32s	3.52s	4.40s	4.84s
$1/1024$	13.13s	14.70s	17.64s	22.06s

Numerical Results - Convergence Factors

Asymptotic convergence factors of resulting V(1,1) cycles

	Problem 1		Problem 2	
h	Standard AMG	Adapted AMG	Standard AMG	Adapted AMG
$1/64$	0.104	0.067	0.209	0.194
$1/128$	0.115	0.073	0.212	0.202
$1/256$	0.124	0.079	0.233	0.243
$1/512$	0.131	0.080	0.290	0.288
$1/1024$	0.137	0.079	0.375	0.376

Test Problems - Scaling

- Standard AMG does fine on these problems
- Break assumption on local character of algebraically-smooth error
- Scale problems nodally by 10^{5r}, where r is chosen uniformly between 0 and 1 for each node

Numerical Results - Solution Time

Total time to setup MG method and then reduce residual by 10^{10}

	Problem 1		Problem 2	
h	Standard AMG	Adapted AMG	Standard AMG	Adapted AMG
$1/64$	*	0.03s	*	0.05s
$1/128$	*	0.22s	*	0.27s
$1/256$	*	0.91s	*	1.22s
$1/512$	*	3.64s	*	5.35s
$1/1024$	*	15.64s	*	28.27s

Numerical Results - Solution Time

Total time to setup MG method and then reduce residual by 10^{10}

or residual reduction after 200 iterations

	Problem 1		Problem 2	
h	Standard AMG	Adapted AMG	Standard AMG	Adapted AMG
$1/64$	3.3×10^{-5}	0.03s	4.5×10^{-5}	0.05s
$1/128$	3.6×10^{-5}	0.22s	2.5×10^{-5}	0.27s
$1/256$	2.5×10^{-5}	0.91s	1.7×10^{-5}	1.22s
$1/512$	1.8×10^{-5}	3.64s	1.2×10^{-5}	5.35s
$1/1024$	1.3×10^{-5}	15.64s	9.3×10^{-6}	28.27s

Numerical Results - Convergence Factors

Asymptotic convergence factors of resulting V(1,1) cycles

	Problem 1		Problem 2	
h	Standard AMG	Adapted AMG	Standard AMG	Adapted AMG
$1/64$	0.991	0.069	0.996	0.187
$1/128$	0.997	0.078	0.996	0.212
$1/256$	0.996	0.077	0.996	0.235
$1/512$	0.996	0.078	0.996	0.292
$1/1024$	0.996	0.079	0.995	0.383

Theoretical Results

- Questions:
  - Convergence of adaptive process
  - Convergence of the resulting V-cycle
- Approached theory in a 2-level, reduction-based AMG setting
- Convergence of solution phase based on A-orthogonal decomposition
- Show convergence of 2-level adaptive process in reducing Rayleigh Quotient of near-null-space approximation

Summary - Upscaling

- Robust variational multigrid methods define useful coarse-scale models
- Solution of these models accurately approximates fine-scale solution
  - Net outflow flux
  - Fine-scale structure of pressure
- Accurate recovery of coarse-scale material properties
- Coarse-scale model is accurate with multiscale basis function, but includes regularization term with coarse-scale basis interpretation

Future Work - Upscaling

- Compare with multiscale finite element method (Hou et al.)
- Coarsen using AMG or adaptive AMG
- Extend to three dimensions
- Apply techniques to nonlinear problem
- Reconcile averaging theory with numerical results

Summary - Adaptive Multigrid

- Assumptions on knowledge of algebraic smoothness in classical algebraic multigrid methods can be relaxed
- Additional work in setup to expose prototypical algebraically-smooth error results in improved convergence behavior
- Prove resulting algorithm is invariant to diagonal scalings
- Theory supports convergence of adaptive process

Future Work - Adaptive Multigrid

- Apply adaptive framework to systems of PDEs
- Extend AMG interpolation to fit multiple prototypes
  - Adaptive smoothed aggregation performs well on linear elasticity
- Adaptive choice of coarse grid
- Improve and extend theory

Conclusions

- Effective multiscale basis functions created naturally in variational multigrid
- Solutions to variational coarse-scale problems accurately predict fine-scale behavior
- Adaptive process creates accurate coarse-scale models through exposure of algebraically-smooth error
- Resulting V-cycle outperforms classical AMG on many problems
- Many interesting questions remain