Algebraic Multigrid Coarsening from theory to practice

Scott MacLachlan
scott.maclachlan@gmail.com

Delft Institute of Applied Mathematics, TU-Delft
Centrum voor Wiskunde en Informatica, Amsterdam

Joint work with
Yousef Saad, University of Minnesota
Luke Olson, University of Illinois at Urbana-Champaign

July 3, 2007

Target Applications

- Fluid flow in porous media
  - Highly heterogeneous media
  - Interested in global properties of the solution
- Coupled fluid-elastic systems
  - Multiple material regimes
  - Different models require very different treatment
- Lattice quantum chromodynamics
  - Highly heterogeneous operator
  - Randomized heterogeneity within Monte Carlo process

Target Discrete Models

- Finite difference/element models of elliptic systems
- Matrices are
  - sparse
  - symmetric
  - positive definite

Solving $Ax = b$

- Sparsity with large bandwidth means that direct methods are not effective for these problems

Algorithmic Goals

Efficient and robust solver for heterogeneous models

Efficient: Optimal scalability, both algorithmic and parallel

- Cost of solve linearly proportional to number of unknowns
- Natural parallelism; most calculations should be data-local

Robust: Consistent performance with few parameters

- Predictable performance based on simple characteristics
- Not expert software; no magic parameters

Efficiency First

Heterogeneity is an added complication, but not fundamental

Still need techniques to handle

- Large problem sizes
- Large condition numbers
- Multiscale structure of operator

Robustness without efficiency is EASY!

Gaussian Elimination

Efficiency First

Heterogeneity is an added complication, but not fundamental

Still need techniques to handle

- Large problem sizes
- Large condition numbers
- Multiscale structure of operator

Robustness without efficiency is EASY!
Gaussian Elimination

- Start with an efficient solver for homogeneous models
  - Geometric multigrid
- Look for where robustness and heterogeneity play a role

Stationary Iterative Methods

- Want to improve approximation, $x^{(0)}$, to $x = A^{-1}b$
- Residual, $r^{(0)}$, is a measure of the error

$$r^{(0)} = b - Ax^{(0)} = Ax - Ax^{(0)} = A(x - x^{(0)})$$

- Choose $M^{-1} \approx A^{-1}$
- Take $x^{(1)} = x^{(0)} + M^{-1}r^{(0)}$

Error propagation form: $e^{(1)} = (I - M^{-1}A)e^{(0)}$

Stationary Iterative Methods

- Want to improve approximation, $x^{(0)}$, to $x = A^{-1}b$
- Residual, $r^{(0)}$, is a measure of the error

$$r^{(0)} = b - Ax^{(0)} = Ax - Ax^{(0)} = A(x - x^{(0)})$$

- Choose $M^{-1} \approx A^{-1}$
- Take $x^{(1)} = x^{(0)} + M^{-1}r^{(0)}$

Error propagation form: $e^{(1)} = (I - M^{-1}A)e^{(0)}$

$$e^{(2)} = (I - M^{-1}A)e^{(1)}$$

Stationary Iterative Methods

- Want to improve approximation, $x^{(0)}$, to $x = A^{-1}b$
- Residual, $r^{(0)}$, is a measure of the error

$$r^{(0)} = b - Ax^{(0)} = Ax - Ax^{(0)} = A(x - x^{(0)})$$

- Choose $M^{-1} \approx A^{-1}$
- Take $x^{(1)} = x^{(0)} + M^{-1}r^{(0)}$

Error propagation form: $e^{(1)} = (I - M^{-1}A)e^{(0)}$

$$e^{(2)} = (I - M^{-1}A)^2 e^{(0)}$$

Stationary Iterative Methods

- Want to improve approximation, $x^{(0)}$, to $x = A^{-1}b$
- Residual, $r^{(0)}$, is a measure of the error

$$r^{(0)} = b - Ax^{(0)} = Ax - Ax^{(0)} = A(x - x^{(0)})$$

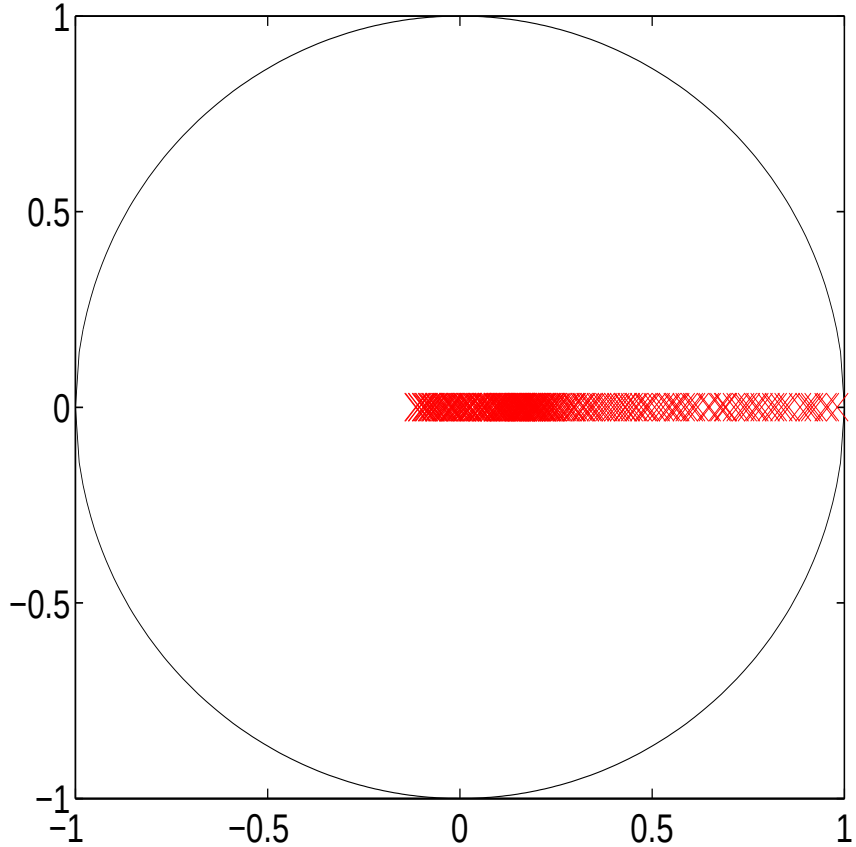
- Choose $M^{-1} \approx A^{-1}$
- Take $x^{(1)} = x^{(0)} + M^{-1}r^{(0)}$

Error propagation form:

$$\begin{aligned} e^{(1)} &= (I - M^{-1}A)e^{(0)} \\ e^{(2)} &= (I - M^{-1}A)^2 e^{(0)} \\ &\vdots \\ e^{(n)} &= (I - M^{-1}A)^n e^{(0)} \end{aligned}$$

Convergence of Stationary Iterations

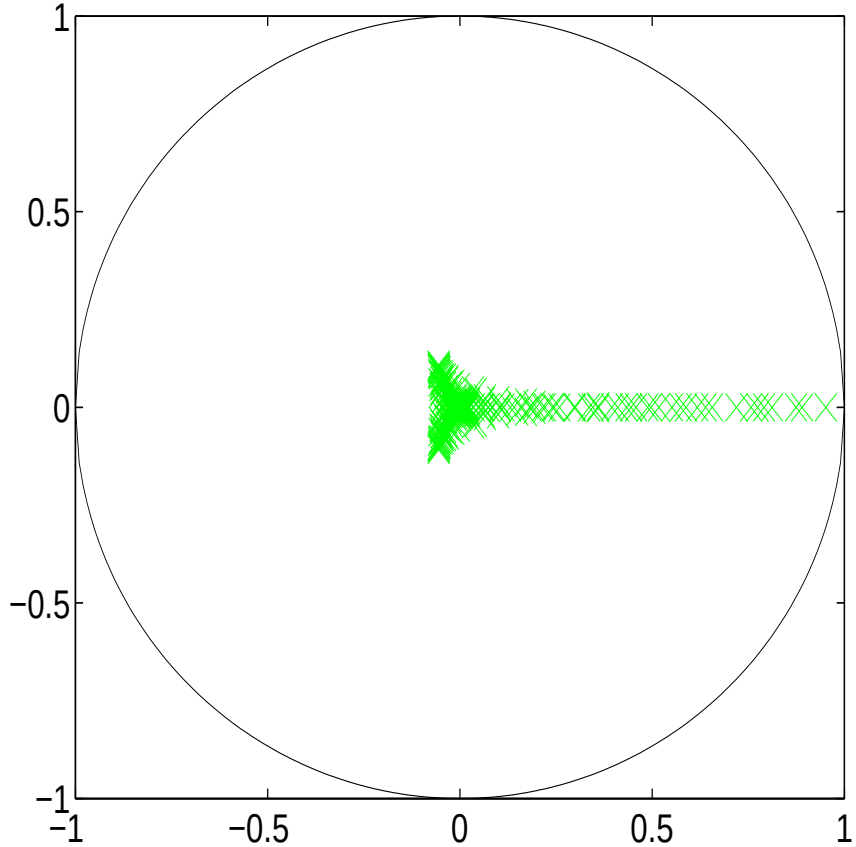
Convergence depends on spectrum of $I - M^{-1}A$



Weighted Jacobi Iteration: $e^{(n)} = (I - \frac{3}{4}D^{-1}A)^n e^{(0)}$

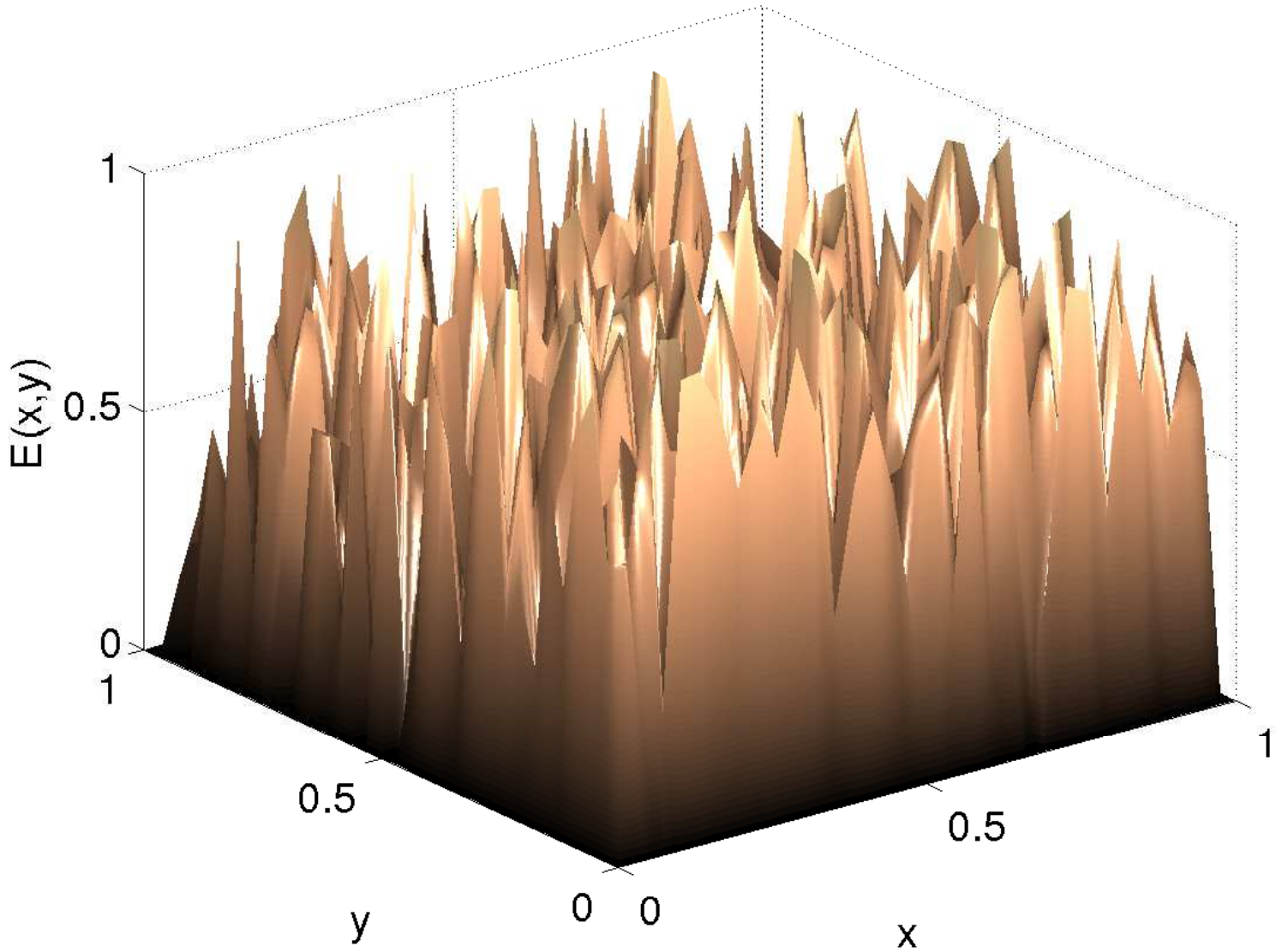
Convergence of Stationary Iterations

Convergence depends on spectrum of $I - M^{-1}A$



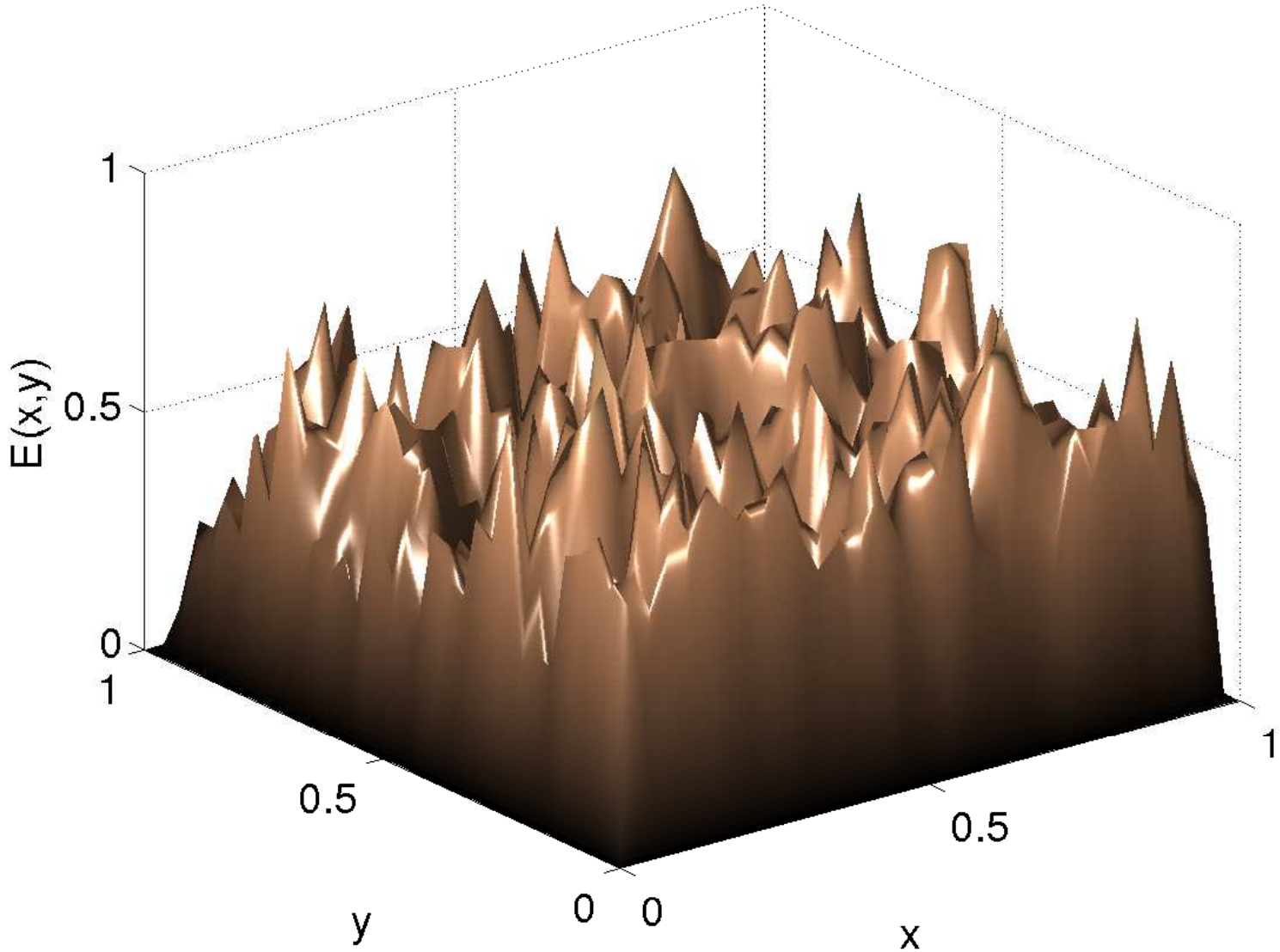
Gauss-Seidel Iteration: $e^{(n)} = (I - L^{-1}A)^n e^{(0)}$

Smoothing Property



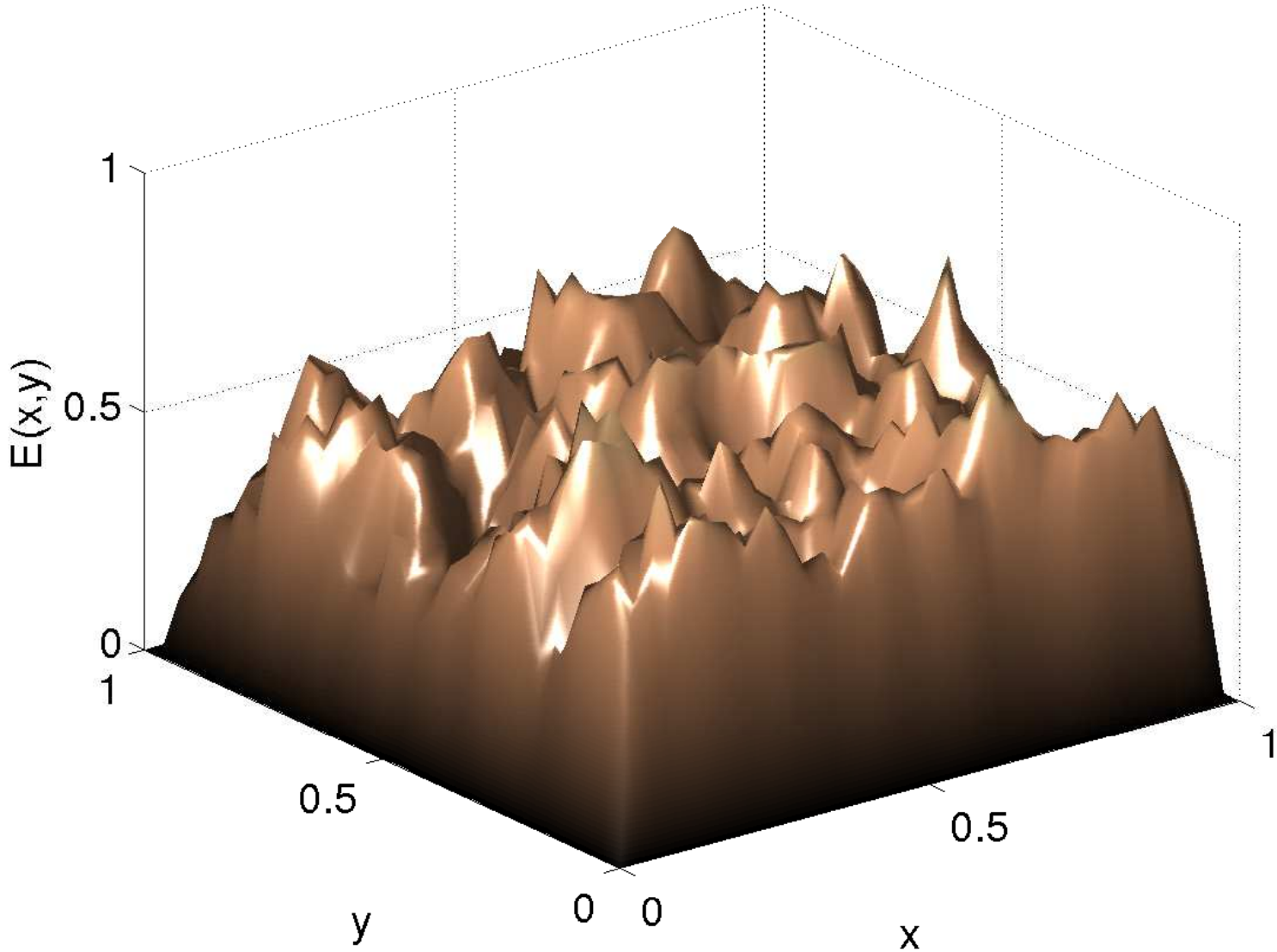
Random initial error

Smoothing Property



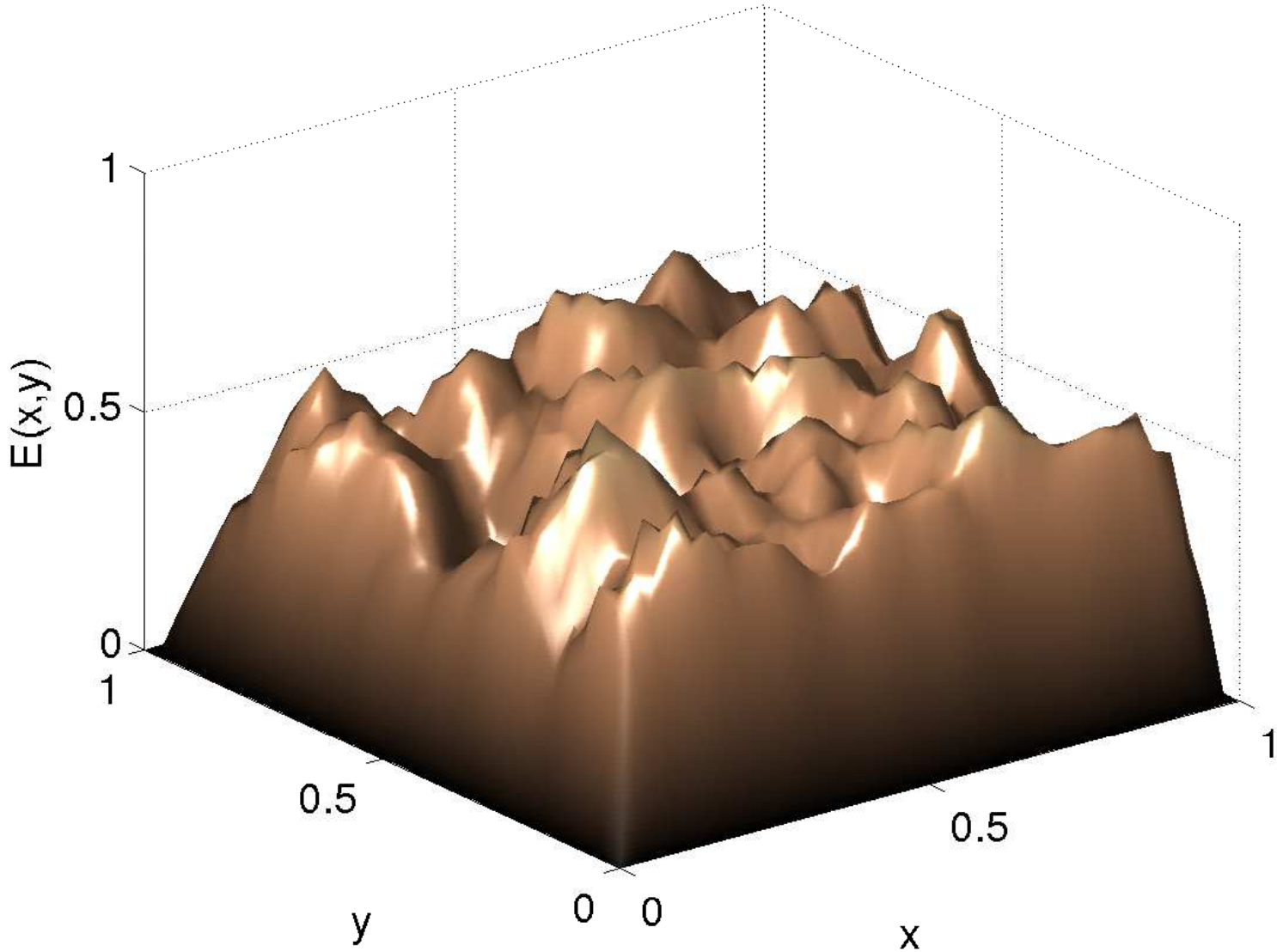
Error after 1 weighted Jacobi iteration

Smoothing Property



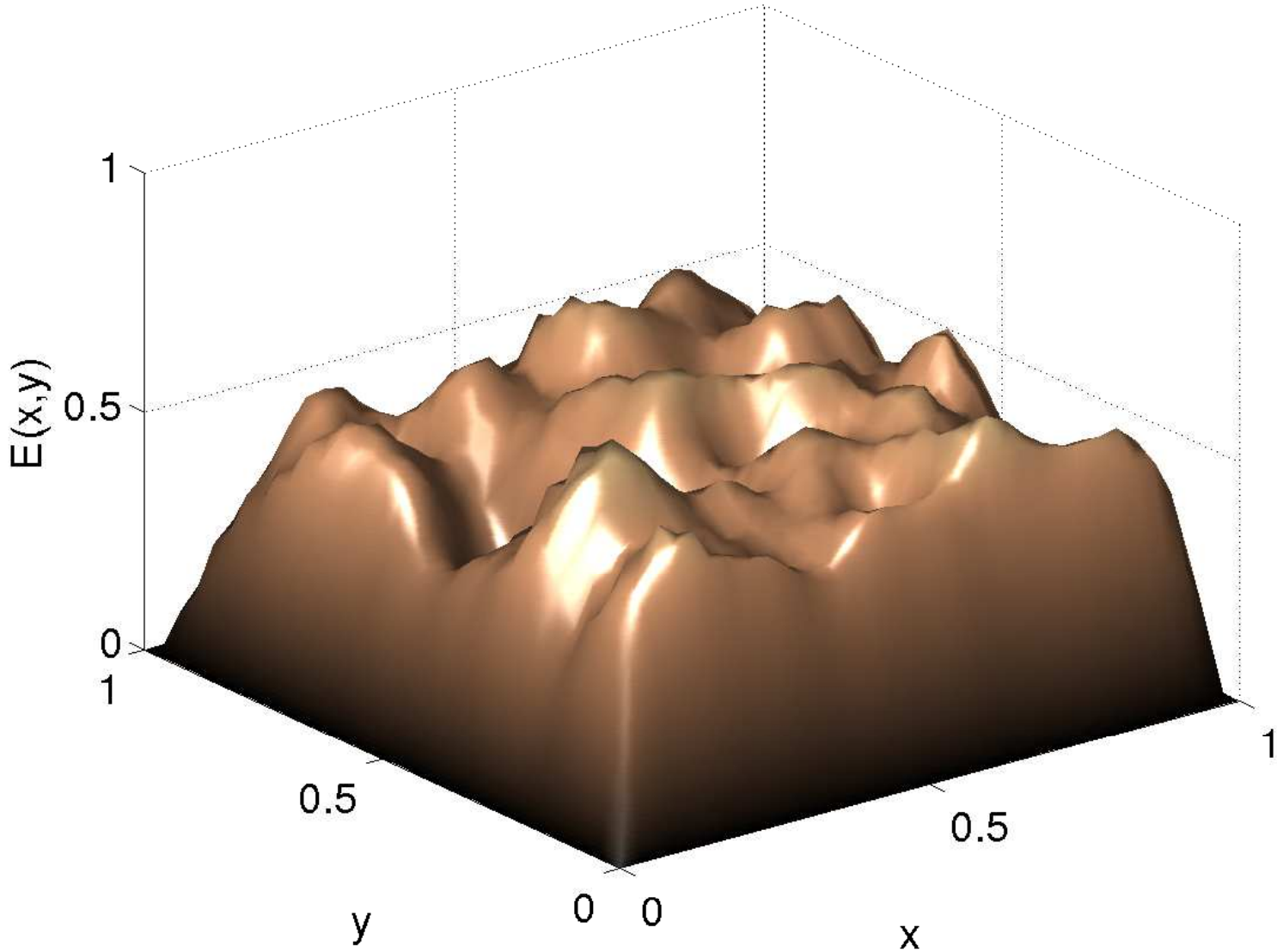
Error after 2 weighted Jacobi iterations

Smoothing Property



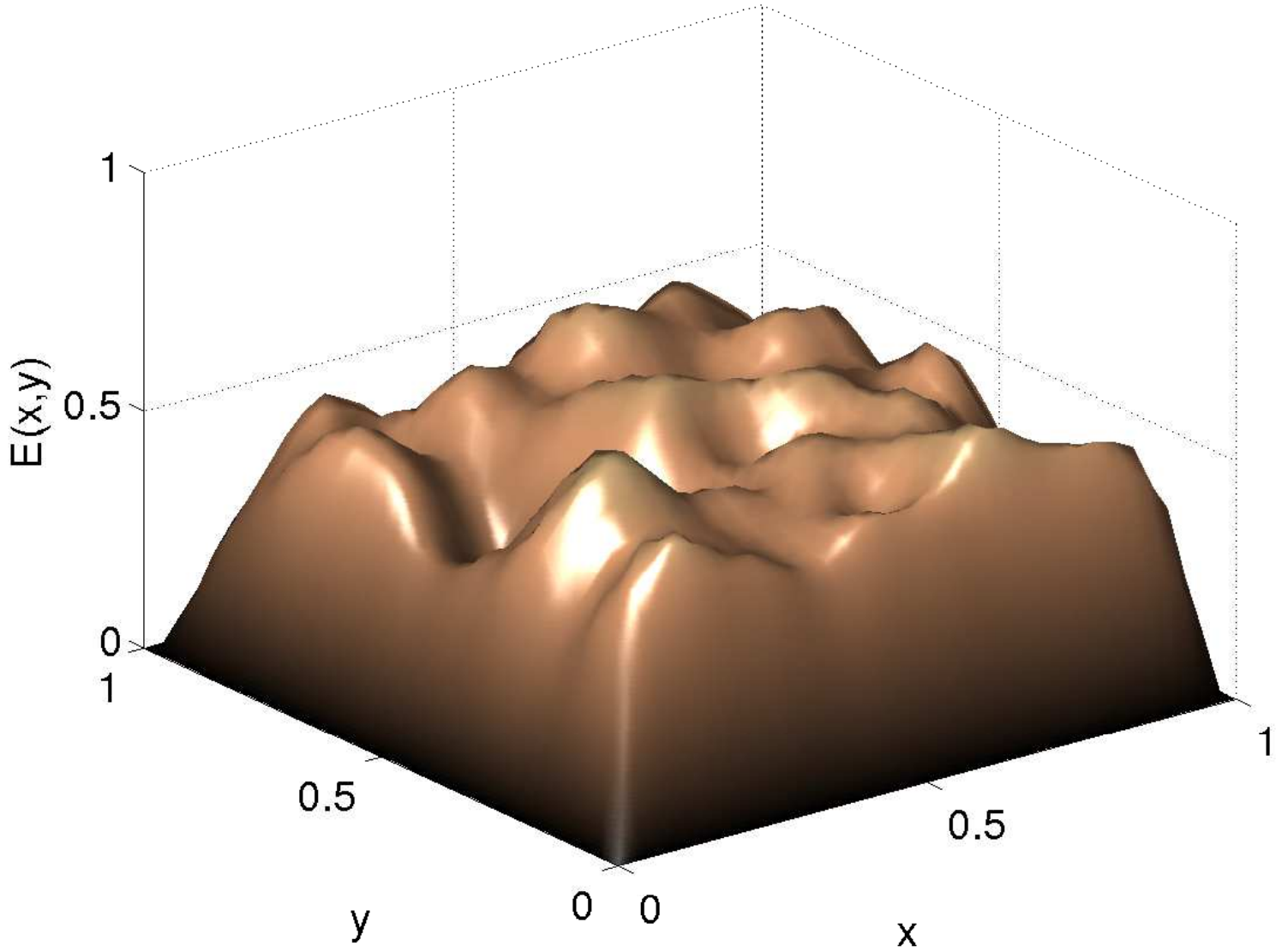
Error after 3 weighted Jacobi iterations

Smoothing Property



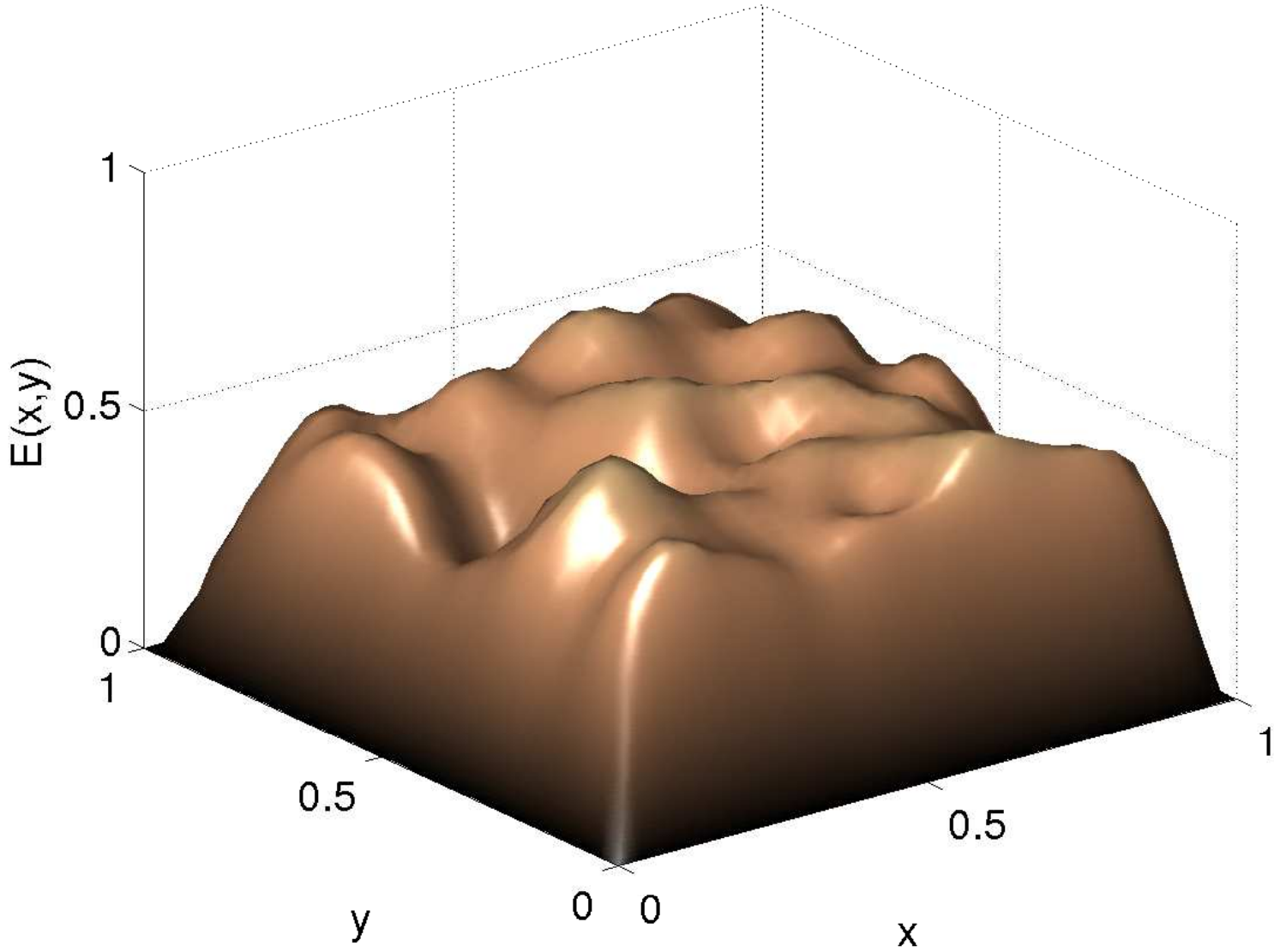
Error after 4 weighted Jacobi iterations

Smoothing Property



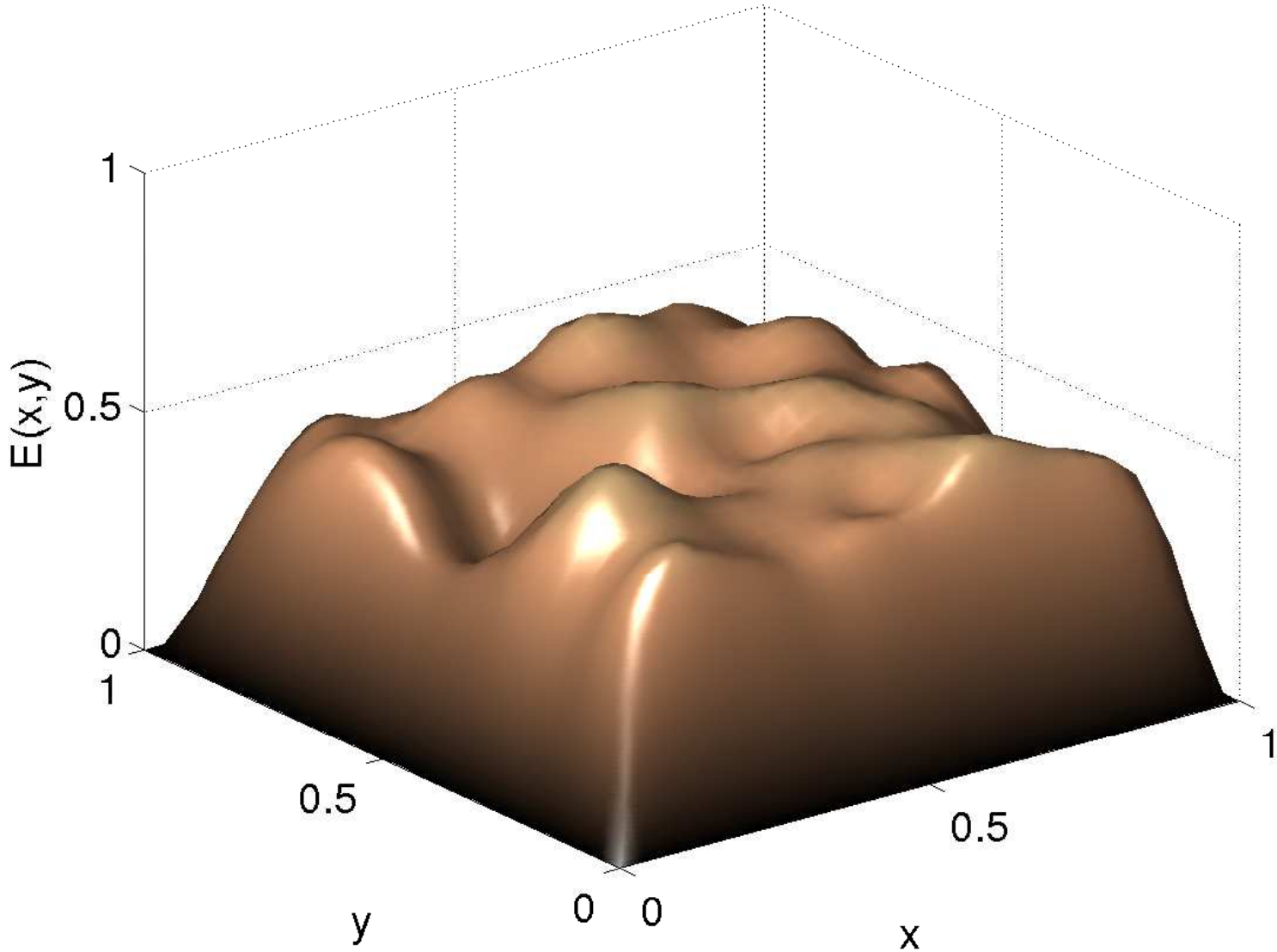
Error after 5 weighted Jacobi iterations

Smoothing Property



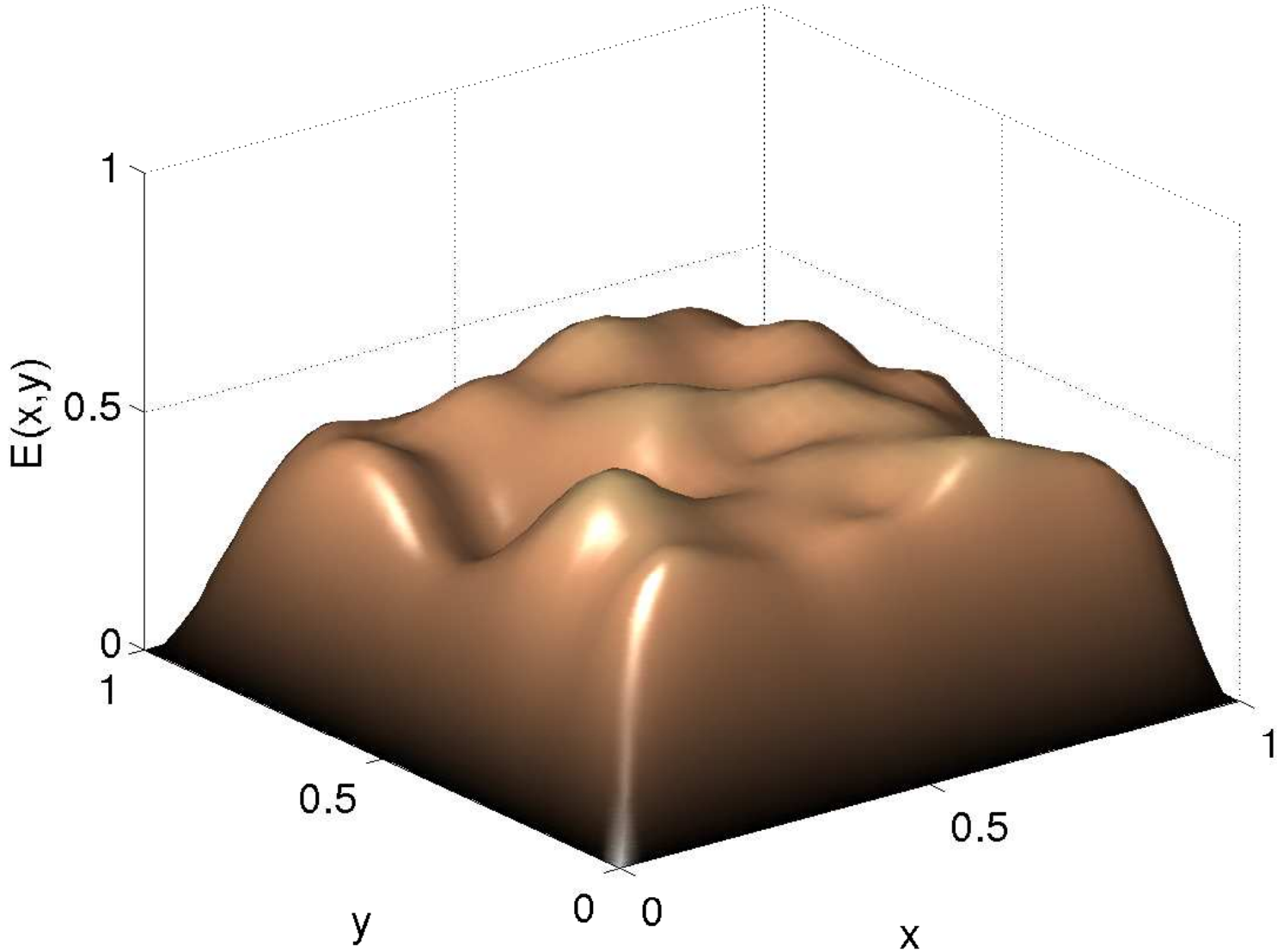
Error after 6 weighted Jacobi iterations

Smoothing Property



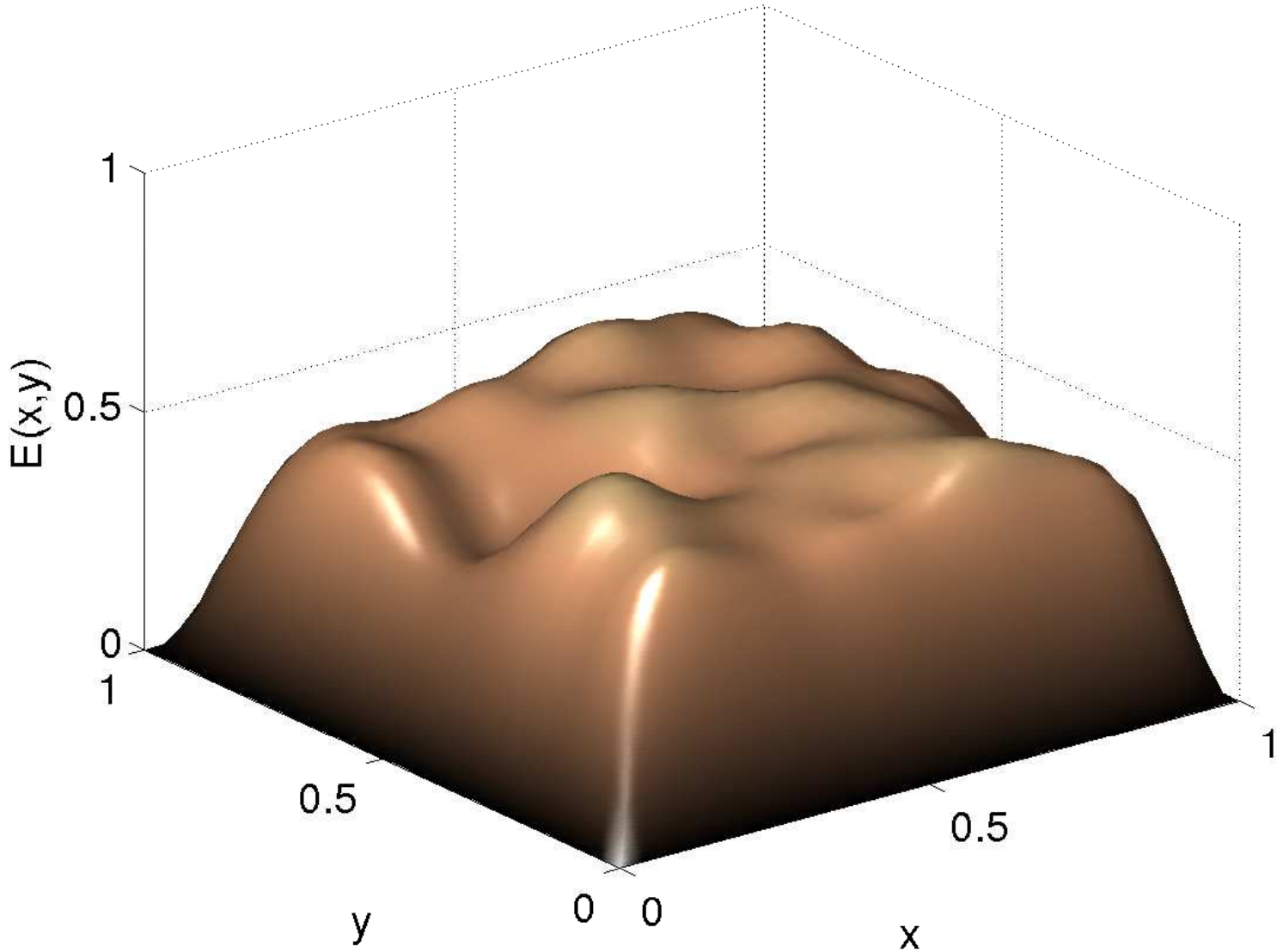
Error after 7 weighted Jacobi iterations

Smoothing Property



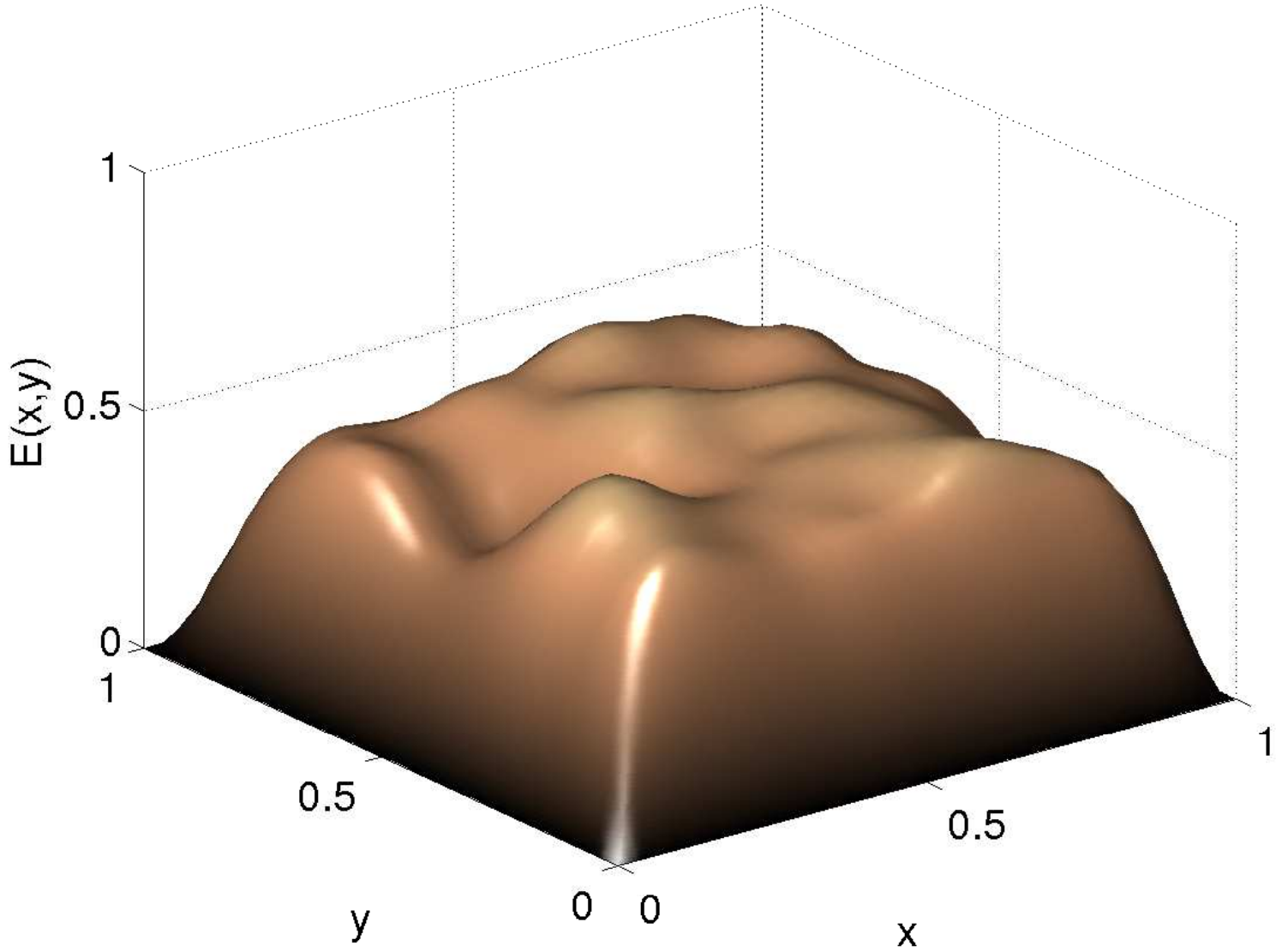
Error after 8 weighted Jacobi iterations

Smoothing Property



Error after 9 weighted Jacobi iterations

Smoothing Property



Error after 10 weighted Jacobi iterations

Complementarity

- Error after a few weighted Jacobi iterations has structure
- Instead of throwing out the method, look to complement its failings

How can we best correct error modes that are slow to be reduced by relaxation?

Complementarity

- Error after a few weighted Jacobi iterations has structure
- Instead of throwing out the method, look to complement its failings

How can we best correct error modes that are slow to be reduced by relaxation?

- Slow-to-converge errors are smooth
- Smooth vectors can be easily represented using fewer degrees of freedom

Coarse Grids

- Sine series representation:

$$f(x) = \sum_{k=1}^{\infty} c_k \sin(k\pi x)$$

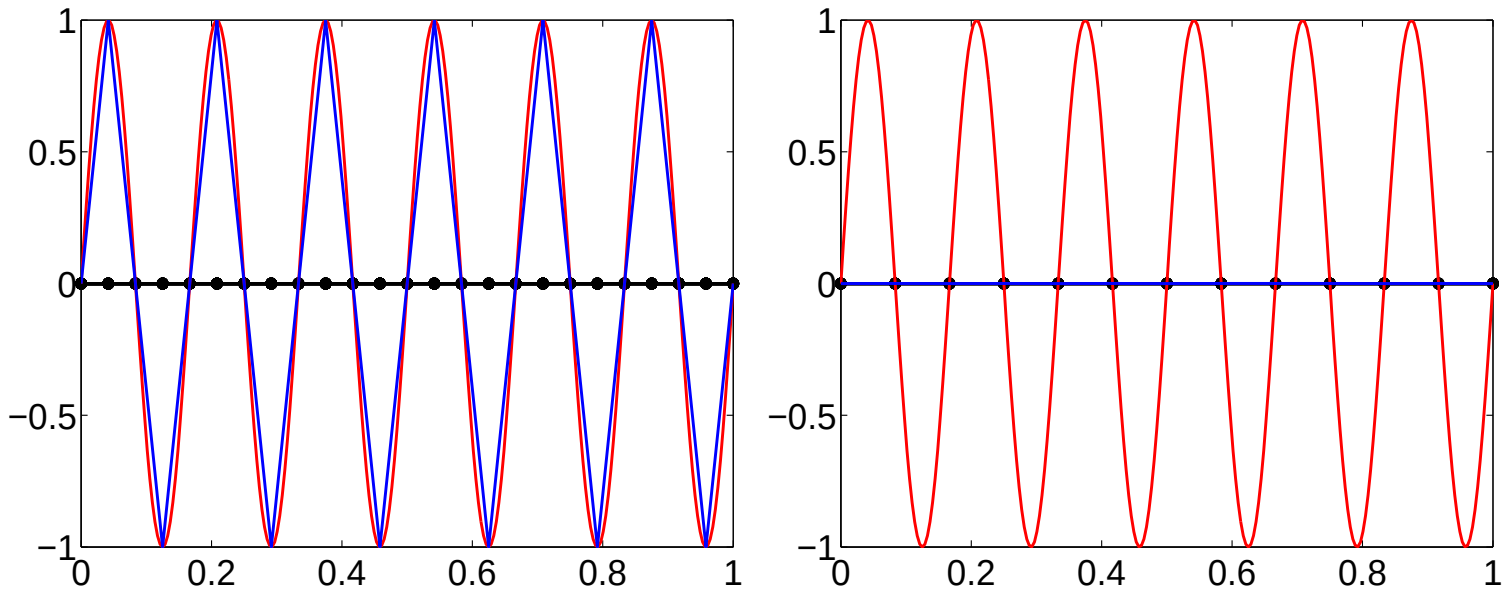
- Discrete problems can only approximate certain modes

Coarse Grids

- Sine series representation:

$$f(x) = \sum_{k=1}^{\infty} c_k \sin(k\pi x)$$

- Discrete problems can only approximate certain modes



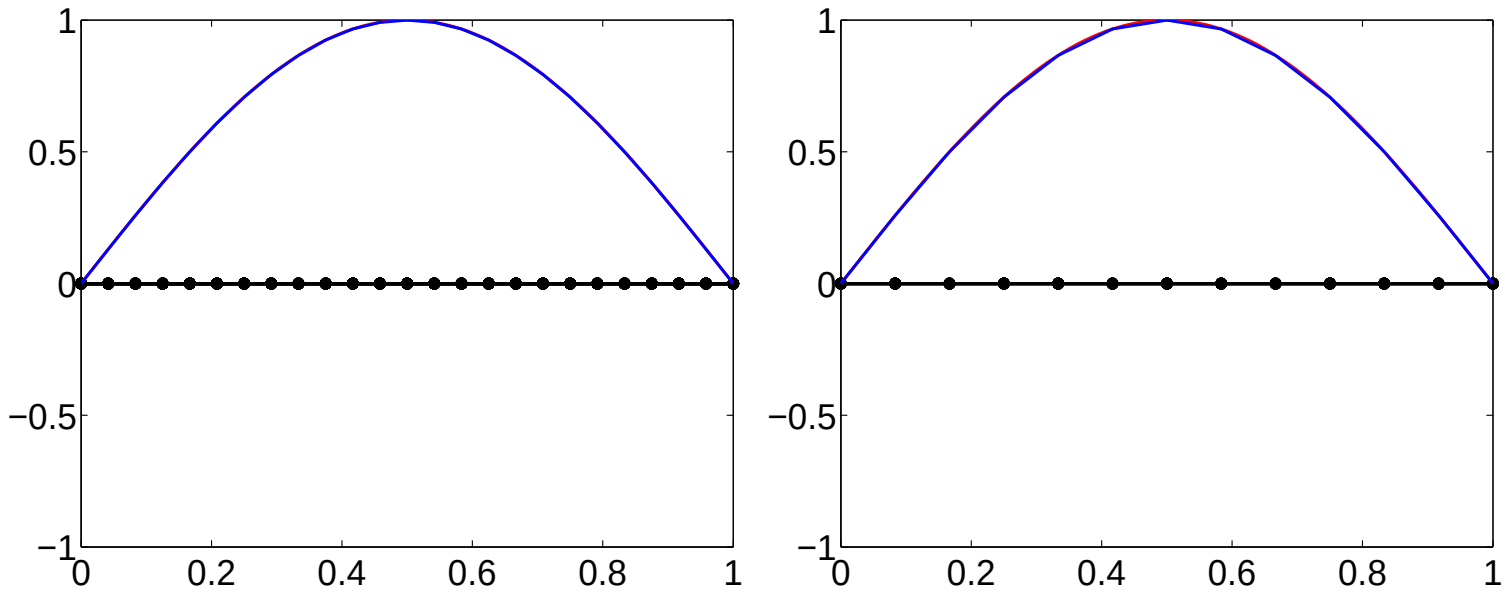
$\sin(12\pi x)$

Coarse Grids

- Sine series representation:

$$f(x) = \sum_{k=1}^{\infty} c_k \sin(k\pi x)$$

- Discrete problems can only approximate certain modes



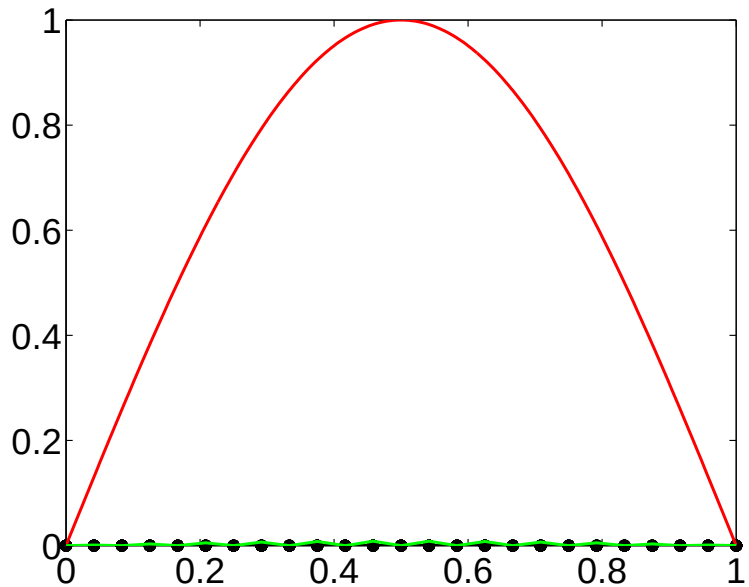
$\sin(\pi x)$

Coarse Grids

- Sine series representation:

$$f(x) = \sum_{k=1}^{\infty} c_k \sin(k\pi x)$$

- Discrete problems can only approximate certain modes



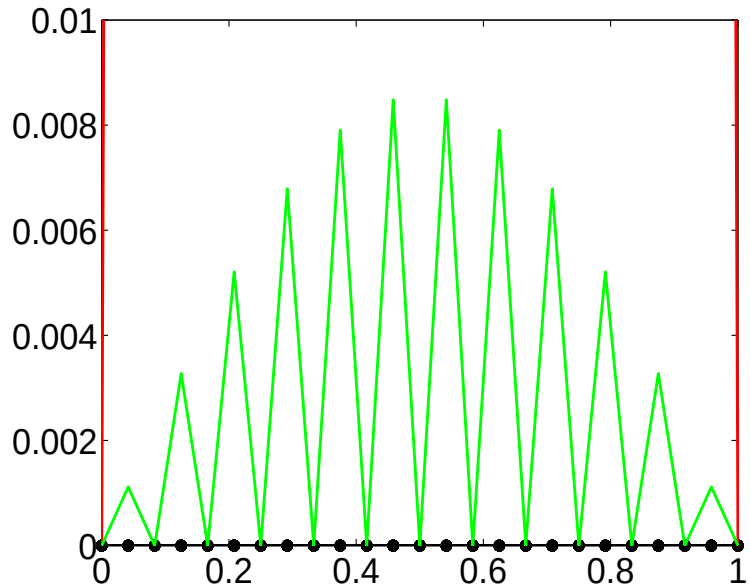
Error in coarse-grid representation of $\sin(\pi x)$

Coarse Grids

- Sine series representation:

$$f(x) = \sum_{k=1}^{\infty} c_k \sin(k\pi x)$$

- Discrete problems can only approximate certain modes



Error in coarse-grid representation of $\sin(\pi x)$

Coarse Grids

- Sine series representation:

$$f(x) = \sum_{k=1}^{\infty} c_k \sin(k\pi x)$$

- Discrete problems can only approximate certain modes

Coarse grids accurately represent low-frequency modes

Natural complement to relaxation

Complementarity

Use two complementary processes to efficiently damp all errors

Relaxation: Damp high-frequency error by stationary iteration

Coarse-grid correction: Eliminate low-frequency error by relaxation on coarse grids

Complementarity

Use two complementary processes to efficiently damp all errors

Relaxation: Damp high-frequency error by stationary iteration

Coarse-grid correction: Eliminate low-frequency error by relaxation on coarse grids

Key realization: Solve for coarse-grid representation of error

- At any stage, error is reflected in residual:

$$r^{(k)} = b - Ax^{(k)} = Ax - Ax^{(k)} = A(x - x^{(k)})$$

- Don't transfer $Ax = b$ to coarse grid, transfer $Ae = r$

The Details

- Correct the approximation after relaxation, $x^{(1)}$, from an auxiliary (coarse-grid) problem
- Need interpolation map, P, from coarse grid to fine grid
- Corrected approximation will be $x^{(2)} = x^{(1)} + Px_c$
- Optimal x_c satisfies $P^T APx_c = P^T r$
  - Optimal means $\|x - x^{(2)}\|_A$ is minimal

Multigrid choices:

- Coarse grid
- Interpolation operator, P
- Restriction, coarse-grid operator given by optimality

The Details

- Correct the approximation after relaxation, $x^{(1)}$, from an auxiliary (coarse-grid) problem
- Need interpolation map, P, from coarse grid to fine grid
- Corrected approximation will be $x^{(2)} = x^{(1)} + Px_c$
- Optimal x_c satisfies $P^T APx_c = P^T r$
  - Optimal means $\|x - x^{(2)}\|_A$ is minimal

Multigrid choices:

- Coarse grid
- Interpolation operator, P
- Restriction, coarse-grid operator given by optimality
- Restriction, coarse-grid operator given by physics

The Details

- Correct the approximation after relaxation, $x^{(1)}$, from an auxiliary (coarse-grid) problem
- Need interpolation map, P, from coarse grid to fine grid
- Corrected approximation will be $x^{(2)} = x^{(1)} + Px_c$
- Optimal x_c satisfies $P^T APx_c = P^T r$
  - Optimal means $\|x - x^{(2)}\|_A$ is minimal

Multigrid choices:

- Coarse grid
- Interpolation operator, P
- Restriction, coarse-grid operator given by optimality
- Restriction, coarse-grid operator given by physics
- Restriction, coarse-grid operator constrained by computation

Multigrid

Multigrid Components

Relax: $x^{(1)} = x^{(0)} + D^{-1} r^{(0)}$

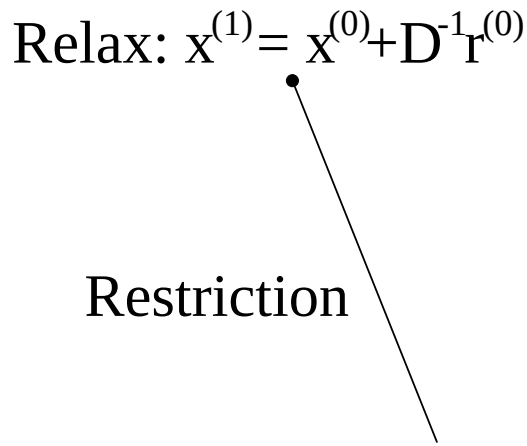
- Relaxation

- Use a smoothing process (such as Jacobi or Gauss-Seidel) to eliminate oscillatory errors
- Remaining error satisfies $Ae^{(1)} = r^{(1)} = b - Ax^{(1)}$

Multigrid

Multigrid Components

- Relaxation
- Restriction

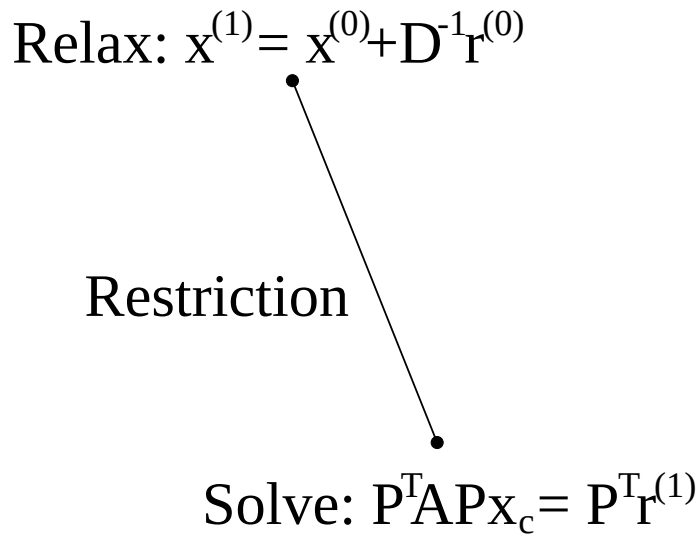


- Transfer residual to coarse grid
- Compute $P^T r^{(1)}$

Multigrid

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction



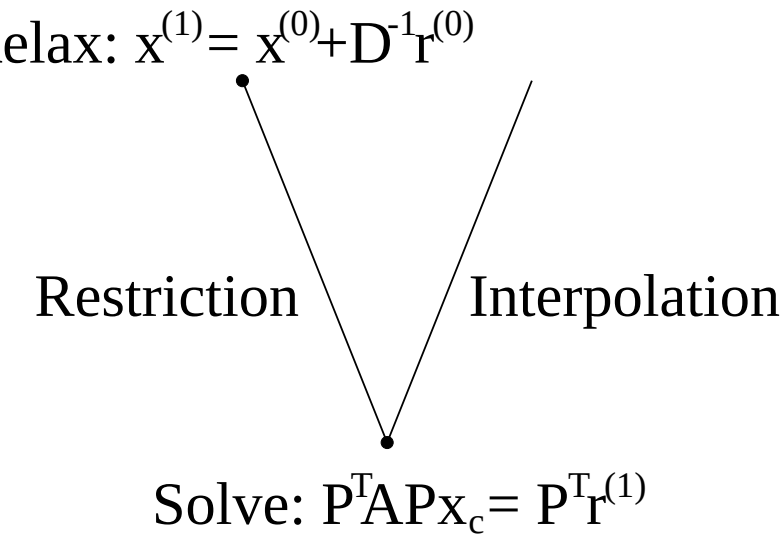
- Use coarse-grid correction to eliminate smooth errors
- Best correction, x_c, in terms of A-norm satisfies

$$P^T A P x_c = P^T r^{(1)}$$

Multigrid

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction
- Interpolation

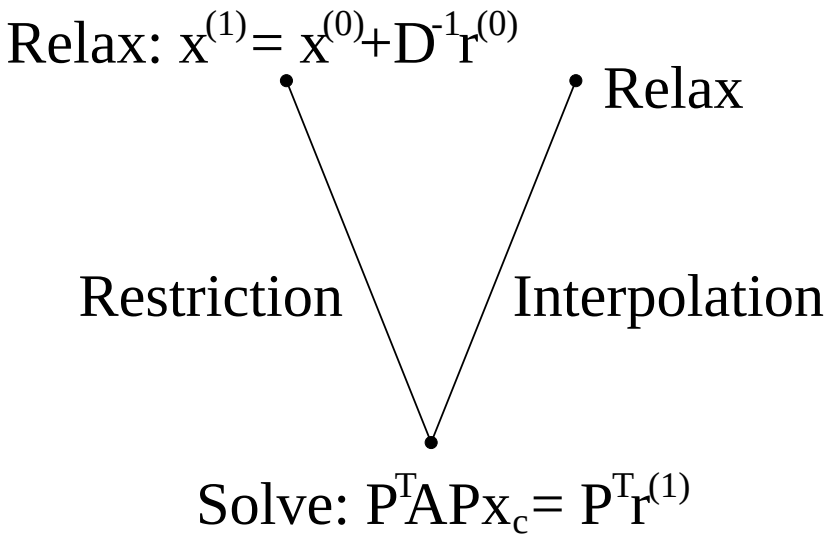


- Transfer correction to fine grid
- Compute $x^{(2)} = x^{(1)} + Px_c$

Multigrid

Multigrid Components

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

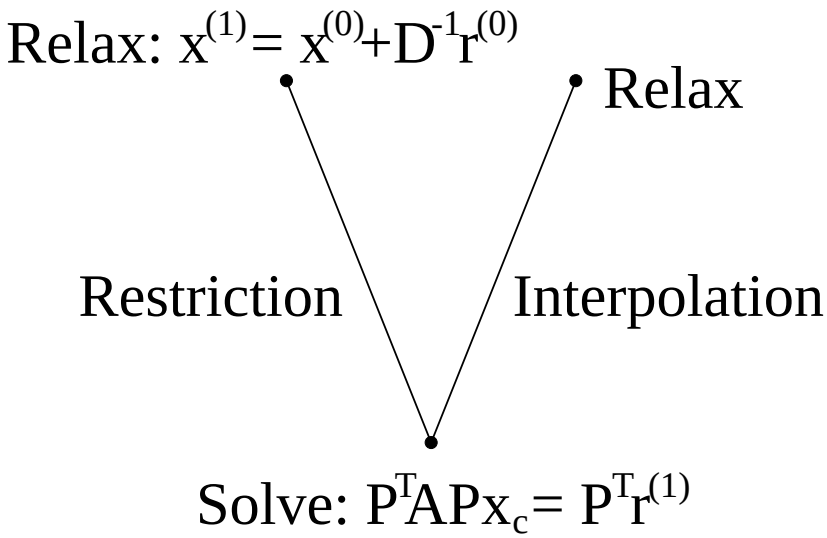


- Relax once again to remove oscillatory error introduced in coarse-grid correction

Multigrid

Multigrid Components

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



Direct solution of coarse-grid problem isn't practical

Recursion!

Apply same methodology to solve coarse-grid problem

Analysing Performance

For simple problems, Fourier analysis predicts performance

Analysis based on invariant subspaces

- Eigenvectors of simple relaxation are Fourier modes
- Aliased pairs make invariant subspaces of interpolation and restriction
- Rediscretized CGO has same eigenvectors

Can directly compute convergence factor for two-level cycle

Analysis and Heterogeneity

Fourier analysis only valid for operators with periodic character

Heterogeneity in A affects entire cycle

Analysis and Heterogeneity

Fourier analysis only valid for operators with periodic character

Heterogeneity in A affects entire cycle

- Convergence properties of relaxation change

Analysis and Heterogeneity

Fourier analysis only valid for operators with periodic character

Heterogeneity in A affects entire cycle

- Convergence properties of relaxation change
- Slowest-to-converge errors reflect heterogeneity

Analysis and Heterogeneity

Fourier analysis only valid for operators with periodic character

Heterogeneity in A affects entire cycle

- Convergence properties of relaxation change
- Slowest-to-converge errors reflect heterogeneity
- Interpolation must adapt to fit these errors

Analysis and Heterogeneity

Fourier analysis only valid for operators with periodic character

Heterogeneity in A affects entire cycle

- Convergence properties of relaxation change
- Slowest-to-converge errors reflect heterogeneity
- Interpolation must adapt to fit these errors
- Coarse-grid operator must account for heterogeneity

Must ensure complementarity is not lost

Types of Heterogeneity

Heterogeneity comes in several forms

Types of Heterogeneity

Heterogeneity comes in several forms

Heterogeneity in the coefficients:

- multiphase flow, variations in material properties
- bubbles in water, oil-reservoir, aquifer modelling

Types of Heterogeneity

Heterogeneity comes in several forms

Heterogeneity in the coefficients:

- multiphase flow, variations in material properties
- bubbles in water, oil-reservoir, aquifer modelling

Heterogeneity in the equations:

- coupled fluid-solid interactions, multiphysics
- blood flow in a vein, multi-species chemistry

Types of Heterogeneity

Heterogeneity comes in several forms

Heterogeneity in the coefficients:

- multiphase flow, variations in material properties
- bubbles in water, oil-reservoir, aquifer modelling

Heterogeneity in the equations:

- coupled fluid-solid interactions, multiphysics
- blood flow in a vein, multi-species chemistry

Heterogeneity in the grid

- Local refinement, unstructured triangulation
- shock waves, irregular geometry

Aim for an algorithm that is robust to all of these

Multigrid Without Grids

The essence of multigrid has nothing to do with grids!

Complementarity is key:

- Fix choice of relaxation
- For any A, some errors are slow to converge
- These errors must be corrected some other way

A. Brandt, S. McCormick, J. Ruge, in *Sparsity and Its Applications*, 1984
J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Multigrid Without Grids

The essence of multigrid has nothing to do with grids!

Complementarity is key:

- Fix choice of relaxation
- For any A, some errors are slow to converge
- These errors must be corrected some other way

Coarse-grid correction:

$$x \leftarrow x + PB_c^{-1}Rr$$
$$e \leftarrow e - PB_c^{-1}Rr$$

A. Brandt, S. McCormick, J. Ruge, in *Sparsity and Its Applications*, 1984
J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Variational Coarsening

Coarse-grid correction,

$$I - PB_c^{-1}RA,$$

can only correct errors in the range of P

Choosing $R = P^T$ and $B_c = P^T AP$ exactly eliminates errors in this space.

Complementarity is key:

- Errors reduced by relaxation and coarse-grid correction
- Errors that relaxation reduces slowly must be in range(P)

R. Nicolaides, Math. Comp. 1977, **31**:892-906

Algebraically Smooth Error

Slow to converge errors of relaxation replace smooth modes within AMG

Design interpolation to accurately represent these modes

- Assume these errors give small residuals, $Ae \approx 0$
- Expand residual equation:

$$a_{ii}e_i = -\sum_{j \in C} a_{ij}e_j - \sum_{k \notin C} a_{ik}e_k$$

- Use assumption on character of these errors to eliminate connections to $k \notin C$

A. Brandt, S. McCormick, J. Ruge, in *Sparsity and Its Applications*, 1984
J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Graph-based coarsening

Goal: Choose coarse-grid nodes to allow easy elimination of fine-fine connections

- Filter to eliminate small matrix entries
- Create graph of filtered matrix
- Greedy algorithm to choose maximal independent subset

Maximal independent subset ensures

- every fine-fine connection is "close" to a coarse-grid point
- coarse grid is small, but not too small

A. Brandt, S. McCormick, J. Ruge, in *Sparsity and Its Applications*, 1984
J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Big Picture Convergence

Convergence is a spectral property

Write error-propagation operator as a matrix:

$$(I - M^{-1}A)(I - PB_c^{-1}P^TA)$$

Convergence factor is the spectral radius

$$\rho_{\mathsf{MG}} = \lambda_{\max}\left((I - M^{-1}A)(I - PB_c^{-1}P^TA)\right)$$

or $$\rho_{\mathsf{MG}} = \left\|(I - M^{-1}A)(I - PB_c^{-1}P^TA)\right\|_A$$

J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Big Picture Convergence

Convergence is a spectral property

Write error-propagation operator as a matrix:

$$(I - M^{-1}A)(I - PB_c^{-1}P^TA)$$

Convergence factor is the spectral radius

$$\rho_{\text{MG}} = \lambda_{\max}\left((I - M^{-1}A)(I - PB_c^{-1}P^TA)\right)$$

or $$\rho_{\text{MG}} = \left\|(I - M^{-1}A)(I - PB_c^{-1}P^TA)\right\|_A$$

Expect: $\rho_{\text{MG}} = \sqrt{1 - \frac{1}{k}}$ for some $k \geq 1$

Goal: Give an upper bound for k and, thus, ρ_{MG}

J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Bounding ρ_{MG}

Abstract bounds on ρ_{MG} come from interaction of relaxation and coarse-grid correction:

Define $G = I - M^{-1}A$, $T = I - P(P^TAP)^{-1}P^TA$,

Assume that $\|Ge\|_A^2 \leq \|e\|_A^2 - \delta\|Te\|_A^2$, then

$$\rho_{\mathsf{MG}} \leq \sqrt{1-\delta}$$

J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Bounding ρ_{MG}

Abstract bounds on ρ_{MG} come from interaction of relaxation and coarse-grid correction:

Define $G = I - M^{-1}A$, $T = I - P(P^TAP)^{-1}P^TA$,

Assume that $\|Ge\|_A^2 \leq \|e\|_A^2 - \delta\|Te\|_A^2$, then

$$\rho_{MG} \leq \sqrt{1-\delta}$$

More commonly, we separate the assumptions:

Smoothing assumption: $\|Ge\|_A^2 \leq \|e\|_A^2 - \alpha\|e\|_{AD^{-1}A}^2$

Approximation property: $\|Te\|_A^2 \leq \beta\|Te\|_{AD^{-1}A}^2$

Then $\rho_{MG} \leq \sqrt{1-\frac{\alpha}{\beta}}$

J. Ruge and K. Stüben, in *Multigrid Methods*, 1987

Smoothing Assumption

Achieving bound on relaxation is "easy":

Weighted Jacobi: For SPD A and $\gamma_0 \geq \rho(D^{-1}A)$, then if $0 < \omega < \frac{2}{\gamma_0}$, $\alpha \leq \omega(2 - \omega\gamma_0)$

Gauss-Seidel: For SPD A, $\alpha \leq \frac{1}{(1+\gamma_-)(1+\gamma_+)}$, for

$$\gamma_- = \sum_{j<i} \frac{|a_{ij}|}{a_{ii}} \quad \text{and} \quad \gamma_+ = \sum_{j>i} \frac{|a_{ij}|}{a_{ii}}$$

Kaczmarz: For any A, bound α as in Gauss-Seidel using $\gamma_\pm(AA^T)$

A. Brandt, Appl. Math. Comput. 1986, **19**:23-56

Approximation Property

Typically bound β indirectly:

$$\beta \leq \max_{e \neq 0} \min_{e_c} \frac{\|e - Pe_c\|_D^2}{\|e\|_A^2}$$

This leads to the Brandt-McCormick principle:

For each eigenvector, v, of A, interpolation must represent v with accuracy proportional to its eigenvalue

Challenge: How do we turn this into an algorithm?

- Partition grid
- Define interpolation coefficients

A. Brandt, Appl. Math. Comput. 1986, **19**:23-56

S. McCormick and J. Ruge, SINUM 1985, **19**:924-929

Generalized Approximation Property

Instead of considering

$$\beta \leq \max_{e \neq 0} \min_{e_c} \frac{\|e - Pe_c\|_D^2}{\|e\|_A^2},$$

incorporate relaxation, $I - M^{-1}A$, into bound:

$$K = \max_{e \neq 0} \min_{e_c} \frac{\|e - Pe_c\|_{M(M+M^T-A)^{-1}M^T}^2}{\|e\|_A^2}$$

Then

$$\rho_{\mathrm{MG}} \leq \left(1 - \frac{1}{K}\right)^{\frac{1}{2}}$$

R. Falgout and P. Vassilevski, SIAM J. Numer. Anal. 2004, **42**:1669-1693

Bounding K

Let

- $A = \begin{bmatrix} A_{ff} & -A_{fc} \\ -A_{cf} & A_{cc} \end{bmatrix}$
- M be symmetric
- $x^T Ax \leq \omega x^T Mx < 2x^T Mx$
- $\rho_f = \|I - M_{ff}^{-1} A_{ff}\|_{A_{ff}}$

Then,

$$\min_P \max_{e \neq 0} \min_{e_c} \frac{\|e - Pe_c\|^2_{M(M+M^T-A)^{-1}M^T}}{\|e\|^2_A} \leq \frac{1}{(2-\omega)(1-\rho_f)}$$

For a given F/C partition, the best possible measure depends on the equivalence between M_{ff} and A_{ff}

R. Falgout and P. Vassilevski, SIAM J. Numer. Anal. 2004, **42**:1669-1693

Compatible Relaxation

"A general measure for the quality of the set of coarse variables is the convergence rate of the compatible relaxation"

One approach:

- Run relaxation on tentative F-set
- Identify points where compatible relaxation is slow
- Choose subset of these points to add to C

This tells us about choosing C, but not about choosing P

A. Brandt, Elect. Trans. Numer. Anal. 2000, **10**:1-20

O. Livne, Numer. Linear Algebra Appl. 2004, **2**:205-227

Reduction-Based AMG

Suppose we can partition the grid, $\Omega = F \cup C$, so that

$$x_f^T M_{ff} x_f \leq x_f^T A_{ff} x_f \leq \lambda_{\max} x_f^T M_{ff} x_f$$

and that $\begin{bmatrix} M_{ff} & -A_{fc} \\ -A_{cf} & A_{cc} \end{bmatrix}$ is positive semi-definite. Choose

Relaxation: $I - \frac{2}{1+\lambda_{\max}} \begin{bmatrix} M_{ff}^{-1} & 0 \\ 0 & 0 \end{bmatrix} A$

Coarse-grid correction: variational with $P = \begin{bmatrix} M_{ff}^{-1} A_{fc} \\ I \end{bmatrix}$

Then

$$\rho_{\mathrm{MG}} \leq \left(1 - \left(\frac{2}{\lambda_{\max} + 1}\right)^2\right)^{\frac{1}{2}}$$

M. Ries, U. Trottenberg, G. Winter, J. Lin. Alg. Applic., 1983
S. MacLachlan, T. Manteuffel, S. McCormick, Numer. Linear Algebra Appl. 2006.

Putting It Together

Goal: Use convergence theory to design algorithm

Pieces:

- Smoothing assumption
- Approximation property
- Generalized approximation property
- Compatible relaxation
  - Measure of quality of coarse-grid set
- AMGr
  - Choice of interpolation given good coarse-grid set

Still need an algorithm for partitioning

Additive Multigrid

Theory for additive preconditioners has same limitations.

Let

- $B = \begin{bmatrix} I & 0 \\ -A_{cf}M_{ff}^{-1} & I \end{bmatrix} \begin{bmatrix} M_{ff} & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} I & -M_{ff}^{-1}A_{fc} \\ 0 & I \end{bmatrix}$
- $\begin{bmatrix} M_{ff} & -A_{fc} \\ -A_{cf} & Acc \end{bmatrix}$ be positive semi-definite
- $x_f^T M_{ff} x_f \leq \lambda_{\min} x_f^T M_{ff} x_f \leq x_f^T A_{ff} x_f \leq \lambda_{\max} x_f^T M_{ff} x_f$
- $\nu_{\min} x_c^T S x_c \leq x_c^T (A_{cc} - A_{cf} A_{ff}^{-1} A_{fc}) x_c \leq \nu_{\max} x_c^T S x_c$

Then,

$$\kappa(B^{-\frac{1}{2}} A B^{-\frac{1}{2}}) \leq \left(1 + \sqrt{1 - \frac{1}{\lambda_{\max}}}\right)^2 \frac{\lambda_{\max}^2 \nu_{\max}}{\min(\nu_{\min}, \lambda_{\min})}.$$

O. Axelsson, *Iterative Solution Methods*, 1994

Y. Saad and B. Suchomel, Numer. Linear Algebra Appl. 2002, **9**:359-378

Y. Notay, Numer. Linear Algebra Appl. 2005, **12**:419-451

Coarse-grid Selection

Key to success in these bounds is in the partitioning of A

$$A = \begin{bmatrix} A_{ff} & -A_{fc} \\ -A_{cf} & A_{cc} \end{bmatrix}$$

Need: Good approximation, M_{ff}, to A_{ff}
Need: Cheap computation of $M_{ff}^{-1} r_f$ and $M_{ff}^{-1} A_{fc}$
Need: Dimension of A_{cc} much smaller than A

Two Observations

1. Cost of $M_{ff}^{-1} r_f$ depends on sparsity structure of M_{ff}
   - Cheapest when M_{ff} is diagonal

S. MacLachlan, Y. Saad, SISC, to appear

Two Observations

1. Cost of $M_{ff}^{-1} r_f$ depends on sparsity structure of M_{ff}
   - Cheapest when M_{ff} is diagonal
2. Diagonally dominant A_{ff} can be approximated by its diagonal
   - More diagonally dominant $\longrightarrow$ better approximation

S. MacLachlan, Y. Saad, SISC, to appear

Two Observations

1. Cost of $M_{ff}^{-1} r_f$ depends on sparsity structure of M_{ff}
   - Cheapest when M_{ff} is diagonal
2. Diagonally dominant A_{ff} can be approximated by its diagonal
   - More diagonally dominant $\longrightarrow$ better approximation

A_{ff} is called θ-dominant if, for each $i \in F$,

$$a_{ii} \geq \theta \sum_{j \in F} |a_{ij}|$$

S. MacLachlan, Y. Saad, SISC, to appear

Two Observations

1. Cost of $M_{ff}^{-1} r_f$ depends on sparsity structure of M_{ff}
   - Cheapest when M_{ff} is diagonal
2. Diagonally dominant A_{ff} can be approximated by its diagonal
   - More diagonally dominant $\longrightarrow$ better approximation

A_{ff} is called θ-dominant if, for each $i \in F$,

$$a_{ii} \geq \theta \sum_{j \in F} |a_{ij}|$$

Coarsening Goal: Find largest set F such that A_{ff} is θ-dominant.

S. MacLachlan, Y. Saad, SISC, to appear

Optimality, Robustness, Complexity

Optimality:

- Choosing diagonal M_{ff} ensures efficient relaxation
- Maximizing $|F|$ so that A_{ff} is θ-dominant minimizes iteration cost

---

S. MacLachlan, Y. Saad, SISC, to appear

Optimality, Robustness, Complexity

Optimality:

- Choosing diagonal M_{ff} ensures efficient relaxation
- Maximizing $|F|$ so that A_{ff} is θ-dominant minimizes iteration cost

Robustness:

- Fixing θ guarantees $\lambda_{\max} \leq \frac{1}{2\theta-1}$, bounds ρ_{MG}

S. MacLachlan, Y. Saad, SISC, to appear

Optimality, Robustness, Complexity

Optimality:

- Choosing diagonal M_{ff} ensures efficient relaxation
- Maximizing $|F|$ so that A_{ff} is θ-dominant minimizes iteration cost

Robustness:

- Fixing θ guarantees $\lambda_{\max} \leq \frac{1}{2\theta-1}$, bounds ρ_{MG}

Complexity:

Finding $\max\{|F| : A_{ff} \text{ is } \theta\text{-dominant}\}$, is NP-complete.

- Constrained 0-1 integer programming problem
- Decision variables, f_i, are indicators of F/C partition
- Constraint function is row-wise θ-dominance criterion

S. MacLachlan, Y. Saad, SISC, to appear

Greedy Algorithm

Instead,

- Initialize $U = \{1, \ldots, n\}$, $F = C = \emptyset$
- For each point in U, compute $\hat{\theta}_i = \dfrac{a_{ii}}{\sum_{j \in F \cup U} |a_{ij}|}$
- Whenever $\hat{\theta}_i \geq \theta$, $i \rightarrow F$
- If $U \neq \emptyset$, then pick $j = \text{argmin}_{i \in U}\{\hat{\theta}_i\}$
  - $j \rightarrow C$
  - Update $\hat{\theta}_i$ for all $i \in U$ with $a_{ji} \neq 0$

Guarantee: A_{ff} is θ-dominant

Look for: largest A_{ff} possible in linear time

S. MacLachlan, Y. Saad, SISC, to appear

A Challenge

AMGr theory requires $P = \begin{bmatrix} M_{ff}^{-1} A_{fc} \\ I \end{bmatrix}$ and $\begin{bmatrix} M_{ff} & -A_{fc} \\ -A_{cf} & Acc \end{bmatrix}$ be SPSD

$\mathcal{K}$	grid	c_A	ρ_{MG}
constant	2048×2048	1.33	0.41
smooth	2048×2048	1.33	0.41
binary	2048×2048	1.98	0.91
anisotropic	32×32	48.16	0.98989

Finite-element discretizations of $-\nabla \mathcal{K} \nabla p$

In practice: Use greedy coarsening algorithm in combination with classical AMG interpolation

Numerical Results

$\mathcal{K}$	Grid	c_A	ρ
constant	512×512	1.33	0.13
	1024×1024	1.33	0.14
	2048×2048	1.33	0.14
smooth	512×512	1.33	0.13
	1024×1024	1.33	0.14
	2048×2048	1.33	0.14
binary	512×512	2.06	0.35
	1024×1024	2.08	0.40
	2048×2048	2.10	0.46
anisotropic	512×512	2.39	0.13
	1024×1024	2.41	0.20
	2048×2048	2.43	0.20

Time-Harmonic Maxwell Equations

Maxwell Equations:

$$\nabla \times H = J + \frac{\partial D}{\partial t}, \qquad \nabla \cdot D = \rho,$$

$$\nabla \times E = -\frac{\partial B}{\partial t}, \qquad \nabla \cdot B = 0,$$

Reduce system by assuming:

- linear constitutive laws

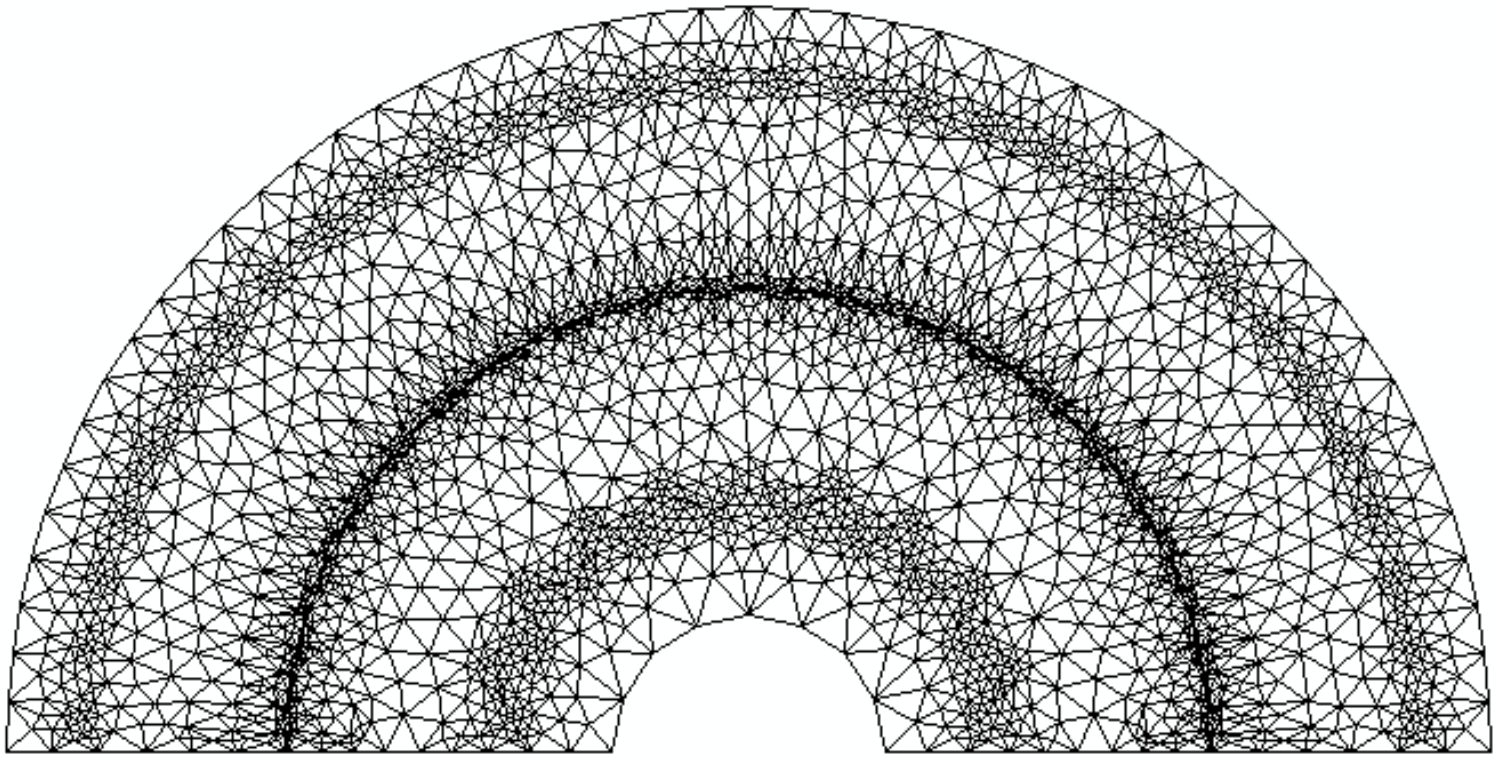
$$B = \mu H, \qquad D = \epsilon E, \qquad J = \sigma E$$

- low-frequency excitation
- Constant 2D cross-section

Induction Motor

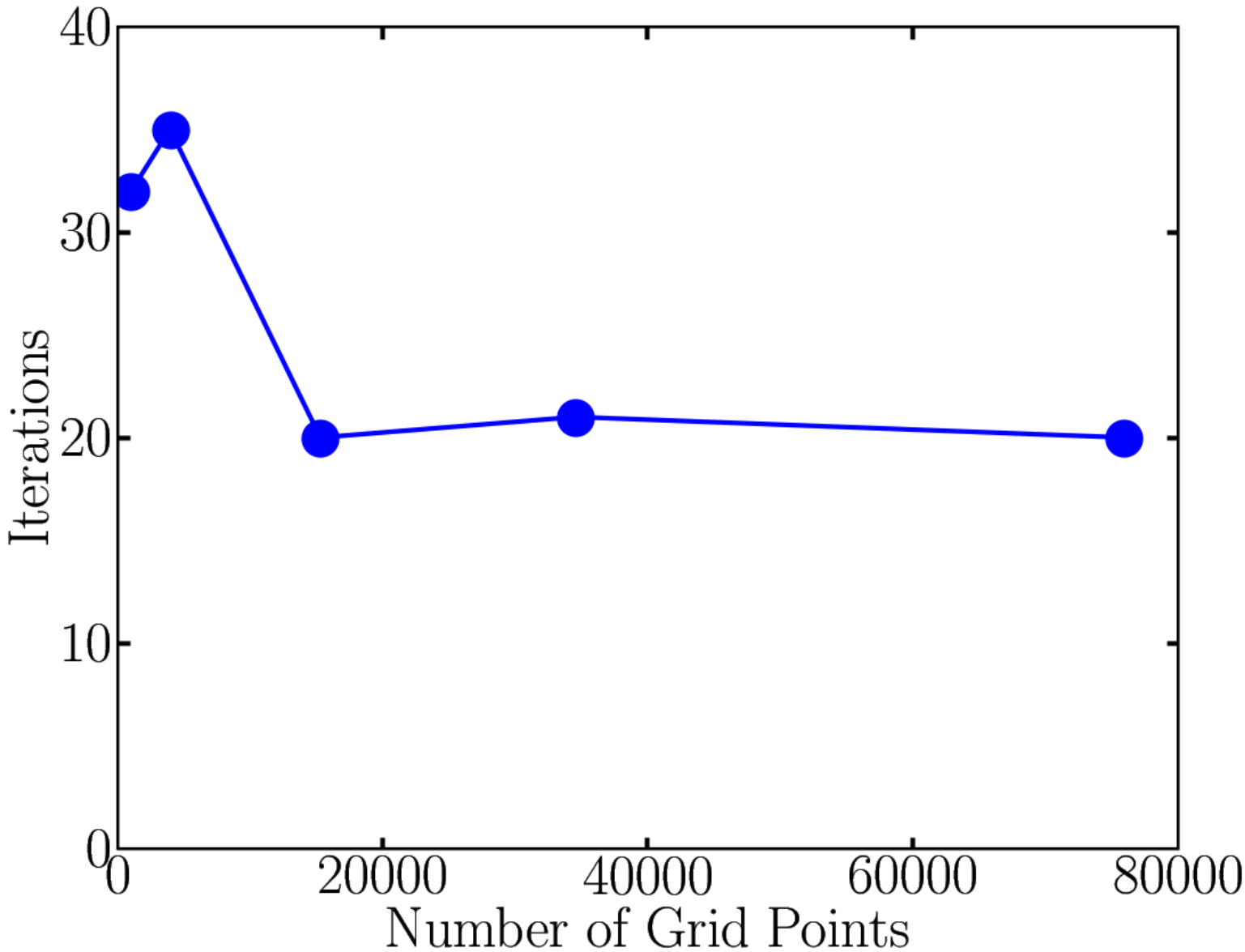
Find Fourier-domain potential, $\hat{A} = (0, 0, \hat{A}_z)^T$, by solving

$$-\nabla \cdot \left(\frac{1}{\mu} \nabla \hat{A}_z \right) + \imath\omega\sigma\hat{A}_z = \hat{J}_{s,z}$$

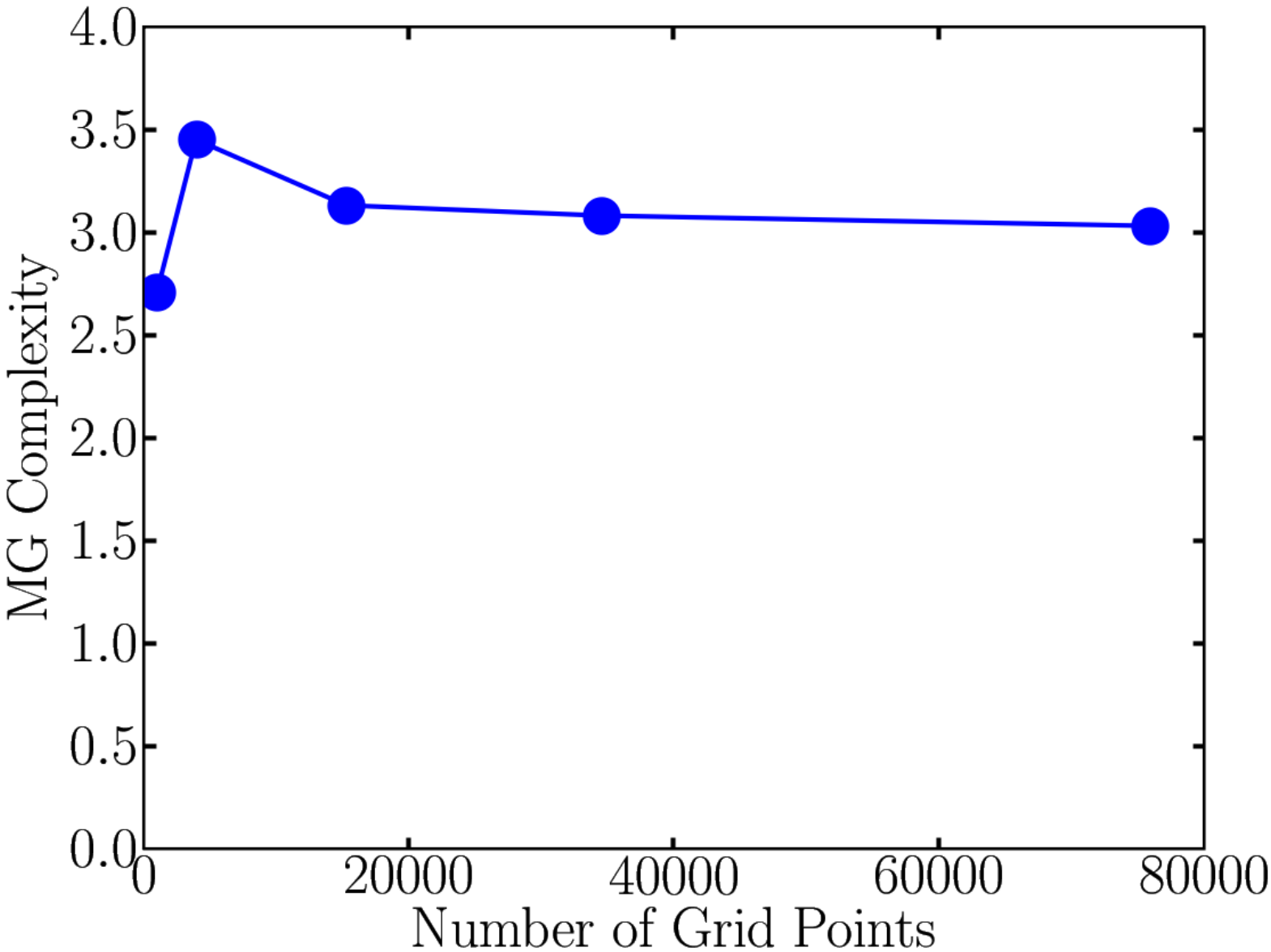


Lahaye et al., IEEE Trans. Magn. 2000, **36**:1535-1538

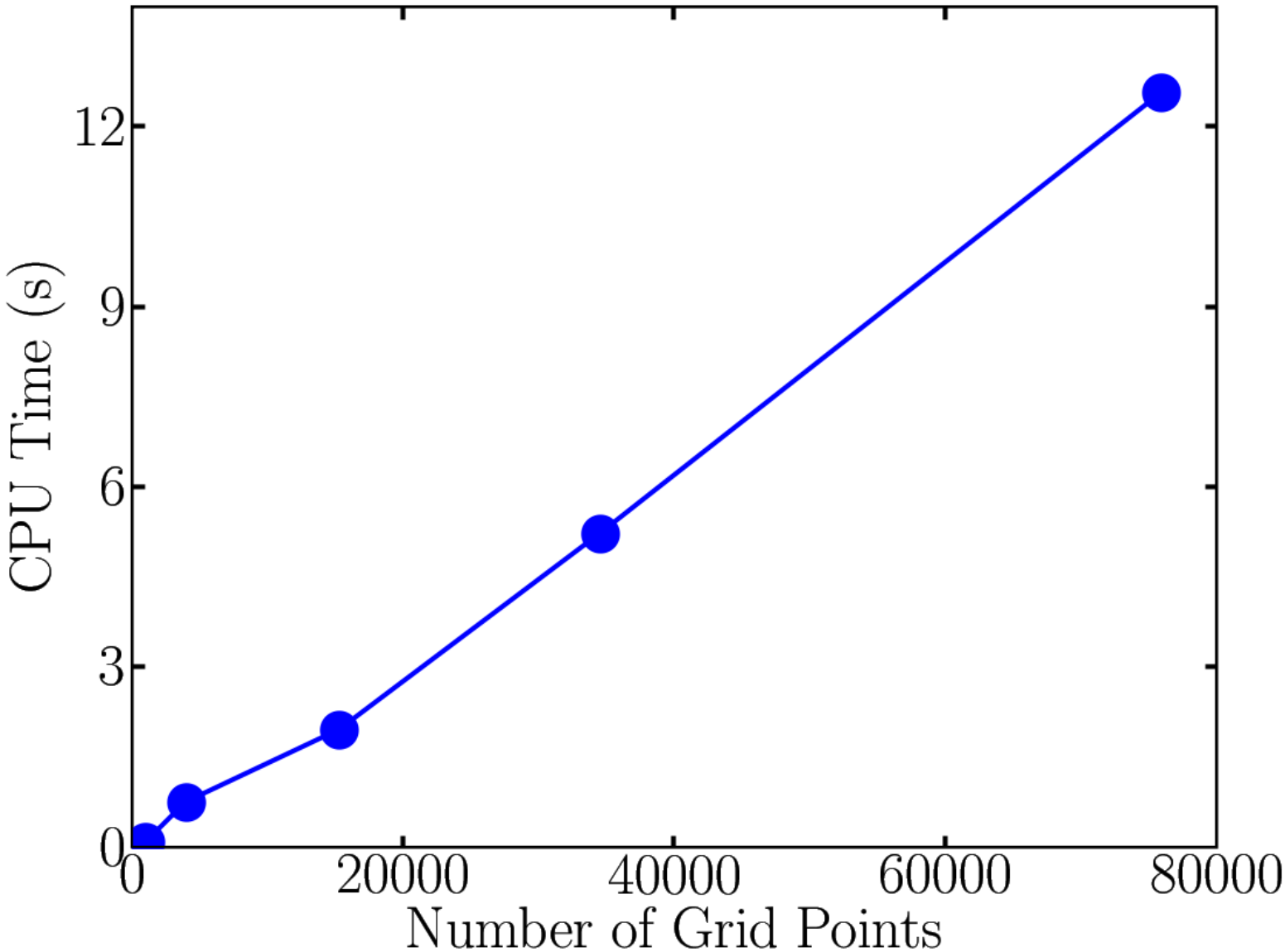
AMG Performance



AMG Performance



AMG Performance



Summary

- Multigrid methods provide effective large-scale solvers
- Algebraic multigrid effective for heterogeneous systems
- Complementarity is key
  - Relaxation and coarse-grid correction

Summary

- Multigrid methods provide effective large-scale solvers
- Algebraic multigrid effective for heterogeneous systems
- Complementarity is key
  - Relaxation and coarse-grid correction
- Convergence theory separates smoothing and approximation
- Compatible relaxation evaluates quality of coarse-grid set
- AMGr defines interpolation based on dominance principles
- Greedy algorithm puts these together

Limitations and Outlook

Predictive AMG theory is very limited

AMGr & Greedy coarsening:

- In some cases, works as well as classical AMG
- Only guaranteed to be effective with diagonal dominance

Goal: push both ways

- Find new algorithms based on old bounds
- Find new, predictive bounds on AMG theory