Adaptive Algebraic Multigrid

Scott MacLachlan

maclachl@colorado.edu

Department of Applied Mathematics, University of Colorado at Boulder

Support and Collaboration

- This work has been supported by 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, and James Brannick at CU-Boulder.

Recent Advances in Multilevel Methods

- Significant interest in simulating complex physical systems with features, and hence solutions, that vary on multiple scales
- Accuracy constraints are often driven by motivating applications, requiring efficient iterative methods to solve the resulting linear (and non-linear) systems
- Multiscale solution techniques, such as multigrid, are often most efficient approach
- Recent advances include
  - new multigrid techniques to broaden applicability of algebraic multigrid solvers/preconditioners
  - improvements in implementation and understanding of multigrid in (massively) parallel environments
  - new approaches to the application of multiscale/multilevel ideas in many application areas

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

- Relaxation

Relax
$A^{(1)}v^{(1)}=f^{(1)}$

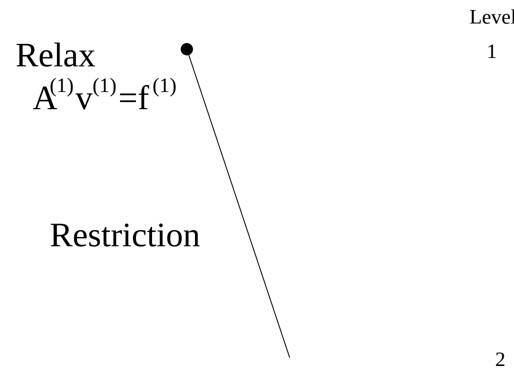
- Use a smoothing process (such as Gauss-Seidel) to eliminate oscillatory errors
- Remaining error satisfies $A\mathbf{e} = \mathbf{r} \equiv \mathbf{f} - A\mathbf{v}$

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

- Relaxation
- Restriction



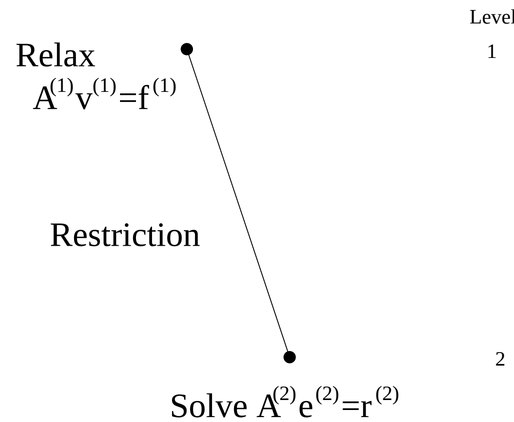
- Transfer residual to coarse grid

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

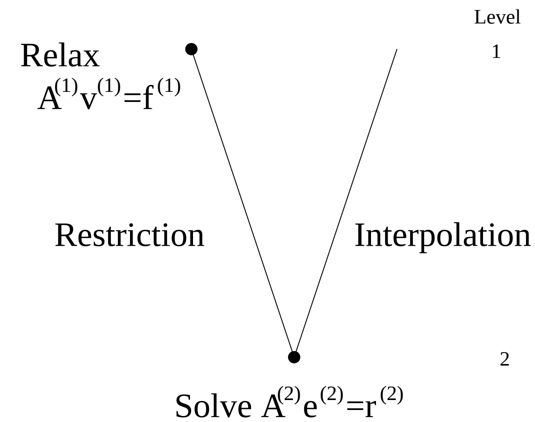
$$A^{(2)}\mathbf{e}^{(2)} = \mathbf{r}^{(2)}$$

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

- Relaxation
- Restriction
- Coarse-Grid Correction
- Interpolation



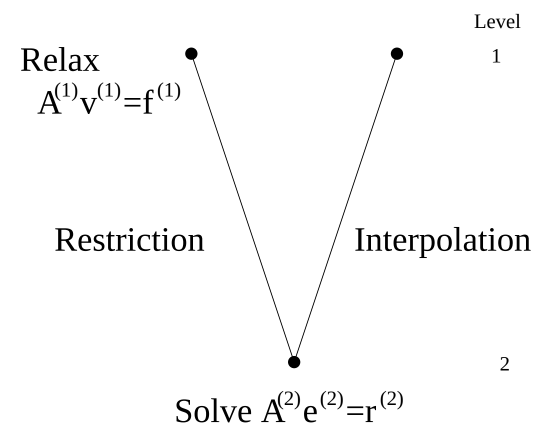
- Transfer correction to fine grid

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

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



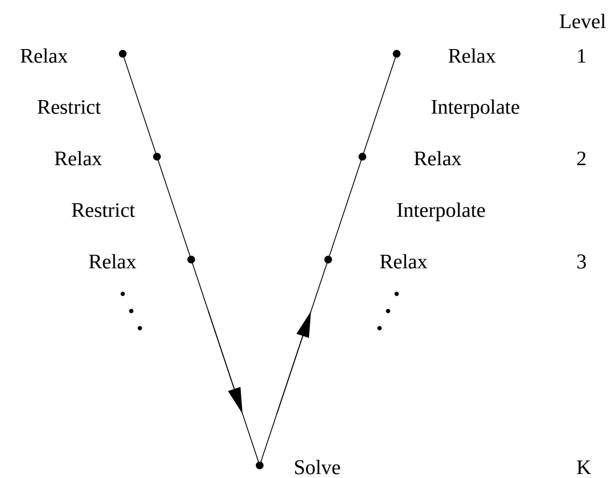
- Relax once again to remove oscillatory error introduced in coarse-grid correction

Multigrid

Multigrid Methods achieve optimality through complementarity

Multigrid Components

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



Obtain optimal efficiency through recursion

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

Classical AMG Coarsening

- Strong Connections based on matrix entries:

$$S_i = \left\{ j : -a_{ij} \geq \theta \max_{k \neq i} \{-a_{ik}\} \right\}$$

- Coarse grid chosen by maximal independent set heuristics

H1: For each $i \in F$, every $j \in S_i$ should be either in C_i or should strongly depend on at least one point in C_i

H2: The set, C, should be a maximal subset of the fine grid, such that no C-point strongly depends on another C-point

Weaknesses

- Definition of strong connections based on "nice" M-matrix properties
- Breaks down if near null space of A is far from the constant
  - Diagonal rescaling,

$$A \to DAD$$

  - Finite element anisotropy,

$$-u_{xx} - \epsilon u_{yy} \to \frac{1}{6}\begin{bmatrix} (-1-\epsilon) & (2-4\epsilon) & (-1-\epsilon) \\ (-4+2\epsilon) & (8+8\epsilon) & (-4+2\epsilon) \\ (-1-\epsilon) & (2-4\epsilon) & (-1-\epsilon) \end{bmatrix}$$

- Even for simple problems, size of a_{ij} may not reflect true connection between i and j

What are Strong Connections?

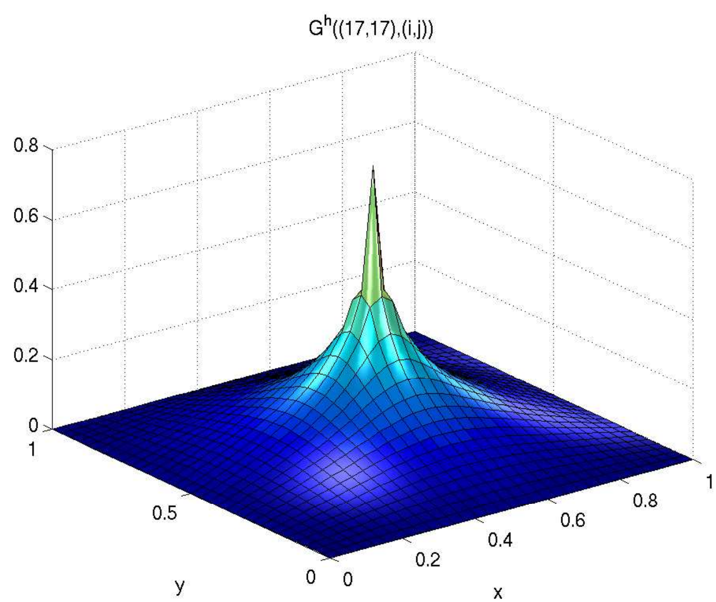
- Point i strongly depends on point j if
  - a change in the right-hand side at point j significantly changes the solution at point i.
  - a change in the residual at point j significantly changes the error at point i
- Good coarse-grid correction depends on identifying strong connections
  - Interpolation to i is most effective from points that it strongly depends on
  - Corrections from weakly connected points have little effect on the error at i

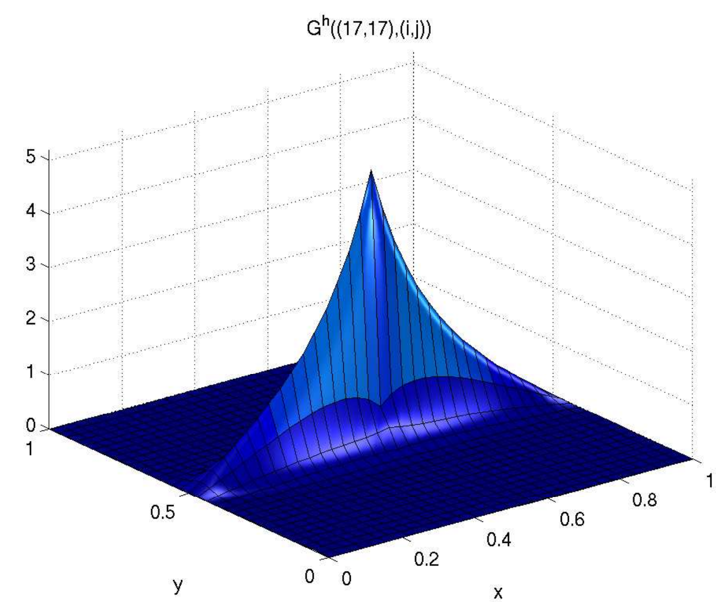
Inverse-based Strength

- For the discrete linear system, $A\mathbf{v} = \mathbf{f}$, the inverse relates changes in $\mathbf{f}$ to changes in $\mathbf{v}$

$$\mathbf{v} = (A)^{-1}\,\mathbf{f}$$

- If a change in f_j causes a significant change in v_i, then $(A)^{-1}_{ij}$ must be large relative to other values of $(A)^{-1}_{ik}$



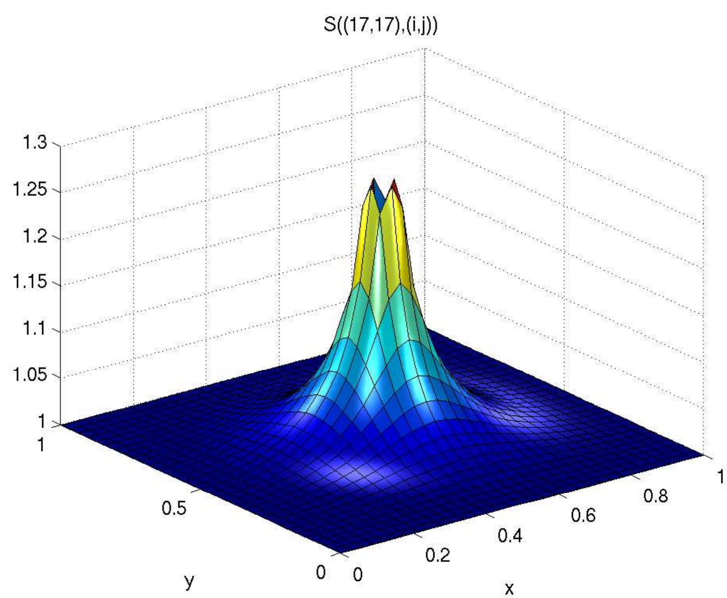


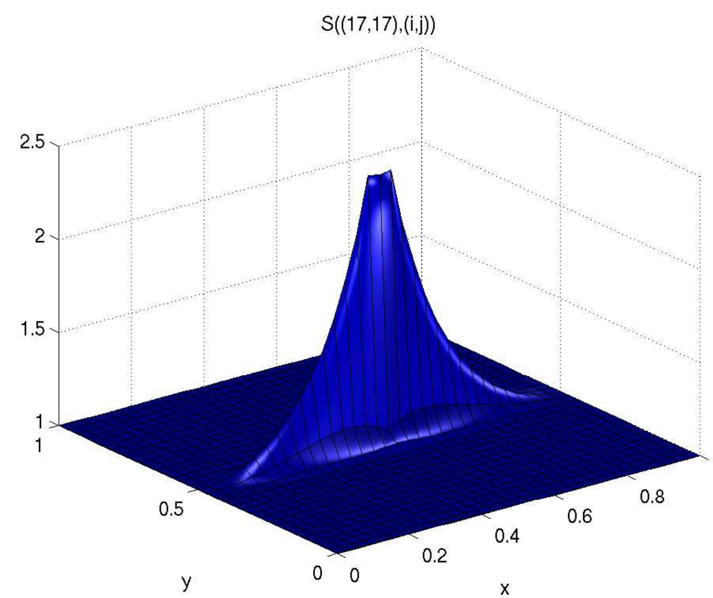
Measures of Strong Connections

- Strength of dependence of i on j depends on size of $(A)^{-1}_{ij}$
- How should we measure this size, relative to $(A)^{-1}_{ik}$?

Measures of Strong Connections

- Strength of dependence of i on j depends on size of $(A)^{-1}_{ij}$
- How should we measure this size, relative to $(A)^{-1}_{ik}$?
- L^2 measure: $(A)^{-1}_{ij} \geq \theta \max_{k \neq i} \left\{ (A)^{-1}_{ik} \right\}$
- Energy measure: Let $G^{(i)}_j = (A)^{-1}_{ij}$, $S_{ij} = \dfrac{\|\mathbf{G}^{(i)} - G^{(i)}_j \mathbf{I}^{(j)}\|_A}{\|\mathbf{G}^{(i)}\|_A}$



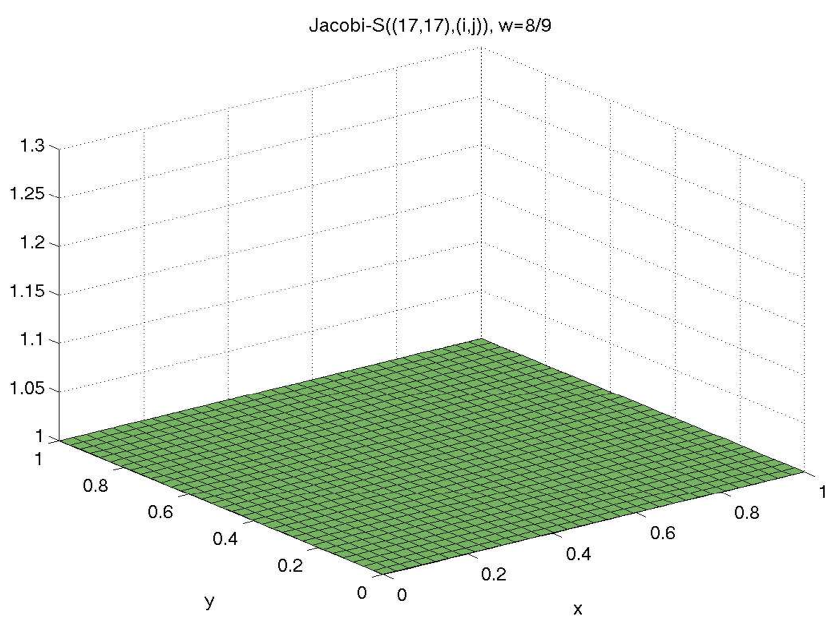


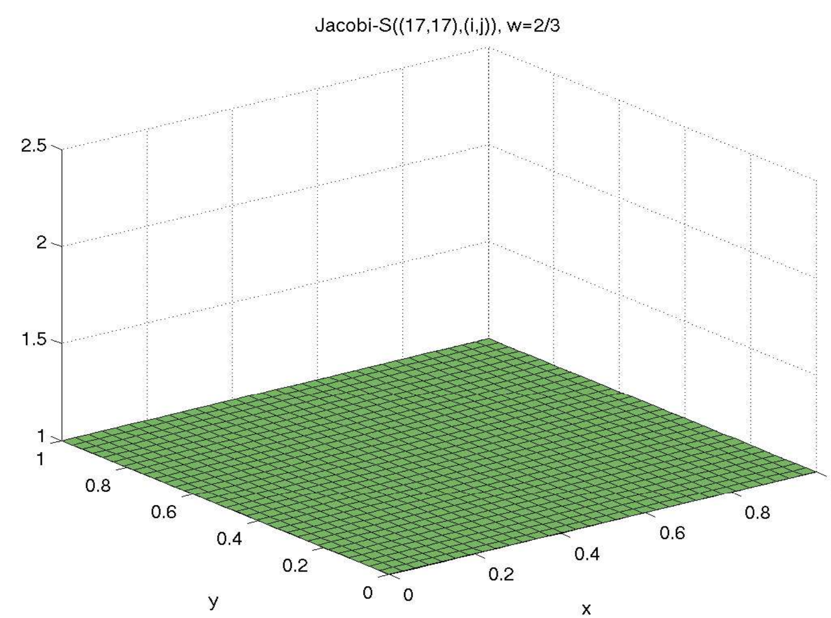
Approximating S_{ij}

- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$

Approximating S_{ij}

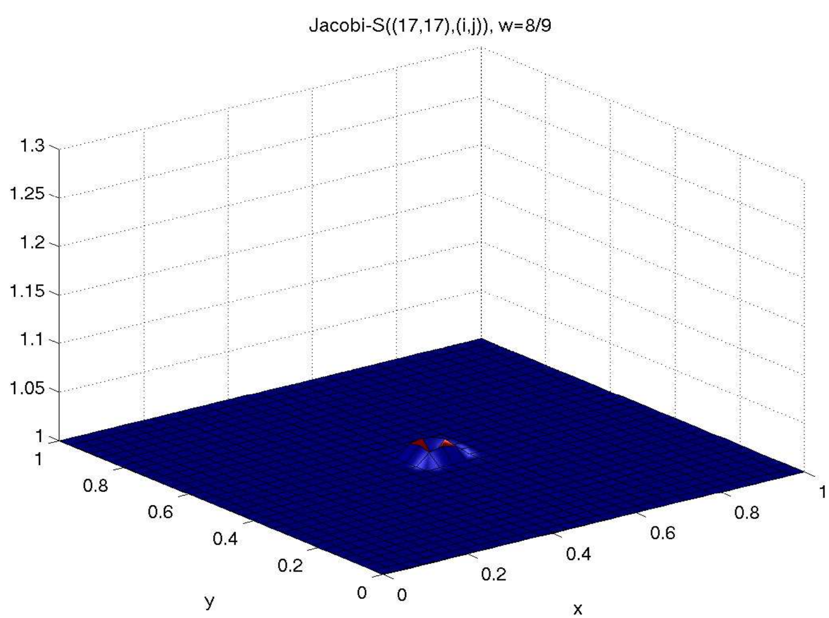
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 1 step:

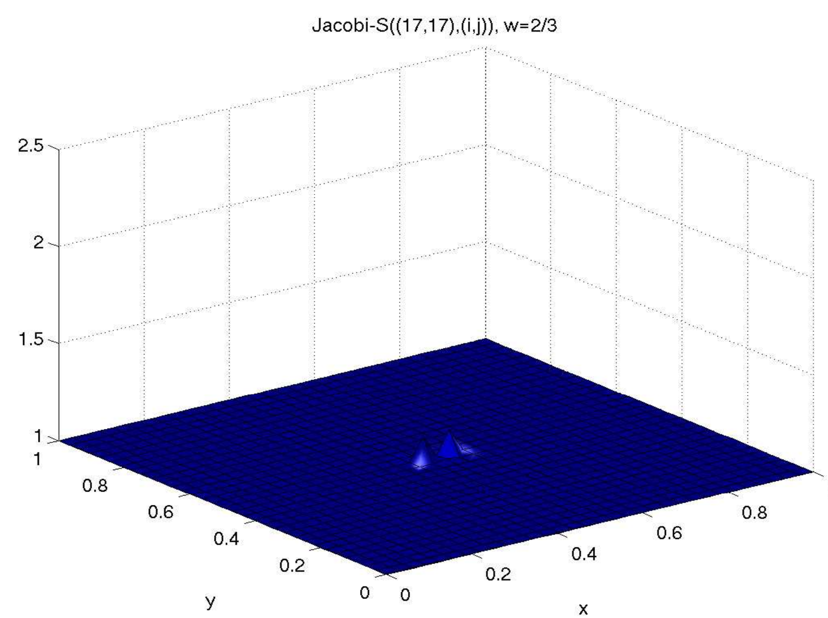




Approximating S_{ij}

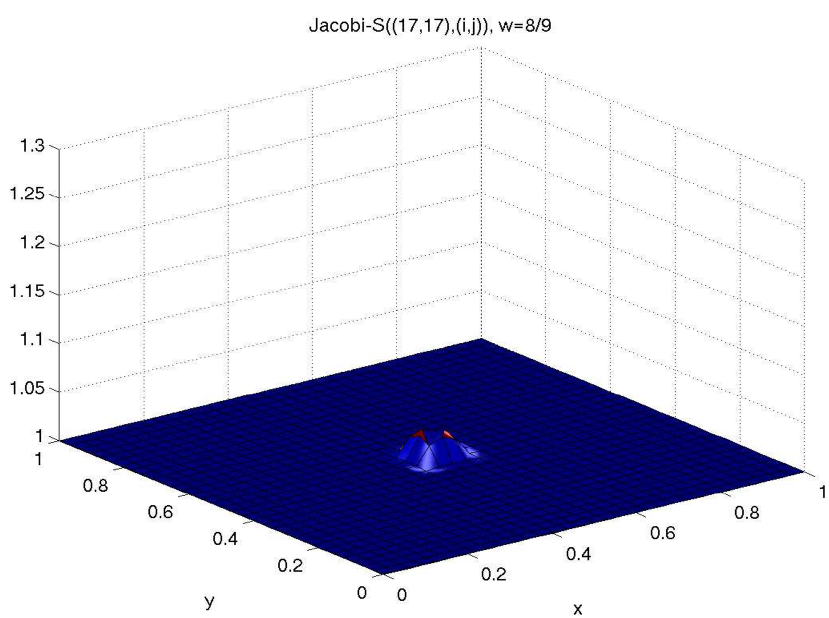
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 2 steps:

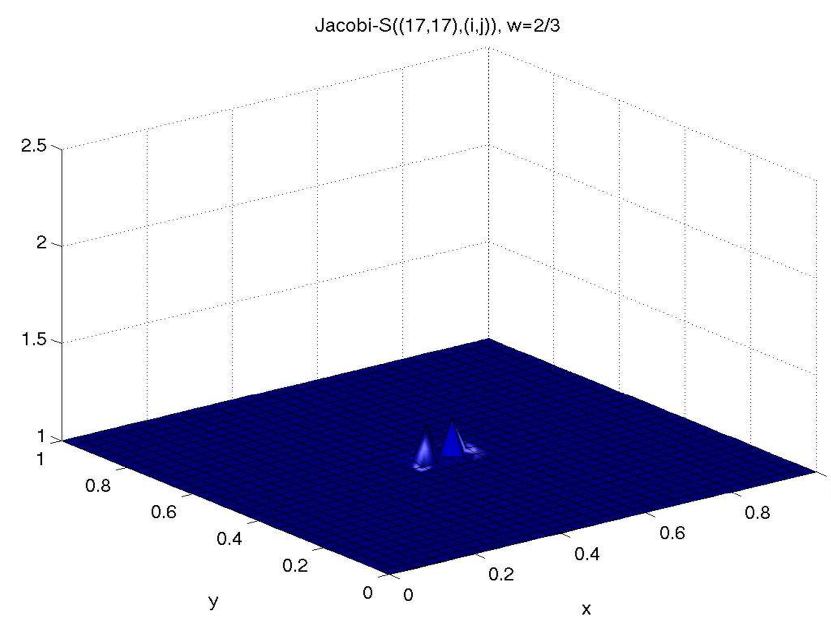




Approximating S_{ij}

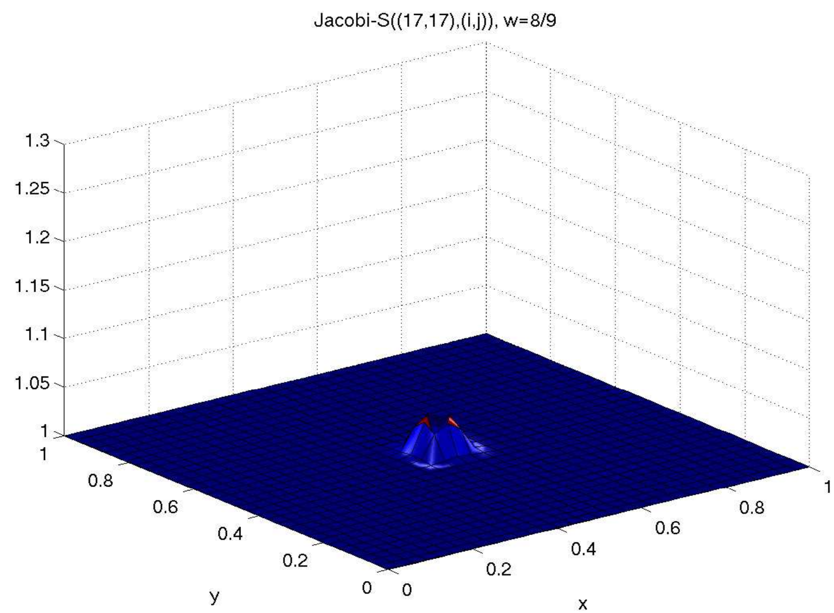
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 3 steps:

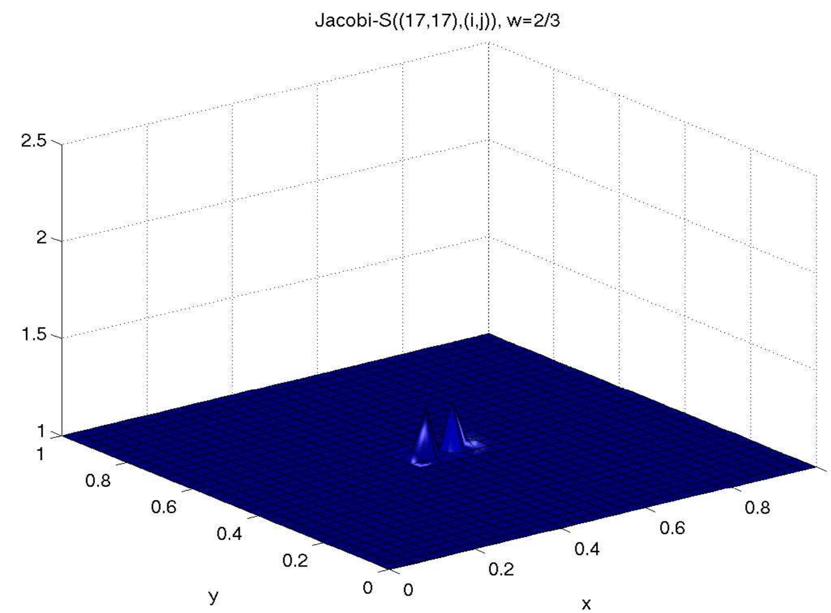




Approximating S_{ij}

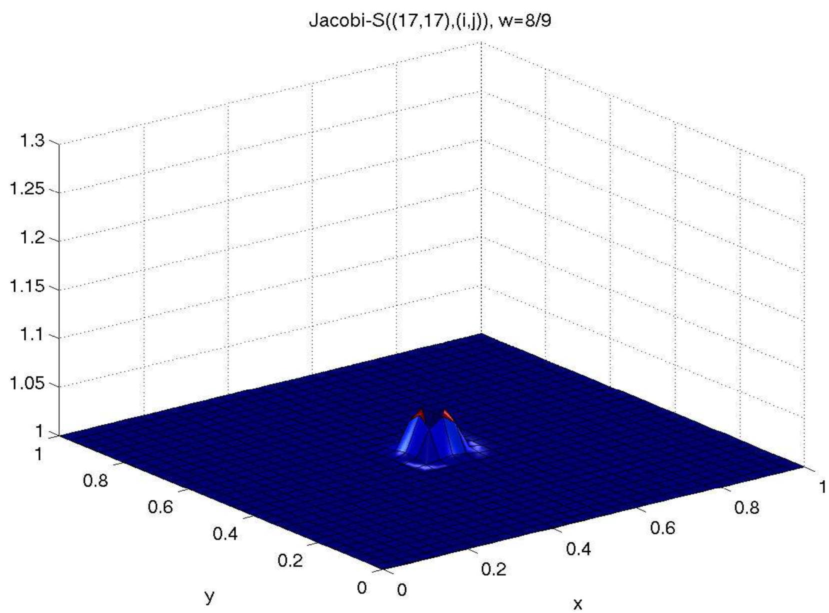
- Can we get useful, local approximations to $(A)_{ij}^{-1}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 4 steps:

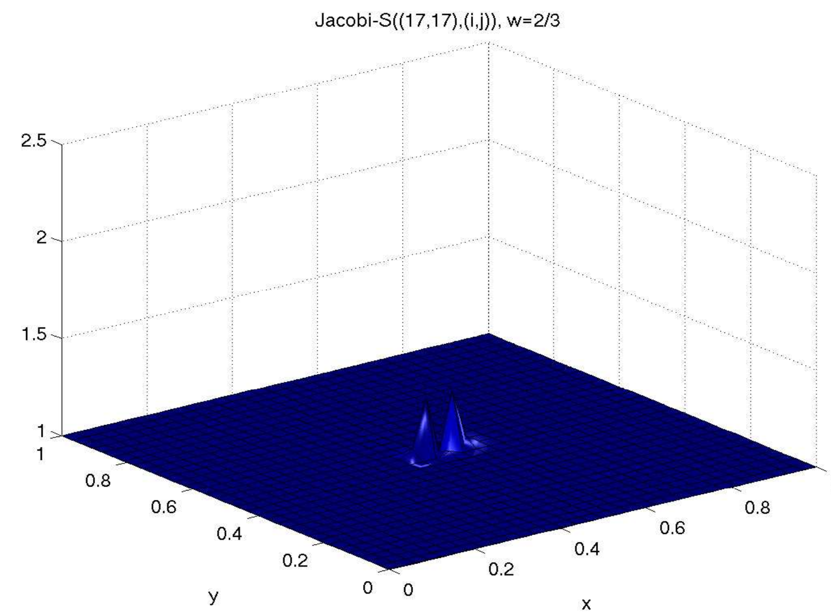




Approximating S_{ij}

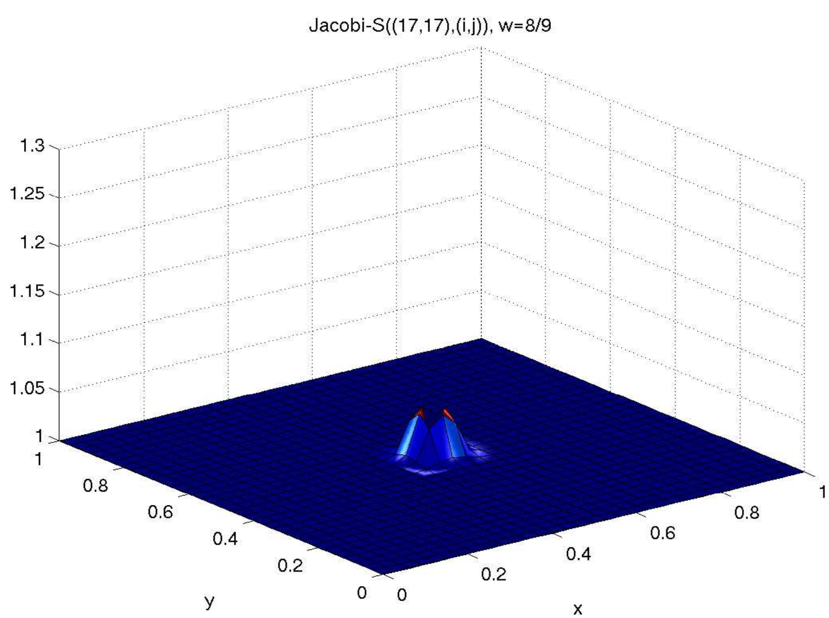
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 5 steps:

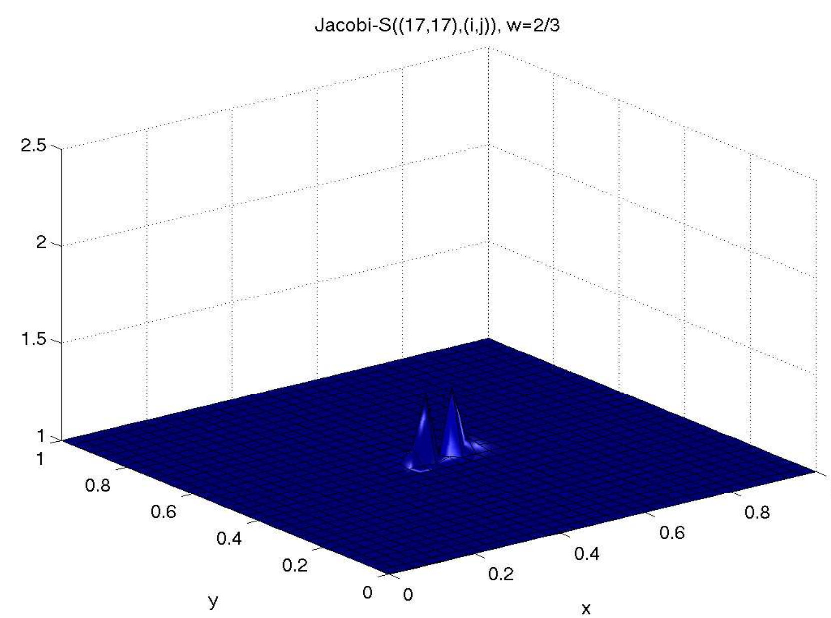




Approximating S_{ij}

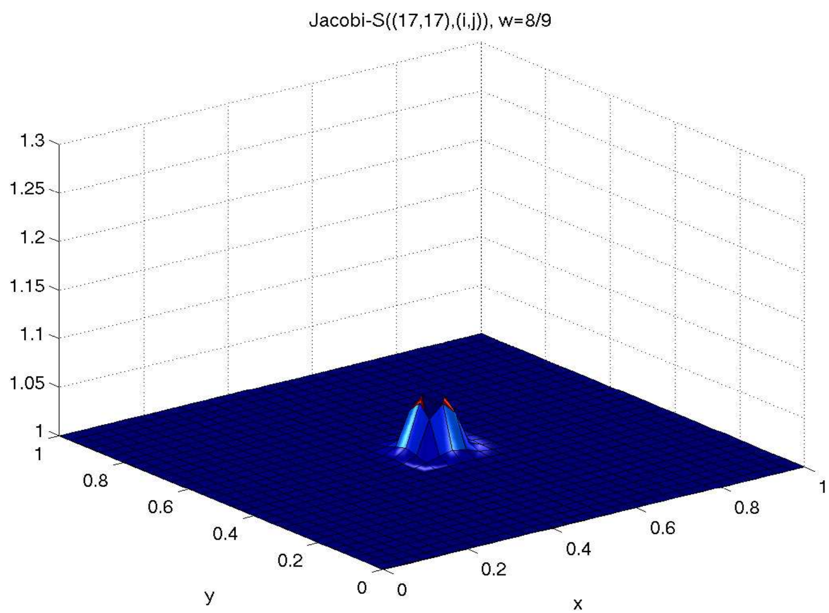
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 6 steps:

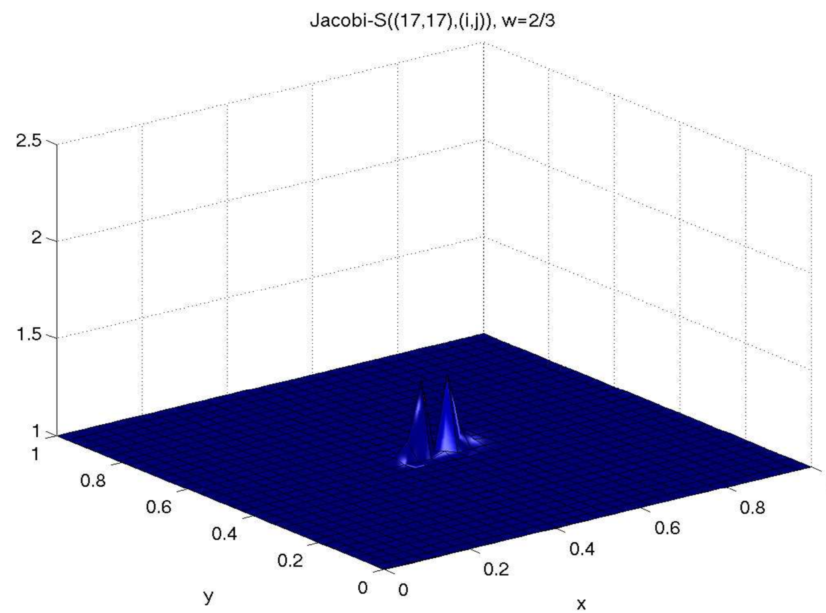




Approximating S_{ij}

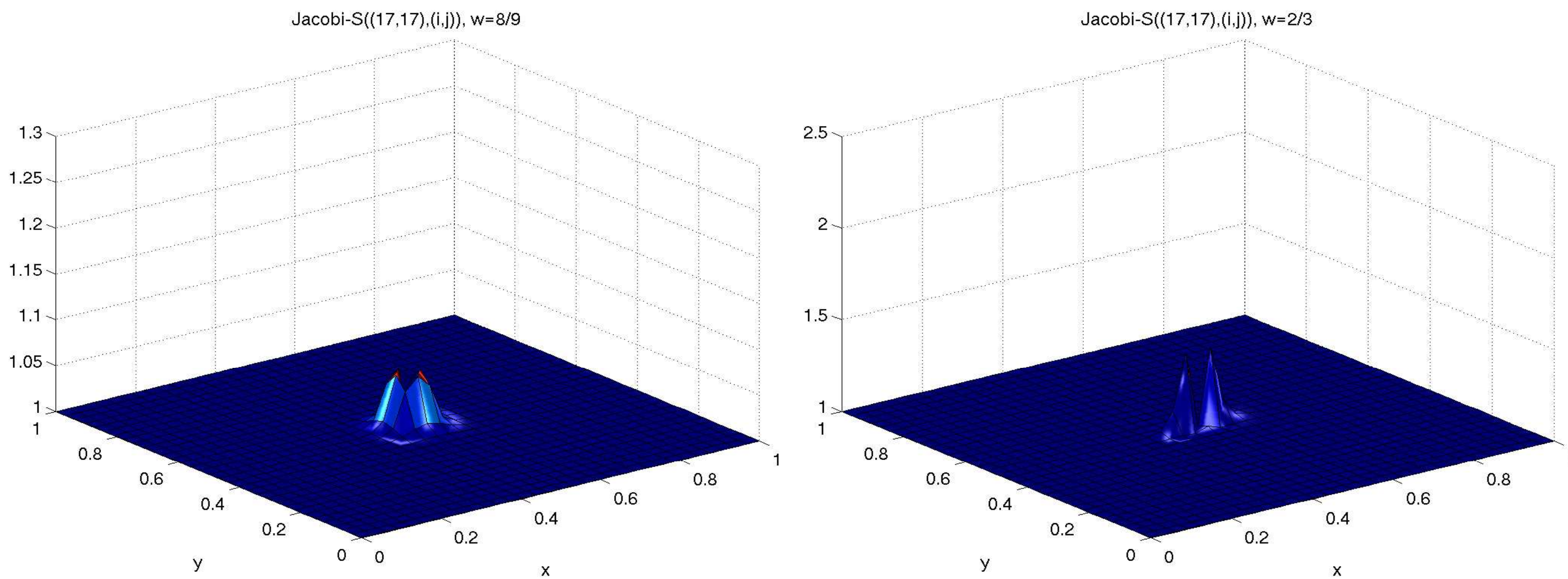
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 7 steps:





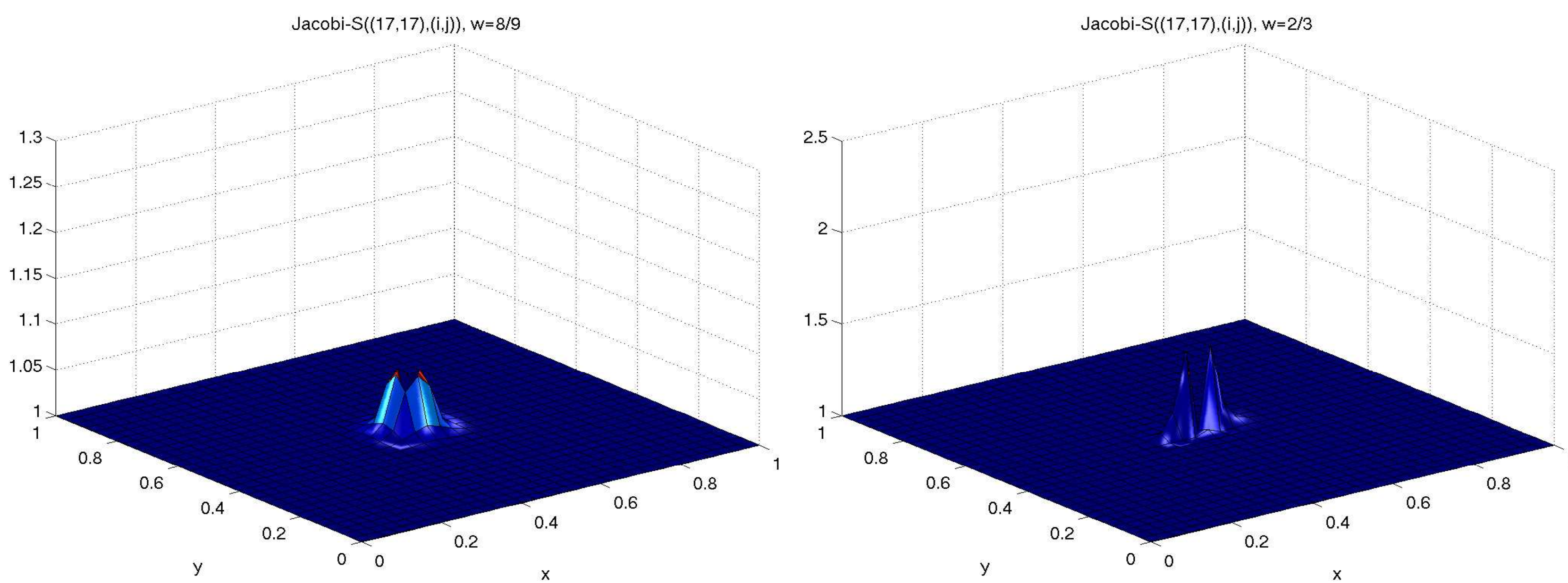
Approximating S_{ij}

- Can we get useful, local approximations to $(A)_{ij}^{-1}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 8 steps:



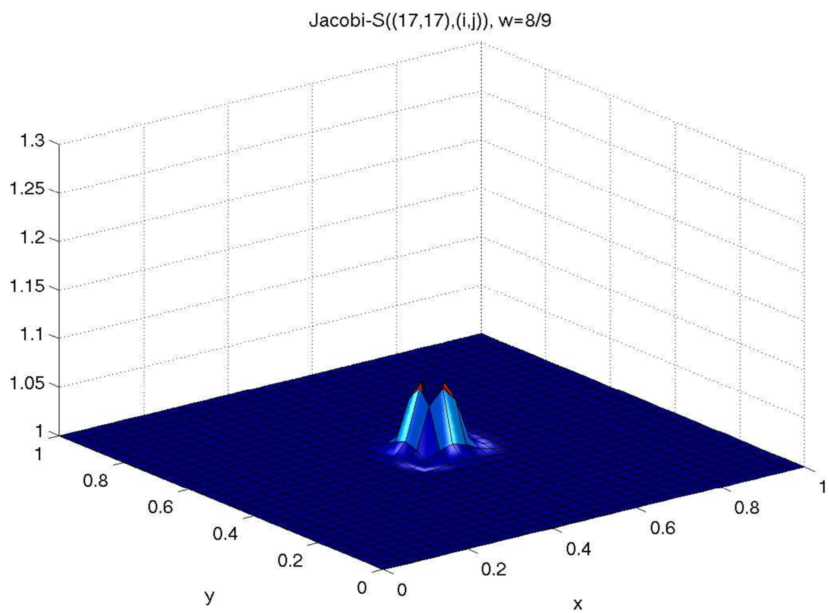
Approximating S_{ij}

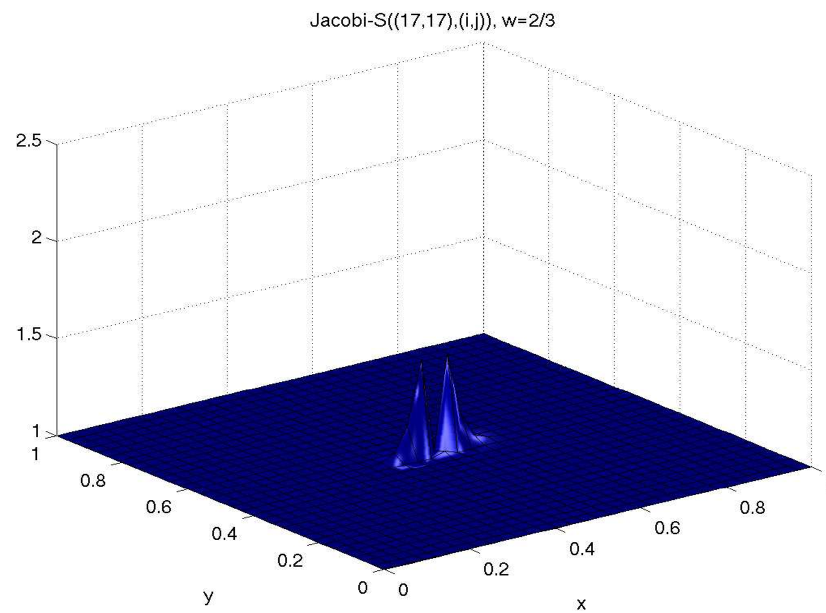
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 9 steps:



Approximating S_{ij}

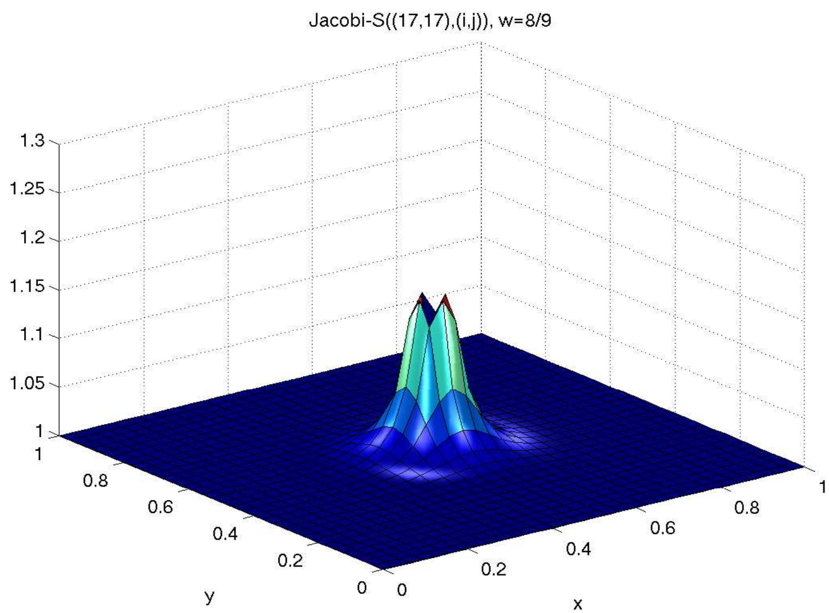
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 10 steps:

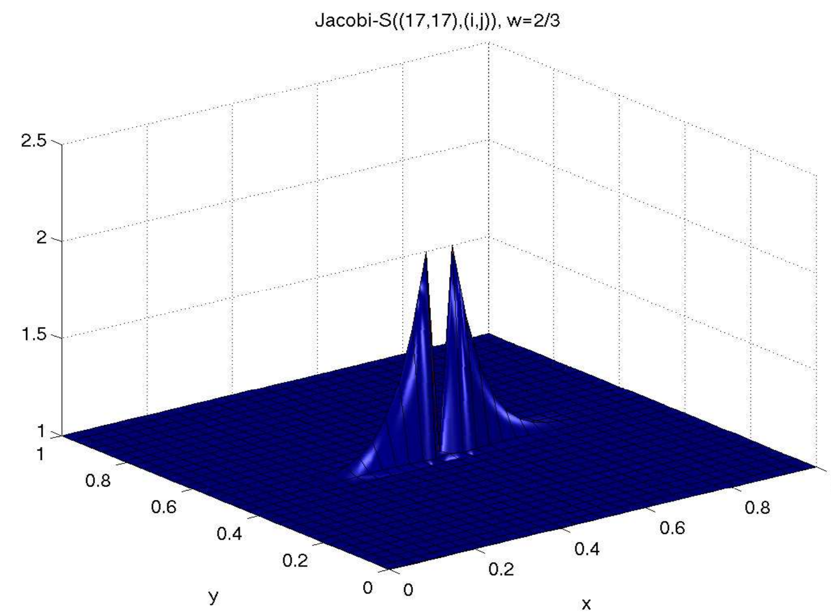




Approximating S_{ij}

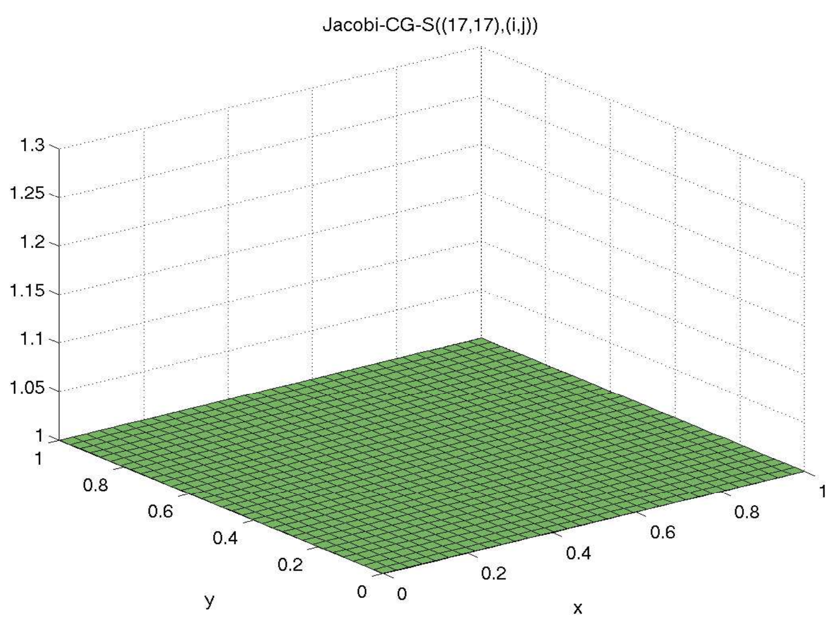
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Weighted Jacobi, 50 steps:

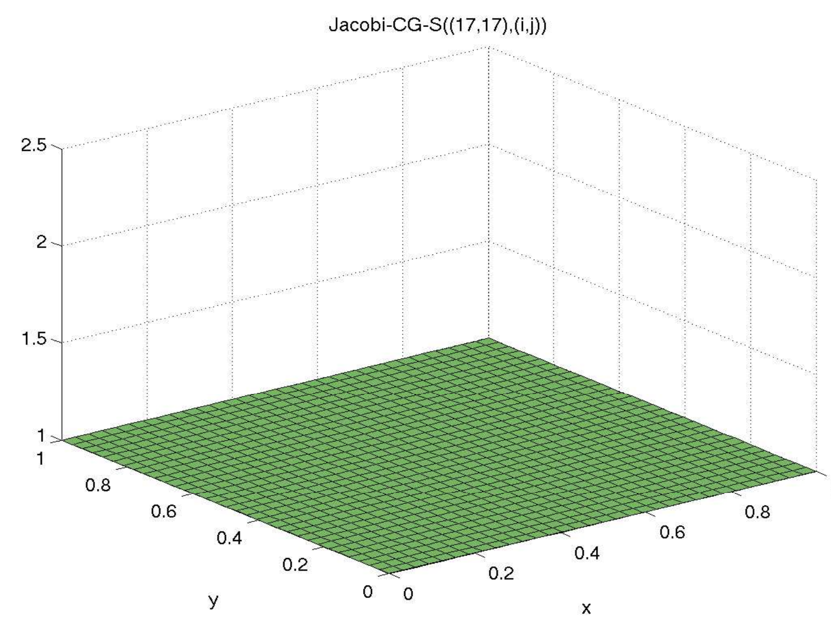




Approximating S_{ij}

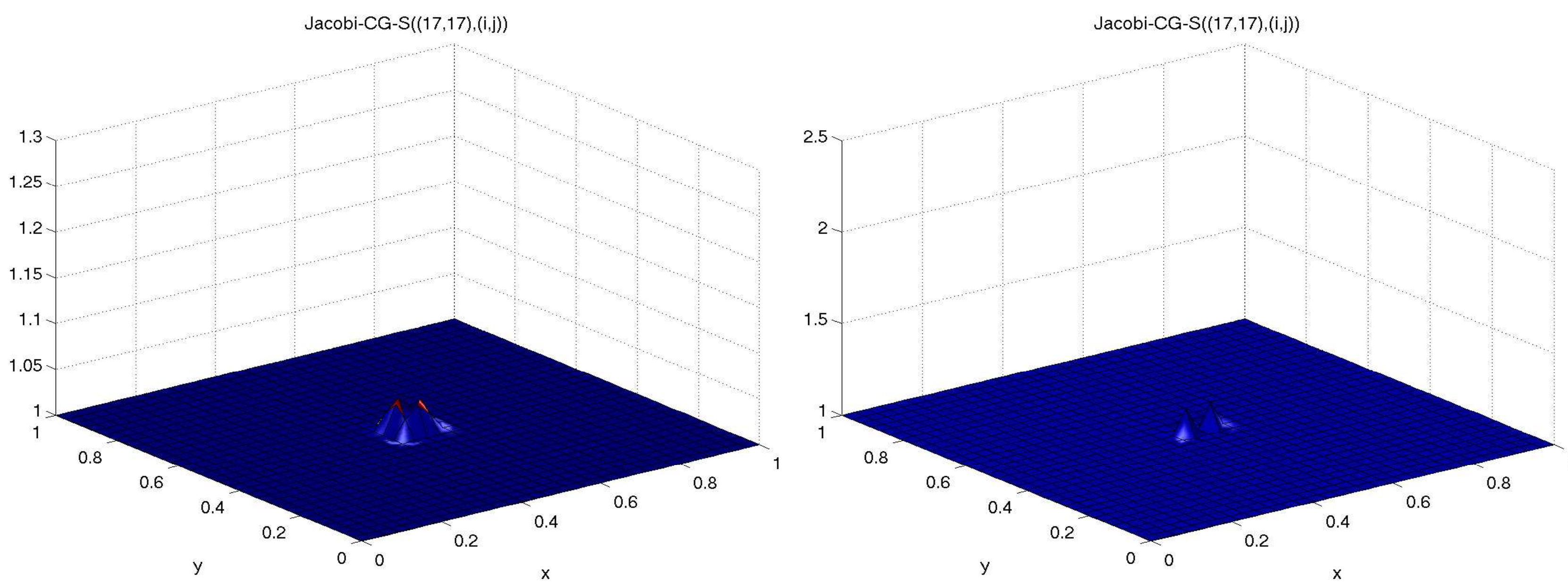
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Jacobi-Preconditioned CG, 1 step:





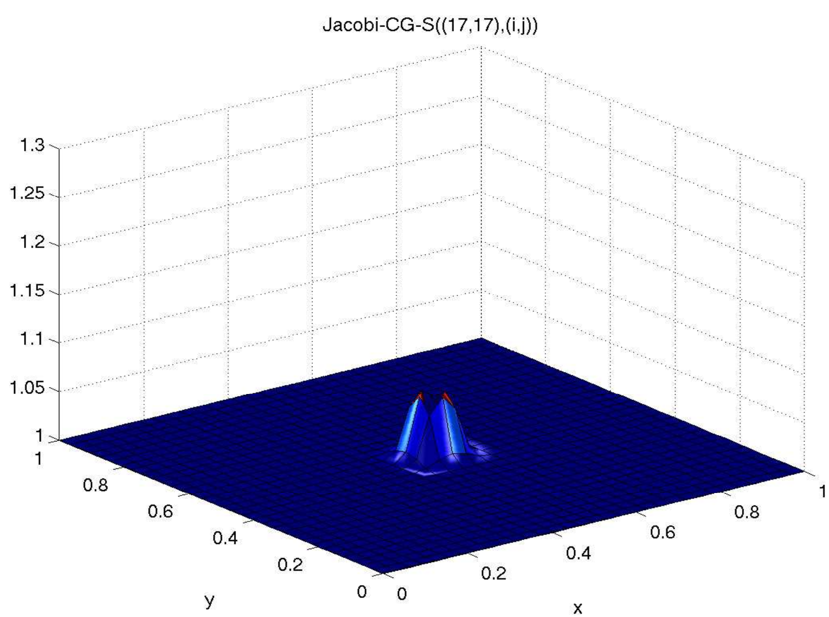
Approximating S_{ij}

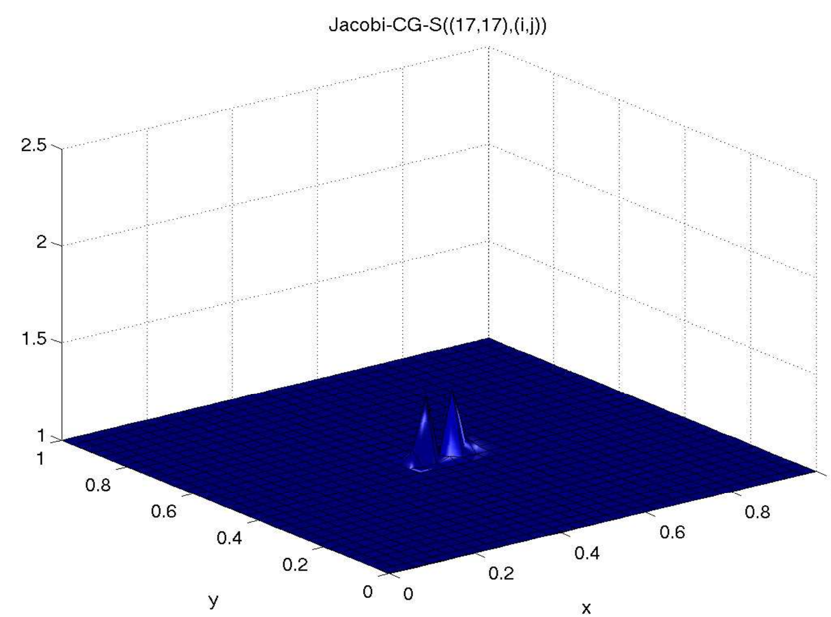
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Jacobi-Preconditioned CG, 2 steps:



Approximating S_{ij}

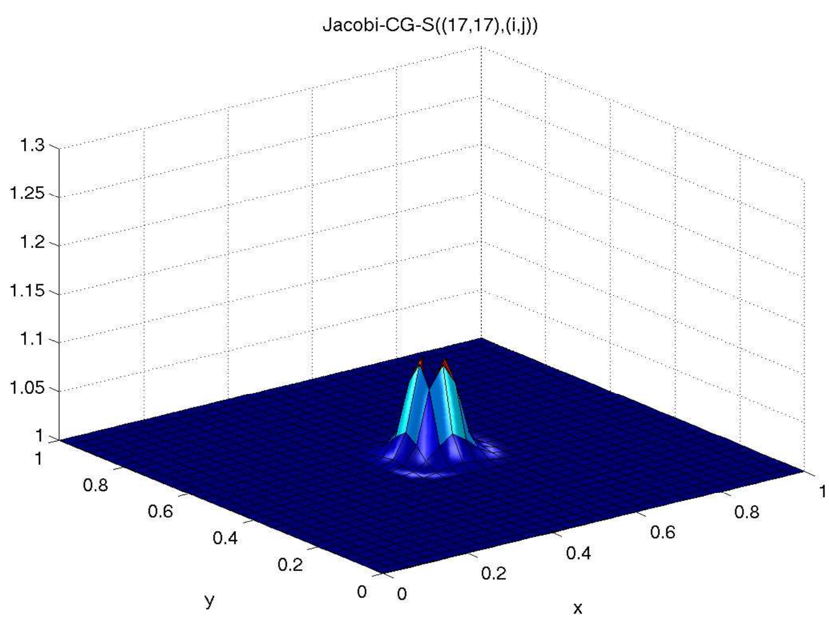
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Jacobi-Preconditioned CG, 3 steps:

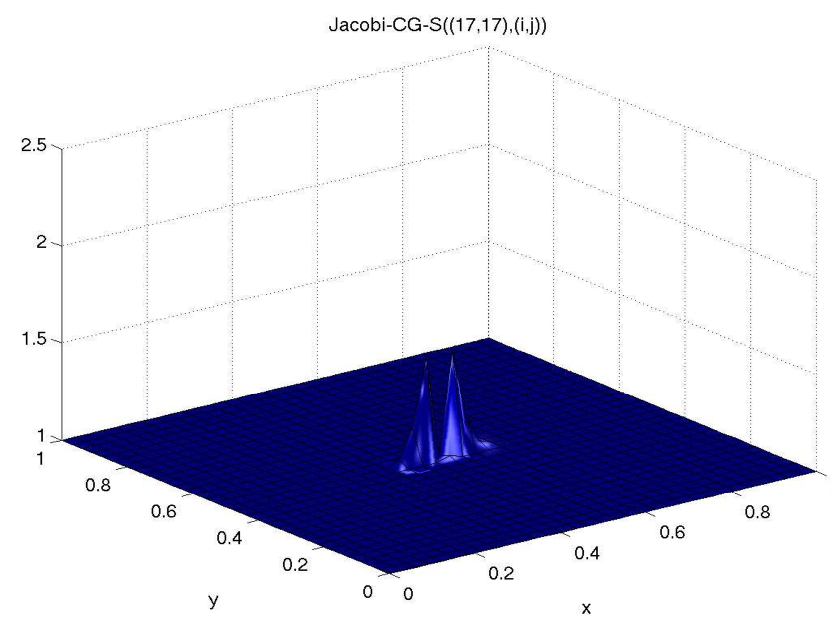




Approximating S_{ij}

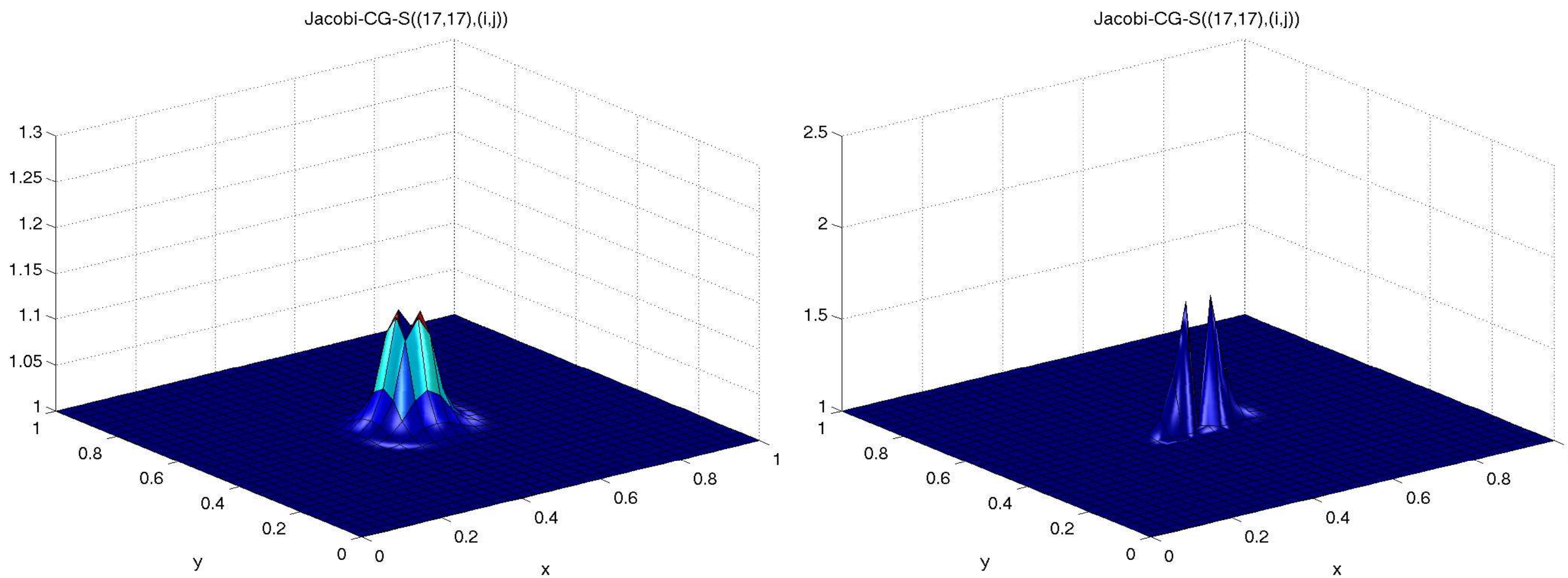
- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Jacobi-Preconditioned CG, 4 steps:





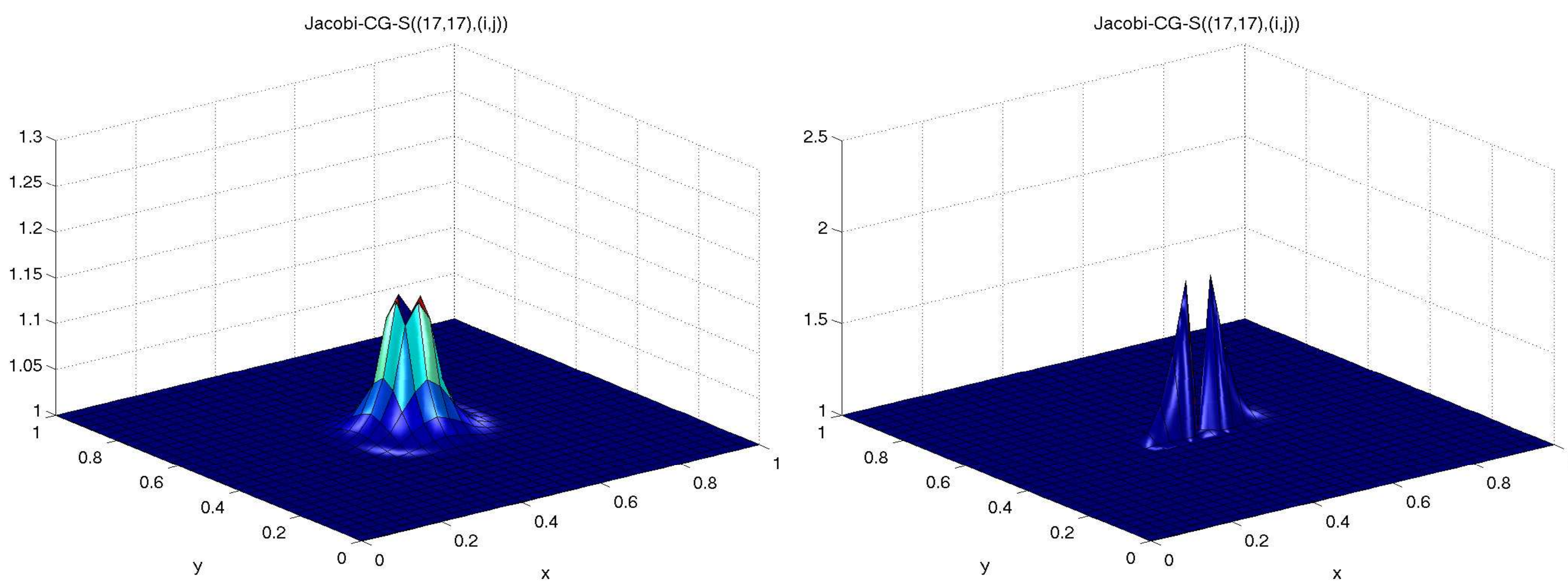
Approximating S_{ij}

- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Jacobi-Preconditioned CG, 5 steps:



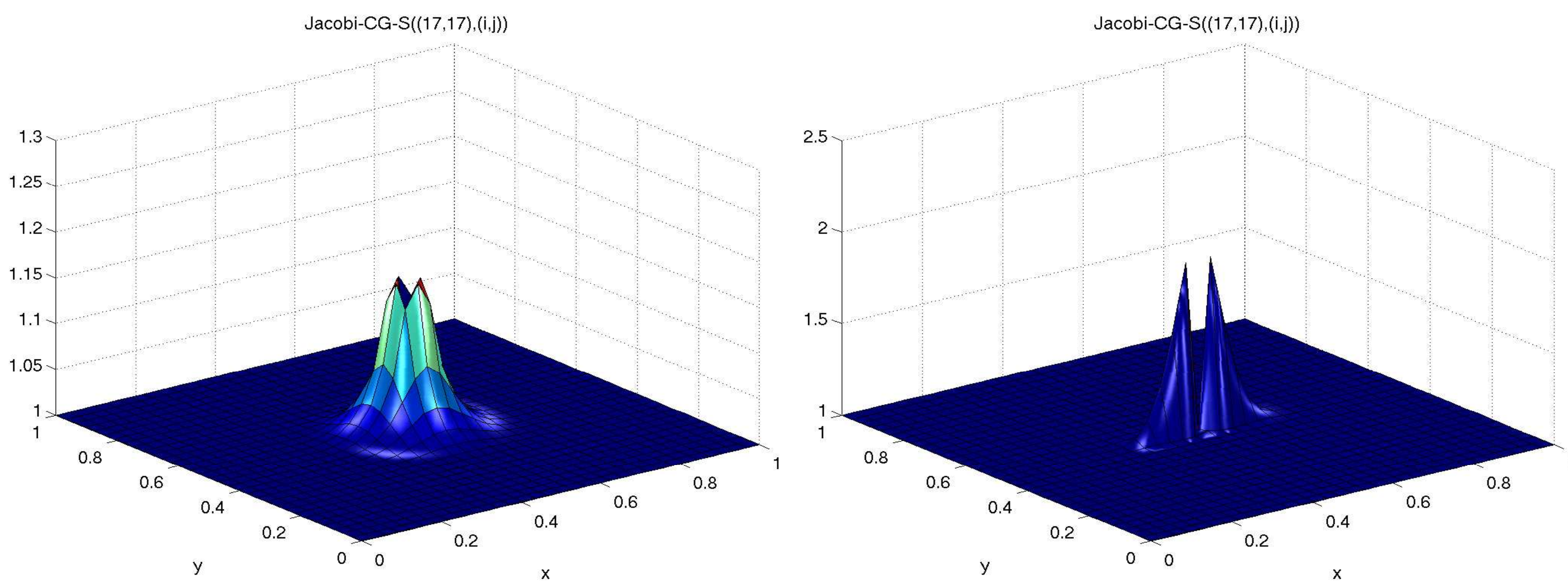
Approximating S_{ij}

- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Jacobi-Preconditioned CG, 6 steps:



Approximating S_{ij}

- Can we get useful, local approximations to $(A)^{-1}_{ij}$ and, thus, S_{ij}?
- Apply (localized) relaxation to $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
- Jacobi-Preconditioned CG, 7 steps:



Choosing C

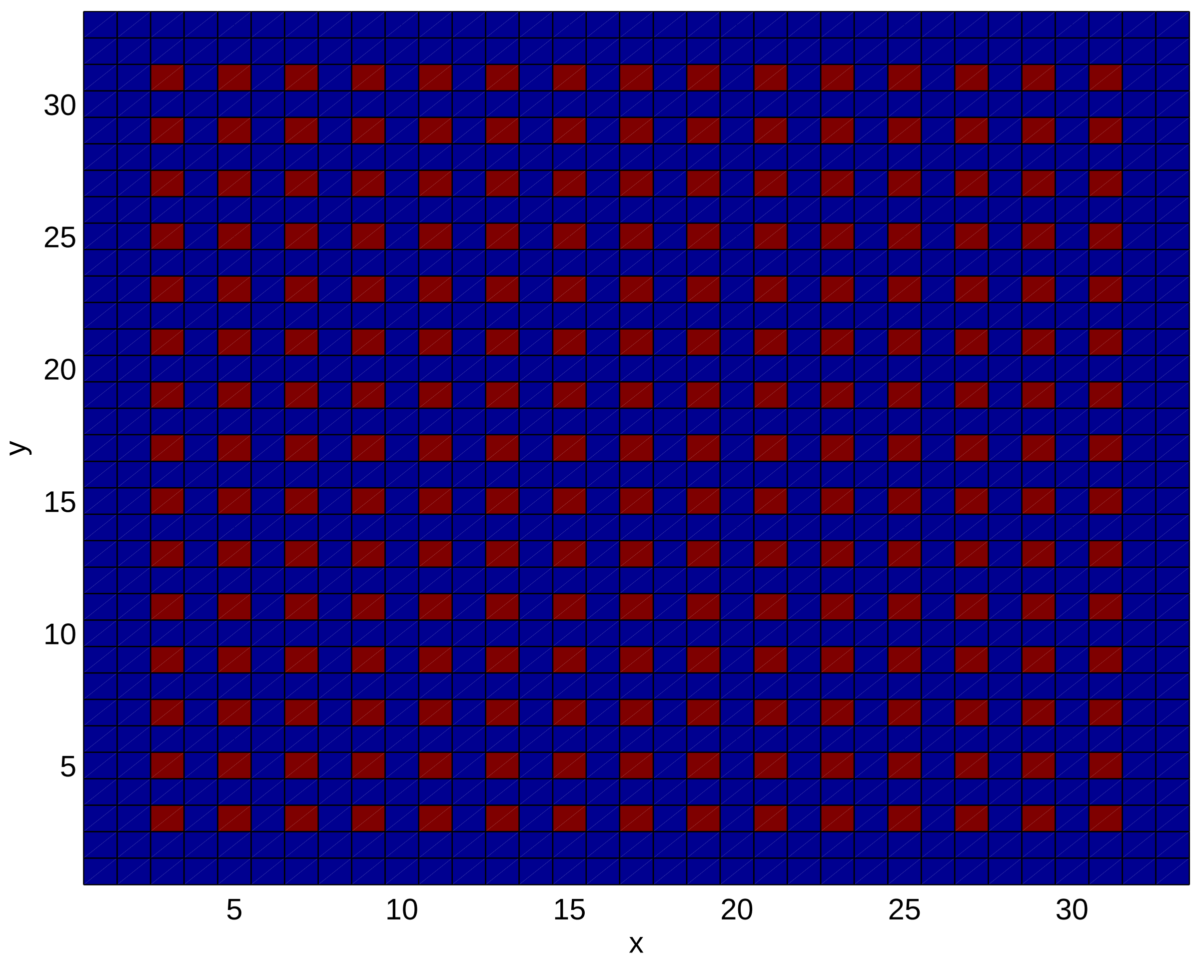
- For point i, $\{S_{ij}\}$ are now measures of strengths of connection
- We now say i strongly depends on j if $(A)_{ij} \neq 0$ and

$$S_{ij} - 1 \geq \theta \max_{k \neq i} \{S_{ik} - 1\}$$

- For now, $\theta = 0.25$ seems to work fine
- Coarse grid selection now accomplished by taking a maximal independent subset of the graph of strong connections

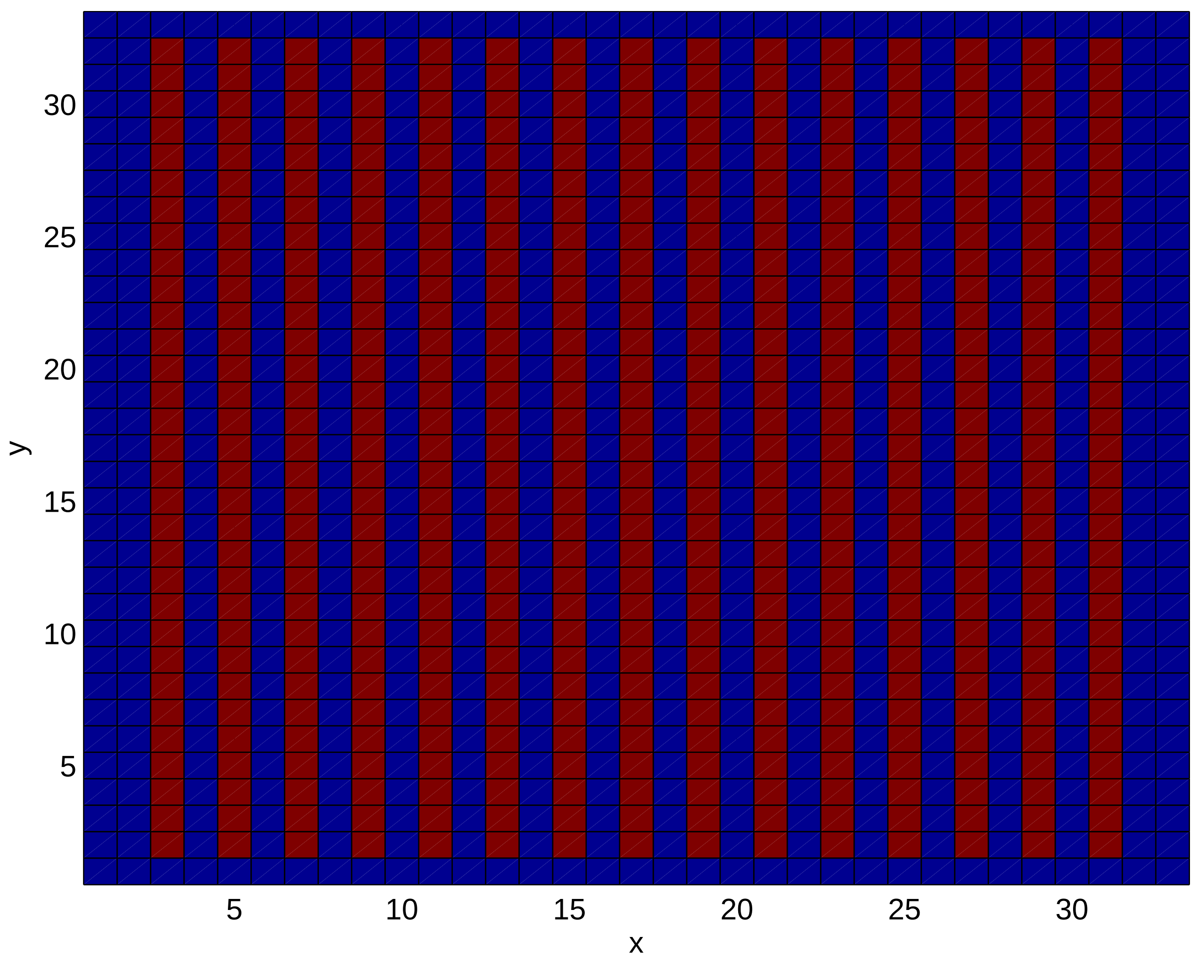
Choices of coarse grids

- $-u_{xx} - u_{yy} = f$, Dirichlet BCs
- 32×32 bilinear finite element grid
- 2 Steps Weighted Jacobi to determine S_i



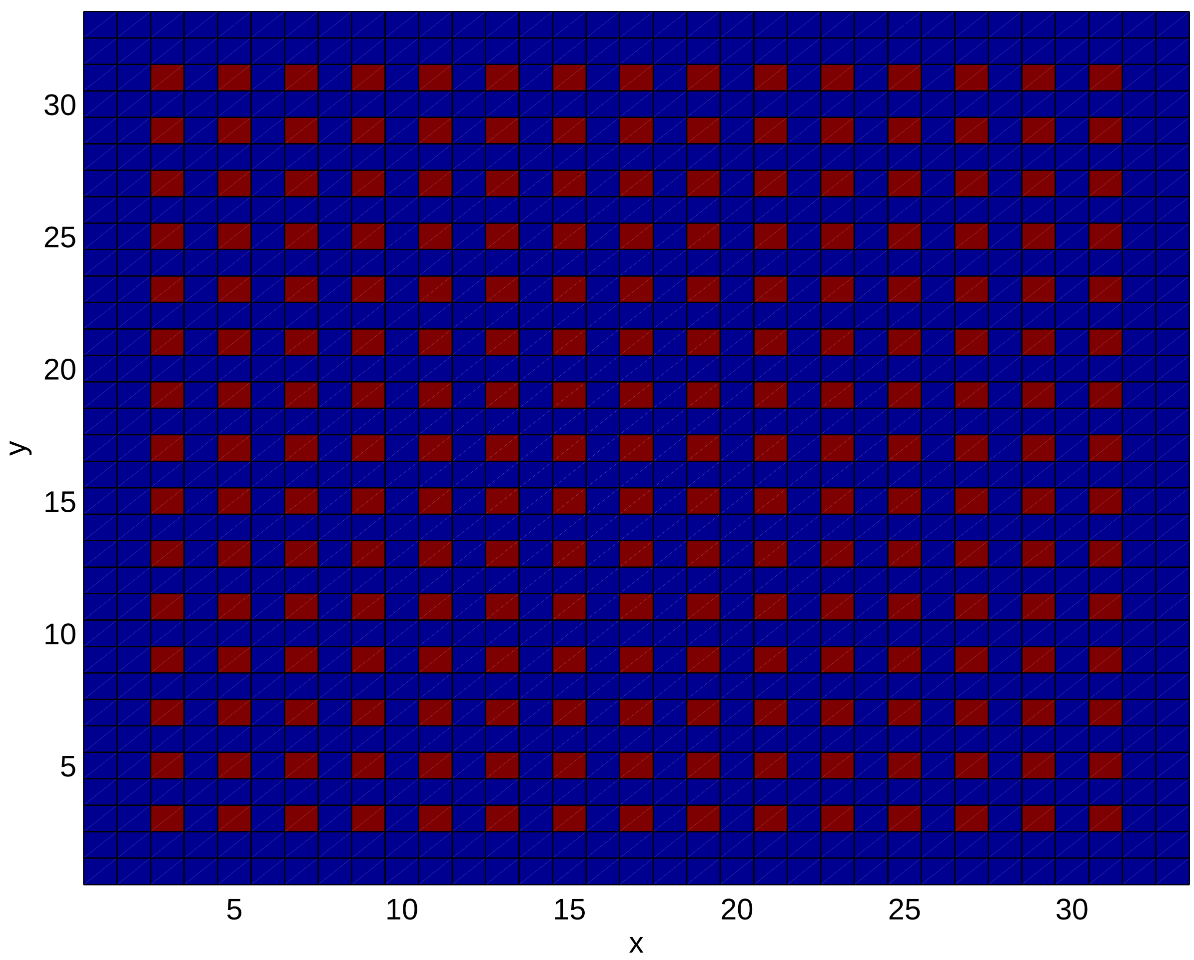
Choices of coarse grids

- $-u_{xx} - 0.01u_{yy} = f$, Dirichlet BCs
- 32×32 bilinear finite element grid
- 2 Steps Weighted Jacobi to determine S_i



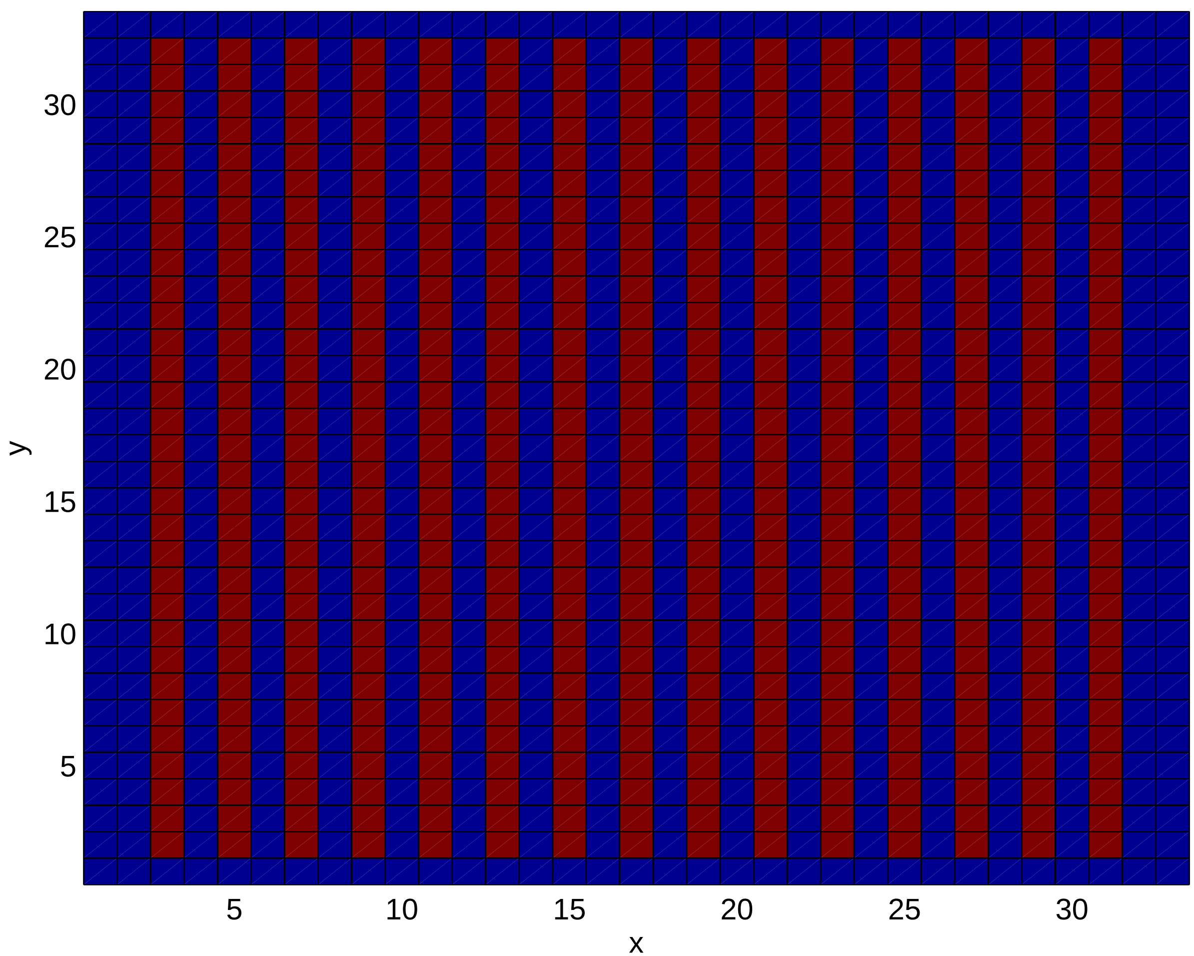
Choices of coarse grids

- $-u_{xx} - u_{yy} = f$, Dirichlet BCs
- 32×32 bilinear finite element grid, $A \rightarrow DAD$, $d_{ii} = 10^{5r_i}$
- 2 Steps Weighted Jacobi to determine S_i



Choices of coarse grids

- $-u_{xx} - 0.01u_{yy} = f$, Dirichlet BCs
- 32×32 bilinear finite element grid, $A \rightarrow DAD$, $d_{ii} = 10^{5r_i}$
- 2 Steps Weighted Jacobi to determine S_i



Classical AMG Interpolation

- Once a coarse grid has been chosen, want to interpolate from strongly connected coarse-grid neighbours, $C_i = S_i \cap C$
- Interpolation must be accurate for algebraically smooth error components, so consider $(A\mathbf{e})_i \approx 0$:

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

- Algebraically smooth error characterised by

H3: Algebraically smooth error varies slowly in the direction of strong connections

Weaknesses

- Assumption on algebraically smooth error based on “nice” M-matrix properties
- Breaks down if near null space of A is far from the constant
  - Diagonal rescaling,

$$A \to DAD$$

  - Finite element anisotropy,

$$-u_{xx} - \epsilon u_{yy} \to \frac{1}{6}\begin{bmatrix} (-1-\epsilon) & (2-4\epsilon) & (-1-\epsilon) \\ (-4+2\epsilon) & (8+8\epsilon) & (-4+2\epsilon) \\ (-1-\epsilon) & (2-4\epsilon) & (-1-\epsilon) \end{bmatrix}$$

- Even for simple problems,
  - Algebraically smooth error difficult to categorise
  - Strong connections difficult to identify

What is Algebraically Smooth Error?

- By definition, any error not efficiently reduced by relaxation
- Easily exposed by relaxation on homogeneous problem, $A\mathbf{x} = \mathbf{0}$, with a random initial guess
- Use this error to characterise variation in general algebraically smooth errors along strong connections

$$e_i = -\sum_{j \in C_i} \frac{a_{ij} + \displaystyle\sum_{k \notin C_i \cup \{i\}} a_{ik} \left(\frac{a_{kj} x_k}{\displaystyle\sum_{j' \in C_i} a_{kj'} x_{j'}} \right)}{a_{ii}} e_j$$

- Need only a local approximation of the variation in algebraically smooth error
- In practice, relax only enough to expose local character, then form interpolation and restrict problem to coarse grid

Algorithm

- Given A,$\mathbf{f}$
- Relax ν_0 times on $A\mathbf{x} = \mathbf{0}$ with a random initial guess
- On each level
  - Determine local strong connections by μ relaxations on $A\mathbf{G}^{(i)} = \mathbf{I}^{(i)}$
  - Choose coarse grid by colouring algorithm
  - Relax ν_1 times on $A\mathbf{x} = \mathbf{0}$ to improve representation of algebraically smooth error
  - Form interpolation, P, based on $\mathbf{x}$
  - Compute $A_c = P^T AP$, inject $\mathbf{x}_c = (\mathbf{x})_c$
- Examples have fixed $\nu_0 = \nu_1 = 15$, $\mu = 2$

Numerical Results

Convergence Factors of Resulting V(1,1) Cycles

grid	Laplace	Scaled Laplace	Anisotropic	Scaled Anisotropic
32×32	0.06	0.06	0.10	0.10
64×64	0.07	0.07	0.10	0.10
128×128	0.07	0.07	0.10	0.10
256×256	0.07	0.07	0.10	0.10
512×512	0.07	0.07	0.10	0.10

Current Research

- Integrated implementation very new
- Computing the strength measures is very computationally intensive
  - Main goal is to improve robustness:

    Get it right, then make it efficient

  - Significant structure to approximations computed for $\{\mathbf{G}^{(i)}\}$, must take advantage of it
- Integration of information computed for coarse-grid selection
  - Computing local low-energy components
  - Computing an approximate inverse for A
- Rigorous testing of algorithm and its parameters
  - What are the "right" choices for θ, μ, ν_0, ν_1?
  - What new problems can we solve?

Summary

- Classical AMG algorithms rely on M-matrix assumptions
- These assumptions can be effectively replaced by probing performance of relaxation
- Algebraic measure of strength of connection
- Relaxation-induced definition of interpolation
- Current work: fully study efficiencies and cost implications
- Future work: develop more efficient AMG algorithms for systems of PDEs