Geometric Multigrid Methods for the Helmholtz equations

Ira Livshits\textsuperscript{1}

\textsuperscript{1}Ball State University

RICAM, Linz, 14, 16 November 2011
Multigrid Methods

- Aim: To understand geometric multigrid for the Helmholtz equations (difficulties and some solutions)
- Give a deeper understanding that led to the various choices made.
- Contents
  - Basics of multigrid
  - Basics of Local Fourier analysis
  - Nonlinear multigrid (FAS)
  - The Helmholtz equation
Helmholtz equation

\[ Lu = -\Delta u(x) - k^2 u(x), \quad x \in \Omega \subset \mathcal{R}^d \]

with large wave numbers \( k: \frac{2\pi}{k} \ll \text{dim}(\Omega) \)
The indefinite Helmholtz equation is

\[ Lu = (-\Delta - k^2)u = f, \]

in contrast to

\[ (\Delta - \eta)u = f, \eta > 0 \]

often also called Helmholtz equation, or Helmholtz equation with the good sign, or positive definite Helmholtz equation.

The subject today is the indefinite Helmholtz equation.
The 2D discrete Poisson equation in the unit square

- We consider a stationary diffusion equation with diffusion in two coordinate directions:

\[- \frac{\partial^2 u(x, y)}{\partial x^2} - \frac{\partial^2 u(x, y)}{\partial y^2} = f(x, y) \quad \text{for} \quad (x, y) \in \Omega = (0, 1)^2\]

with boundary conditions:

\[u(x, y) = g(x, y) \quad \text{for} \quad (x, y) \in \partial \Omega\]

- The functions \(f \in C(\overline{\Omega})\) and \(g \in C(\partial \Omega)\) are known.
- The “Laplace Operator” is commonly written as:

\[\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\]
With mesh size $h = 1/n$ a grid $G$ is defined as

$$\Omega = \{(x_i, y_i) | (x_i, y_j) = (ih, jh) \subset \mathbb{R}^2; \ i, j, = 0, 1, \ldots, n\}$$

In analogy to the 1D case we approximate the partial derivatives in $y$-direction. For the second derivative in $y$-direction we obtain:

$$\frac{\partial^2 u}{\partial y^2}_{i,j} \approx \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h_y^2}$$

The accuracy in $x$- and $y$-direction is $O(h_x^2)$, $O(h_y^2)$ or, with $h_x = h_y = h$, $O(h^2)$. 
Gauss-Seidel and Jacobi for the Poisson equation

- The discrete Laplace operator reads:

\[
\Delta_h u_h = \frac{1}{h^2} [u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1} - 4u_{i,j}] \quad \text{or:}
\]

\[
\Delta_h u_h(x, y) = \frac{1}{h^2} [u_h(x - h, y) + u_h(x + h, y) + u_h(x, y - h) + u_h(x, y + h) - 4u_h(x, y)]
\]

- Jacobi iteration reads

\[
u_{h}^{m+1}(x_i, y_j) = \frac{1}{4} \left[ h^2 f_h(x_i, y_j) + u_h^{m}(x_i - h, y_j) + u_h^{m}(x_i + h, y_j) + u_h^{m}(x_i, y_j - h) + u_h^{m}(x_i, y_j + h) \right],
\]

(with \((x_i, y_j) \in \Omega_h\))

- Gauss-Seidel iteration reads

\[
u_{h}^{m+1}(x_i, y_j) = \frac{1}{4} \left[ h^2 f_h(x_i, y_j) + u_h^{m+1}(x_i - h, y_j) + u_h^{m}(x_i + h, y_j) + u_h^{m+1}(x_i, y_j - h) + u_h^{m}(x_i, y_j + h) \right],
\]
Linear systems of equations in matrix form

- For the solution of a linear system of equations $Au = b$,
- Iterative method:

$$u^{m+1} = Qu^m + s, \ (m = 0, 1, 2, \ldots),$$

so that the system $u = Qu + s$ is equivalent to the original problem.
- Matrix $Q$ is usually not computed explicitly!
From the matrix equation \( Au = b \) we construct a splitting \( A = B + (A - B) \), so that \( B \) is “easily invertible”.

With the help of a general nonsingular \( N \times N \) matrix \( B \) we obtain such iteration procedures from the equation

\[
Bu + (A - B)u = b.
\]

If we put

\[
Bu^{m+1} + (A - B)u^m = b,
\]

or, rewritten for \( u^{m+1} \);

\[
u^{m+1} = u^m - B^{-1}(Au^m - b) = (I - B^{-1}A)u^m + B^{-1}b.
\]

We will see, that it is important, to have the eigenvalues of \( Q = I - B^{-1}A \) possibly small. This is the more likely, the better \( B \) resembles \( A \).
Definition: An iterative process \( u^{m+1} = Qu^m + s \) is called consistent, if matrix \( I - Q \) is not singular and if \( (I - Q)^{-1} s = A^{-1} b \).

If the iteration is consistent, the equation \( (I - Q)u = s \) has exactly one solution \( u^{m \to \infty} \equiv u \).

Definition: An iteration is called convergent, if the sequence \( u^0, u^1, u^2, \ldots \) converges to a limit, independent of \( u^0 \).
The spectral radius

- **Definition**: The spectral radius \( \rho(Q) \) of a \((N, N)\) matrix \( Q \) is

\[
\rho(Q) := \max \left\{ |\lambda| : \lambda \text{ eigenvalue of } Q \right\}
\]

- **Lemma**: The spectral radius is the infimum over all (vector norm related) matrix norms:

\[
\rho(Q) = \inf \left\{ \| Q \| \right\}
\]

- **Theorem**: The iterative method \( u^{m+1} = Q u^m + s \ m = 0, 1, 2, \ldots \) yields a convergent sequence \( \{u^m\} \) for a general \( u^0, s \) (“the method converges”), if

\[
\rho(Q) < 1.
\]

- In the case \( \rho(Q) < 1 \), matrix \( I - Q \) is regular, there is for each \( s \) exactly one fix point of the iterative scheme.
Error reduction

- General splitting of $A$ (another way of writing):

$$Bu^{m+1} + (A - B)u^m = b = Au^* = Bu^* + (A - B)u^*$$

- or:

$$B(u^{m+1} - u^*) = (B - A)(u^m - u^*),$$

- So,

$$(u^{m+1} - u^*) = Q(u^m - u^*).$$

- By induction, we can obtain:

$$(u^m - u^*) = Q^m(u^0 - u^*).$$

- Assume $\rho(Q) < 1$. 

Ira Livshits (BSU) - November 2011, Linz
Efficient solution methods for discrete equations – Hierarchy

- One level methods, like Jacobi, Gauss-Seidel iterative methods are far from optimal: $O(N^\alpha)$, $\alpha > 1$.
- $O(N)$ complexity requires hierarchical approaches.
- **Multigrid** is the classical hierarchical approach
  - Approximation on different scales
  - Compute only those “parts” of a solution on fine scales that really require the fine resolution.
  - Compute/correct the remaining parts on coarser scales
  - Gives $O(N)$ complexity
- Advantages increase with increasing problem size
- The idea behind multigrid is very general
Relaxation of $\Delta u = 0$

(a) Random initial error, (b) Error after 10 GS iterations, (c) Error after 10 more GS iterations.
Convergence factor vs smoothing factor

- **Convergence factor** (GS, Jacobi, wJacobi, SOR) is $\rho = 1 - O(h^2)$

- **Smoothing factor** is a convergence factor for oscillatory components, and it seems to be good.

- Remaining error components are smooth.
Motivation to use coarse grids

The remaining after relaxation, aka smoothing, error has a good coarse grid representation.
Detour to local Fourier analysis
with a hint of trouble brewing for the Helmholtz

- The basic idea of any multigrid algorithm is to annihilate the high frequency error components efficiently on a fine grid by relaxation (smoothing) while the low frequencies are approximated from coarser grids.

- Smoothing analysis is used to analyze quantitatively this reduction of the high frequency error components on the fine grid.

- Assuming the ideal situation in which
  - the relaxation does not affect the low frequencies
  - all low frequencies are approximated more or less exactly on the coarse grid
  - there is no interaction between high and low frequency error components

it already leads to an “estimate” or prediction of the 2–level convergence factor.
Assumptions

For a discretized PDE:

\[ L_h U_h = F_h \]

With the basic assumptions of any local Fourier analysis:

- linearize operator \( L_h \) and/or freeze coefficients to local values
- neglect boundaries and boundary conditions
- consider the frozen, linearized operator on an infinite grid \( G_h \)
- apply the linear constant coefficient local Fourier analysis
Components, the 1D case

- Infinite uniform mesh, meshsize $h$
  \[ G_h = \{ k \cdot h : k \in \mathbb{Z} \} \]

- 1D Fourier components: $e^{i\omega x} = \cos \omega x + i \sin \omega x$

- On $G_h$ only Fourier components $e^{i\theta x/h}$ with $\theta \in (-\pi, \pi]$ are “visible”.

- **Definition:** A component $e^{i\theta x/h}$ is “visible” on $G_h$ if there is no frequency $\theta_0$ with $|\theta_0| < |\theta|$ such that
  \[ e^{i\theta_0 x/h} = e^{i\theta x/h} \text{ for all } x \in G_h. \]

- A visible component $e^{i\theta x/h}$ does not coincide with any lower frequency on grid $G_h$.
- $\theta$ “frequency”, $|\frac{2\pi h}{\theta}|$ “period”
Components, the 2D case

- Infinite uniform mesh, meshsize $h = (h_x, h_y)$ in x- and y- direction:

$$G_h = \{(kh_x, lh_y) : k, l \in \mathbb{Z}\}$$

- $\theta = (\theta_1, \theta_2)$ “frequency”

- On $G_h$, only Fourier components

$$e^{i\theta x / h} := e^{i\theta_1 x / h_x} \cdot e^{i\theta_2 y / h_y}$$

with

$$|\theta| := \max(|\theta_1|, |\theta_2|) \leq \pi$$

are “visible”.

- For higher frequencies $|\theta| > \pi$, the Fourier components coincide with a lower frequency $\in (-\pi, \pi]$.

- For higher dimensions: analogously
Difference operators

- Let \( G_h \) be an infinite uniform mesh, meshsize \( h \);
- \( u_h \): grid function on \( G_h \);
- \( I = \{-\nu, \ldots, 0, \ldots, \nu\} \): set of indices.

Difference operators \( L_h \):

2D \( \mathbf{x} = (x, y) \in G_h \)

\[
L_h u_h(\mathbf{x}) = \sum_{\mu_1 \in I} \sum_{\mu_2 \in I} a_{\mu_1 \mu_2} u_h(\mathbf{x} + \mu_1 h_x, y + \mu_2 h_y)
\]
The stencil notation, examples

2D finite differences for $\Delta u$, \( l = \{-1, 0, 1\} \)

\[
L_h u_h(x) = \begin{bmatrix}
a_{-11} & a_{01} & a_{11} \\
a_{-10} & a_{00} & a_{10} \\
a_{-1-1} & a_{0-1} & a_{1-1}
\end{bmatrix} \begin{bmatrix}
u_h(x)
\end{bmatrix}
\]

\[
= \frac{1}{h^2} \begin{bmatrix}
1 & -4 & 1 \\
1 & 4 & 1
\end{bmatrix} u_h(x) \quad \text{5-point stencil}
\]

or for the biharmonic operator $\Delta^2$, \( l = \{-2, -1, 0, 1, 2\} \)

\[
\Delta_h \Delta_h u_h(x) = \frac{1}{h^4} \begin{bmatrix}
2 & 1 & 2 \\
1 & -8 & 20 & -8 & 1 \\
2 & -8 & 2 \\
1
\end{bmatrix} u_h(x)
\]

13-point stencil

Here: \( x = (x, y) \in G_h, \ h_x = h_y = h \).
Motivation to consider Biharmonic equation is Kaczmarz relaxation

Kaczmarz relaxation (equivalent to GS applied to $AA^*$)
Implementation: Consider $i^{th}$ equation: $A_i x = b_i$ and a current approximation $\tilde{x}$, here $A_i$ is the $i^{th}$ row of $A$.
A new approximation is given by

$$\tilde{x} \leftarrow \tilde{x} + r_i A_i,$$

where the normalized residual is computed as $r_i = \frac{b_i - A_i \tilde{x}}{A_i^* A_i}$.

Kaczmarz relaxation always converges (indeed, $AA^t$ is always SPD) but it is often slow (much slower than the GS on original matrix) and thus should be used only as the last reserve ($\mu = 0.8$ for Laplace vs GS’s $\mu = 0.25$).
Difference operators applied to Fourier components

- On an infinite grid $G_h$, the Fourier components $e^{i\theta x/h}$ are eigenfunctions of any scalar difference operator $L_h$ with constant coefficients.

**1D:**

$$L_h e^{i\theta x/h} = \left( \sum_{\mu \in I} a_{\mu} e^{i\theta \mu} \right) \cdot e^{i\theta x/h}$$

$\tilde{L}_h(\theta)$ is the eigenvalue, also called the Fourier symbol of $L_h$.

**2D:**

$$L_h e^{i\theta x/h} = \left( \sum_{\mu_1 \in I} \sum_{\mu_2 \in I} a_{\mu_1 \mu_2} e^{i\theta_{1\mu_1}} e^{i\theta_{2\mu_2}} \right) \cdot e^{i\theta x/h}$$

$\tilde{L}_h(\theta)$ is the eigenvalue, also called the Fourier symbol of $L_h$. 
Examples

1D: \[ L_h = \frac{1}{h^2} \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} \]

\[ \implies L_h e^{i\theta x/h} = \frac{1}{h^2} (e^{-i\theta} - 2 + e^{i\theta}) \cdot e^{i\theta x/h} = \frac{2}{h^2} (\cos \theta - 1) \cdot e^{i\theta x/h} \]

2D: \[ \theta = (\theta_1, \theta_2) \]

\[ L_h = \Delta_h = \frac{1}{h^2} \begin{bmatrix} 1 & -4 & 1 \\ 1 & 1 & 1 \end{bmatrix} \]

\[ \implies \tilde{L}_h(\theta) = \frac{2}{h^2} (\cos \theta_1 + \cos \theta_2 - 2) \]

or for the biharmonic operator \( L_h = \Delta_h \Delta_h \) (13-point stencil):

\[ \tilde{L}_h(\theta) = \frac{4}{h^4} (\cos \theta_1 + \cos \theta_2 - 2)^2 \]
Smoothing analysis; a general class of relaxations:

- Consider a linear problem

\[ L_h U_h(x) = F_h \text{ on grid } G_h \]

- Then: Many (not all !) relaxation schemes for this problem are based on an operator splitting

\[ L_h = A_h + B_h \]

where \( A_h \) and \( B_h \) are again difference operators.

- Then: A relaxation sweep starting from the initial approximation \( u_h \) ("old values") produces a new approximation \( \bar{u}_h \) ("new values") by solving

\[ A_h u_h(x) + B_h \bar{u}_h(x) = F_h(x) \]

at each gridpoint \( x \in G_h \).
Smoothing analysis; a general class of relaxations:

- What does that mean for the errors before and after the relaxation sweep?
- Let $e_h = U_h - u_h$ be the error before and $\bar{e}_h = U_h - \bar{u}_h$ after the sweep.

Then:

$$A_h e_h(x) + B_h \bar{e}_h(x) = 0$$

at each gridpoint $x \in G_h$.

- How do these relaxations act on Fourier components?
- Let $e_h = A e^{i\theta x/h}$: error before relaxation; then $\bar{e}_h(x) = \bar{A} e^{i\theta x/h}$ error after relaxation.

$A, \bar{A} \in \mathbb{R}$ are the error amplitudes before and after relaxation.
Smoothing analysis; a general class of relaxations:

From

\[ A_h e_h(x) + B_h \bar{e}_h(x) = 0 \]

follows

\[ \tilde{A}_h(\theta)A + \tilde{B}_h(\theta)\bar{A} = 0 \]

and therefore

\[ \bar{A} = - \left( \frac{\tilde{A}_h(\theta)}{\tilde{B}_h(\theta)} \right) A \]

where \( \tilde{A}_h(\theta), \tilde{B}_h(\theta) \in \mathbb{C} \) are the Fourier symbols of \( A_h, B_h \).

\[ \mu(\theta) := \left| \frac{\tilde{A}_h(\theta)}{\tilde{B}_h(\theta)} \right| \]

is the amplification factor of the component \( \theta \).
1D Example: Gauss-Seidel relaxation

- Corresponding splitting of $L_h$:

$$
\frac{1}{h^2} \begin{bmatrix} 1 & -2 & 1 \\ \end{bmatrix} = \frac{1}{h^2} \begin{bmatrix} 0 & 0 & 1 \\ \end{bmatrix} + \frac{1}{h^2} \begin{bmatrix} 1 & -2 & 0 \\ \end{bmatrix}
$$

- Then: at gridpoint $x \in G_h$ for the errors

$$
\overline{A} e^{i\theta(x-h)/h} - 2\overline{A} e^{i\theta x/h} + \overline{A} e^{i\theta(x+h)/h} = 0
$$

$$
\left( e^{-i\theta} - 2 \right) \overline{B}_h(\theta) + \left( e^{i\theta} \right) \overline{A}_h(\theta) = 0.
$$

- This yields:

$$
\overline{A} = \frac{-e^{i\theta}}{e^{-i\theta} - 2} \overline{A}
$$
Questions

- What is smoothing?
- How can the smoothing property of a given relaxation scheme be measured quantitatively?
- How does the smoothing property (or its quantitative measure) influence the performance of multigrid cycling?
- Can the performance of a multigrid cycle be predicted in terms of the smoothing measures?

Note: In a multigrid cycle, relaxations are used to “smooth” the highly oscillating error components which cannot be approximated on coarser grids.

Therefore, smoothing roughly means “convergence of high frequency error components”.

Ira Livshits (BSU)
November 2011, Linz
What are high and low frequencies?

- Let $G_h$ be a fine grid and $G_H$ a coarse grid.
- Then: A Fourier component $e^{i\theta x/h}$ on grid $G_h$ is called a high frequency (with respect to $G_H$) if its restriction (injection) to the coarse grid $G_H$ is not “visible” there. Otherwise it is called a low frequency.
- Take one-dimensional standard coarsening:

$$G_h = \{kh : k \in \mathbb{Z}\}, \quad G_H := G_{2h} = \{2kh : k \in \mathbb{Z}\}$$

- Let $|\theta| \leq \pi$; then by injection from $G_h$ to $G_{2h}$:

$$e^{i\theta x/h} \overset{\text{inject}}{\rightarrow} e^{i2\theta x/2h}$$

- This component $e^{i2\theta x/2h}$ is “visible” on $G_{2h}$ only if

$$|2\theta| \leq \pi, \quad \text{i.e.} \quad |\theta| \leq \frac{\pi}{2}$$
The high frequencies on $G_h$ (with respect to $G_{2h}$) are those with
$\pi \geq |\theta| \geq \frac{\pi}{2}$.

Fourier components in 2D: $e^{i\theta x/h} = e^{i\theta_1 x/h} e^{i\theta_2 y/h}$,
$\theta = (\theta_1, \theta_2)$ with
$|\theta| := \max(|\theta_1|, |\theta_2|) \leq \pi$

Standard coarsening

$G_h = \{(kh, lh) : k, l \in \mathbb{Z}\}$

$G_H = G_{2h} = \{(2kh, 2lh) : k, l \in \mathbb{Z}\}$

High frequencies: $\frac{\pi}{2} \leq |\theta| \leq \pi$
Low frequencies: $|\theta| < \frac{\pi}{2}$

Note: The definition of high and low frequencies depends on the coarsening.
“Smoothing” stands for convergence of high frequency error components which cannot be approximated from the coarse grids in a multigrid cycle.

For a relaxation scheme with amplification factors $\mu(\theta)$ of Fourier components $e^{i\theta x/h}$, the smoothing rate $\mu$ is defined by:

$$\mu := \max\{|\mu(\theta)| : \theta = \text{high frequencies}\}$$

Note:  
- The above definition of $\mu$ depends on the coarsening.  
- The amplification or reduction factors $\mu(\theta)$ are also called “smoothing factors.”
Example: Gauss–Seidel relaxation for Poisson’s equation:

\[
\Delta_h U_h = \frac{1}{h^2} \begin{bmatrix}
1 & 1 & 1 \\
1 & -4 & 1 \\
\end{bmatrix} U_h = F_h
\]

- Relaxation in error terms:

\[
\frac{1}{h^2} \begin{bmatrix}
1 & 0 & 0 \\
1 & -4 & 0 \\
\end{bmatrix} \bar{e}_h(x) + \frac{1}{h^2} \begin{bmatrix}
0 & 1 & 1 \\
0 & 0 & 1 \\
\end{bmatrix} e_h(x) = 0
\]

- Symbols:

\[\theta = (\theta_1, \theta_2)\]

\[
B_h e^{i\theta x/h} = \frac{1}{h^2} \left( e^{-i\theta_1} + e^{-i\theta_2} - 4 \right) \cdot e^{i\theta x/h}
\]

\[\tilde{B}_h(\theta)\]

\[
A_h e^{i\theta x/h} = \frac{1}{h^2} \left( e^{i\theta_1} + e^{i\theta_2} \right) \cdot e^{i\theta x/h}
\]

\[\tilde{A}_h(\theta)\]
Amplification factors:

$$\mu(\theta) = -\frac{\tilde{A}_h(\theta)}{\tilde{B}_h(\theta)} = -\frac{e^{i\theta_1} + e^{i\theta_2}}{e^{-i\theta_1} + e^{-i\theta_2} - 4}$$

Smoothing rate:

$$\mu = \max\{|\mu(\theta)| : \frac{\pi}{2} \leq |\theta| \leq \pi\} \approx 0.5$$

The high frequency error components are reduced by a factor of (at least) $\mu$ per relaxation sweep.

Note: For low frequencies, the reduction per relaxation sweep is much worse:

$$|\mu(\theta)| \rightarrow 1 \quad \text{if} \quad \theta \rightarrow 0$$
Now to Helmholtz: the stencils and the relaxations

1D finite differences for $Lu = u'' + k^2 u$, $L_h = \frac{1}{h^2} [1 - 2 + (kh)^2 \ 1]$

2D finite differences for $Lu = \Delta u + k^2 u$, $L_h = \frac{1}{h^2} \begin{bmatrix} 1 & 1 \\ 1 & -4 + (kh)^2 & 1 \\ 1 & 1 \end{bmatrix}$ 5-point stencil

Here: $x = (x, y) \in G_h$, $h_x = h_y = h$. 

Ira Livshits (BSU) - November 2011, Linz 36 / 83
Fourier symbols

1D: \[ L_h = \frac{1}{\hbar^2} \begin{bmatrix} 1 & -2 & + k^2 \hbar^2 & 1 \end{bmatrix} \]

\[ \implies L_h e^{i\theta x/\hbar} = \frac{1}{\hbar^2} (2 \cos \theta - 2 + k^2 \hbar^2) \cdot e^{i\theta x/\hbar} \]

\[ \tilde{L}_h(\theta) \]

2D: \[ \theta = (\theta_1, \theta_2) \]

\[ \implies L_h e^{i\theta x/\hbar} = \frac{1}{\hbar^2} (2 \cos \theta_1 + 2 \cos(\theta_2) - 4 + k^2 \hbar^2) \cdot e^{i\theta x/\hbar} \]

\[ \tilde{L}_h(\theta) \]

or for the biharmonic operator (think Kaczmarz!)

\[ \tilde{L}_h(\theta) = \frac{1}{\hbar^4} (2 \cos \theta_1 + 2 \cos \theta_2 - 4 + k^2 \hbar^2)^2 \]
1D Example:

Gauss-Seidel relaxation of

$$\frac{1}{h^2} \begin{bmatrix} 1 & -2 + k^2 h^2 & 1 \end{bmatrix} U_h(x) = F_h(x)$$

Relaxation:

$$u_h \Longrightarrow \bar{u}_h$$

$$\frac{\bar{u}_h(x - h) - (2 - k^2 h^2)\bar{u}_h(x) + u_h(x + h)}{h^2} = F_h(x)$$

- Corresponding splitting of $L_h$:

$$\frac{1}{h^2} \begin{bmatrix} 1 & -2 + k^2 h^2 & 1 \end{bmatrix} = \frac{1}{h^2} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} + \frac{1}{h^2} \begin{bmatrix} 1 & -2 + k^2 h^2 & 0 \end{bmatrix}$$

- Then: at grid point $x \in G_h$ for the errors

$$\bar{A} e^{i\theta(x-h)/h} - (2 - k^2 h^2)\bar{A} e^{i\theta x/h} + A e^{i\theta (x+h)/h} = 0$$
\[
\left( e^{-i\theta} - 2 + k^2 h^2 \right) \tilde{B}_h(\theta) \bar{A} + \left( e^{i\theta} \right) \tilde{A}_h(\theta) A = 0.
\]

This yields:
\[
\bar{A} = \frac{-e^{i\theta}}{e^{-i\theta} - (2 - k^2 h^2)} A
\]

What is the difference compared to Laplace?

- What happens when \( \theta = 0 \)?
- What happens when \( \theta \) is between \( \pi/2 \) and \( \pi \)?
- What happens when \( \theta \approx kh \)?
- Happens when \( kh \) grows, i.e., \( k \) is fixed and \( h \) is coarsened?
Answers

- If $kh \ll 1$ i.e., $\exp(\pm ikx)$ are smooth:
  - other smooth components $\theta \approx 0$ remain almost unchanged by relaxation;
  - For $|\theta| = \pm kh$ there is a \textit{slight} divergence;
  - Oscillatory components $\pi/2 < |\theta| \leq \pi$ converge almost as fast as in Laplace
    (convergence does depend on $kh$);
- If $kh = O(1)$ divergence worsens!
- Results for $kh = 0.1$ and $kh = 0.5$
More examples: 2D Helmholtz equation:

- Gauss-Seidel Relaxation in error terms:

\[
\begin{bmatrix}
\frac{1}{h^2} & 1 & -4 + k^2 h^2 & 0 \\
0 & 1 & 0 & 0 \\
\end{bmatrix}
\bar{e}_h(x) + \begin{bmatrix}
\frac{1}{h^2} & 0 & 1 \\
0 & 0 & 1 \\
\end{bmatrix}
\tilde{e}_h(x) = 0
\]

- Symbols:

\[\theta = (\theta_1, \theta_2)\]

\[B_h e^{i\theta x/h} = \frac{1}{h^2} \left( e^{-i\theta_1} + e^{-i\theta_2} - 4 + k^2 h^2 \right) \cdot e^{i\theta x/h}
\]

\[\tilde{B}_h(\theta)\]

\[A_h e^{i\theta x/h} = \frac{1}{h^2} \left( e^{i\theta_1} + e^{i\theta_2} \right) \cdot e^{i\theta x/h}
\]

\[\tilde{A}_h(\theta)\]
Amplification factors:

\[ \mu(\theta) = -\frac{\tilde{A}_h(\theta)}{\tilde{B}_h(\theta)} = -\frac{e^{i\theta_1} + e^{i\theta_2}}{e^{-i\theta_1} + e^{-i\theta_2} - 4 + k^2 h^2} \]

Smoothing rate (what is smooth here?)�:

- \( \mu \approx \mu_L \) for \( kh \ll 1 \);
- \( \mu \) deteriorates (grows) as \( kh \) grows;
Questions

- What is smooth? Is $\theta = 0$ smooth?
- What is the smoothing factor?
- What do we want from relaxation?
Kaczmarz relaxation $c = -4 + k^2 h^2$

$$\Delta_h \Delta_h u_h(x) = \frac{1}{h^4} \begin{bmatrix} \frac{1}{h^4} & 1 & 2c & 4 + c^2 & 1 \\ 1 & 2c & 2 & 2c & 2 \\ 0 & 0 & 0 & 2c & 2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} u_h(x)$$

$$\frac{1}{h^4} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 2c & 4 + c^2 & 0 \\ 2 & 2c & 2 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \bar{e}_h(x) + \frac{1}{h^4} \begin{bmatrix} 1 \\ 2 & 2c & 2 \\ 0 & 2c & 0 \\ 0 & 0 & 0 \end{bmatrix} e_h(x) = 0$$
Symbols and amplification

Symbol \( B_h e^{i\theta x/h} = \)
\[
\frac{1}{h^4} \left( (e^{-2i\theta_1} + e^{-2* i\theta_2}) + 2(e^{-i\theta_1-i\theta_2} + e^{-i\theta_1+i\theta_2}) + 2c(e^{-i\theta_1} + e^{-i\theta_2}) + (4 + c^2) \right) \cdot e^{i\theta x/h}
\]
\[\tilde{B}_h(\theta)\]

Symbol \( A_h e^{i\theta x/h} = \)
\[
\frac{1}{h^4} \left( (e^{2i\theta_1} + e^{2* i\theta_2}) + 2(e^{i\theta_1-i\theta_2} + e^{i\theta_1+i\theta_2}) + 2c(e^{i\theta_1} + e^{i\theta_2}) \right) \cdot e^{i\theta x/h}
\]
\[\tilde{A}_h(\theta)\]
Good news: $\mu(\theta) \leq 1$ for any $\theta$

Q: But for which is works the best?

Results for $kh = 1$ (high frequencies and overall convergence).
More on convergence

Consider ordered eigen pairs of $A$: $(\lambda_j, v_j): Av_j = \lambda_j v_j$ such that $|\lambda_1| \leq |\lambda_2| \cdots \leq |\lambda_N|$.

For errors representation $e^{m+1} = \sum \alpha^{m+1} v_n$ and $e^m = \sum \alpha^m v_n$, Due to linearity (of everything),

$$\sum \alpha^{m+1} v_n = \sum \alpha^m Qv_n$$

and thus overall convergence is defined by

$$\max_n \left| \frac{\alpha^{m+1}}{\alpha^m} \right|$$

- The smoothing factor is a convergence factor for $\alpha_n/2, \ldots, \alpha_n$, i.e., for the high energy eigenfunctions $v_{n/2}, \ldots, v_n$
- Remaining non reduced low energy eigenfunctions $v_1, \ldots, v_{n/2-1}$ are smooth.
Reason for using coarse grids

Coarse grids can be used to compute an improved initial guess for the fine-grid relaxation. This is advantageous because:

- Relaxation on the coarse-grid is much cheaper ($1/2$ as many points in $1D$, $1/4$ in $2D$, $1/8$ in $3D$);
- Relaxation on the coarse grid has a marginally better convergence rate, for example $1 - O(4h^2)$ instead of $1 - O(h^2)$;
- Smooth eigenfunctions on the finest grid become more oscillatory on the coarse grid.
Multigrid components

- Relaxation (smoother): efficiently dumps high-energy (eigenfunctions with large eigenvalues) error components; leaves unchanged low eigenmodes;
- A coarse-grid system operator: accurately resolve the low eigenmodes;
- Restriction operator (averaging): transfers fine-grid information (residual) to coarse grid;
- Prolongation operator (interpolation): transfers coarse-grid correction back to the fine grid; has to be accurate for low eigenmodes.
Nice example: \[ \Delta u = f \]

- Relaxation: Gauss-Seidel.
- A coarse-grid operator: discretization of the differential Laplace operator on coarse scale \((2h, 4h, \ldots)\)
- Restriction operator: full weighting of the residual (annihilates remaining oscillatory components), transferring only smooth ones to coarser grids;
- Interpolation: polynomial (linear) which is accurate for smooth components.

Works great because the lowest eigenmodes are physically smooth.
One picture is worth a thousand words

Geometric Multigrid (MG)

Relax on $A^h u^h = f^h$

Restrict residual

Relax on $A^{2h} e^h = r^{2h}$

Interpolate correction
Back to 2d Helmholtz : discretization error

Discrete wave number vs exact wave number:

\[ |k^h| \approx k \left(1 + \frac{k^2 h^2}{\gamma}\right), \quad 24 \leq \gamma \leq 48. \]

The total (accumulated) phase error is

\[ E(kd, kh) = kd \frac{k^2 h^2}{2\pi \gamma}. \]

For the discrete solutions to be accurate approximation to the differential solutions need

\[ E(kd, kh) \ll 1. \]

For fixed \( k, d \) this means small \( h \) and large problem size, \( N \).
Bottlenecks for multigrid

- Dominant phase discretization error;
- Quality of standard coarse-grid and prolongation operators deteriorates on coarse grids;
- Poor multigrid relaxation - standard "fast" relaxation scheme diverge;
Lowest eigenmodes for $k = 6\pi$ and $k = 18\pi$. 
More bottlenecks for multigrid

- Oscillatory low eigenmodes;
- There are many of them;
- They are different from each other;
- On sufficiently coarse grids, there is no single prolongation or coarse grid operator of a reasonable sparsity that works for all such eigenmodes.
One-dimensional Helmholtz equation: slow to converge components, aka lowest eigenmodes

(a) a random initial error; (b)-(d) examples of remaining errors, obtained by applying three multigrid V-cycles to the homogenous Helmholtz equation (periodic boundary conditions)
Assumption on the error

The error unreduced by standard multigrid

\[ e(x) = \sum_{|w| \approx k} \exp(iwx) = \sum_{w \approx k} \alpha_w \exp(iwx) + \sum_{w \approx k} \beta_w \exp(-iwx) \]

where the value of \( w \) is determined by the relaxation history. One can drive it to be

\[ (1 - \gamma_1)k \leq w \leq (1 + \gamma_2)k, \]

where typical are values \( \gamma_j \approx 0.3 \)
Dual representation

If we agreed on the previous page, we can also agree that

\[ e(x) = e_1(x) \exp(ikx) + e_2(x) \exp(-ikx), \]

where \( e_1(x) = \sum_{w \approx k} \alpha_w \exp(i(w - k)x) \) and \( e_2(x) = \sum_{w \approx k} \beta_w \exp(-i(w - k)x) \).

Obviously, envelope functions \( e_1 \) and \( e_2 \) are smooth (compared to \( \exp(\pm ikx) \)). Indeed the highest frequencies are \( |w - k| \approx \gamma j k \)

Remark: Representation is not unique - one has to work to guarantee the statement above.

The idea here is simple: Let’s try approximate smooth \( e_1(x) \) and \( e_2(x) \) instead of oscillatory \( e(x) \) (which we cannot do well anyway) and hope for the best.
Defect correction

If we assume the error in the form

\[ e(x) = e_1(x)\exp(i k x) + e_2(x)\exp(-i k x), \]

then naturally

\[ r(x) = Le(x) = L(e_1(x)\exp(i k x) + e_2(x)\exp(-i k x)) = \]

\[ L(e_1(x)\exp(i k x)) + L(e_2(x)\exp(-i k x)) = \exp(i k x)L_1 e_1(x) + \exp(-i k x)L_2 e_2(x) \]
Two conclusions from the previous page:
Note that we use the differential Helmholtz operators and its kernel components, resulting in differential equations for the envelope functions \( e_1 \) and \( e_2 \), \( L_1 \) and \( L_2 \):

\[
L_1 u = u_{xx} + 2iku_x \quad \text{and} \quad L_2 u = u_{xx} - 2iku_x.
\]

Trick! Operator \( L \) has two kernel components \( \exp(\pm ikx) \) and so do \( L_j \) \((1 \text{ and } \exp(\mp 2ikx))\) But we consider grids where only one (a constant) is visible!

Also, given the form of \( e_j \) and \( L_j \), we can claim that

\[
r(x) = r_1(x)\exp(ikx) + r_2(x)\exp(ikx)
\]

where \( r_j \) are smooth (as smooth as \( e_j \) in fact).
From Waves to Rays

To reduce the *irreducible wave* error $e(x): Le = r$
we approximate two *ray* errors

$$L_1e_1 = r_1 \quad \text{and} \quad L_2e_2 = r_2$$

and then reconstruct back to the *wave* $e(x) = e_1(x)\exp(ikx) + e_2(x)\exp(-ikx)$

Where is the gain:

- To compute the ray functions one should start on the grid with $kH \approx \pi$ (very coarse)
- To accelerate, one may use standard multigrid on each $L_je_j = r_j$ if wishes to go coarser.
Residual separation and why on $kH \approx \pi$

Consider the wave (Helmholtz) residual with $r_1$ and $r_2$ are smooth

$$r(x) = r_1(x)\exp(ikx) + r_2(x)\exp(-ikx).$$

- Clearly, such representation is not unique and some others allow

$$r(x) = \tilde{r}_1(x)\exp(ikx) + \tilde{r}_2(x)\exp(-ikx)$$

with oscillatory $\tilde{r}_j$.

- We want to approximate smooth ray residuals (on coarse grids).
Separation (implementation considered on grids)

- Compute fine-grid wave residual \( r^h = f^h - L^h u^h \) (\( h \) is fine enough for small phase error \( kd(kh)^2 \ll 1 \))
- Transfer it to the grid with \( kH_s \approx \pi/2 \) using full weighting (annihilates physically oscillatory error components), resulting in
  \[
  r^{H_s} = r^{H_s}_1 \exp(ikx) + r^{H_s}_2 \exp(-ikx)
  \]
- To approximate \( r_1 \): Multiply by \( \exp(-ikx) \)
  \[
  \exp(-ikx)r^{H_s} = r^{H_s}_1 + r^{H_s}_2 \exp(-2ikx)
  \]
  Note that first term is smooth and the second one is extremely oscillatory \( 2kH_s \approx \pi \). Applying one more full weighting results in the desired ray residual
  \[
  i^{2H_s}(\exp(-ikx)r^{H_s}) = i^{2H_s}(r^{H_s}_1) + i^{2H_s}(r^{H_s}_2 \exp(-2ikx)) \approx r^{2H_s}_1
Ray operator

The discrete operators for the ray functions are derived from the differential operators:

\[ L_1 u = u'' + 2iku' = f \]

considered on the grid with \( kH \approx \pi \).

The choice of discretization is defined by the principality of terms:

- The second-order term: \( \frac{1}{H^2} \)
- The first-order term: \( \frac{2k}{H} \)
- If \( kH \geq \pi \) the first-order term is principal and should be accurately discretized;
- Thus using staggered grids.

\[ L^H_j = \begin{bmatrix} \frac{1}{H^2} & -\frac{2ik}{H} - \frac{1}{H^2} & \frac{2ik}{H} - \frac{1}{H^2} & \frac{1}{H^2} \end{bmatrix} \]
Wave-ray cycle

- Wave cycle:
  - Reduces all but problematic characteristic components;
  - Produces residual for the ray cycles (on one of the fine grids with small phase error);
  - Finest grid: \((kd)(k^2h^2) \ll 1\)
  - Coarsest grid: \(kH > \pi\) (Used to effectively wipe out constants and alikes, stencil \([1 \ (k^2H^2 - 2) \ 1]\)
  - Correction Scheme;
  - GS one pre- and one post relaxation on fine grid and on the coarsest grid;
  - Kaczmarz four pre- and one post relaxation sweeps everywhere else;
  - Linear interpolation and full weighting;

- Two ray cycles to reduce the characteristic components.
Ray cycles

- Staggered grid, short central differences for $u_x$;
- Marching scheme in the propagation direction;
- CS or FAS multigrid scheme depending on interpretation;
- Natural introduction of boundary conditions in terms of rays.
Before we proceed, detour to FAS

Full Approximation Scheme

- Allows coarse grid approximation of the solution \( u^H \) rather than correction \( e^H \);
- Considers full equation \( L^H u^H = f^H \) residual equation \( L^H e^H = r^h \);
- Developed for solving non linear equations.
FAS details

Consider $\tilde{u}^h$ a current fine-grid solution and $r^h = f^h - L^h \tilde{u}^h$ is its corresponding residual;

- **CS:**
  - Coarse grid variable correction: $v^H$
  - Coarse grid system: $L^H v^H = R^H = I_h^H r^h$
  - Coarse grid correction: $\tilde{u}_{NEW}^h = \tilde{u}^h + I_H^H v^H$

- **FAS:**
  - Coarse grid variable solution: $\tilde{u}^H = I_h^H \tilde{u}^h + v^H$
  - Coarse grid system: $L^H u^H = f^H = L^H(I_h^H \tilde{u}^h) + R^H$
  - Coarse grid correction: $\tilde{u}_{NEW}^h = u^h + I_H^H(\tilde{u}_{NEW}^H - I_h^H \tilde{u}^h)$
  - If $L^H$ resolves all solution components then $\tilde{u}_{NEW}^h = I_H^H(\tilde{u}_{NEW}^H)$
Consider the c.g. equation as

\[ L^H u^H = f^H + \tau_h^H \]

where

\[ f^H = I_h^H f^h, \quad \tau_h^H = L^H(\tilde{I}_h^H u^h) - I_h^H(L^h u^h) \]

Term \( \tau_h^H \) is a fine-to-coarse defect correction.

**Nonlinear equations:**

- The correction equation is \( L^h(\tilde{u}^h + v^h) - L^h\tilde{u}^h = r^h \).
- On coarse grid its reads: \( L^H(\tilde{I}_h^H \tilde{u}^h + v^H) - L^H(\tilde{I}_h^H \tilde{u}^h) = I_h^H r^h \) which is exactly the top equation.
Rays, directions and BC

\[ \exp(ikx) \]

Sommerfeld BC or Interior one-sided eq.

Entrance

Dirichlet (homegenous or not)

Exit
Back to Boundary conditions

Assumptions:

- Away from the source (near the boundary) the solution has a pure ray character so can be described using ray operators;
- The typical boundary conditions are the radiation BC, sometimes combined with the Dirichlet boundary conditions.
- The rays can enter the computational domain from the infinity and leave the computational domain freely unless there is an obstacle, media change, etc.
- The entering part is to be defined by the Dirichlet boundary conditions at the entrance (for each ray); the exiting part is controlled by the Sommerfeld boundary conditions at the exit (the latter can be replaced by an interior equation appropriately discretized).
- If no rays (waves) enter from infinity, the Dirichlet boundary conditions are homogenous.
Full ray formulation (for two-sided Sommerfeld BC)

Exiting Sommerfeld BC: At $x = 0$: $u' + 2iku = 0$ and at $x = d$: $u' - 2iku = 0$

- **Positively** propagating ray (corresponding to $\exp(ikx)$)
  
  $$L_1 a(x) = a_1''(x) + 2ika_1'(x) = f_1(x), \quad a_1(0) = a_{1,0}, \quad a_1'(d) = 0$$

- **Negatively** propagating ray (corresponding to $\exp(-ikx)$)
  
  $$L_2 a(x) = a_2''(x) + 2ika_2'(x) = f_2(x), \quad a_2(d) = a_{2,0}, \quad a_1'(0) = 0$$

- Two systems are solved independently;

- If there is a Dirichlet BC on one side, the two has two be combined (incident wave defines the amplitude of the reflected wave);
Wave-ray interaction

- Because of BC the ray cycles are implemented in FAS (solution representation);
- At the interior, regular FAS correction:
  \[ \tilde{u}^h_{NEW} = \tilde{u}^h + \exp(ikx)(I^h_H(\tilde{a}^H_{1,NEW} - \tilde{a}^H_1)) + \exp(-ikx)(I^h_H(\tilde{a}^H_{2,NEW} - \tilde{a}^H_2)) ; \]
- At the boundary, direct replacement
  \[ \tilde{u}^h_{NEW} = \exp(ikx)(I^h_H \tilde{a}^H_{1,NEW}) + \exp(-ikx)(I^h_H \tilde{a}^H_{2,NEW}) ; \]
Conclusions on 1d

- The solver is a combination of the wave and the ray approaches;
- The complexity is similar to one for Laplace, it is $O(N)$.
- Worked for jumping and slightly varying wave numbers;
Wave-Ray algorithm

Geometric solver for the 2D Helmholtz equation with constant wave numbers

*Brandt, Livshits: Wave-ray multigrid method solving standing wave equations, 1997*

*Livshits, Brandt: Accuracy Properties of the Wave-Ray Multigrid Algorithm for Helmholtz Equations, 2006*

**Special feature:** Separate treatment of oscillatory kernel $e^{ikx}$, $|k| = k$ on coarse grids.
Directions of propagation $e^{ikx}$
Discretization of the propagation directions: few chosen direction $k_\ell$
Wave-ray approach

- Components not changed (at best) by the standard multigrid:

\[ \exp(iwx), \quad |w| \approx k, \quad w, x \in \mathbb{R}^2 \]

- Ray representation for such components \( u(x) \) and residual \( r(x) \)

\[
\begin{align*}
  u(x) &= \sum_{\ell} u_\ell(x) \exp(ik_\ell x) \quad \text{and} \quad r(x) = \sum_{\ell} r_\ell(x) \exp(ik_\ell x)
\end{align*}
\]

- Instead of oscillatory \( u(x) \) compute smooth \( u_\ell(x) \);
- Differential operators for \( u_\ell(x) \) are obtained directly from \( L \).
Phase error: Wave vs Rays

- Waves (as discussed): \( E(kd, kh) \approx kd \frac{k^2 h^2}{2\pi \gamma} \), \( 24 \leq \gamma \leq 48 \)

- Rays: \( E(kd, L) \approx kd \frac{\beta}{2\pi L^2} \), \( \beta \approx 0.12 \)

This is assuming very aggressive coarsening with \( L_n = 2^{n+2} \):
- Propagation direction \( \xi \): \( h^n_{\xi} = (L^n)^2/(16k) \);
- Orthogonal direction \( \eta \): \( h^n_{\eta} = CL^n/4k \)
- For \( L^1 = 8 \), \( kh_{\xi} \approx 4 \) and \( kh_{\eta} \approx 2 \)
Main features

- Smaller wave computational domain;
- Larger ray domains; grids aligned with propagation directions;
- Aggressive ray coarsening;
- Ray equations $a_{\xi\xi} + a_{\eta\eta} + 2ika_{\xi}, \xi = k_\ell$
- Ray residual separation using consecutive full weighting;
- FAS for both waves and rays - boundary values for waves come directly from rays;
More about rays
Algorithm

- Repeat until happy
  - CS standard wave \( \bigtriangledown \) cycle
  - FAS ray cycle:
    - computation of wave residual \( kh \ll 1 \);
    - separation into ray residuals starting \( kh \approx 1 \);
    - forming FAS ray rhs, line relaxation \( kh_\xi \approx 4, \; kh_\eta \approx 2 \);
    - reconstructing to the wave form \( kh \approx 1 \);
    - interpolating (with relaxation) to the finest wave grid \( kh \ll 1 \).
Requirements and advantages

- **Requirements**
  Helmholtz PDE
  Structured grids
  Analytical knowledge of $Lu \approx 0$

- **Advantages**
  Local processes - wave representation on fine grids
  Geometrical optics processes - ray representation on coarse grids and larger domains
  Natural intro of radiation boundary conditions