Multilevel Projection-based Krylov Methods for solving a class of PDE's

Yogi A. Erlangga
Institut für Mathematik – TU Berlin

Reinhard Nabben – TU Berlin

with support from
Deutsche Forschungsgemeinschaft (DFG)

Outline of the Talk

- Introduction: Linear system, $P_N = I - AZE^{-1}Y^T + \lambda_N ZE^{-1}Y^T$, etc.
- Spectral properties
- Implementation aspects
- Numerical examples:
  - SPD case: 2D Poisson equation
  - Nonsymmetric case: 2D convection-diffusion equation
  - Indefinite case: 2D Helmholtz equation
- Conclusion

Introdcution

The linear system:

$$Au = b, \qquad A \in \mathbb{C}^{N\times N}, \quad u, b \in \mathbb{C}^N.$$

(A is in general nonsymmetric, sparse and large)

Problems:

- Diffusion problem (symmetric)
- Convection-diffusion equation (nonsymmetric)
- Helmholtz equation (symmetric, indefinite)

Preconditioned system:

$$M_1^{-1} A M_2^{-1} \widetilde{u} = M_1^{-1} b, \quad \widetilde{u} = M_2 u, \qquad M_1, M_2 \text{ nonsingular.}$$

For generality,

$$\widehat{A}\widehat{u} = \widehat{b}, \qquad \widehat{A} := M^{-1}A, \quad \widehat{u} := u, \quad \widehat{b} := M^{-1}b.$$

Introduction

Consider "the second level preconditioner":

$$P_N = P_D + \lambda_N Z E^{-1} Y^T, \qquad \widehat{E} = Y^T \widehat{A} Z,$$

where

$$P_D = I - \widehat{A} Z \widehat{E}^{-1} Y^T, \qquad \text{(Deflation)}$$

and solve the system

$$P_N \widehat{A} \widehat{u} = P_N \widehat{b}.$$

- $\lambda_N = \max_{x \neq 0} (x^T \widehat{A} x)/(x^T x)$
- $\widehat{E}$: Galerkin product
- $Z, Y \in \mathbb{R}^{n \times r}$ are full rank

- P_N is derived from generalized Wielandt's deflation, with P_D a special case [E., Nabben, 2007]

Introduction

Right preconditioning version:

With

$$Q_N = Q_D + \lambda_N Z E^{-1} Y^T, \qquad \widehat{E} = Y^T \widehat{A} Z,$$

where

$$Q_D = I - Z \widehat{E}^{-1} Y^T \widehat{A},$$

solve

$$\widehat{A} Q_N \widehat{u} = \widehat{b}, \qquad u = Q_N \widehat{u}, \quad \widehat{b} = b.$$

For theory, we focus on $P_N \widehat{A}$. In the implementation, $\widehat{A} Q_N$.

Spectral properties of $P_N \widehat{A}$

For $P_N \widehat{A} \widehat{u} = P_N \widehat{b}$:

Denote the spectrum of $\hat{A}$: $\sigma(\hat{A}) = \{\lambda_1, \ldots, \lambda_N\}$, $\lambda_i \leq \lambda_j$ for $i < j$.

Theorem 1 *Columns of $Z, Y \in \mathbb{R}^{n \times r}$ are right and left eigenvectors of $\hat{A}$. Thus, $\hat{E}$ is the eigenvalue matrix of $\hat{A}$ and*

$$\sigma(P_N \hat{A}) = \{\lambda_N, \ldots, \lambda_N, \lambda_{r+1}, \ldots, \lambda_N\}.$$

- $P_N \hat{A}$ is not symmetric, even if $\hat{A}$ is symmetric.
  For symmetric $\hat{A}$, $\kappa = \lambda_N / \lambda_{r+1}$ is not the condition number.
- But, $\kappa_{eff} := \lambda_N / \lambda_{r+1} \leq \lambda_N / \lambda_1 =: \kappa$.
  $P_N \hat{A}$ is more clustered than $\hat{A}$ $\longrightarrow$ Favorable for Krylov methods

Compare: (from Nabben's talk)

Deflation: $\sigma(P_D \hat{A}) = \{0, \ldots, 0, \lambda_{r+1}, \ldots, \lambda_N\}$.

Spectral properties of $P_N\widehat{A}$

Spectral relation between $P_D\widehat{A}$ and $P_N\widehat{A}$.

Theorem 2 Z, Y *are "arbitrary" rectangular matrices with rank* r*.*

$$\sigma(P_D\widehat{A}) = \{0, \ldots, 0, \mu_{r+1}, \ldots, \mu_N\} \quad \Longrightarrow \quad \sigma(P_N\widehat{A}) = \{\lambda_N, \ldots, \lambda_N, \mu_{r+1}, \ldots, \mu_N\}.$$

- $\sigma(P_N\widehat{A})$ is similar to $\sigma(P_D\widehat{A})$
- convergence is likely very similar (if $\lambda_N \sim \mu_N$, and σ convergence sole criterion)
- Since Z and Y are arbitrary, we can choose, e.g.,

$$Z = I_H^h = (I_h^H)^T, \quad Y^T = I_h^H, \qquad \widehat{E} = I_h^H \widehat{A}_h I_H^h. \qquad (Z \text{ an interpolation matrix})$$

Spectral properties of $P_N\widehat{A}$

Deflation:

- $P_D^2 = P_D$ (Projection)
- $P_D\widehat{A} = \widehat{A}Q_D$
- If $\widehat{A}$ is symmetric, then $P_D\widehat{A}$ is also symmetric

In contrast:

- $P_N^2 \neq P_N$
- $P_N\widehat{A} \neq \widehat{A}Q_N$. However, $\sigma(P_N\widehat{A}) = \sigma(\widehat{A}Q_N)$
- $P_N\widehat{A}$ is not symmetric even if $\widehat{A}$ is symmetric.

Furthermore,

- P_NA can not be expressed in terms of iteration matrix in the Richardson method (?)
- Consequence: P_NA has to be seen only from Krylov subspace method context.

Spectral properties of $P_N\widehat{A}$

Spectral sensitivity w.r.t. inexact coarse grid solves.

Proposition 3 *Z eigenvectors. In $P_N = I - \widehat{A}Z\widetilde{E}^{-1}Y^T + \lambda_N Z\widetilde{E}^{-1}Y^T$, assume*

$$\widetilde{E}^{-1} = \textit{diag}\left(\frac{1-\epsilon_1}{\lambda_1} \dots \frac{1-\epsilon_r}{\lambda_r}\right)$$

where $|\epsilon_i|_{i=1,r} \ll 1$. Then,

$$\sigma(P_N\widehat{A}) = \{(1-\epsilon_1)\lambda_N + \lambda_1\epsilon_1, \dots, (1-\epsilon_r)\lambda_N + \lambda_r\epsilon_r, \lambda_{r+1}, \dots, \lambda_N\}.$$

(Recall Nabben's talk) : $\sigma(P_D\widehat{A}) = \{\lambda_1\epsilon_1, \dots, \lambda_r\epsilon_r, \lambda_{r+1}, \dots, \lambda_N\}$.

$\rightarrow$ $P_N\widehat{A}$ is less sensitive than $P_D\widehat{A}$ w.r.t. inexact coarse grid solves.

- r can be chosen very large (large projection subspace)
- E^{-1} can be computed only approximately (by an inner iteration)

Implementation (1): two level

Two-grid (two-level) notations:

- $\widehat{A}_h = M_h^{-1} A_h, \quad P_N = I_h - \widehat{A}_h Z \widehat{E}^{-1} Y^T + \lambda_N Z \widehat{E}^{-1} Y^T$
- $\widehat{E} = Y^T \widehat{A} Z = Y^T M_h^{-1} A_h Z =: \widehat{A}_H.$

Preconditioning step in a Krylov subspace method:

$$x_h = \widehat{A}_h v_h$$
$$w_h = P_N x_h$$

In an expanded form:

$$\begin{aligned} w_h &= (I - \widehat{A}_h Z \widehat{E}^{-1} Y^T + \lambda_N Z \widehat{E}^{-1} Y^T) x_h \\ &= x_h - (\widehat{A}_h - \lambda_N I) Z \widehat{A}_H^{-1} Y^T x_h & (\widehat{E} \equiv \widehat{A}_H) \\ &= x_h - (\widehat{A}_h - \lambda_N I) Z \widehat{A}_H^{-1} x_H, & (x_H = Y^T x_h) \end{aligned}$$

$x_H := Y^T x_h$ a fine-to-coarse projection of x_h.

$\widehat{A}_H^{-1} x_H =: \widehat{x}_H \Rightarrow \widehat{A}_H \widehat{x}_H = x_H$ is solved only approximately by a ($inner$) Krylov method.

Implementation (2): two level

Notes on $\widehat{A}_H \widehat{x}_H = x_H$.

- $\widehat{A}_H$ is inverted exactly $\longrightarrow$ the fastest convergence of the *outer* iteration.
  It means inner iterations with a severe termination criterion.
- P_N is a "stable" projection method.
  Inner iteration with less thight termination criterion (e.g., $tol = 10^{-2}$).
- Residual/error in the inner iteration can be fast reduced by applying P_N at the "second" level, i.e.,

  Solve: $P_{N,H} \widehat{A}_H \widehat{x}_H = P_{N,H} x_H$ (instead of $\widehat{A}_H \widehat{x}_H = x_H$)
- With inner Krylov iterations, P_N is in general not constant
  Use flexible Krylov subspace method (FGMRES, FQMR, ...)

Implementation (3): multilevel projection algorithm

- Initialization. With given $u_0^{(1)}$,

  Set $Z_{i,i+1}$, $i = 1, \ldots, m-1$ ($m > 1$ the maximum level) and $Y_{i,i+1} = Z_{i,i+1}$

  Compute $\widehat{A}^{(i)} = Y_{i-1,i}^T \widehat{A}^{(i-1)} Z_{i-1,i}$, for $i = 2, \ldots, m$, and $\lambda_N^{(i)}$
- At $i = 1$, solve $P_N^{(1)} \widehat{A}^{(1)} u^{(1)} = P_N^{(1)} b$ with a Krylov method until convergence using

  $x^{(1)} = \widehat{A}^{(1)} v^{(1)} = A^{(1)} (M^{(1)})^{-1} v^{(1)}$

  Restriction: $x^{(2)} = Y_{1,2}^T x^{(1)}$

  At $i = 2$, solve $P_N^{(2)} \widehat{A}^{(2)} d^{(2)} = P_N^{(2)} x^{(2)}$ with a Krylov subspace method using

  $x^{(2)} = \widehat{A}^{(2)} v^{(2)} = A^{(2)} (M^{(2)})^{-1} v^{(2)}$

  Restriction: $x^{(3)} = Y_{2,3}^T x^{(2)}$

  $\ldots$

  At $i = i+1$. If $i = m$, $x^{(m)} = (\widehat{A}^{(m)})^{-1} d^{(i)}$. Else, solve $P_N^{(i)} \widehat{A}^{(i)} d^{(i)} = P_N^{(i)} x^{(i)}$.

  $\ldots$

  Interpolation: $\widehat{w}^{(2)} = Z_{2,3} d^{(3)}$

  $w^{(2)} = x^{(2)} - (\widehat{A}^{(2)} - \lambda_N^{(2)} I) \widehat{w}^{(2)}$

  Interpolation: $\widehat{w}^{(1)} = Z_{1,2} d^{(2)}$

  $w^{(1)} = x^{(1)} - (\widehat{A}^{(1)} - \lambda_N^{(1)} I) \widehat{w}^{(1)}$

Implementation (4): some other issues

- The choice of Z and Y

  Sparsity of Z and Y;

  May be the same as interpolation and restriction matrices in multigrid (e.g., piece-wise constant, bi-linear interpolation, etc.);

  But not eigenvectors;

  $Y = Z$;

- About λ_N

  Expensive too compute, but an approximate is sufficient:

  $\longrightarrow$ by Gerschgorin's theorem.

  $\longrightarrow$ by other means (in case of the Helmholtz equation)

  Because of this approximation:

$$\lambda_N \leftarrow \omega \cdot \lambda_N, \qquad 0 < \omega \leq 1, \qquad (\omega\text{: a “correction” factor})$$

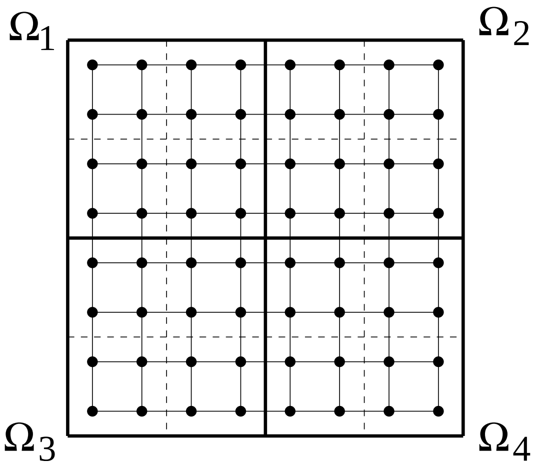
Numerical example: 2D Poisson equation (1/2)

The 2D Poisson equation:

$$
\begin{aligned}
-\nabla \cdot \nabla u &= g, && \text{in } \Omega \in (0,1)^2, \\
u &= 0, && \text{on } \Gamma = \partial\Omega.
\end{aligned}
$$

Discretization: finite differences.

Z: Piecewise linear interpolation, $Y = Z$



Ω with index set $\mathcal{I} = \{i | u_i \in \Omega\}$.
Ω is partitioned into non-overlapping subdomain Ω_j, $j = 1, \ldots, l$, with respective index $\mathcal{I}_j = \{i \in \mathcal{I} | u_i \in \Omega_j\}$.
Then, $Z = [z_{ij}]$:

$$
z_{ij} = \begin{cases} 1, & i \in \mathcal{I}_j, \\ 0, & i \notin \mathcal{I}_j. \end{cases}
$$

In $\widehat{A} = M^{-1}A$, $M = I$.

Numerical example: 2D Poisson equation (2/2)

Convergence results: relative residual $\leq 10^{-6}$, $\omega = 1$
Gerschgorin estimate for λ_N

N	MP(2,2,2,2)	MP(4,2,2,2)	MP(6,2,2,2)	MP(4,3,3,3)	MG
32^2	15	14	14	14	11
64^2	16	14	14	14	11
128^2	16	14	14	14	11
256^2	16	14	14	14	11

Notation:

- MP(4,2,2,2): Multilevel Projection with 4,2,2 and 2 FGMRES iterations at level no. 1,2,3 and 4. Etc.
- MG: Multi Grid (here, V-cycle, one pre- and post RB-GS smoothing, bilinear interpolation)

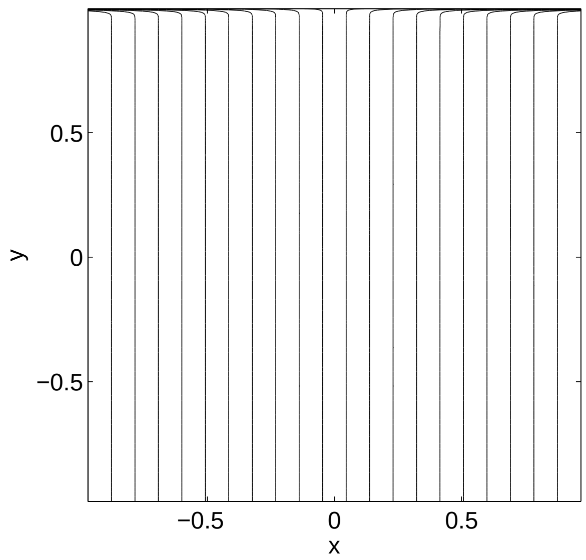
Obersvation:

- Level $i = 2$ is important!
- h-indepedent convergence
- Convergence of MP is comparable with MG.

Numerical example: 2D Convection-diffusion equation (1/2)

The 2D convection-diffusion equation with vertical winds:

$$
\begin{aligned}
\frac{\partial u}{\partial y} - \frac{1}{Pe}\nabla \cdot \nabla u &= 0, \quad \text{in } \Omega = (-1,1)^2, \\
u(-1,y) &\approx -1, \quad u(1,y) \approx 1, \\
u(x,-1) &= x, \quad u(x,1) = 0.
\end{aligned}
$$



Discretization: Finite volume, upwind discretization for convective term

Z: piece-wise constant interpolation, $Y = Z$

In $\widehat{A} = M^{-1}A$, $M = diag(A)$

Numerical example: 2D Convection-diffusion equation (2/2)

Convergence results: relative residual $\leq 10^{-6}$
MP(4,2,2,2), $\omega = 0.8$, Gerschgorin estimate for λ_N

	Pe:			
Grid	20	50	100	200
128^2	16	16	18	24
256^2	16	16	16	17
512^2	15	16	16	15

- In MP, FGMRES is used
- MG (with V-cycle, one pre- and post RB-GS smoothing and bilinear interpolation) does not converge

Observation:

- Level $i = 2$ is important!
- Almost h- and Pe-independent convergence

Numerical example: 2D Helmholtz equation (1/8)

The 2D Helmholtz equation:

$$\mathcal{A}u := -\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} - k^2 u = f, \quad \text{in } \Omega = (0,1),$$

with radiation conditions on $\Gamma = \partial\Omega$.

Preconditioner operator: 2D shifted Laplacian [E., Oosterlee & Vuik, SISC (2006)]:

$$\mathcal{M} := -\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - (1 - 0.5\hat{i})k^2.$$

In $\widehat{A} = M^{-1}A$, A and M are the discrete form of $\mathcal{A}$ and $\mathcal{M}$.

Theorem 4 $\lambda_N(M^{-1}A) \to 1$ *(or* $\lambda_N(AM^{-1}) \to 1$*).*

M is approximately inverted by multigrid with F-cycle, one Jacobi pre- and postsmoothing.
AM^{-1} is not explicitly known!

Numerical example: 2D Helmholtz equation (2/8)

For $\widehat{A} = AM^{-1}$: Recall that $\hat{E} = Z^T \hat{A} Z = Z^T M^{-1} A Z$ $(Y = Z)$.

In two-level projection:

$$\begin{aligned} \hat{E} &= Z^T A_h M_h^{-1} Z \\ &\approx (Z^T A_h Z)(Z^T M_h Z)^{-1} Z^T Z = A_H M_H^{-1} B_H, \end{aligned}$$

where

$$A_H = Z^T A_h Z, \qquad M_H = Z^T M_h Z, \qquad B_H = Z^T I Z.$$

In multi-level projection:

At level $j = 1$, $A^{(1)} := A$, $M^{(1)} := M$, $B^{(1)} := I$, $\hat{A}^{(1)} = A^{(1)}(M^{(1)})^{-1}$ and $Q_N^{(1)} = Q_N$.
For $j = 2, \ldots, m$,

$$\begin{aligned} A^{(j)} &= Z_{(j-1,j)}^T A^{(j-1)} Z_{(j-1,j)}, \\ M^{(j)} &= Z_{(j-1,j)}^T M^{(j-1)} Z_{(j-1,j)}, \\ B^{(j)} &= Z_{(j-1,j)}^T B^{j-1} Z_{(j-1,j)}, \\ \hat{A}^{(j)} &= A^{(j)} (M^{(j)})^{-1} B^{(j)}, \\ Q_N^{(j)} &= I - Z_{(j-1,j)} (\hat{A}^{(j)})^{-1} Z_{(j-1,j)}^T \left(\hat{A}^{(j-1)} - \omega \lambda_N^{(j)} I \right). \end{aligned}$$

Numerical example: 2D Helmholtz equation (3/8)

Algorithm: Multilevel projection, with $u^{(1)} = (M^{(1)})^{-1}\widetilde{u}^{(1)}$

At $j = 1$, solve $A^{(1)}(M^{(1)})^{-1}\widetilde{u}^{(1)} = b$ with Krylov method by computing:

$v_M^{(1)} = (M^{(1)})^{-1}v^{(1)}$ with multigrid;

$s^{(1)} = A^{(1)}v_M^{(1)}$; $t^{(1)} = s^{(1)} - \omega\lambda_N^{(1)}v^{(1)}$;

Restriction: $(v_R')^{(2)} = Z_{(1,2)}^T t^{(1)}$;

If $j = m$: solve exactly $v_R^{(m)} = (\hat{A}^{(m)})^{-1}(v_R')^{(m)}$;

else

At $j = 2$, solve $A^{(2)}(M^{(2)})^{-1}B^{(2)}v_R^{(2)} = (v_R')^{(2)}$ with Krylov iterations by computing:

$v_M^{(2)} = (M^{(1)})^{-1}B^{(2)}v^{(2)}$ with multigrid;

$s^{(2)} = A^{(2)}v_M^{(2)}$; $t^{(2)} = s^{(2)} - \omega\lambda_N^{(2)}v^{(2)}$;

Restriction: $(v_R')^{(3)} = Z_{(2,3)}^T t^{(2)}$;

If $j = m$: solve exactly $v_R^{(m)} = (\hat{A}^{(m)})^{-1}(v_R')^{(m)}$;

else

$\ldots$

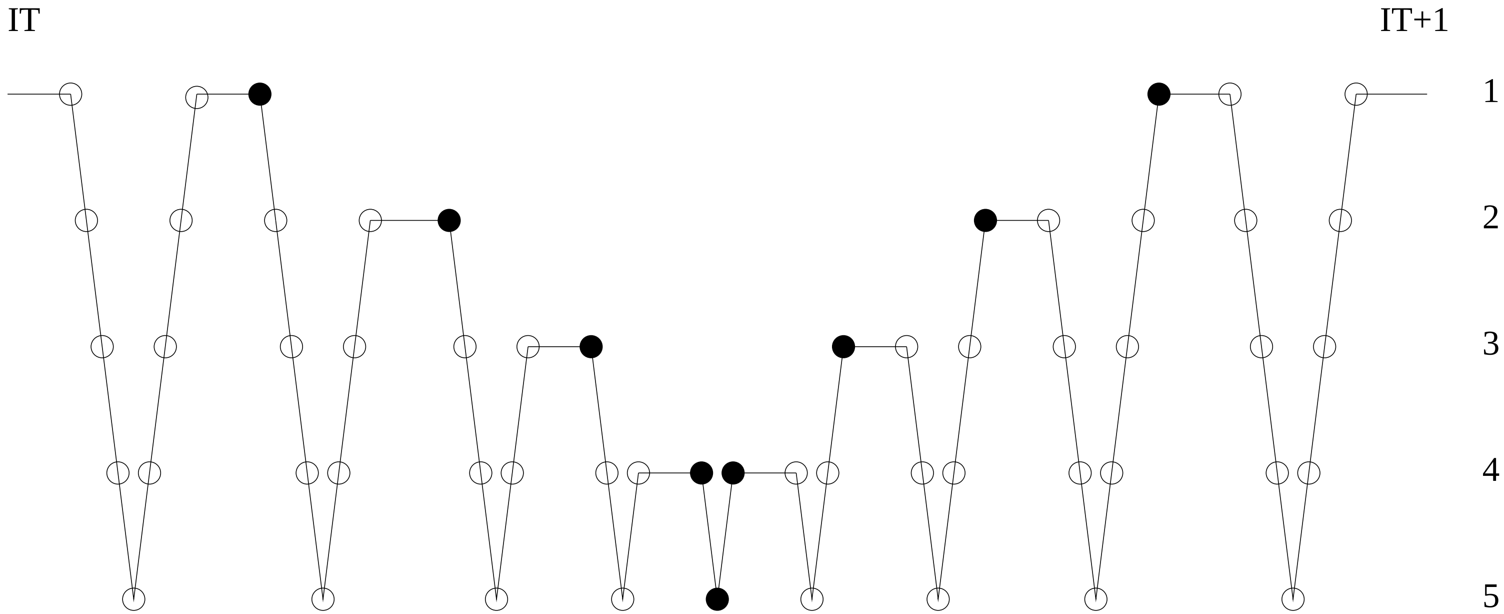
Interpolation: $q^{(1)} = v^{(I)} - Z_{(1,2)}^{(T)}v_R^{(2)}$;

$w^{(1)} = (M^{(1)})^{-1}q^{(1)}$ with multigrid;

$p^{(1)} = A^{(1)}w^{(1)}$;

Numerical example: 2D Helmholtz equation (4/8)

Multigrid–Multilevel projection (MG–MP) cycle:



Black circle: Projection step

White circle: Multigrid step (shown as V-cycle, but can also be with other cycles).

Numerical example: 2D Helmholtz equation (5/8)

Convergence results: relative residual $\leq 10^{-6}$
Multilevel with MG–MP(4,2,1), $\omega = 1$, $\lambda_N = 1$

	k:						
g/w	20	40	60	80	100	120	200
15	11	14	15	17	20	22	39
20	12	13	15	16	18	21	30
30	11	12	12	13	13	15	24

- Constant wavenumber k
- "g/w" : number of gridpoints per wavelength
- In MP, FGMRES is used
- In MG, F-cycle with one Jacobi pre- and postsmoothing is used

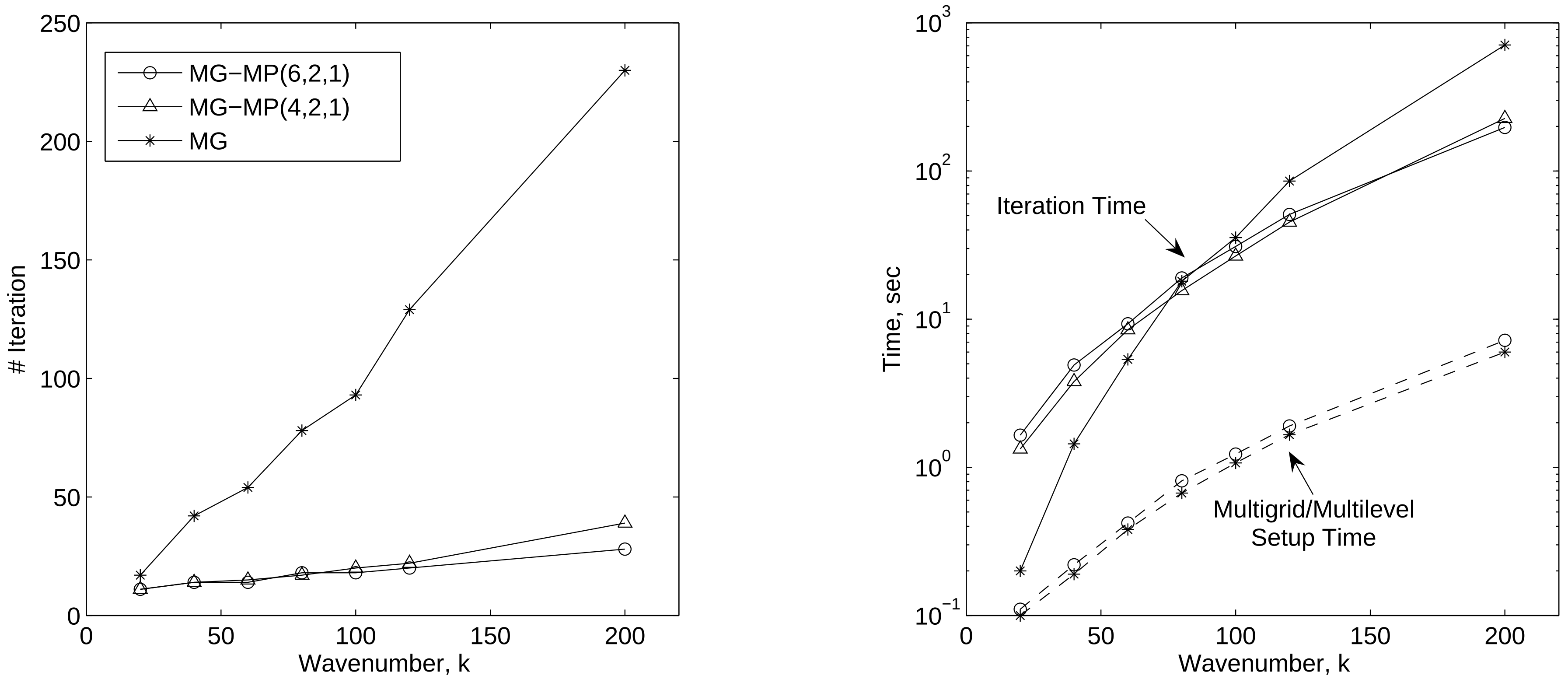
Numerical example: 2D Helmholtz equation (6/8)

Convergence results: relative residual $\leq 10^{-6}$
Multilevel with MG–MP(6,2,1), $\omega = 1$, $\lambda_N = 1$

	k:						
g/w	20	40	60	80	100	120	200
15	11	14	14	18	18	20	28
20	12	13	15	15	16	17	25
30	11	12	12	13	13	15	16

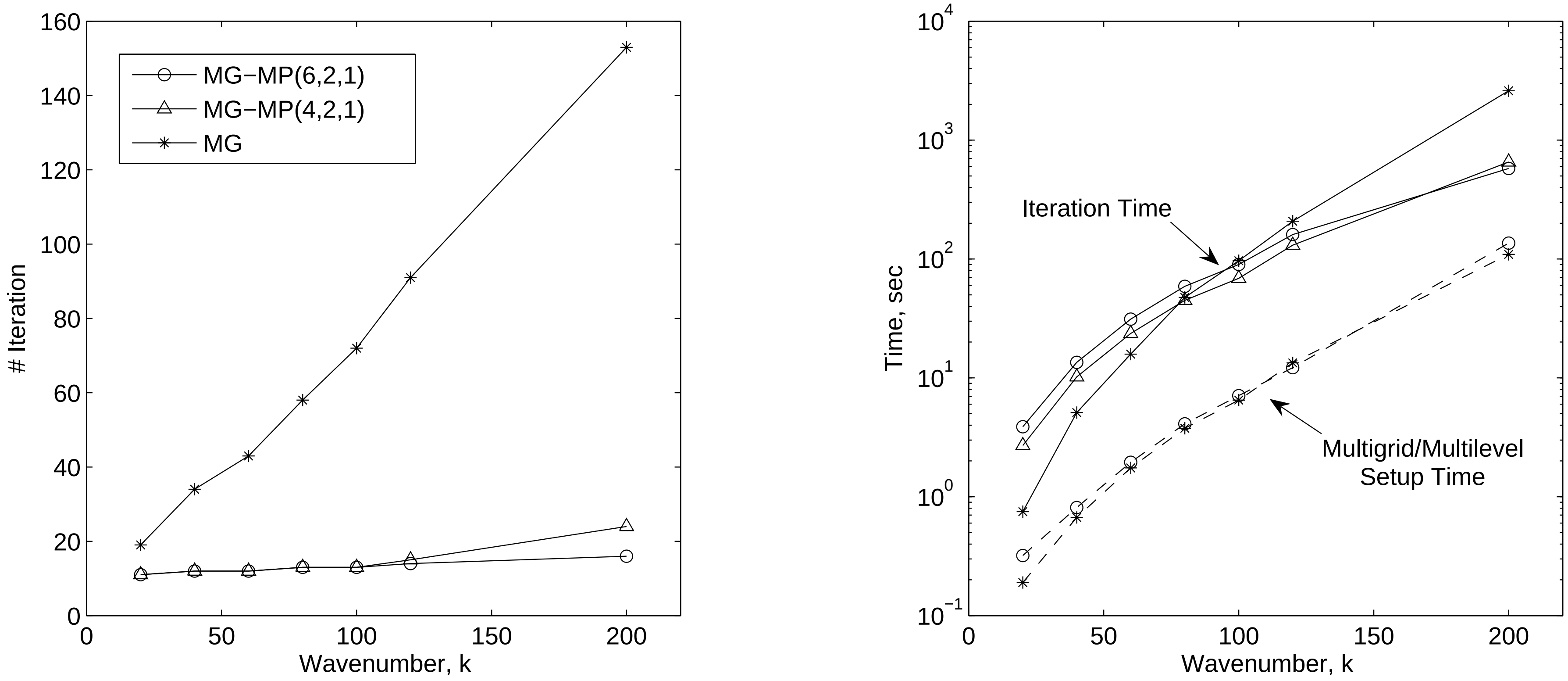
Numerical example: 2D Helmholtz equation (7/8)

Convergence and CPU time for 15 gridpoints/wavelength



Numerical example: 2D Helmholtz equation (8/8)

Convergence and CPU time for 30 gridpoints/wavelength



Conclusion

- We discussed a multilevel projection-based iteration based on shifting small eigenvalues to the max eigenvalue.
- Theoretical aspects of the method had been shown
- The stability of the projection operator allows the use of inner iterations to handle the coarse grid problem with low accuracy.
- Parameter λ_N can be determined algebraically (using Gerschgorin's theorem, e.g.) or analytically (in the case of the Helmholtz equation).
- Coarse grid (preconditioned) matrices are approximated by a product of coarse grid matrices
- Numerical experiments were shown for a different class of problems:

  $\longrightarrow$ Poisson equation: h-independent convergence (multigrid-like)

  $\longrightarrow$ Convection-diffusion equation: h- and nearly Pe-independent convergence

  $\longrightarrow$ Helmholtz equation: combination of multigrid and multilevel projection iterations

  h- and nearly k-independent convergence

  gain in CPU time at high wavenumbers

Bibliography

- Y.A. Erlangga, R. Nabben, Multilevel projection-based nested Krylov iteration for boundary value problems, (2007) submitted
- Y.A. Erlangga, R. Nabben, On the projection method for the preconditioned Helmholtz linear system, (2007) submitted

http://www.math.tu-berlin.de/∼erlangga