Fourier Analysis of Iterative Methods for the Helmholtz problem

Hanzhi Diao

MSc Thesis Defence
supervisor: Prof. dr. ir. C. Vuik

December 20, 2012

Outline

Helmholtz Problem

Helmholtz equation

$$-\Delta\, u(\mathbf{x}) - k^2\, u(\mathbf{x}) = f(\mathbf{x}) \text{ in } \Omega \in \mathbb{R}^3$$

Boundary condition

Dirichlet / Neumann / Sommerfeld

Discretization

finite difference method / finite element method

Linear system

- sparse
- symmetric but non-Hermitian

Thesis Work

Objective

spectral properties $\Longrightarrow$ convergence behaviour

Task

1. Preconditioning techniques
   - shifted Laplacian preconditioner M
   - deflation operator P and Q
2. Iterative solver
   - multigrid method for M^{-1}
   - Krylov subspace method for $Ax = b$ / $AM^{-1}x = b$ / $AM^{-1}Qx = b$
3. Fourier analysis
   - spectrum distribution
   - convergence factor
4. Numerical solution

Thesis Work

Objective

spectral properties $\Longrightarrow$ convergence behaviour

Task

1. Preconditioning techniques
   - shifted Laplacian preconditioner M
   - deflation operator P and Q
2. Iterative solver
   - multigrid method for M^{-1}
   - Krylov subspace method for $Ax = b$ / $AM^{-1}x = b$ / $AM^{-1}Qx = b$
3. Fourier analysis
   - spectrum distribution
   - convergence factor
4. Numerical solution

Model Problem

1D dimensionless Helmholtz problem with homogeneous Dirichlet boundary condition

$$\begin{cases} -\Delta\, u(x) - k^2\, u(x) = f(x) \text{ for } x \in (0,1), \\ u(0) = u(1) = 0. \end{cases}$$

The resulting linear system

$$Ax = b \quad \text{where } A = \frac{1}{h^2}\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} - k^2 I$$

Wave resolution

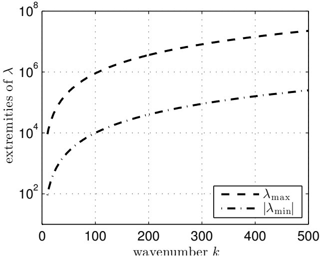
$$gw \cdot h = \frac{2\pi}{k}$$

Model Problem

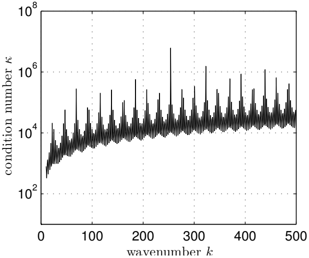
Eigenvalue

$$\lambda_l = \frac{4}{h^2}\sin^2(l\pi h/2) - k^2 \quad \text{for } l = 1, 2, \cdots, n$$

Difficulty in solving Helmholtz problem



indefiniteness



condition number

Multigrid Method

The solver for inverting the shifted Laplacian preconditioner M

- The coarsening strategy is done by doubling the mesh size, i.e. $\Omega_h \to \Omega_{2h}$.
- The smoother is ω-Jacobi iteration operator.
  - ω is chosen as the optimal one ω_{opt}.
- The intergrid transfer
  - restriction by full weighting operator
  - prolongation by linear interpolation operator

The failure of MG in solving $Ax = b$

- *The coarse grid cannot cope with high wavenumber problem.*
- *The ω-Jacobi iteration does not converge.*

Multigrid Method

The solver for inverting the shifted Laplacian preconditioner M

- The coarsening strategy is done by doubling the mesh size, i.e. $\Omega_h \to \Omega_{2h}$.
- The smoother is ω-Jacobi iteration operator.
  - ω is chosen as the optimal one ω_{opt}.
- The intergrid transfer
  - restriction by full weighting operator
  - prolongation by linear interpolation operator

The failure of MG in solving $Ax = b$

- *The coarse grid cannot cope with high wavenumber problem.*
- *The ω-Jacobi iteration does not converge.*

Krylov Subspace Methods

The solver for solving the linear system $Ax = b$

- **GMRES**, used in the thesis work
- CG
- BiCGStab
- GCR,IDR(s),...

Approximated Inversion

Given the iteration operator G, there is the approximated inversion

$$\mathbb{A}^{-1} = (I - G)A^{-1}.$$

For the stationary iteration, there is

$$\mathbb{A}_m^{-1} = (I - G^m)A^{-1}.$$

For the multigrid iteration, there is

$$\mathbb{A}_{\mathrm{MG}}^{-1} = (I - T_1^m)A^{-1}.$$

Shifted Laplacian Preconditioner

$$M := -\Delta_h - \underbrace{(\beta_1 + \iota\beta_2)}_{\text{shift}} k^2 I$$

- preconditioned system

$$\hat{A} := AM^{-1} = M^{-1}A \quad \text{and} \quad \sigma(AM^{-1}) = \sigma(M^{-1}A)$$

- preservation of symmetry

$$(AM^{-1})^T = AM^{-1}$$

- circular spectrum distribution

$$(\lambda_r - \frac{1}{2})^2 + (\lambda_i - \frac{\beta_1 - 1}{2\beta_2})^2 = \frac{\beta_2^2 + (1 - \beta_1)^2}{(2\beta_2)^2}$$

Shifted Laplacian Preconditioner

$$M := -\Delta_h - \underbrace{(\beta_1 + \iota\beta_2)}_{\text{shift}} k^2 I$$

- preconditioned system

$$\hat{A} := AM^{-1} = M^{-1}A \quad \text{and} \quad \sigma(AM^{-1}) = \sigma(M^{-1}A)$$

- preservation of symmetry

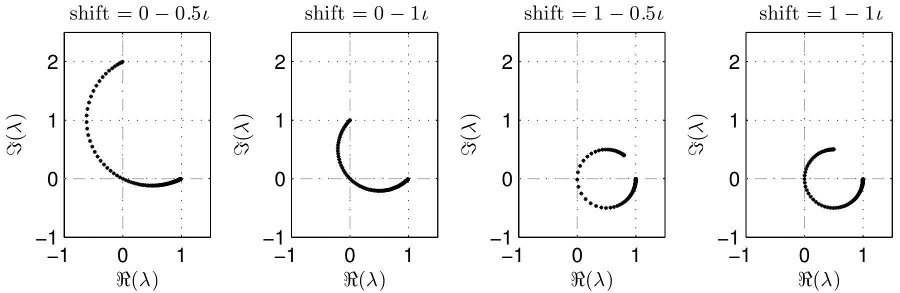
$$(AM^{-1})^T = AM^{-1}$$

- circular spectrum distribution

$$(\lambda_r - \frac{1}{2})^2 + (\lambda_i - \frac{\beta_1 - 1}{2\beta_2})^2 = \frac{\beta_2^2 + (1 - \beta_1)^2}{(2\beta_2)^2}$$

Shifted Laplacian Preconditioner

$\beta_1 = 1 \Longrightarrow$ the most compact distribution

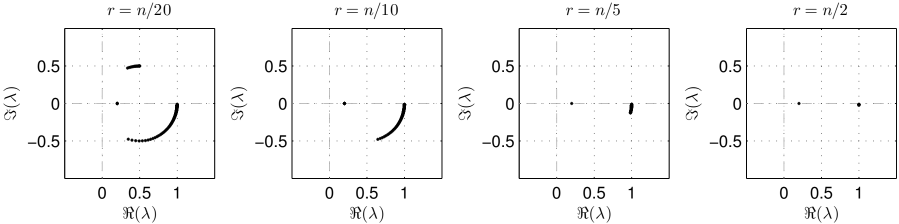


The spectrum distributions of the preconditioned matrix AM^{-1} with respect to several typical shifts when $k = 100$

Deflation Operator

For an invertible $\hat{A}$, take any $n \times r$ full rank matrices Y and Z

$$\begin{cases} \text{left} & P := I - \hat{A}\, Z\hat{E}^{-1}Y^T + \lambda_d\, Z\hat{E}^{-1}Y^T, \\ \text{right} & Q := I - Z\hat{E}^{-1}Y^T\, \hat{A} + \lambda_d\, Z\hat{E}^{-1}Y^T, \end{cases} \quad \text{where } \hat{E} = Y^T \hat{A} Z.$$



The spectrum distributions of the deflated matrix $\hat{A}Q$ towards $\lambda_d = 0.2$ where $k = 100$, $\text{shift} = 1 - \iota 1$, $A\,Z = Z\,\Lambda_r$ and $Y^T\,\hat{A} = \Lambda_r\,Y^T$

Deflation Operator

$\sigma(\hat{A}) = \{\lambda_1, \cdots, \lambda_n\}$ with $|\lambda_1| \leqslant \cdots \leqslant |\lambda_n|$

- Projector in case of $\lambda_d = 0$

$$P_D \cdot P_D = P_D \quad \text{and} \quad Q_D \cdot Q_D = Q_D.$$

- Preservation of symmetry in case of $\lambda_d = 0$,
- Spectrum distribution

$$\sigma(P\hat{A}) = \sigma(\hat{A}Q) = \{\lambda_d, \cdots, \lambda_d, \ \mu_{r+1}, \cdots, \mu_n\}.$$

- Condition number

$$\kappa(P\hat{A}) = \frac{|\mu_n|}{\min\{|\lambda_d|, |\mu_{r+1}|\}} \quad \text{in case of } \lambda_d \neq 0,$$

Deflation Operator

Inaccuracy in $\hat{E}^{-1}$

Assume $A\,Z = Z\,\Lambda_r$ and $Y^T\,\hat{A} = \Lambda_r\,Y^T$, then $\hat{E} = Y^T\hat{A}Z = \Lambda_r$.

$$\hat{\mathbb{E}}^{-1} = diag(\frac{1-\epsilon_1}{\lambda_1}, \cdots, \frac{1-\epsilon_r}{\lambda_r})$$

Deflation Operator

Inaccuracy in $\hat{E}^{-1}$

Assume $A\,Z = Z\,\Lambda_r$ and $Y^T\,\hat{A} = \Lambda_r\,Y^T$, then $\hat{E} = Y^T\hat{A}Z = \Lambda_r$.

$$\hat{\mathbb{E}}^{-1} = diag(\frac{1-\epsilon_1}{\lambda_1}, \cdots, \frac{1-\epsilon_r}{\lambda_r})$$

$$\Downarrow$$

$$\sigma(\mathbb{P}\hat{A}) = \{(1-\epsilon_1)\lambda_d + \lambda_1\epsilon_1, \cdots, (1-\epsilon_r)\lambda_d + \lambda_r\epsilon_r,\ \lambda_{r+1}, \cdots, \lambda_n\}$$

Deflation Operator

Inaccuracy in $\hat{E}^{-1}$

Assume $A\,Z = Z\,\Lambda_r$ and $Y^T\,\hat{A} = \Lambda_r\,Y^T$, then $\hat{E} = Y^T\hat{A}Z = \Lambda_r$.

$$\hat{\mathbb{E}}^{-1} = diag(\frac{1-\epsilon_1}{\lambda_1}, \cdots, \frac{1-\epsilon_r}{\lambda_r})$$

$$\Downarrow$$

$$\sigma(\mathbb{P}\hat{A}) = \{(1-\epsilon_1)\lambda_d + \lambda_1\epsilon_1, \cdots, (1-\epsilon_r)\lambda_d + \lambda_r\epsilon_r,\ \lambda_{r+1}, \cdots, \lambda_n\}$$

$$\Downarrow$$

$$\begin{cases} \lambda_d = 0 & \Rightarrow \sigma(\mathbb{P}\hat{A}) = \{\underbrace{\lambda_1\epsilon_1}_{\neq 0}, \cdots, \underbrace{\lambda_r\epsilon_r}_{\neq 0},\ \lambda_{r+1}, \cdots, \lambda_n\}, \\ \lambda_d = \lambda_n & \Rightarrow \sigma(\mathbb{P}\hat{A}) \approx \{\lambda_n, \cdots, \lambda_n,\ \lambda_{r+1}, \cdots, \lambda_n\}. \end{cases}$$

Multilevel Krylov Multigrid Method

A recursive Krylov solution of $\hat{E}^{-1}$

1. Use the approximation

$$M^{-1} \approx Z(Y^T M Z)^{-1} Y^T.$$

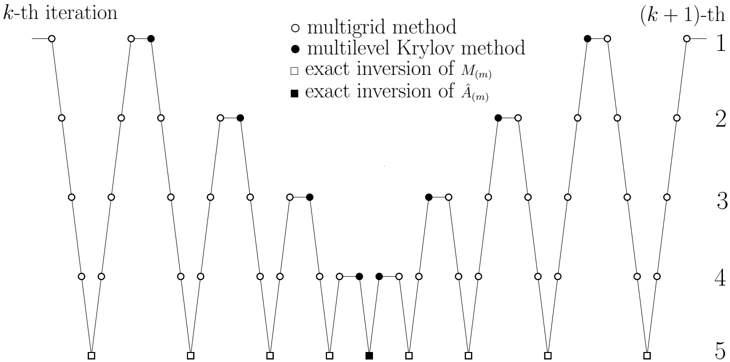
2. Take the replacement

$$\hat{E} := Y^T \hat{A} Z = Y^T A M^{-1} Z \approx \underbrace{Y^T A Z}_{A_{(2)}} (\underbrace{Y^T M Z}_{M_{(2)}})^{-1} \underbrace{Y^T Z}_{B_{(2)}}$$

$$\hat{E}^{-1} \approx \left(A_{(2)} M_{(2)}^{-1} B_{(2)} \right)^{-1}$$

3. Solve $A_{(2)}^{-1}$ in the same way as A^{-1}

Multilevel Krylov Multigrid Method



The illustration of multilevel Krylov multigrid method in a five-level grid

Principles of Fourier Analysis

Find out a subspace $E = \mathrm{span}\{\phi_1, \cdots, \phi_m\}$ such that

$$KE \subset E \Longrightarrow K\,\Phi = \Phi\,\tilde{K}.$$

For any $v = \Phi c \in E$, there is

$$K\,v = K\,\Phi c = \Phi \tilde{K}\,c \quad \text{where } \tilde{K} \text{ amplifies } c.$$

Assume E is the union of several disjoint subspaces. Then, there is a diagonal block matrix

$$K :\stackrel{\triangle}{=} [\tilde{K}^l] \quad \text{with } l \text{ as the block index.}$$

Principles of Fourier Analysis

Find out a subspace $E = \mathrm{span}\{\phi_1, \cdots, \phi_m\}$ such that

$$KE \subset E \Longrightarrow K\,\Phi = \Phi\,\tilde{K}.$$

For any $v = \Phi c \in E$, there is

$$K\,v = K\,\Phi c = \Phi \tilde{K}\,c \quad \text{where } \tilde{K} \text{ amplifies } c.$$

Assume E is the union of several disjoint subspaces. Then, there is a diagonal block matrix

$$K :\stackrel{\wedge}{=} [\tilde{K}^l] \quad \text{with } l \text{ as the block index.}$$

Fourier Analysis for Multigrid Analysis

In a two-level grid, the invariance subspace is given by

$$E_h^l := \text{span}\{\phi_h^l, \phi_h^{n-l}\} \text{ in } \Omega_h \quad \Longrightarrow \quad E_{2h}^l := \text{span}\{\phi_{2h}^l\} \text{ in } \Omega_{2h}.$$

$$\left.\begin{aligned} A_1, M_1, S &: E_h^l \to E_h^l \\ A_2, M_2 &: E_{2h}^l \to E_{2h}^l \\ R_1^2 &: E_h^l \to E_{2h}^l \\ P_2^1 &: E_{2h}^l \to E_h^l \end{aligned}\right\} \Longrightarrow T_1^2 : \; E_h^l \to E_h^l$$

In a multilevel grid, there is

$$\tilde{T}_k^m = \tilde{S}_k^{\nu_2} \left(I - \tilde{P}_{k+1}^k (I - \tilde{T}_{k+1}^m) \tilde{M}_{k+1}^{-1} \tilde{R}_k^{k+1} \tilde{M}_k \right) \tilde{S}_k^{\nu_1} \quad \text{with } \tilde{T}_m^m = 0,$$

Fourier Analysis for Multigrid Analysis

In a two-level grid, the invariance subspace is given by

$$E_h^l := \text{span}\{\phi_h^l, \phi_h^{n-l}\} \text{ in } \Omega_h \quad \Longrightarrow \quad E_{2h}^l := \text{span}\{\phi_{2h}^l\} \text{ in } \Omega_{2h}.$$

$$\left.\begin{aligned} A_1, M_1, S : E_h^l &\to E_h^l \\ A_2, M_2 : E_{2h}^l &\to E_{2h}^l \\ R_1^2 : E_h^l &\to E_{2h}^l \\ P_2^1 : E_{2h}^l &\to E_h^l \end{aligned}\right\} \Longrightarrow T_1^2 : E_h^l \to E_h^l$$

In a multilevel grid, there is

$$\tilde{T}_k^m = \tilde{S}_k^{\nu_2} \left(I - \tilde{P}_{k+1}^k \,(I - \tilde{T}_{k+1}^m) \tilde{M}_{k+1}^{-1} \, \tilde{R}_k^{k+1} \tilde{M}_k \right) \tilde{S}_k^{\nu_1} \quad \text{with } \tilde{T}_m^m = 0,$$

Application to Preconditioning

Shifted Laplacian preconditioner

$$\tilde{\hat{A}} = \tilde{A}\tilde{M}^{-1}$$

Deflation operator

$$\tilde{Q} = I - \tilde{P}_2^1 \tilde{\hat{E}}_2 \tilde{R}_2^1 (\lambda_n I - \tilde{\hat{A}}) \quad \text{with } \tilde{\hat{E}}_2 = \tilde{R}_1^2 \tilde{\hat{A}} \tilde{P}_2^1$$

Advantage of Fourier analysis

- computational time
- memory requirement
- accuracy

Application to Preconditioning

Shifted Laplacian preconditioner

$$\tilde{\hat{A}} = \tilde{A}\tilde{M}^{-1}$$

Deflation operator

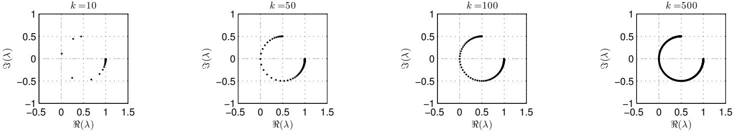
$$\tilde{Q} = I - \tilde{P}_2^1 \tilde{\hat{E}}_2 \tilde{R}_2^1 (\lambda_n I - \tilde{\hat{A}}) \quad \text{with } \tilde{\hat{E}}_2 = \tilde{R}_1^2 \tilde{\hat{A}} \tilde{P}_2^1$$

Advantage of Fourier analysis

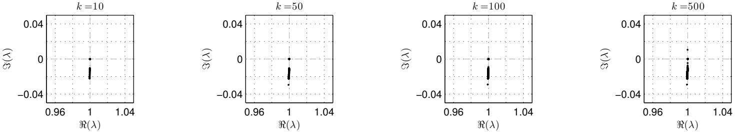
- computational time
- memory requirement
- accuracy

Basic Preconditioning Effect

- The spectrum of AM^{-1} is restricted to a circular distribution.

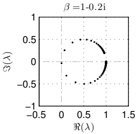


- The spectrum of $AM^{-1}Q$ is clustered around $(1, 0)$.

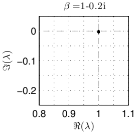


Choice of Shift $\beta_1 + \iota\beta_2$

1. β_2 determines the length of arc on which the eigenvalues of AM^{-1} are located.

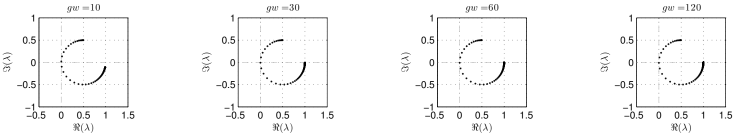


2. β_2 has the indirect influence on the tightness of spectrum distribution of $AM^{-1}Q$.

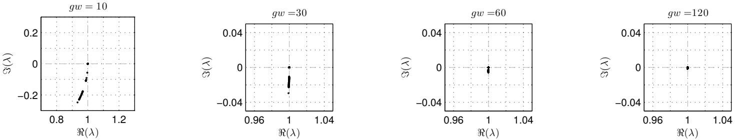


Influence of Wave resolution gw

1. High resolution exerts little negative influence on the spectrum distribution of AM^{-1}.



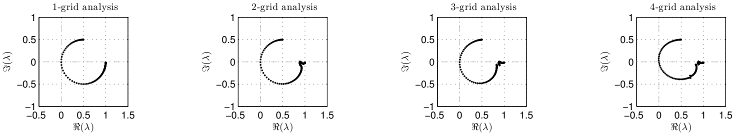
2. High resolution results in a more favourable spectrum distribution of $AM^{-1}Q$.



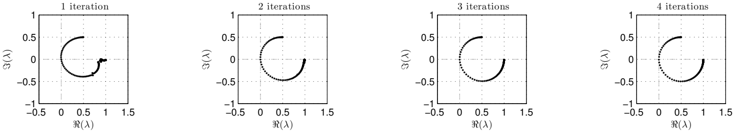
Approximated Shifted Laplacian Preconditioning

$A\mathbb{M}^{-1} = A(I - T_1^m)M^{-1}$

- The multigrid introduces disturbance to the preconditioning effect.



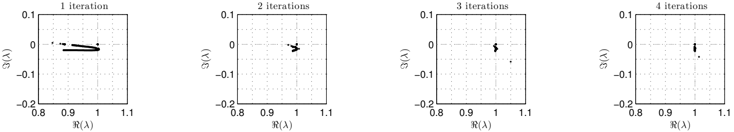
- The disturbance can be easily corrected by several iterations at a cheap cost.



Approximated Deflation Preconditioning

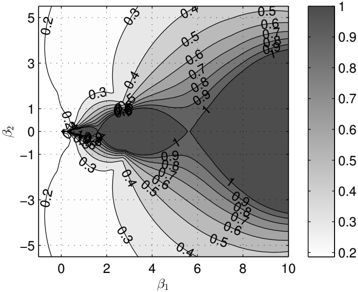
$A\mathbb{M}^{-1}\mathbb{Q}$ where the construction of $\mathbb{Q}$ is based on the $\mathbb{M}^{-1}$

- The preconditioning $AM^{-1}Q$ is much more sensitive to the accuracy in the approximation of M^{-1}.

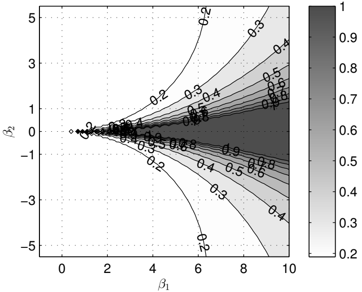


Multigrid Convergence Factor

- independence of k
- independence of the sign of β_2



$k = 30$ and $gw = 30$



$k = 30$ and $gw = 120$

- High resolution is favourable for the convergence.

Optimal Shift for the Preconditioner

- A small shift is favourable for the Krylov convergence of AM^{-1}.
- A large shift is favourable for the multigrid convergence of M^{-1}.

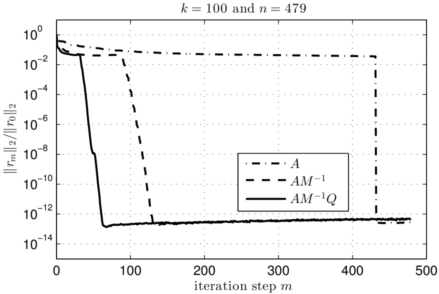
To find out

$$(\beta_1 + \iota\beta_2)_{\text{opt}} := \arg\min\{|\beta_1 + \iota\beta_2| : \max_{1\leqslant l\leqslant n-1} \mathcal{G}(l, \beta_1, \beta_2) \leqslant c < 1\}.$$

	$gw = 10$	$gw = 30$	$gw = 60$	$gw = 120$	$gw = 240$
$m = 2$	0.1096	0.0126	0	0	0
$m = 3$	0.3228	0.0616	0.0150	0	0
$m = 4$	0.3931	0.2002	0.0632	0.0155	0
$m = 5$	0.3931	0.2886	0.2012	0.0636	0.0156

The optimal β_2 in the shift $1 + \iota\beta_2$ for $\rho(T_1^m) \leqslant c = 0.9$

Basic Convergence Behaviour

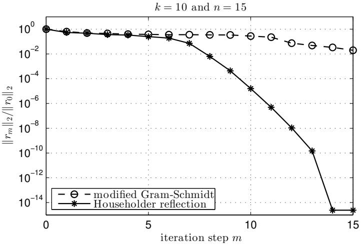


Overview of the convergence behaviour by different preconditioning

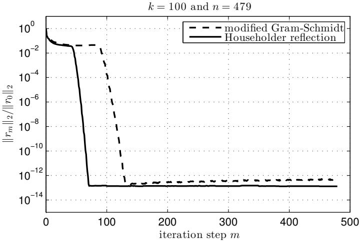
- verify
  - the different preconditioning effect
  - the advantage of $AM^{-1}Q$ over AM^{-1}
  - the influence of wavenumber and wave resolution

Influence of Orthogonalization

- Householder reflection outperforms modified Gram-Schmidt in convergence behaviour.
- GMRES using modified Gram-Schmidt fails to converge in the very small system.



very small system



normal size system

Influence of Approximated Preconditioning

- The inaccuracy in $\mathbb{M}^{-1}$ has little influence on the convergence behaviour

gw	$k=10$	$k=50$	$k=100$	$k=200$	$k=300$	$k=400$	$k=500$
10	11/11	36/36	60/60	108/105	153/149	193/188	265/258
30	12/12	36/36	60/58	114/108	161/152	209/196	255/240
60	12/12	36/36	63/62	113/111	161/158	207/204	255/250

Number of iterations with respect to different degrees of approximations i.e. AM^{-1} / $A\mathbb{M}^{-1}$

- The inaccuracy in $\mathbb{Q}$ slows down the convergence.

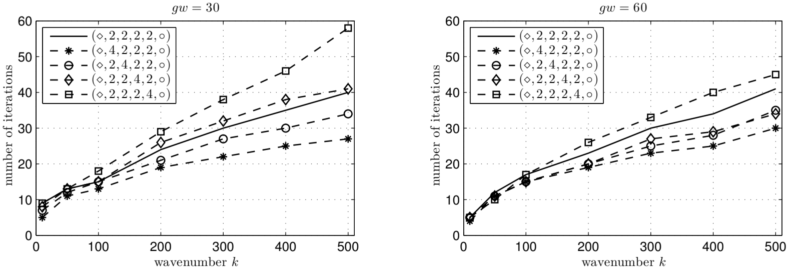
gw	$k=10$	$k=50$	$k=100$	$k=200$	$k=300$	$k=400$	$k=500$
10	6/9/9	11/18/18	14/26/27	21/43/44	28/59/61	33/71/70	39/98/111
30	4/5/5	6/13/13	6/15/17	8/36/37	9/54/55	10/73/75	11/92/ 94
60	3/4/4	4/ 5/ 8	4/ 7/ 9	5/12/16	6/18/22	6/24/27	6/32/ 34

Number of iterations with respect to different degrees of approximations i.e. $AM^{-1}Q$ / $AM^{-1}\mathbb{Q}$ / $A\mathbb{M}^{-1}\mathbb{Q}$

Internal Iteration in MKMG

$(\diamond, \#_2, \cdots, \#_{m-1}, \circ)$

- It is worth doing more iterations on the higher levels.
- The convergence behaviour will be slowed by more iterations on the lower levels.



Number of iterations with respect to different MKMG setup in a six-level grid

Suggestion on Future Work

- higher dimensional problems
- local Fourier analysis
- different Krylov solvers

Thank you for watching !