Deflated and Multilevel Krylov Methods

Reinhard Nabben

Outline

How to solve $Ax = b$

when

- A is large (of dimension $n > 10^6$)?
- A is sparse?
- A inherits properties from the underlying problem?

Why to solve $Ax = b$?

- many applications lead to $Ax = b$.
- bottleneck of many complex algorithms.

Krylov subspace methods - CG method

A symmetric positive definite

x_0 start vector, $r_0 = b - Ax_0$

$$\mathcal{K}_j(A, r_0) = span\{r_0, Ar_0, \ldots, A^{j-1}r_0\}$$

j-th step of CG method: find x_j such that

$$x_j \in x_0 + \mathcal{K}_j(A, r_0)$$

$$r_j = b - Ax_j \perp \mathcal{K}_j(A, r_0)$$

Krylov subspace methods - CG method

CG method for Ax = b

- Select arbitrary x_0
- $r_0 := b - Ax_0 \quad p_0 := r_0$
- FOR $j := 0, 1, \ldots,$ until convergence
- $\alpha_j := (r_j, r_j)/(p_j, Ap_j)$
- $x_{j+1} := x_j + \alpha_j p_j$
- $r_{j+1} := r_j - \alpha_j A p_j$
- $\beta_j := (r_{j+1}, r_{j+1})/(r_j, r_j)$
- $p_{j+1} := r_{j+1} + \beta_j p_j$
- ENDFOR

$$
\begin{aligned}
||x - x_j||_A &\leq 2||x - x_0||_A \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^j, \\
\kappa &= \kappa(A) = \frac{\lambda_n}{\lambda_1}
\end{aligned}
$$

Classical Preconditioners

Classical choices for M include

- $M^{-1} = \alpha I$, for $\alpha \approx \|A\|$ (Richardson)
- $M^{-1} = \alpha \text{diag}(A)$ (Jacobi)
- $M^{-1} = \text{tril}(A)$ (Gauss-Seidel)
- $M^{-1} = \hat{L}\hat{U}$, for $A = LU$, $\hat{L} \approx L$, $\hat{U} \approx U$ (ILU)

Deflated CG

Nicolaides 1987, Mansfield 1988, 1990, Kolotilina 1998, Vuik, Segal, and Meijerink 1999, Morgan 1995, Saad, Yeung, Erhel, and Guyomarch 2000, Frank and Vuik 2001, Blaheta 2006

Deflation and restarted GMRES

Morgan 1995, Erhel, Burrage, and Pohl 1996, Chapman and Saad 1997, Eiermann, Ernst, and Schneider 2000, Morgan 2002

Clemens et al. 2003,2004, de Sturler et al. 2006, Aksoylu, H. Klie, and M.F. Wheeler 2007

Deflation with eigenvectors

$$Au_i = \lambda_i u_i \qquad Z = [u_1, \ldots, u_r] \qquad u_i^T u_j = \delta_{ij}$$

$$P_D = I - AZ(Z^T AZ)^{-1} Z^T, \quad Z \in \mathbb{R}^{n \times r},$$

$$spectrum(P_D A) = \{0, \ldots, 0, \lambda_{r+1}, \ldots \lambda_n\}$$

Deflation with general vectors

A symmetric positive definite

$$Z = [z_1, \ldots, z_r] \qquad rankZ = r \quad E = Z^T AZ$$

$$P_D = I - AZE^{-1}Z^T, \quad Z \in \mathbb{R}^{n \times r},$$

$$P_D AZ = 0$$

$$spectrum(P_D A) = \{0, \ldots, 0, \mu_{r+1}, \ldots \mu_n\}$$

Deflated CG

$$Z \in \mathbb{R}^{n \times r} \quad Z = [z_1, \ldots, z_r] \quad rankZ = r$$

$$Ax = b \qquad P_D = I - AZE^{-1}Z^T$$

We have: $x = (I - P_D^T)x + P_D^T x$ Compute both!

1. $(I - P_D^T)x = Z(Z^T AZ)^{-1}Z^T b$
2. Solve $P_D A\tilde{x} = P_D b$ preconditioner M^{-1}:
   $M^{-1}P_D A\tilde{x} = M^{-1}P_D b$
3. Build $P_D^T \tilde{x} = P_D^T x$

Deflation for linear systems

cg for singular system

cg works if $b \subseteq rangeA$
Speed of convergence

$$cond_{eff}(P_D A) = \frac{\lambda_n(P_D A)}{\lambda_{r+1}(P_D A)}$$

instead of

$$cond(A) = \frac{\lambda_n(A)}{\lambda_1(A)}$$

How to choose the $z_1, \ldots, z_r$?

- best choice: eigenvectors
- approximate eigenvectors, Ritz vectors
- information from similar systems
  - several right hand sides
  - within a non linear method

Deflation

M^{-1} preconditioner, ILU $\quad Z$ appro. eigenvectors

$ZE^{-1}Z^T$

Domain decomposition

M^{-1} add. Schwarz $\quad Z$ grid transfer operator

$ZE^{-1}Z^T$ coarse grid correction

Multigrid

M^{-1} smoother $\quad Z$ grid transfer operator

$ZE^{-1}Z^T$ coarse grid correction

Name	Method	Operator
PREC	Traditional Preconditioned CG	M^{-1}
AD	Additive Coarse Grid Correc.	$M^{-1} + Q$
DEF1	Deflation Variant 1	$M^{-1}P_D$
DEF2	Deflation Variant 2	$P_D^T M^{-1}$
A-DEF1	Adapted Deflation Variant 1	$M^{-1}P_D + Q$
A-DEF2	Adapted Deflation Variant 2	$P_D^T M^{-1} + Q$
BNN	Abstract Balancing	$P_D^T M^{-1} P_D + Q$
R-BNN1	Reduced Balancing Variant 1	$P_D^T M^{-1} P_D$
R-BNN2	Reduced Balancing Variant 2	$P_D^T M^{-1}$

$$Q = ZE^{-1}Z^T = Z(Z^T AZ)^{-1}Z^T$$

Name	Method	Operator
PREC	Traditional Preconditioned CG	M^{-1}
AD	Additive Coarse Grid Correc.	$M^{-1} + Q$
DEF1	Deflation Variant 1	$M^{-1}P_D$
DEF2	Deflation Variant 2	$P_D^T M^{-1}$
A-DEF1	Adapted Deflation Variant 1	$M^{-1}P_D + Q$
A-DEF2	Adapted Deflation Variant 2	$P_D^T M^{-1} + Q$
BNN	Abstract Balancing	$P_D^T M^{-1} P_D + Q$
R-BNN1	Reduced Balancing Variant 1	$P_D^T M^{-1} P_D$
R-BNN2	Reduced Balancing Variant 2	$P_D^T M^{-1}$

$$Q = ZE^{-1}Z^T = Z(Z^TAZ)^{-1}Z^T$$

Nabben, Vuik 04 , Nabben, Vuik 06, Nabben Vuik 08
Tang, Nabben, Vuik, Erlangga 07
Tang, MacLachlan, Nabben, Vuik 08

Theorem

Tang, Nabben, Vuik, Erlangga 07
The spectrum of the systems preconditioned by DEF1, DEF2, R-BNN1 or R-BNN2 is given by

$$\{0, \ldots, 0, \mu_{r+1}, \ldots, \mu_n\}.$$

The spectrum of the systems preconditioned by A-DEF1, A-DEF2, BNN is given by

$$\{1, \ldots, 1, \mu_{r+1}, \ldots, \mu_n\}.$$

Non-symmetric Problems

- Erlangga, Nabben 06:

$$Z^T \to Y^T \qquad E \to Y^T AZ$$

$$P_D = I - AZE^{-1}Y^T$$

$$P_D^T \to Q_D = I - ZE^{-1}Y^T A$$

$$P_{BNN} = Q_D M^{-1} P_D + ZE^{-1}Y^T$$

$$\|M^{-1}(b - Au_{k,D})\|_2 \leq \|M^{-1}(b - Au_{k,BNN})\|_2.$$

Multilevel Krylov methods (MK methods)

Erlangga, Nabben 07/08

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

A is in general nonsymmetric, sparse and large

Problems:

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

Preconditioned system:

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

Here,

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

Consider

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

where

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

and solve the system

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

- $Z, Y \in \mathbb{R}^{n \times r}$ are full rank
- $\widehat{E}$: Galerkin product
- λ_N Approximation of largest eigenvalue of $\widehat{A}$.

Properties of $P_N\widehat{A}$

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

Theorem

Z, Y are arbitrary rectangular matrices with rank r.

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

- $\sigma(P_N\widehat{A})$ is similar to $\sigma(P_D\widehat{A})$

Multilevel Krylov method

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

Need to solve the coarse system with $\widehat{A}_H := \widehat{E}$.

- P_N is stable w.r.t. inexact solves.
- Applying P_N at the “second” level, i.e. use $P_{N,H}$

  instead of $\widehat{A}_H \widehat{x}_H = b_H$

  solve: $P_{N,H} \widehat{A}_H \widehat{x}_H = P_{N,H} b_H$

  using a Krylov method
- With inner Krylov iterations, P_N is i.g. not constant

  Use flexible Krylov subspace method (FGMRES, FQMR, ...)

Multilevel Krylov method

- The choice of Z and Y

  Sparsity of Z and Y;

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

  But not eigenvectors;

  $Y = Z$;
- About λ_N

  Expensive to compute, but an approximate is sufficient:

  $\rightarrow$ by Gerschgorin's theorem.

Numerical example: 2D Poisson equation

The 2D Poisson equation:

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

Discretization: finite differences.

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, $Z = [z_{ij}]$:

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

$Y = Z$

Numerical example: 2D Poisson equation

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

N	MK(2,2,2)	MK(4,2,2)	MK(6,2,2)	MK(4,3,3)	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

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

Observation:

- *h*-independent convergence
- Convergence of MK is comparable with MG.

2D Convection-diffusion equation

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, $Y = Z$

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

2D Convection-diffusion equation

Convergence results: relative residual $\leq 10^{-6}$
MK(4,2,2,2), 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 MK, FGMRES is used
- MG (with V-cycle, one pre- and post RB-GS smoothing and bilinear interpolation) does not converge

Observation:

- Almost h- and Pe-independent convergence

Conclusion - so far

Comparison of deflation methods

Multilevel Krylov method (MK method)

- preconditioner M
- flexible Krylov method
- multilevel structure (subspace systems)
- restrictions, prolongations, deflation vectors etc.
- estimates for λ_N.

Conclusion - so far

Comparison of deflation methods

Multilevel Krylov method (MK method)

- preconditioner M
- flexible Krylov method
- multilevel structure (subspace systems)
- restrictions, prolongations, deflation vectors etc.
- estimates for λ_N.
- h- and Pe-independent convergence

Conclusion - so far

Comparison of deflation methods

Multilevel Krylov method (MK method)

- preconditioner M
- flexible Krylov method
- multilevel structure (subspace systems)
- restrictions, prolongations, deflation vectors etc.
- estimates for λ_N.
- *h*- and *Pe*-independent convergence

Next:

- Helmholtz equation
- algebraic construction of restrictions, prolongations

  algebraic MK methods, AMK methods

MK for Helmholtz equation: MKMG method

$$\mathcal{A}u(\mathbf{x}) := -\left(\nabla \cdot \nabla - (1 - \hat{i}\alpha)\left(\frac{\omega}{c}\right)^2\right) u(\mathbf{x}) = f(\mathbf{x}) \qquad (1)$$

$$(\mathbf{x} \in \Omega = (0, L)^2),$$

equipped with radiation condition:

$$\frac{du}{dn} - \hat{i}\frac{\omega}{c}u = 0 \qquad (\Gamma = \partial\Omega).$$

- $\hat{i} = \sqrt{-1}$
- $0 \leq \alpha \leq 0.1$, attenuative (damping) factor
- $\omega = 2\pi f$ the angular frequency, with f the temporal frequency
- $c = c(\mathbf{x})$ the velocity data

Applications: aero- and marineacoustics, electromagnetics, seismics, etc.

MKMG method

Preconditioner for the Helmholtz equation:
Erlangga, Vuik, Oosterlee, 2004:

$$\mathcal{M} := -\nabla \cdot \nabla - (1 - \beta\hat{i}) \left(\frac{\omega}{c}\right)^2, \quad \beta = (0, 1].$$

Discretization of $\mathcal{M} \rightarrow M$.

M is inverted approximately by one multigrid iteration

MKMG method

Preconditioner for the Helmholtz equation:
Erlangga, Vuik, Oosterlee, 2004:

$$\mathcal{M} := -\nabla \cdot \nabla - (1 - \beta\hat{i}) \left(\frac{\omega}{c}\right)^2, \quad \beta = (0, 1].$$

Discretization of $\mathcal{M} \rightarrow M$.

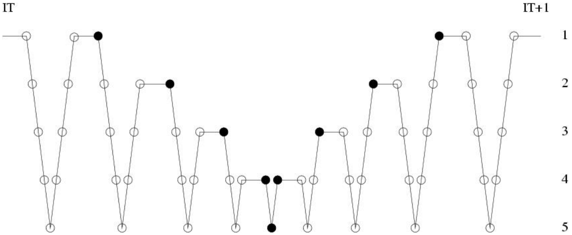
M is inverted approximately by one multigrid iteration

$$\widehat{A}_2 \quad := \quad RA_1 M_1^{-1} R^T \approx RA_1 R^T (RM_1 R^T)^{-1} RR^T.$$

$$R = Z^T = Y \qquad R^T = P = Z$$

MKMG method

Multilevel Krylov-Multigrid (MKMG) Cycle



•: projection steps, ○: multigrid cycle

MKMG method

constant $k := \omega L/c$.
g/w means "#grid points per wavelength".

MKMG(4,2,1)

	k:							
g/w	20	40	60	80	100	120	200	300
15	11	14	15	17	20	22	39	64
20	12	13	15	16	18	21	30	45
30	11	12	12	13	13	15	24	39

MKMG(8,2,1)

	k:							
g/w	20	40	60	80	100	120	200	300
15	11	14	14	17	18	21	27	39
20	12	13	15	14	15	16	20	28
30	11	12	12	12	13	14	15	19

- convergence almost independent of k (with MK(8,2,1))

AMK methods

So far geometric restrictions, prolongations, coarse grids use AMG techniques in the MK method

algebraic MK methods, AMK methods

We used

- Ruge-Stüben technique
- agglomeration-based technique

to build R and P, coarse grid matrix : RAP

AMK for 2D Convection-diffusion equation

2D Convection-diffusion equation with rotating flow

$$\nabla \cdot \left(\vec{c}(x,y)u(x,y)\right) - \Delta u(x,y) = f(x,y), \qquad \Omega \in (0,1)^2$$

homogeneous Dirichlet boundary conditions, $\vec{c}(x,y)$ is the prescribed velocity vector field

$$c_1(x,y) = -Cxy(1-x), \qquad c_2(x,y) = Cxy(1-y),$$

where $C = 80$.

Discretization: cell-centered finite volumes, uniform mesh

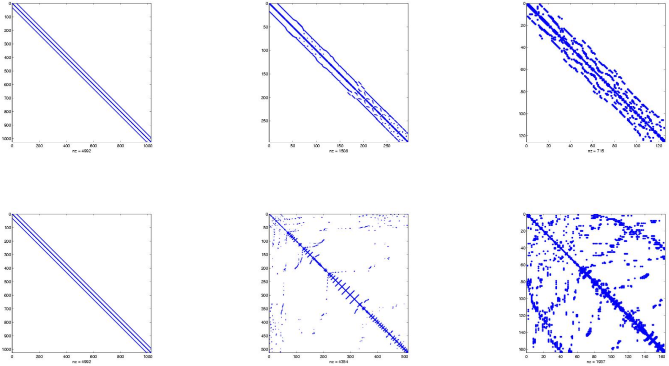
AMK for 2D Convection-diffusion equation

	# iterations for mesh:			
	16^2	32^2	64^2	128^2
AMK(4,2,1)-AG	16	18	19	21
AMK(4,2,1)-RS	11	13	20	37
AMK(4,2,1)-AG-M	16	18	19	21
AMK(4,2,1)-RS-M	11	13	19	35
AMG-RS	16	39	87	154

$$M = diag(A_l)$$

AMK for 2D Convection-diffusion equation

Aggregation and Ruge-Stüben techniques



Conclusion

Multilevel Krylov methods (MK methods)

Algebraic Multilevel Krylov methods (AMK methods)

- preconditioner M
- flexible Krylov method
- multilevel structure (subspace systems)
- restrictions, prolongations, deflation vectors etc.
- estimates for λ_N.

http://www.math.tu-berlin.de/˜nabben