GMRES approach with preconditioning

September 07 2017

Plan

Generalized Minimum Residual method[1]

Problem: solve $Ax = b$ with $A \in \mathbb{C}^{n\times n}$, $b \in \mathbb{C}^n$.

Method: GMRES (Krylov subspace iteration):

- Krylov subspace associated to A and $r_0 = b - Ax_0$

$$K_m(A, r_0) = \text{span}\{r_0, Ar_0, ..., A^{m-1}r_0\}.$$

- Approximate solution $x_m \in x_0 + K_m(A, r_0)$.
- Residual norm $||r_m|| = ||b - Ax_m||$ minimized over $x_0 + K_m(A, r_0)$.

[1] Y. Saad and M.H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput, 7 (1986), pp. 856–869.

Algorithm

Arnoldi algorithm:

- Orthogonal basis $V_m = [v_1, ..., v_m]$ of $K_m(A, r_0)$ by modified Gram-Schmidt process.
- Arnoldi relation:

$$AV_m = V_{m+1}\bar{H}_m$$

  with $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$ Hessenberg.
- $||r_m|| = ||r_0 - AV_m y_m|| = ||V_{m+1}(\beta e_1 - \bar{H}_m y_m)||.$

Least squares problem:

- Find $y \in \mathbb{R}^m$ such that

$$||r_m|| = \min_{y\in\mathbb{R}^m} ||\beta e_1 - \bar{H}_m y||.$$

- solved by QR factorisation of $\bar{H}_m = QR$.
- $y_m = R^{-1}Q^T\beta e_1$

Convergence properties

- GMRES converges in at most n steps with exact solution.
- $K_m \subset K_{m+1} \Rightarrow ||r_{m+1}|| \leq ||r_m||$.

- Convergence bound:
  If A est diagonalizable with $A = ZDZ^{-1}$ then

$$||r_m|| \leq ||r_0||||Z||\,||Z^{-1}|| \min_{q \in \mathbb{P}_m^0} \max_{\lambda \in \mathrm{sp}(A)} |q(\lambda)|.$$

- Breakdown $\Leftrightarrow$ convergence with exact solution
  $h_{j+1,j} = 0$ $\qquad$ $Ax_j - b = 0$

- Very large systems $\Rightarrow$ n steps impracticable:
  expensive computational cost and the storage requirements.

Restarting

Restarted GMRES(m):

- Dimension of the subspace is restricted to $m \ll n$.
- Algorithm restarted with updated initial guess $x_0 = x_m$.

- Orthogonal basis V_m is discarded.
- New basis of the Krylov subspace $K_m(A, r_m)$ generated from scratch.
- $\Rightarrow$ restarting slows down the convergence.

Accelerate convergence:

Implicitly Restarting Arnoldi[2], Augmented subspace method[3], Deflation, Block GMRES...

[2] R.B. Morgan, *Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1112–1135.
[3] R.B. Morgan, *A Restarted GMRES Method Augmented with Eigenvectors*, SIAM J. Matrix Ana. Appl, 16 (1995), pp. 1154–1171.

Block GMRES

- Solve $AX = B$ with $A \in \mathbb{C}^{n \times n}$ and $B \in \mathbb{C}^{n \times p}$.
- Block Krylov space

$$K_m(A, R_0) = \text{span}\{R_0, AR_0, ..., A^{m-1}R_0\} \subset \mathbb{R}^{n \times p}$$

  with $R_0 = B - AX_0 \in \mathbb{C}^{n \times p}$ and initial block solution $X_0 \in \mathbb{C}^{n \times p}$.
  $\Rightarrow$ richer Krylov subspace.
- Arnoldi relation:

$$A\mathcal{V}_m = \mathcal{V}_{m+p}\bar{\mathcal{H}}_m$$

  with $\bar{H}_m \in \mathbb{C}^{(m+p) \times m}$ band-Hessenberg.
- Block operation $\Rightarrow$ reduce memory access, improve performance.

Plan

Deflated Krylov subspace method

Deflations:

- Deflation with vectors: $A - \sigma u_1 v^T$ where $v^T u_1 = 1$.
  $\Rightarrow$ eigenvalues $\lambda_1 - \sigma, \lambda_2, ..., \lambda_p$.
- Deflation by projection: $PAx = Pb$.
- Augmented Krylov subspace method[4].

GMRES(m,k):

- k vectors $u_1, ..., u_k$ retained from GMRES(m).
- Augment the subspace with k vectors:

$$\text{span}\{r_0, Ar_0, ..., A^{m-k-1} r_0, u_1, ..., u_k\},$$

with structure of Krylov subspace[2].

Augmentation:

- Many choices possible for vectors u_i, $i = 1, ..., k$.
- Eigenvectors of A, Ritz-vectors (eigenvectors of H)...
- $\Rightarrow$ corresponding eigenvalues deflated from the spectrum.

[4] A. Chapman and Y. Saad, *Deflated and augmented Krylov subspace techniques*, Numer. Linear Algebra Appl., 4 (1997), pp. 43–66.

Approximate eigenvectors

- Small eigenvalues slows down the convergence
  $\Rightarrow$ deflation of k eigenvectors corresponding to k smallest eigenvalues.
- Rayleigh–Ritz procedure from Arnoldi:
  - extract Ritz vectors (eigenvectors of H_m).
  - effective at extracting exterior eigenvalues.
  - effective at finding well-separated eigenvalues.
- Interior Rayleigh–Ritz[5] (Shift and invert operator):
  eigenvalues shifted to the exterior of the spectrum.
  $\Rightarrow$ Interior Arnoldi method: eigenpairs (u_j, λ_j), $1 \leq j \leq k$, solutions of

$$(H_m + |h_{m+1,m}|^2 H_m^{-T} e_m e_m^T) u_j - \lambda_j u_j = 0.$$

- Clustered eigenvalues.

[5] R.B. Morgan and M.Zeng, *Harmonic projection methods for large non-symmetric eigenvalue problems*, Numer. Linear Algebra Appl., 5 (1998), pp. 33–55.

Algorithm: GMRES(m,k) with deflation

- Determine and compute k appropriated eigenpairs (u_i, λ_i).
- Orthogonalize $u_1, ..., u_k$.
- Orthogonalize $u_{k+1} = V_{m+1}^T r_m - \bar{H}_m y_m$ against previous $(v_i)_{i=1}^k$ to form v_{k+1}.
- Define V_{k+1} and $\bar{H}_k$.
- Loss of orthogonality for V_{k+1}.
  - v_{k+1} re-orthogonalized.
  - V_{k+1} re-orthogonalized[6].
- Approximative Arnoldi relation.
  $\Rightarrow$ Balance between orthogonality and Arnoldi relations.
- Arnoldi iteration: complete to V_{m+1} and $\bar{H}_m$.
- Solve least squares problem $||V_{m+1}^T r_m - \bar{H}_m y||$.

[6] S. Röllin and W. Fichtner, *Improving the accuracy of GMRES with deflated restarting*, SIAM J. Sci. Comput., 30 (2007), pp. 232–245.

Plan

Preconditioning

Left-preconditioning: $M^{-1}Ax = M^{-1}b$, with M the 'preconditioner'.
Right-preconditioning: system $Ax = b$ written as

$$\begin{aligned} AM^{-1}y &= b, \\ y &= Mx. \end{aligned}$$

Effective preconditioning:

- M easy to inverse.
- Ideally AM^{-1} close to the identity matrix.
- $\text{spec}(AM^{-1})$ compressed around 1, well clustered, away from 0.

Various preconditioners:

- From matrix splitting to iterative method...
- Jacobi preconditioner: diagonal of A, i.e. $M_J = \text{diag}(A)$.
- SSOR preconditioner: if $A = D - L - U$, then $M_\omega = (D - \omega L)D^{-1}(D - \omega U)$, with $\omega = 1$ typical.

Algorithm 1: GMRES(m) with right Preconditioning

Set $r_0 = b - Ax_0$, $\beta = ||r_0||$ and $v_1 = r_0/\beta$.
for $j = 1, ..., m$ **do**
  Compute $z_j := \boldsymbol{M^{-1}} v_j$
  Compute $w := Az_j$
  for $i = 1, ..., j$ **do**
    $h_{i,j} := (w, v_i)$
    $w := w - h_{i,j} v_i$
  Compute $h_{j+1,j} = ||w||$ and $v_{j+1} = w/h_{j+1,j}$.
Define $V_{m+1} := [v_1, ..., v_{m+1}]$.
Compute $x_m = x_0 + \boldsymbol{M^{-1}} V_m y_m$ where $y_m = \text{argmin}_y ||\beta e_1 - \bar{H}_m y||$
Restart: If satisfied stop, else set $x_0 \leftarrow x_m$ and goto 2.

- Orthogonal basis of the preconditioned subspace

$$\text{span}\{v_1, A\boldsymbol{M^{-1}} v_1, ..., (A\boldsymbol{M^{-1}})^{m-1} v_1\}.$$

- Residual unchanged (right-preconditioning).

Plan

Incomplete LU factorization (ILU)

- $M = L_{in}U_{in}$ where L_{in} and U_{in} are approximations of L and U of the LU factorization of A.
- Incomplete LU factorization:
  L_{in} and U_{in} pattern based on lower and upper pattern of A.
- $A = L_{in}U_{in} - R$ where R are element outside the pattern of A.

ILU with no fill: ILU(0)

- L_{in} and U_{in} have the same pattern as the lower and upper parts of A.
- Gaussian elimination dropping fill-in element outside the pattern of A:
  Let $NZ(A) = \{(i,j) \mid A_{i,j} \neq 0\}$, and $f_{i,j}$ fill-in element:
  if $(i,j) \notin NZ(A)$ then $f_{i,j} = 0$ dropped.
- SPD matrices: Incomplete Cholesky factorization IC(0).

Incomplete LU with level-of-fill: ILU(p)

- Augmented pattern with level of fill-in.
- Fill-in element $f_{i,j}$, level of fill-in *lev*.
- *lev* with respect to the pattern of A, is defined by induction.
  - Initially: for all $(i,j) \in NZ(A)$, $lev(A_{i,j}) = 0$,
  - *lev* assigned to each fill-in element during the Gaussian elimination.

$$lev(f_{i,j}) = \min(lev(f_{i,j}), \max(lev(f_{i,k}), lev(f_{k,j})) + 1)$$

$$\begin{pmatrix} 3 & & -1 & -1 & & -1 \\ & 2 & & -1 & & \\ -1 & & 3 & & -1 & \\ -1 & -1 & & 2 & & -1 \\ & & -1 & & 3 & -1 \\ -1 & & & -1 & -1 & 4 \end{pmatrix} \overset{lev}{\Rightarrow} \begin{pmatrix} 0 & & 0 & 0 & & 0 \\ & 0 & & 0 & & \\ 0 & & 0 & \mathbf{1} & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 0 & \mathbf{2} & 0 \\ & & 0 & \mathbf{2} & 0 & 0 \\ 0 & & \mathbf{1} & 0 & 0 & 0 \end{pmatrix}$$

- Gaussian elimination dropping fill-in element:
  Let $NZ_p(A) = \{(i,j) \mid lev(f_{i,j}) \le p\}$:
  if $(i,j) \notin NZ_p(A)$, then $f_{i,j} = 0$ dropped.

- Zero/small pivots $\Rightarrow$ instability ($||(LU)^{-1}||$ large).
- Forward/backward substitutions $\Rightarrow$ rounding errors.
- A ill-conditioned $\Rightarrow$ L and U more ill-conditioned.

- Indicator of instability[4]: $||(LU)^{-1}\epsilon||_\infty$, where e is a vector of all ones.
- Preprocessing coefficient of A by permutations, and scalings
  $\Rightarrow$ improve structure, augment diagonal dominance...

- **ILU variants:** drop tolerance, dual parameter, pivoting, diagonal compensation...

[4] E. Chow and Y. Saad, *Approximate inverse preconditioners for general sparse matrices*, Technical Report UMSI 94–101, University of Minnesota Supercomputer Institute, Minneapolis, 1994.

Incomplete LU with threshold: ILUT

Drop tolerance: ILUT(ϵ)

- Dropping element relies on magnitude rather than position.
- $|a_{i,j}| < \epsilon ||a_{i,:}|| \Rightarrow a_{i,j} = 0$, where $a_{i,:}$ is the ith row.
- Value of the drop tolerance?

Dual threshold: ILUT(k,ϵ)[5]

- Drop tolerance ϵ and fill-in k parameters.
- Limit number of fill-in in each row of incomplete L_{in} and U_{in} factors.
-
  - Drop elements as ILUT(ϵ),
  - then keep the largest k elements in each row of L_{in} and U_{in}.
- SPD matrices: Incomplete Cholesky ICT.

[5] Y. Saad, *ILUT: a dual threshold incomplete ILU factorization*, Technical Report, Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, 1992.

Variants of ILU

ILUT with pivoting: ILUTP

- ILUT can encounter zero pivot, whole row of zeros.
- Permutations and new ordering.

Modified ILU: MILU

- Threshold dropping ILUT and diagonal compensation.
- Dropped fill-ins are summed up and added to the diagonal of U.

MILU with pivoting: MILUP

ILUS (sparse), ILQ, ILUC (Crout)...

Plan

Block LU:

- If $A = L + D + U$ is block tridiagonal.
- Then $M = (L+\Delta)\Delta^{-1}(\Delta + U)$, where Δ is block-diagonal.

Polynomial preconditioning:

- $M^{-1} = p(A)$.

Normal equation:

- Consider the normal equation $A^T Ax = A^T b$ (NE).
- Precondition (NE) by Incomplete Choleski (IC), SSOR...
- IC factorization on the "shifted" matrix $A^T A + \alpha I$.
  - If lack of diagonal dominance.
  - issue: value of α?

Sparse Approximate inverse

Frobenius norm minimization:

- Find M_F such that $M_F = \min_{M \in S} ||AM - I||_F$ where S sparse matrix.
- Choice of the sparsity pattern? Sparse pattern of A, of A^k with $k \geq 2$.

Factorized approximate inverse:

- Factorization $A = LDU \Rightarrow$ factorized approximate inverse:

$$M^{-1} = \bar{Z}\bar{D}^{-1}\bar{W}^T \simeq A^{-1}$$

  with $\bar{Z}$ and $\bar{W}$ sparse approximations of U^{-1} and L^{-1}
- Inverses approximated by 'biconjugation' process: $W^T AZ = D$.

...

- More expensive and require more storage than ILU.
- Not prone to instability/pivot breakdown, easier implemented in parallel than ILU.
- ... ILU-type/SPAI complementary.

Block preconditioner:

- row operations $\rightarrow$ block operations $\Rightarrow$ faster.
- Block-Jacobi.
- **Block variants of ILU:**
  - BILU.
  - Block-fill ILU(p): BFILU(p)
  - Block ILUT: BILUT(ϵ).
  - Multilevel block ILU: BILUM.

Parallel preconditioner:

- **Parallel ILU:**
  - Graph coloring $\Rightarrow$ reordering $Q^T AQ$.
  - Multi-elimination: ILUM.
- **Multilevel techniques:**
  - Domain decomposition.
  - Multigrid method.

Adaptively preconditioner

- Preconditioner varies at each restart.
- Preconditioner with deflation.
- Eigenvalue translation[8]:

$$M^{-1} = (I_n + u_1 v_1^T)\dots(I_n + u_k v_k^T),$$

where u_i and v_i are associated with small eigenvalues to translate.

- Invariant subspace[9,10]:
  - $M^{-1} = I_n + U(|\lambda_n| T^{-1} - I_r)U^T$, where U is an invariant subspace corresponding to the r smallest eigenvalues of A and $T = U^T AU$.
    $\Rightarrow$ eigenvalues of AM^{-1} are $\lambda_{r+1}, ..., \lambda_n, |\lambda_n|$.
  - $M^{-1} = VH^{-1}V^T + WW^T$, where V span an invariant subspace of A associated with eigenvalues $\lambda_1, ..., \lambda_k$.
    $\Rightarrow$ eigenvalues of AM^{-1} are $\lambda_{k+1}, ..., \lambda_n, 1$.

[8] S.A. Kharchenko and A.Yu. Yeremin, *Eigenvalue translation based preconditioners for the GMRES(k) method*, Num. Lin. Alg. with Appli., 51–77, 1995.
[9] J. Erhel, K. Burrage and B. Pohl, *Restarted GMRES Preconditioned by Deflation*, J. Comput. Appl. Math., 69 (1996), pp. 303–318
[10] J. Baglama, D. Calvetti, L. Reichel and G.H. Golub, *Adaptively Preconditioned GMRES Algorithms*, SIAM J. Sci. Comput., 1996.

Plan

Flexible GMRES[11]

Algorithm 2: Flexible GMRES

Compute $r_0 = b - Ax_0$, $\beta = ||r_0||$ and $v_1 = r_0/\beta$.
for $j = 1, ..., m$ **do**
  Compute $z_j := \boldsymbol{M_j^{-1}} v_j$
  Compute $w := Az_j$
  for $i = 1, ..., j$ **do**
    $h_{i,j} := (w, v_i)$
    $w := w - h_{i,j} v_i$
  Compute $h_{j+1,j} = ||w||$ and $v_{j+1} = w/h_{j+1,j}$.
Define $\boldsymbol{Z_m := [M_1^{-1} v_1, ..., M_m^{-1} v_m]}$,
Compute $x_m = x_0 + \boldsymbol{Z_m} y_m$, where $y_m = \text{argmin}_y ||\beta e_1 - \bar{H}_m y||$
If satisfied stop, else set $x_0 \leftarrow x_m$ and goto 1.

[11] Y. Saad, *A Flexible Inner-outer Preconditioned GMRES Algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469.

- Preconditioner M_j varies at every step j.
- Iterative method as a preconditioner: SSOR, GMRES(inner-outer) ...
- $\Rightarrow$ Krylov subspaces with a higher dimension.

Arnoldi:

- Basis of the preconditioned Krylov subspace

$$\text{span}\{r_0, AM_1^{-1} r_0, ..., (AM_{m-1}^{-1})^{m-1} r_0\}.$$

- Arnoldi relation: $A\boldsymbol{Z_m} = V_{m+1}\bar{H}_m$.
- x_m minimizes $||r_m||$ over $x_0 + \text{span}\{\boldsymbol{Z_m}\}$.

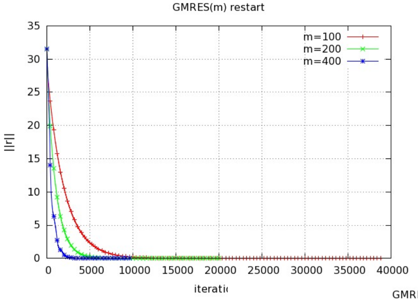
Plus:

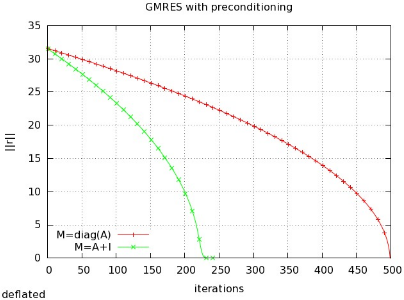
- Flexibility + spectral deflation: flexible GMRES-DR[12].
- Precondition the inner GMRES: LU-SSOR...

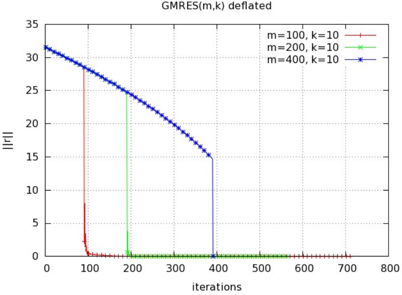
[12] L. Giraud, S. Gratton, X. Pinel and X. Vasseur, *Flexible GMRES with deflated restarting*, SIAM J. Sci. Comp., 32 (2010), pp. 1858–1878.

Plan

Discret Laplace operator ($n = 1000$, $\epsilon_{\text{stop}} = 10^{-7}$)







Plan

- M. Benzi, *Preconditioning Techniques for Large Linear Systems: A Survey*, J. Comput. Phys., 182 (2002), pp 418–477.
- A. Chapman, Y. Saad and L. Wigton, *High-order ILU preconditioners for CFD problems*, Technical Report UMSI 96/14, Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, 1996.
- O. Coulaud, L. Giraud, P. Ramet and X. Vasseur, *Deflation and augmented techniques in Krylov subspace methods for the solution of linear systems*, Research report, 2013
- R.B. Lehoucq and D.C. Sorensen, *Deflation techniques for an implicity restarted Arnoldi iteration*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 798–812.
- R.B. Morgan, *A restarted GMRES method augmented with eigenvectors*, SIAM J. Matrix Anal. Appl. 16 (1995), pp. 1154–1171.
- R.B. Morgan, *On restarting the Arnoldi method for large nonsymmetric eigenvalue problems*, Math. Comp., 65 (1996), pp. 1213–1230.
- R.B. Morgan, *GMRES with deflated restarting*, SIAM J. Sci. Comput., 24 (2002), pp. 20–37.
- Y. Saad, *Numerical methods for large eigenvalue problems*, SIAM, 2011.
- Y. Saad, *Preconditioned Krylov subspace methods for CFD applications*, W.G. Habashi, Wiley (1995), pp. 139–158.
- Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing, Boston, MA, 1996.
- Y. Saad, *Further analysis of minimum residual iterations*, Technical Report UMSI-97-14, University of Minnesota Supercomputer Institute, 1997.
- V. Simoncini and D.B. Szyld, *Flexible inner-outer Krylov subspace methods*, SIAM J. Num. Anal., 40 (2002), pp. 2219–2239.