THEORY OF RESIDUAL KRYLOV METHODS

Che Rung Lee
G. W. Stewart
Department of Computer Science
University of Maryland
BACKGROUND

- We will be concerned with Krylov sequence methods for finding eigenpairs of a matrix $A$ of order $n$.
- Typically such methods produce a sequence of orthonormal matrices
  $$U_k = (u_1 \ u_2 \ \cdots \ u_k)$$
  spanning the $k$th Krylov subspace $\mathcal{K}_k(A, u_1)$.
- Eigenpair approximations are recovered by the Rayleigh–Ritz procedure.
  - Form the Rayleigh quotient $B = U_k^* A U_k$.
  - Compute selected eigenpairs $(\mu, y)$ of $B$.
  - Test the Ritz pair $(\mu, z = U y)$ for convergence.
THE EFFECTS OF ERRORS

- Errors in the $u_k$ are fatal.

- For example, a single error in $u_2$ of size $\epsilon$ will cause the Ritz pairs to stagnate at a level proportional $\epsilon$.
  - Heuristically, the reason is that the error destroys the polynomial character of the Krylov sequence.

- An arbitrary starting subspace can be regarded as a big error.
  - If $U_2$ is not a Krylov subspace, the iteration will stagnate.
  - Even if the desired eigenvector $x$ is well represented in $U_2$. 

Expansion by Residuals
EXPANSION BY RESIDUALS

• In the Arnoldi process, $U_k$ is expanded by orthogonalizing $Au_k$.

• In the residual Krylov method, we compute a candidate Ritz pair $(\mu, z)$ approximating a target eigenpair $(\lambda, x)$ and orthogonalize the residual

$$r = A z - \mu z.$$

• Without error this is the same as the Arnoldi expansion.
  ○ In fact the residuals all line up with $u_{k+1}$.

• With error the two methods behave differently.
SHIFT-INVERT ENHANCEMENT

- In the Arnoldi method, one frequently iterates with 
  \((A - \sigma I)^{-1}\), where \(\sigma\) is a shift near the desired eigenpairs.
- In this case the system
  \[(A - \sigma I)w = u_k\]
  must be solved to full accuracy.
- The RK method orthogonalizes the vector
  \[w = (A - \sigma I)^{-1}(Az - \mu z),\]
  where \((z, \mu)\) is a candidate Ritz pair (wrt \(A\)).
  - This method is related to the Cayley transform method investigated by Lehoucq and Meerbergen.
- These mathematically equivalent methods behave differently in the presence of error.
SOME EXAMPLES

• The matrix is

\[ A = X \text{diag}(1, .95, .95^2, \ldots, .95^{99}) X^{-1}, \]

where \( X \) consists of random normal deviates.

• Four experiments.
  - Arnoldi with and without error.
  - Residual Krylov with target 1.0.
  - Residual Krylov switching targets midstream.
  - Residual Krylov with shift-invert.

• The relative error was \( 10^{-3} \).
THE RESULTS

1. Krylov with and without error

2. Residual Krylov: target 1.0

3. Residual Krylov: target 1.0 then 0.95

4. IPM and Residual Krylov
SOME HISTORY

- In writing my book on eigensystems, I introduced the Jacobi–Davidson method as a variant of Newton's method.
  - Jacobi–Davidson with error is an inexact Newton method and converges linearly.
- During a visit (spring 2001) to Utrecht, I tried some experiments and observed the Krylov-like superlinear convergence.
- Henk van der Vorst and I showed that the same held for simple shift-invert with errors.
- Gerard Sleijpen suggested we try residual expansion on simple Arnoldi without inverses.
ANALYSIS: A PREVIEW

- The problem is to explain the Krylov-like convergence.

- The strategy is to show that the $k$th RK subspace is an exact Krylov subspace of $\tilde{A}_k = A + E_k$.

- We will argue that that the $\tilde{A}_k$ contain increasingly accurate approximations to the target $x$.

- If the the Ritz vectors $\tilde{x}_k$ of the $\tilde{A}_k$ corresponding to $x$ exhibit typical convergence, then they converge to $x$. 
THE RESIDUAL KRYLOV EQUATION I

• When error is present, we orthogonalize

\[ r_k = \hat{r}_k + f_k = (AU_ky_k - \mu_kU_ky_k) + f_k \]

against \( U_k \) to get \( u_{k+1} \).

• Let \( g_k = U_k^*\hat{r}_k \). Then

\[ \gamma_ku_{k+1} = \hat{r}_k - U_kg_k + f_k^\perp = AU_ky_k - \mu_kU_ky_k - U_kg_k + f_k^\perp, \]

where \( f_k^\perp \) is the projection of \( f_k \) onto the orthogonal complement of \( U_k \).
THE RESIDUAL KRYLOV EQUATION II

\[ \gamma_k u_{k+1} = AU_k y_i - \mu_k U_k y_k - U_k g_k + f_k^\perp, \]

- Let
  - \( \hat{g}_i = (g_i^* \ \gamma_i \ 0_{k-i-1})^* \) and \( G_k = (\hat{g}_1 \cdots \hat{g}_{k-1} \ g_k) \).
  - \( Y_k \) be the upper triangular matrix consisting of the \( y_i \).

Then

\[ AU_k Y_k = U_k (Y_k M_k + G_k) + \gamma_k u_{k+1} e_k^* + F_k^\perp \]

where \( F_k^\perp = (f_1^\perp \cdots f_k^\perp) \) and \( M_k = \text{diag}(\mu_1, \ldots, \mu_k) \).

- Multiplying by \( Y_k^{-1} \) we get

\[ AU_k = U_k (Y_k M_k + G_k) Y_k^{-1} + \frac{\gamma_k}{\eta_k} u_{k+1} e_k^* + F_k^\perp Y_k^{-1} \]

where \( \eta_k \) is the last component of \( y_k \).
THE CANDIDATE RITZ VECTOR

\[ AU_k = U_k (Y_k M_k + G_k) Y_k^{-1} + \frac{\gamma_k}{\eta_k} u_{k+1} e_k^* + F_k^\perp Y_k^{-1}. \]

- The vector \( y_k \) is an eigenvector of the Rayleigh quotient

\[ U_k^T A U_k = (Y_k M_k + G_k) Y_k^{-1} + U_k^* F_k^\perp Y_k^{-1}. \]

- \( U_k^* F_k^\perp Y_k^{-1} \) is strictly lower triangular.

- The last column of \( U_k^* F_k^\perp \) is zero. Hence

\[ [(Y_k M_k + G_k) Y_k^{-1} + U_k^* F_k^\perp Y_k^{-1}] y_k = (Y_k M_k + G_k) Y_k^{-1} y_k, \]

so that \( y_k \) is also an eigenvector of \( (G_k + Y_k D_k) Y_k^{-1} \).
THE BACKWARD ERROR

\[ AU_k = U_k(Y_k M_k + G_k)Y_k^{-1} + \frac{\gamma_k}{\eta_k} u_{k+1} e_k^* + F_k^\perp Y_k^{-1}. \]

If we set

\[ E_k = -F_k^\perp Y_k^{-1} U_k^\ast \]

Then

\[ \tilde{A}_k U_k = (A + E_k) U_k = U_k (Y_k M_k + G_k) Y^{-1} + \frac{\gamma_k}{\eta_k} u_{k+1} e_k^*. \]

Thus \( \mathcal{R}(U_k) \) is a Krylov subspace of a perturbed \( A \).

\( \circ \) However, the perturbation is not small.

Note that the primitive Ritz vector \( \tilde{y}_k \) of \( \tilde{A}_k \) and \( y_k \) of \( A_k \) are the same.
PROPERTIES OF \( E \)

\[
E_k = -F_k^\perp Y_k^{-1} U_k^* 
\]

\(\diamondsuit\)

- We now make the assumption that the error \( f_k \) satisfies
  \[
  \|f_k\| \leq \epsilon \|r_k\|. 
  \]

- To analyze the residual Krylov method, we need to make the following assumption.
  
  There is a constant \( C_1 \) such that
  \[
  \|E_k\| \leq C_1 \epsilon 
  \]

- We are not able to prove this, but empirically it appears to hold.

- The problem is that \( Y_k^{-1} \) blows up as the process converges.
\[ E_k = -F_k^\perp Y_k^{-1}U_k^* \]

- \( E_k \) consists of two parts: \( F_k^\perp \) and \( Y^{-1} \).

- The \( i \)th column of the matrix \( F \) is bounded by \( \epsilon \| r_i \| \equiv \epsilon \rho_i \).

Hence we can represent the norms of the \( f_i \) by

\[ \epsilon e^T \text{diag}(\rho_i). \]
HEURISTIC JUSTIFICATION II

\[ E_k = -F_k \perp Y_k^{-1} U_k^* \]

- We have \( y_{ij} \rightarrow u_i^T x \equiv \beta_i \)

- Suppose (by some miracle) the convergence is immediate. Then

\[ Y_k = \text{diag}(\beta_k) \ast \text{triu(ones(k))}. \]

Hence

\[ Y_k^{-1} = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ \cdots & \cdots \end{pmatrix} \text{diag}(\beta_i)^{-1}. \]
HEURISTIC JUSTIFICATION III

\[ E_k = -F_k^\perp Y_k^{-1} U_k^* \]

\[ \epsilon \mathbf{e}^T \begin{pmatrix} \rho_1/\gamma_1 & -\rho_1/\gamma_2 \\ \rho_2/\gamma_2 & -\rho_2/\gamma_3 \\ \rho_3/\gamma_3 & -\rho_3/\gamma_4 \\ \vdots & \vdots \end{pmatrix} \]

- Hence \( F_k^\perp \) should have the structure

- If the \( \rho_i \) and \( \gamma_i \) are proceeding apace to zero, \( E_k \) will be of order \( \epsilon \).

- The actual \( Y_k^{-1} \) has elements that shrink as one proceeds to the northeast.
HITTING THE TARGET

• Let the target be \((\lambda, x)\), where \(\lambda\) is simple.

• By a rather technical argument, we can show that there are constants \(c\) and \(C_2\) such that if \(\epsilon \leq c\) then there is a unique eigenpair \((\tilde{\lambda}_k, \tilde{x}_k)\) of \(\tilde{A}\) satisfying

\[
\|\tilde{x}_{k+1} - x_k\| \leq C_2\|r_k\|\epsilon
\]

• This says that if \(C_2\epsilon < 1\) then \(\tilde{A}_{k+1}\) contains a better approximation \(\tilde{x}_{k+1}\) to \(x\) than \(\|r_k\|\) indicates.
UNIFORM CONVERGENCE OF ARNOLDI

- Unfortunately, we are working with the Ritz vector \( \tilde{z}_k \) of \( A_k \), not \( \tilde{x}_k \). To handle this we will assume the following result.

- Let \( \hat{A} = A + E \). Then there are constants \( C_3 \) and \( \kappa_i \to 0 \) such that if \( \| E \| \leq C_3 \), then:
  - There is a unique eigenpair \( (\hat{\lambda}, \hat{x}) \) approximating \( (\lambda, x) \).
  - Moreover, the corresponding Ritz vectors \( \hat{z}_k \) satisfy  
    \[ \| \hat{z}_k - \hat{x} \| \leq \kappa_k. \]
PUTTING IT TOGETHER

For $\epsilon$ sufficiently small
\[
\|x - z_{k+1}\| = \|x - \tilde{z}_{k+1}\| \leq \|x - \tilde{x}_{k+1}\| + \|\tilde{x}_{k+1} - \tilde{z}_{k+1}\|
\]
\[
= C\|r_k\| \epsilon + \kappa_k
\]

Hence
\[
\|r_{k+1}\| \leq 6(C\|r_k\| \epsilon + \kappa_k)
\]
It follows that if $6C\epsilon < 1$ and $\sum_{k=1}^{\infty} \kappa_k < \infty$, then
\[
\|r_k\| \to 0.
\]

- The rate is essentially the slower of the rates of approach of $(6C\epsilon)^k$ and $\kappa_k$ to zero.
  - For epsilon small, we get essentially Krylov-like convergence.
THE SHIFT-INVERT ALGORITHM

- Orthogonalize \((A - \sigma I)^{-1}(Az_k - \mu_k z_k)\) against \(U_k\).
  - The residual \(Az_k - \mu_k z_k\) must be calculated to full accuracy.
  - However, errors can be tolerated in the solves.
- A similar convergence result holds for the shift-invert algorithm.
- Once the target eigenpair has been found, one can switch to another target.
  - While the first target is being found, an approximation to the second is emerging—but only to the level of \(\epsilon\).
  - Convergence to the old target has not been observed.
INITIAL SUBSPACES

• Switching targets is an example of an initial subspace that is not Krylov.
  ○ However, it is a well prepared subspace.

• An arbitrary subspace may not initially work, even when it contains a modest approximation to the target vector.
  ○ The problem seems to be that the Rayleigh–Ritz process may not reproduce the approximation.
  ○ After a while the typical residual Krylov behavior sets in.

• This area needs further consideration.