Intervals, Tridiagonal Matrices and the Lanczos Method

Wolfgang Wülling

August, 21th 2007

HARRACHOV 2007
Computational Methods with Applications
August 19 - 25, 2007, Czech Republic
harrachov@cs.cas.cz

1. Lanczos Algorithm
   - Exact Arithmetic
   - Finite Precision Arithmetic
   - The residual quantity
2. Clustered Ritz values
   - The Conjecture
   - Estimates for Residual Quantity
3. Conclusion

Symmetric Eigenvalue Problem

Given a large, sparse and symmetric matrix $A \in \mathbb{R}^{N \times N}$, find (approximations to) eigenvalues λ and eigenvectors u, i.e.

$$Au = \lambda u$$

Lanczos Method extracts solution / approximations (θ, z) to eigenpair (λ, u) via orthogonal projection from Krylov subspace(s)

$$Az - \theta z \quad \perp \quad \mathcal{K} = \text{span}(q, Aq, A^2 q, ..)$$

- Ritz value $\theta \in \mathbb{R}$: $\theta \approx \lambda$.
- Ritz vector $z \in \mathcal{K}$, $z \neq 0$: $z \approx u$

Lanczos Recursion, Matrix Notation

$$
\begin{aligned}
AQ_k &= Q_k T_k + \beta_{k+1} q_{k+1} e_k^* & (1)\\
Q_k^* Q_k &= I_k & (2)
\end{aligned}
$$

$$
T_k = \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \beta_k \\ & & & \beta_k & \alpha_k \end{bmatrix} \in \mathbb{R}^{k \times k}
$$

- $Q_k = [q_1, \ldots, q_k] \in \mathbb{R}^{N \times k}$ orthogonal Lanczos vectors
- algorithm terminates with a $\beta_d = 0$, $d \leq N + 1$.

Ritz approximations

Spectral decomposition of T_k

$$T_k s_\bullet = \theta_\bullet s_\bullet$$

$$T_k = S_k \text{diag}(\theta_1, ..., \theta_k) S_k^*, \quad S_k^* S_k = S_k S_k^* = I_k$$

- Ritz values at step k are eigenvalues $\theta_\bullet$ of T_k
- Ritz vectors are $z_\bullet = Q_k s_\bullet$, where $s_\bullet$ eigenvector of T_k.

Qualitiy of Ritz approximations, $\delta_{k,\bullet}$

Quality of Ritz approximation to eigenvalue can be controlled by easily computable residual quantity:

$$\delta_{k,\bullet} := \beta_{k+1}|s_{k,\bullet}|$$

where $s_\bullet = \begin{pmatrix} \vdots \\ s_{k,\bullet} \end{pmatrix}$, $s_{k,\bullet}$ bottom element of eigenvector of T_k.

$$\min|\lambda - \theta_\bullet| \leq \frac{||Az_\bullet - \theta_\bullet z_\bullet||}{||z_\bullet||} = ||Az_\bullet - \theta_\bullet z_\bullet|| = \delta_{k,\bullet}$$

Stabilized Ritz values

Theorem (C. Paige)

Persistence Theorem

$$\min_{\mu} |\theta_\bullet - \mu| \leq \delta_{k,\bullet},$$

μ Ritz value in subsequent step.

- **Definition:** A Ritz value $\theta_\bullet$ is called stabilized to within $\delta_{k,\bullet}$. If $\delta_{k,\bullet}$ is small $\theta_\bullet$ is called stabilized.
- Ritz value can stabilize only close to an eigenvalue

Perturbed Lanczos Recursion

Rounding errors lead to perturbed Lanczos recursion

$$AQ_k = Q_k T_k + \beta_{k+} q_{k+1} e_k^* + \text{rounding errors} \quad (3)$$

$$Q_k^* Q_k = I_k + \text{rounding errors} \quad (4)$$

- Lanczos vectors q_k may lose orthogonality – even for small number of iterations, $k \ll N$.
- It is not guaranteed that Algorithm terminates (with $\beta_d = 0$), might run *forever*

Quality of Ritz approximations, $\delta_{k,\bullet}$

Quality of Ritz approximations in f.p.computations:

$$\delta_{k,\bullet} = \beta_{k+1}|s_{k,\bullet}|$$

controls convergence also in f.p. computations – thanks to

Theorem (C. Paige)

At any step k of the (f.p.) Lanczos algorithm the following is valid

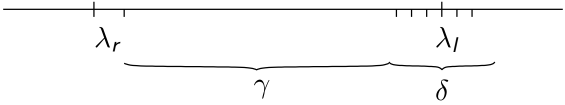
$$\min|\lambda - \theta_\bullet| \quad \leq \quad \max\{2.5(\delta_{k,\bullet} + \textit{small in } \epsilon), \textit{small in } \epsilon\} \qquad (5)$$

Interim Conclusion

Residual Quantity

The residual quantity $\delta_{k,\bullet} = \beta_{k+1}|s_{k,\bullet}|$ controls convergence in exact and finite precision computations!

Tight, well separated cluster



If $\delta \ll \gamma$: tight, well separated cluster.

The Conjecture

- Rounding errors: Multiple copies of Ritz approximations to single eigenvalue are generated.
- Ritz values cluster closely to eigenvalues of A.

Conjecture (Strakoš, Greenbaum, 1992)

For any Ritz value $\theta_\bullet$ being part of a tight, well separated cluster, the value of $\delta_{k\bullet}$ is small, i.e $\delta_{k,\bullet} \ll 1$.
Furthermore, for any Ritz value in the cluster $\delta_{k\bullet}$ is proportional to

$$\sqrt{\frac{\delta}{\gamma}} = \sqrt{\frac{\text{cluster diameter}}{\text{gap in spectrum}}}$$

Strakoš Matrix, Example 1

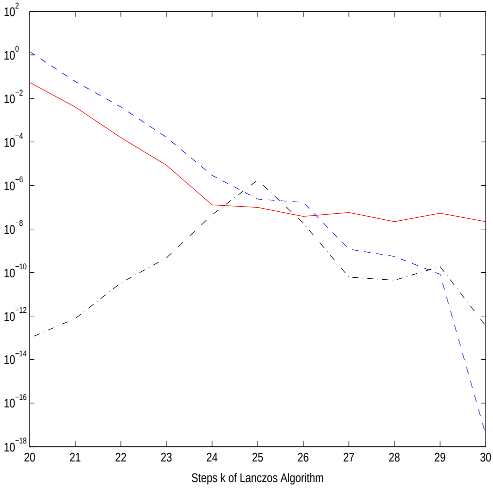
Example 1(Strakoš) Apply Lanczos Algorithm to diagonal matrix A with eigenvalues

$$\lambda_\nu = \lambda_1 + \frac{\nu - 1}{N - 1}(\lambda_N - \lambda_1)\rho^{N-\nu}, \quad \nu = 2, ..., N-1, \tag{6}$$

where $\lambda_1 = 0.1$, $\lambda_N = 100$, $\rho = 0.7$ and $N = 24$. We look at cluster of two Ritz values close to $\lambda_{22} \approx= 44.7944$.

Strakoš Matrix, Example 1

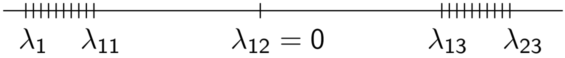
Example 1: Cluster of two Ritz values close to $\lambda_{22} \approx 44.7944$.



Counterexample

Consider diagonal matrix $A \in \mathbb{R}^{23}$ with eigenvalues

$$\lambda_1 = -100, \;\; \lambda_{j+1} = \lambda_j + 0.1 \text{ for } j = 1, \ldots, 10$$
$$\lambda_{12} = 0 \text{ and } \lambda_{12+j} = -\lambda_{12-j} \text{ for } j = 1, \ldots, 11.$$



Starting vector $q_1 = \frac{1}{\sqrt{23}}(1, 1, ..., 1)^*$:

Counterexample

k	$\delta_{k,\text{first}}$	$\delta_{k,\text{second}}$	$\sqrt{\frac{\delta}{\gamma}}$
9	4.6614×10^{-7}	-	-
10	7.0356×10^{1}	7.0356×10^{1}	9.7023×10^{-05}
11	2.3067×10^{-9}	-	-
12	7.0365×10^{1}	7.0347×10^{1}	6.8258×10^{-06}
13	1.0724×10^{-11}	-	-
14	6.7432×10^{1}	7.3162×10^{1}	4.6544×10^{-07}
15	5.6899×10^{-14}	-	-
16	5.3119×10^{1}	6.0079×10^{1}	3.0700×10^{-08}
17	4.1795×10^{-14}	-	-
18	3.7987×10^{-1}	3.4773×10^{-1}	1.8764×10^{-09}
19	6.5049×10^{-12}	-	-
20	1.5658×10^{-3}	4.8528×10^{-4}	2.0156×10^{-10}

- Last example: Clustered Ritz values close to eigenvalue, but $\delta_{k,\bullet} \gg \sqrt{\frac{\delta}{\gamma}}$ for all Ritz value in the cluster.
- But in (Strakoš, Greenbaum, 1992), for some particular cases estimates obtained with $\delta_{k,\bullet} \leq \sqrt{\frac{\delta}{\gamma}}\mathcal{O}(||A||)$.
- Idea from (Strakoš, Greenbaum, 1992): Estimate $\delta_{k,\bullet}$ using only information about Ritz values in several consecutive steps.
- Since

$$\delta_{k,\bullet} = \underbrace{\beta_{k+1}}_{=\mathcal{O}(||A||)} \underbrace{|s_{k,\bullet}|}_{\text{eigenvector element}}$$

concentrate on bottom elements of T_k's eigenvectors.

Constant Cluster Size: $k-1$, k

Theorem

Suppose that number of Ritz values in a tight, well separated cluster is constant at steps $k-1$ and k. Then:

$$\sum_{cluster} (s_{k,\bullet})^2 < 3\frac{\delta}{\gamma - \frac{\delta}{\gamma}}.$$

Outline of the proof

Proof.

Observe, at the kth step, for any Ritz value θ (in the cluster) we have

$$\frac{\psi_k(\theta)}{\psi_{k-1}(\theta)} = 0, \quad \frac{\psi_k'(\theta)}{\psi_{k-1}(\theta)} \neq 0.$$

□

Outline of the proof

Proof.

Observe, at the kth step, for any Ritz value θ (in the cluster) we have

$$\frac{\psi_k(\theta)}{\psi_{k-1}(\theta)} = 0, \quad \frac{\psi_k'(\theta)}{\psi_{k-1}(\theta)} \neq 0.$$

Hence, use formula for $s_{k,\bullet}$ and apply Residue Theorem:

$$\sum_{\text{cluster}} (s_{k,\bullet})^2 = \sum_{\text{cluster}} \frac{\psi_{k-1}(\theta)}{\psi_k'(\theta)} = \frac{1}{2\pi i} \int_\Gamma \frac{\psi_{k-1}(z)}{\psi_k(z)} \, dz.$$

and estimate line integral. □

Constant Cluster Size: k, $k+1$

Theorem

Suppose that number of Ritz values in a tight, well separated cluster is constant at steps k and $k+1$. Then:

$$\sum_{cluster} (s_{k,\bullet})^2 < 3 \, \frac{\mathcal{O}(||A||)}{\beta_{k+1}^2} \, \frac{\delta}{\gamma - \frac{\delta}{2}}$$

Decreasing or Increasing Cluster Size

Theorem

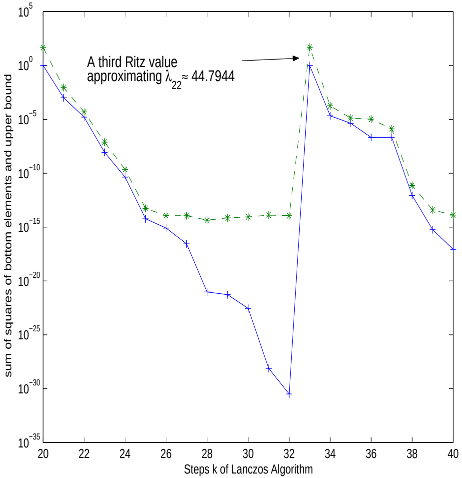
a) Decreasing: Suppose c Ritz values at step $k-1$ and $c-1$ Ritz values at step k form cluster, then:

$$\sum_{cluster} (s_{k,\bullet})^2 < \frac{4}{3}\sqrt{3}\frac{\delta^2}{\left(\gamma - \frac{\delta}{2}\right)^2}.$$

b) Increasing: Let c Ritz values at step k and $c+1$ Ritz values at step $k+1$ form cluster, then:

$$\sum_{cluster} (s_{k,\bullet})^2 < \frac{4}{3}\sqrt{3}\,\frac{\mathcal{O}(||A||)}{\beta_{k+1}^2}\,\frac{\delta^2}{\left(\gamma - \frac{\delta}{2}\right)^2}.$$

Example 1, continued



Stabilized Cluster

Corollary

For a tight well separated cluster with diameter δ and spectral gap γ we have

$$\delta_{k,\min} \leq \mathcal{O}(||A||) \begin{cases} \textit{no estimate} & , \quad c_k = c_{k-1} + 1 \\ \frac{1}{\sqrt{c_k}} \quad \sqrt{\frac{\delta}{\gamma - \frac{\delta}{2}}} & , \quad c_k = c_{k-1} \\ \frac{1}{\sqrt{c_k}} \quad \frac{\delta}{\gamma - \frac{\delta}{2}} & , \quad c_k = c_{k-1} - 1 \end{cases}$$

or

$$\delta_{k,\min} \leq \mathcal{O}(||A||) \begin{cases} \textit{no estimate} & , \quad c_k = c_{k+1} + 1 \\ \frac{1}{\sqrt{c_k}} \quad \sqrt{\frac{\delta}{\gamma - \frac{\delta}{2}}} & , \quad c_k = c_{k+1} \\ \frac{1}{\sqrt{c_k}} \quad \frac{\delta}{\gamma - \frac{\delta}{2}} & , \quad c_k = c_{k+1} - 1 \end{cases}$$

What we have seen ...

- Tight, well separated cluster must approximate an eigenvalue of A.
- At least one Ritz value will always remain close to an eigenvalue (persistence theorem).
- Any tight well separated cluster stabilized.
- in most cases: $\delta_{k,\bullet}$ small for all Ritz values in a tight, well separated cluster; only exception is an alternating number of Ritz values in the cluster.
- Intermediate cluster can comprise at most one Ritz value, see also (Knizhnerman, 1995).

... and the main ingredients of the proofs:

- **Idea** to express / estimate anything in terms of Ritz values – taken from (Strakoš, Greenbaum (1992))
- Residue Theorem
- Interlacing of Ritz values (not on these slides, but needed for the proofs).