On the convergence of the Arnoldi process for computing eigenvalues

Gérard MEURANT

June 2015

The Ritz values are the eigenvalues of H_k which is defined by

$$AV_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^T$$

Let us assume that H_k is diagonalizable. Let θ_i and $y^{(i)}$ be a Ritz value and a corresponding eigenvector of H_k (not necessarily normalized). Then,

$$AV_k y^{(i)} = \theta_i V_k y^{(i)} + h_{k+1,k} y_k^{(i)} v_{k+1}$$

Let $z^{(i)} = V_k y^{(i)}$ be the approximate eigenvector of A

$$Az^{(i)} - \theta_i z^{(i)} = h_{k+1,k} y_k^{(i)} v_{k+1}$$

It is orthogonal to the Krylov subspace of dimension k. Hence

$$\theta_i = (z^{(i)})^* A z^{(i)} = (y^{(i)})^* H_k y^{(i)}$$

since $z^{(i)}$ is in the Krylov subspace of dimension k

Assume A normal, $A = X\Lambda X^*$
We have

$$H_k = U_k C^{(k)} U_k^{-1}$$

with U_k upper triangular, $C^{(k)}$ a companion matrix and $K_k = V_k U_k$. Then

$$\begin{aligned}
0 = \det(H_k - \theta I) &= \det(V_k^* A V_k - \theta I), \\
&= \det(U_k^{-*} K_k^* A K_k U_k^{-1} - \theta I), \\
&= \det(U_k^{-*} [K_k^* A K_k - \theta U_k^* U_k] U_k^{-1}), \\
&= |\det(U_k^{-1})|^2 \det(K_k^* X \Lambda X^* K_k - \theta M_k), \\
&= \frac{1}{\det(M_k)} \det(\mathcal{V}_k^* D_c^* \Lambda D_c \mathcal{V}_k - \theta \mathcal{V}_k^* D_c^* D_c \mathcal{V}_k).
\end{aligned}$$

with $M_k = U_k^* U_k = K_k^* K_k$, D_c diagonal matrix, $c = X^* v$, $v = V_k e_1$, $\mathcal{V}_k$ Vandermonde matrix

Let $B = D_c \mathcal{V}_k$

$$\det(B^* \Lambda B - \theta B^* B) = \det(B^*(\Lambda - \theta I)B) = 0$$

Let $G = (\Lambda - \theta I)B$, we have $\det(B^* G) = 0$
We use the Cauchy-Binet formula

$$\det(B^* G) = \sum_{I_k} \overline{\det(B_{I_k,:})} \det(G_{I_k,:}) = 0$$

$I_k = \{(i_1, \ldots, i_k) | 1 \le i_1 < \cdots < i_k \le n\}$. We have

$$\det(B_{I_k,:}) = c_{i_1} \cdots c_{i_k} \prod_{i_1 \le i_p < i_q \le i_k} (\lambda_{i_q} - \lambda_{i_p})$$

and, since $\Lambda - \theta I$ is diagonal,

$$\det(G_{I_k,:}) = (\lambda_{i_1} - \theta) \cdots (\lambda_{i_k} - \theta)\, c_{i_1} \cdots c_{i_k} \prod_{i_1 \le i_p < i_q \le i_k} (\lambda_{i_q} - \lambda_{i_p})$$

Therefore, we have the new polynomial

$$\sum_{I_k} (\lambda_{i_1} - \theta) \cdots (\lambda_{i_k} - \theta) \, |c_{i_1}|^2 \cdots |c_{i_k}|^2 \prod_{i_1 \le i_p < i_q \le i_k} |\lambda_{i_q} - \lambda_{i_p}|^2 = 0$$

whose roots are the Ritz values

The product $(\lambda_{i_1} - \theta) \cdots (\lambda_{i_k} - \theta)$ is a polynomial of degree k in θ whose roots are k of the eigenvalues of A

We have

$$(\lambda_{i_1} - \theta) \cdots (\lambda_{i_k} - \theta) = \sum_{j=0}^{k} (-1)^{k-j} e_{(j)}(\lambda_{i_1}, \ldots, \lambda_{i_k}) \theta^{k-j}$$

where $e_{(j)}$ is a symmetric elementary polynomial

For the characteristic polynomial of H_k we have to divide by the coefficient of θ^k which is

$$(-1)^k \sum_{I_k} |c_{i_1}|^2 \cdots |c_{i_k}|^2 \prod_{i_1 \le i_p < i_q \le i_k} |\lambda_{i_q} - \lambda_{i_p}|^2 = (-1)^k \det(M_k)$$

If we want a value 1 at the origin (the FOM residual polynomial) we have to divide by the constant coefficient which is

$$\sum_{I_k} \lambda_{i_1} \cdots \lambda_{i_k} \, |c_{i_1}|^2 \cdots |c_{i_k}|^2 \prod_{i_1 \le i_p < i_q \le i_k} |\lambda_{i_q} - \lambda_{i_p}|^2$$

The coefficient of θ^{k-j}, $j = 0, \ldots, k$ is

$$(-1)^{k-j} \sum_{I_k} e_{(j)}(\lambda_{i_1}, \ldots, \lambda_{i_k}) \, |c_{i_1}|^2 \cdots |c_{i_k}|^2 \prod_{i_1 \le i_p < i_q \le i_k} |\lambda_{i_q} - \lambda_{i_p}|^2$$

The coefficients can also be obtained when A is not normal but diagonalizable but the formulas are more intricate

We can do the same thing for the harmonic Ritz values

When A is normal the polynomial in ζ is

$$\sum_{I_k} (|\lambda_{i_1}|^2 - \zeta \overline{\lambda_{i_1}}) \cdots (|\lambda_{i_k}|^2 - \zeta \overline{\lambda_{i_k}}) \, |c_{i_1}|^2 \cdots |c_{i_k}|^2 \prod_{i_1 \le i_p < i_q \le i_k} |\lambda_{i_q} - \lambda_{i_p}|^2 = 0$$

The coefficient of ζ^{k-j}, $j = 0, \dots, k$ is

$$(-1)^{k-j} \sum_{I_k} e_{(j)}(\lambda_{i_1}, \dots, \lambda_{i_k}) \, \overline{\lambda_{i_1}} \cdots \overline{\lambda_{i_k}} \, |c_{i_1}|^2 \cdots |c_{i_k}|^2 \prod_{i_1 \le i_p < i_q \le i_k} |\lambda_{i_q} - \lambda_{i_p}|^2$$

From this we can obtain the characteristic polynomial and the GMRES residual polynomial if we divide by the appropriate coefficient

We are interested in finding bounds for the distance of an eigenvalue of A to the Ritz values at a given iteration of the Arnoldi algorithm

We would like to stress that these bounds are of purely theoretical interest to be able to gain more insights in Arnoldi's convergence. They cannot be used to stop the iterations since they contain quantities which are either unknown or not computable for practical problems

To obtain bounds for the distances of Ritz values to eigenvalues we have to do a change of variable in the characteristic polynomial

For the moment we consider a monic polynomial

$$p(z) = z^k + \alpha_{k-1} z^{k-1} + \cdots + \alpha_1 z + \alpha_0$$

and its roots θ_j
Let σ_j, $j = 1, \ldots, k$ be the singular values of the companion matrix. Then

$$\sigma_n \leq |\theta_j| \leq \sigma_1, \quad \forall j = 1, \ldots, k$$

Since we can compute the singular values of $C^{(k)}$, it yields

$$\left[\frac{1}{2}\left(\|\alpha\|^2 + 1 - ((\|\alpha\|^2 + 1)^2 - 4|\alpha_0|^2)^{\frac{1}{2}}\right)\right]^{\frac{1}{2}} \leq |\theta_j|$$

$$|\theta_j| \leq \left[\frac{1}{2}\left(\|\alpha\|^2 + 1 + ((\|\alpha\|^2 + 1)^2 - 4|\alpha_0|^2)^{\frac{1}{2}}\right)\right]^{\frac{1}{2}}$$

with $\|\alpha\|^2 = \sum_{i=0}^{k-1} |\alpha_i|^2$

- Cauchy's bounds

$$\frac{|\alpha_0|}{\max_{i=1,\dots,k-1}(1, |\alpha_0| + |\alpha_i|)} \le |\theta_j| \le \max_{i=1,\dots,k-1}(|\alpha_0|, 1 + |\alpha_i|)$$

- Montel's bounds,

$$\frac{|\alpha_0|}{\max(|\alpha_0|, 1 + |\alpha_1| + \cdots + |\alpha_{k-1}|)} \le |\theta_j| \le \max(1, |\alpha_0| + \cdots + |\alpha_{k-1}|)$$

- Carmichael and Mason's bounds,

$$\frac{|\alpha_0|}{(1 + |\alpha_0|^2 + \cdots + |\alpha_{k-1}|^2)^{\frac{1}{2}}} \le |\theta_j| \le (1 + |\alpha_0|^2 + \cdots + |\alpha_{k-1}|^2)^{\frac{1}{2}}$$

- De Terán, Dopico and Pérez's bounds

$$\frac{|\alpha_0|}{\max_{i=2,\ldots,k-1}(1+|\alpha_1|, |\alpha_0|(1+|\alpha_i|))} \le |\theta_j|$$

$$|\theta_j| \le \max_{i=1,\ldots,k-2}\left(1+\frac{|\alpha_i|}{|\alpha_0|}, |\alpha_0|+|\alpha_{k-1}|\right)$$

We do a change of variable $z \to z+\lambda_i$. The roots of the new polynomial $\hat{p}_i(z)$ are $\theta_j - \lambda_i$

$$\hat{p}_i(z) = z^k + \gamma_{k-1}z^{k-1} + \cdots + \gamma_1 z + \gamma_0$$

In the normal case the coefficients γ_ℓ are

$$\frac{1}{d_k}\sum_{j=1}^{k}(-1)^j\binom{j}{\ell}\lambda_i^{j-\ell}\qquad\sum_{I_k} e_{(k-j)}(\lambda_{i_1},\ldots,\lambda_{i_k})\,|c_{i_1}|^2\cdots|c_{i_k}|^2\times$$

$$\prod_{i_1\le i_p<i_q\le i_k}|\lambda_{i_q}-\lambda_{i_p}|^2$$

with

$$d_k=\sum_{I_k}|c_{i_1}|^2\cdots|c_{i_k}|^2\prod_{i_1\le i_p<i_q\le i_k}|\lambda_{i_q}-\lambda_{i_p}|^2$$

and the coefficient γ_0 is

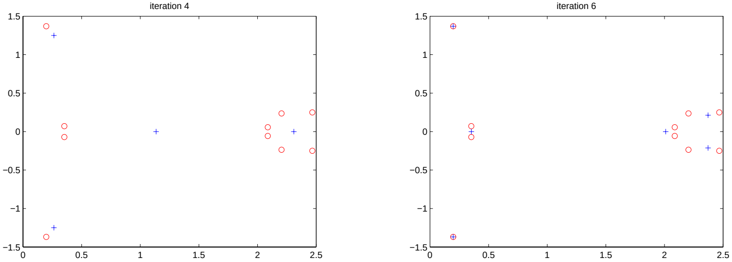
$$\alpha_0+\frac{1}{d_k}\sum_{j=1}^{k}(-1)^j\lambda_i^j\qquad\sum_{I_k} e_{(k-j)}(\lambda_{i_1},\ldots,\lambda_{i_k})\,|c_{i_1}|^2\cdots|c_{i_k}|^2\times$$

$$\prod_{i_1\le i_p<i_q\le i_k}|\lambda_{i_q}-\lambda_{i_p}|^2$$

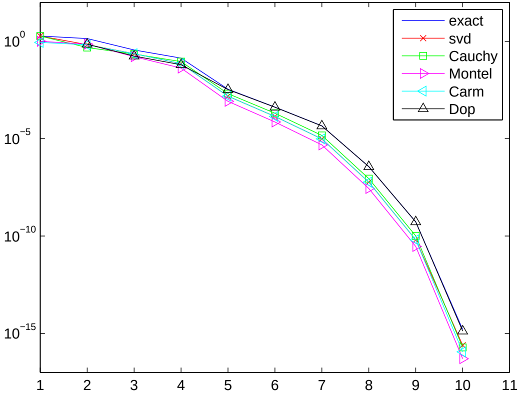
Numerical experiments with the lower bounds

We consider a random normal matrix of order 10 with some eigenvalues which are clustered

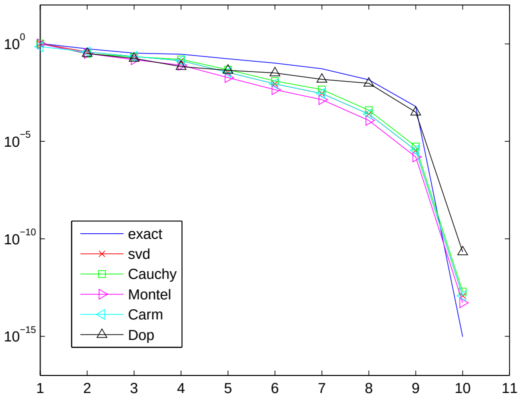
Ritz values converge first to λ_1 and λ_2 which are $0.19872 \pm \imath 1.3691$



Eigenvalues (circles) and Ritz values $(+)$, iterations 4 and 6



Minimum distance to λ_1 and lower bounds



Minimum distance to $\lambda_6 = 2.4677 - \imath 0.25014$ and lower bounds

Upper bounds

The upper bounds we saw are useless because they bound the distance to any of the Ritz values

Let λ_i be a simple eigenvalue and x_i the corresponding eigenvector, $\mathcal{P}_k$ be the orthogonal projector on the Krylov subspace
Then, at iteration k there exists a Ritz value θ such that (6.1)

$$|\theta - \lambda_i| \leq 2\gamma_k \,\mathrm{cond}(X_k) \frac{\|(I - \mathcal{P}_k)x_i\|}{\|\mathcal{P}_k x_i\|}$$

where X_k is the matrix of the eigenvectors of H_k and $\gamma_k = \|\mathcal{P}_k A(I - \mathcal{P}_k)\| \leq \|A\|$

We have closed-form expressions of the distance of the eigenvector to its projection on the Krylov subspace

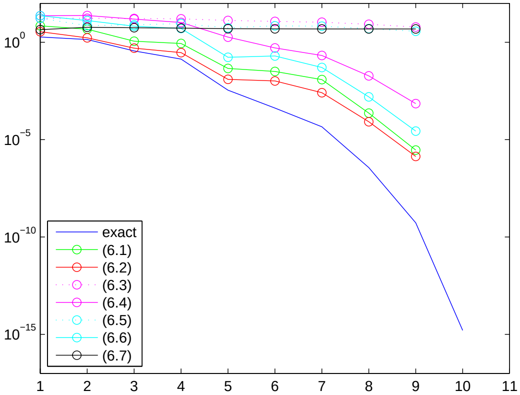
For a simple eigenvalue we also have the bound (6.2)

$$|\theta - \lambda_i| \leq \gamma_k \, \|P^{(k)}\| \frac{\|(I - \mathcal{P}_k)x_i\|}{\|\mathcal{P}_k x_i\|} + O\left(\left(\frac{\|(I - \mathcal{P}_k)x_i\|}{\|\mathcal{P}_k x_i\|}\right)^2\right)$$

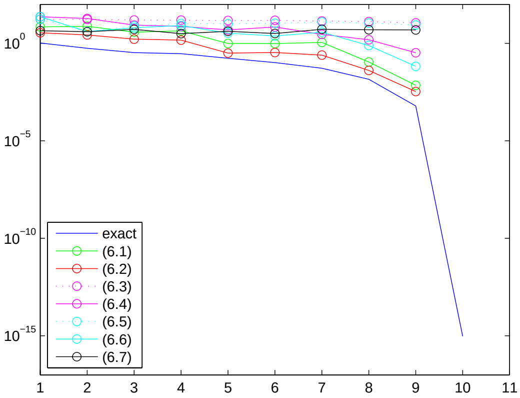
where $P^{(k)}$ is the spectral projector associated to θ. This projector may be computed analytically

Other bounds can be found by perturbation arguments (see Elsner, Stewart)

Numerical experiments with the upper bounds



Minimum distance to λ_1 and upper bounds



Minimum distance to λ_6 and upper bounds

Conclusion

We have seen that we can find upper and lower bounds for the distances of Ritz values to eigenvalues

These bounds are only of theoretical interest since they involve the eigenvalues, eigenvectors of A as well as the Arnoldi initial vector and, for some of them, the Ritz values

We hope they can help understanding the convergence of the Arnoldi algorithm