Misconvergence of Arnoldi Eigenvalue Iterations

Mark Embree

Computational and Applied Mathematics
Rice University

19th June 2006
Overview

We seek to understand when restarted Arnoldi eigenvalue iterations with exact shifts (the basis for ARPACK) will converge when applied to non-Hermitian matrices.

- Summary of convergence theory for Arnoldi eigenvalue iterations
- Computational examples illustrating conventional and pathological behavior
- Demonstration of algorithm failure for small examples
- Theoretical argument explaining circumstances that lead to such behavior
Setting

Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a (non-Hermitian) matrix and let $\mathbf{v}$ be some nonzero vector. The Arnoldi algorithm constructs an orthonormal basis for the Krylov subspace

$$\mathcal{K}_k(\mathbf{A}, \mathbf{v}) = \{\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^{k-1}\mathbf{v}\}.$$

The basis vectors form the columns of $\mathbf{V}_k$, giving the Arnoldi factorization

$$\mathbf{A}\mathbf{V}_k = \mathbf{V}_k \mathbf{H}_k + h_{k+1,k} \mathbf{v}_{k+1} \mathbf{e}_k^*,$$

where $\mathbf{H}_k$ is upper Hessenberg and $\mathbf{v}_{k+1} \perp \text{Ran}\mathbf{V}_k$.

Eigenvalues of $\mathbf{A}$ are approximated by Rayleigh–Ritz eigenvalue estimates drawn from $\mathcal{K}_k(\mathbf{A}, \mathbf{v})$:

$$\{\text{Ritz values}\} = \sigma(\mathbf{V}_k^* \mathbf{A}\mathbf{V}_k) = \sigma(\mathbf{H}_k).$$

In many cases the Ritz values give accurate estimates to eigenvalues on the ‘exterior’ of the spectrum.
Restarted Arnoldi Iterations

Suppose we wish to compute $m$ eigenvalues.

Run Arnoldi for $k = m + p$ iterations.

From the current set of Ritz values, estimate the location of the desired and undesired eigenvalues.

Use this to create a new starting vector, $v_+$, and run Arnoldi again.

Polynomial restarting: $v_+ = \Psi(A)v$ for some polynomial $\Psi$.

Choices for $\Psi$ include the following.

- Construct $v_+$ as the weighted sum of the desired Ritz eigenvector estimates [Saad 1980]
- Take $\Psi$ to be a Chebyshev polynomial small on undesired eigenvalues [Saad 1984; Ho, Chatelin, Bennani 1990]
- Take $\Psi$ to have the $p$ undesired Ritz values as its roots (exact shifts) [Sorensen 1992]
Exact Shifts

Compute $k = m + p$
Exact Shifts

Compute $k = m + p$
**Exact Shifts**

Compute $k = m + p$
Exact Shifts

Compute $k = m + p$

Shade of gray denotes $\log_{10}|\Psi(z)|$. 
The Containment Gap

- For some non-Hermitian matrices, the Ritz pair \((\theta, u)\) may give a small residual
  \[\|Au - \theta u\| \leq \varepsilon \|u\| \ll 1\]
even when \(|\lambda - \theta|\) is large for all \(\lambda \in \sigma(A)\).

- As a robust alternative, track the largest angle the approximating Kylov subspace makes with the desired invariant subspace, \(U_g\). We measure this via the containment gap
  \[
  \delta(U_g, K_k(A, v_1)) = \max_{u \in U_g} \min_{v \in K_k(A, v_1)} \frac{\|u - v\|}{\|v\|}.
  \]
The containment gap measures the sine of the angle.
See, e.g., [Kato 1980], [Chatelin 1983], [Beattie, E., Rossi 2004].

- This measure makes no specifications about how the eigenvalues and eigenvectors are extracted from the approximation subspace.
Two Bounds for the Gap

Assume none of the good eigenvalues are derogatory, and \( v_1 \) isn’t deficient.

**Theorem [Beattie, E., Rossi, 2004]**

\[
\delta(\mathcal{U}_g, \mathcal{K}_k(A, v_1)) \leq \sqrt{2} C_1 \kappa(\Omega_b) \kappa(\Omega_g) \min_{\phi \in \mathcal{P}_{k-m}} \frac{\max\{|\phi(z)| : z \in \Omega_b\}}{\min\{|\phi(z)| : z \in \Omega_g\}}
\]

**Theorem [Beattie, E., Sorensen, 2005]**

\[
\delta(\mathcal{U}_g, \mathcal{K}_k(A, v_1)) \leq C_1 \kappa(\Omega_b) \min_{\phi \in \mathcal{P}_{k-2m}} \max_{z \in \Omega_b} |1 - \phi(z)\alpha(z)|.
\]

- \( \Omega_g, \Omega_b \): compact subsets of \( \mathbb{C} \) containing ‘good’ and ‘bad’ eigenvalues.
- \( C_1 \): constant describing bias in the starting vector.
- \( \kappa(\Omega_g), \kappa(\Omega_b) \): constants describing nonnormality.
- \( \alpha(z) = (z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_m) \).
SUMMARY OF ARNOLDI CONVERGENCE THEORY AND SHIFT STRATEGIES

BOUNDS FOR POLYNOMIAL RESTARTING

BER/BES bounds can be adapted to the restarted setting.

- Recall the BES bound:

\[ \delta(U_g, K_k(A, v_1)) \leq C_1 \kappa(\Omega_b) \min_{\phi \in P_{k-2m}} \max_{z \in \Omega_b} |1 - \phi(z) \alpha(z)|. \]

- Fix \( k = 2m \Rightarrow \) apply \( m \) shifts per outer iteration.

\[ \delta(U_g, K_k(A, \Psi(A)v_1)) \leq C_1 \kappa(\Omega_b) \max_{z \in \Omega_b} |1 - \Phi(z) \alpha(z)|, \]

where \( \Phi \) interpolates \( 1/\alpha \) at the roots of the restart polynomial \( \Psi \).

- Hence \( \Phi \) should mimic the optimal polynomial \( \phi \) in the original bound.
Linear Stability Analysis for Fluid Flow Problem

2D incompressible, viscous flow over a backward facing step. Generated by the IFISS software [Silvester, Elman, Ramage 2005].

Discretize, search for a steady-state solution, and linearize about that solution. Does the linearized system have any eigenvalues in the right half plane?

The generalized eigenvalue problem $Ax = \lambda Bx$ has the structure

$$
\begin{pmatrix}
K & C \\
C^* & 0
\end{pmatrix}
\begin{pmatrix}
u \\
p
\end{pmatrix}
= \lambda
\begin{pmatrix}
M & 0 \\
0 & 0
\end{pmatrix}
\begin{pmatrix}
u \\
p
\end{pmatrix}.
$$

If the rightmost eigenvalue is also smallest in magnitude, we can compute the largest magnitude eigenvalues of $A^{-1}B$; cf. [Cliffe, Garratt, Spence 1994].

Example of flow over a backward facing step (computations of [Elman & E.]):

![Flow over a backward facing step](image-url)
Linear Stability Analysis for Backward Facing Step

Outer Iteration 1:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue indicates $\log_{10} |\Psi(z)|$ over $z \in \mathbb{C}$. 
Linear Stability Analysis for Backward Facing Step

Outer Iteration 2:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue indicates $\log_{10} |\psi(z)|$ over $z \in \mathbb{C}$.
Linear Stability Analysis for Backward Facing Step

Outer Iteration 4:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue indicates \( \log_{10} |\psi(z)| \) over \( z \in \mathbb{C} \).
Linear Stability Analysis for Backward Facing Step

Outer Iteration 6:

• five desired Ritz values from this iteration
• all exact shifts used up to this point

Shade of blue indicates $\log_{10}|\Psi(z)|$ over $z \in \mathbb{C}$. 
Linear Stability Analysis for Backward Facing Step

Outer Iteration 8:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue indicates $\log_{10} |\Psi(z)|$ over $z \in \mathbb{C}$. 
Linear Stability Analysis for Backward Facing Step

Outer Iteration 10:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue indicates $\log_{10} |\Psi(z)|$ over $z \in \mathbb{C}$. 
Linear Stability Analysis for Backward Facing Step

Outer Iteration 12:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue indicates $\log_{10}|\Psi(z)|$ over $z \in \mathbb{C}$. 
Linear Stability Analysis for Backward Facing Step

Outer Iteration 14:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue idicates $\log_{10}|\Psi(z)|$ over $z \in \mathbb{C}$. 
Linear Stability Analysis for Backward Facing Step

Outer Iteration 16:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue indicates $\log_{10} |\Psi(z)|$ over $z \in \mathbb{C}$. 

Summary of Arnoldi Convergence Theory and Shift Strategies
Linear Stability Analysis for Backward Facing Step

Outer Iteration 18:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue indicates $\log_{10}|\Psi(z)|$ over $z \in \mathbb{C}$. 
Linear Stability Analysis for Backward Facing Step

Outer Iteration 20:

- five desired Ritz values from this iteration
- all exact shifts used up to this point

Shade of blue indicates $\log_{10} |\Psi(z)|$ over $z \in \mathbb{C}$. 
Exact Shifts for Non-Hermitian Eigenvalue Problems

An example from [Beattie, E., Sorensen 2005].

We seek the three rightmost eigenvalues (marked $x$).
Unwanted eigenvalues randomly cover the arrow-shaped region.

$$\varepsilon\text{-pseudospectra for } A|_{U_b} \text{ with } \varepsilon = 10^{-1}, 10^{-2}, \ldots, 10^{-20}$$

$$\sigma_\varepsilon(A) = \{z \in \mathbb{C}: z \text{ is an eigenvalue of } A + E \text{ for some } E \text{ with } \|E\| < \varepsilon\}$$
Comparison of restarting strategies

Compare convergence behavior of four Krylov subspaces:

- **Full Krylov subspace**
  No restarts performed.

- **Arnoldi exact shifts**
  Build out to a degree-23 Krylov subspace. Compute Arnoldi Ritz values. Use 20 leftmost Ritz values as roots of the restart polynomial.

- **Optimal interpolation points for the arrow**
  Build out to a degree-23 Krylov subspace. Use degree 20 Fejér points as roots of the restart polynomial.

- **Optimal interpolation points for the $10^{-1}$-pseudospectrum of $A_{|U_b}$**
  Build out to a degree-23 Krylov subspace. Use degree 20 Fejér points as roots of the restart polynomial.
Poor choices for the shifts can lead to slow/no convergence
Restarting at optimal interpolation points

- Color at any given point indicates effective asymptotic convergence rate expected if that point was the nearest desired eigenvalue.
- White dots mark the shifts.
- Solid black lines denote the boundaries of the $10^{-1}$- and $10^{-2}$-pseudospectra.
- Dashed lines are level-curves of convergence rate through nearest eigenvalue.
Restarting at optimal interpolation points

- Color at any given point indicates effective asymptotic convergence rate expected if that point was the nearest desired eigenvalue.
- White dots mark the shifts.
- Solid black lines denote the boundaries of the $10^{-1}$- and $10^{-2}$-pseudospectra.
- Dashed lines are level-curves of convergence rate through nearest eigenvalue.
Location of Ritz Values

We have seen that exact shifts (Ritz values) tend to generate polynomials that generate a good ‘potential’, but what more can be said?

▶ Interlacing of Ritz values for Hermitian A.

Eigenvalues: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$, Ritz values: $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$

Interlacing gives:

$$\theta_1 \in [\lambda_1, \lambda_{N-k+1}], \quad \theta_2 \in [\lambda_2, \lambda_{N-k+2}], \quad \cdots, \quad \theta_k \in [\lambda_k, \lambda_N].$$

Thus the number of Ritz values between exterior eigenvalues is limited.

\[
\begin{array}{ccccccc}
(\lambda_1, \lambda_2) & (\lambda_2, \lambda_3) & (\lambda_3, \lambda_4) & \cdots & (\lambda_{N-3}, \lambda_{N-2}) & (\lambda_{N-2}, \lambda_{N-1}) & (\lambda_{N-1}, \lambda_N) \\
1 & 2 & 3 & \cdots & 3 & 2 & 1
\end{array}
\]

▶ Interlacing is crucial to Sorensen’s proof of restarted Arnoldi convergence with exact shifts for Hermitian matrices.
Location of Ritz Values

- Generalizations of interlacing to non-Hermitian matrices are elusive; cf. literature on divide and conquer generalizations.

- By construction, Ritz values are contained in the numerical range (field of values) of $A$:

$$W(A) = \{x^*Ax : \|x\| = 1\}.$$ 

- Pseudospectra can be used to bound Arnoldi Ritz values. For the Arnoldi decomposition

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

we have

$$\sigma(H_k) \subseteq \sigma_{\varepsilon}(A)$$

for all $\varepsilon > h_{k+1,k}$, where

$$\sigma_{\varepsilon}(A) = \{z \in \mathbb{C} : z \text{ is an eigenvalue of } A + E \text{ for some } E \text{ with } \|E\| < \varepsilon\}.$$
Localization of Ritz Values

- **Polynomial numerical hulls** provide further insight.
  
  For $k = 1, \ldots, N$, define
  
  $$G_k(A) = \left\{ z \in \mathbb{C} : \min_{p \in P_k, p(0) = 1} \| p(z - A) \| = 1 \right\};$$

  see [Nevanlinna 1993; Greenbaum 2002].

- Note that $G_1(A) = W(A)$ and $G_N(A) = \sigma(A)$. 
Localization of Ritz Values

- **Polynomial numerical hulls** provide further insight. For $k = 1, \ldots, N$, define

$$G_k(A) = \left\{ z \in \mathbb{C} : \min_{p \in P_k, p(0) = 1} \| p(z - A) \| = 1 \right\};$$

see [Nevanlinna 1993; Greenbaum 2002].

- Note that $G_1(A) = W(A)$ and $G_N(A) = \sigma(A)$.

- If $z \in G_k(A)$ and if Ideal GMRES = True GMRES for $z - A$, then there exists some unit starting vector $v_1 \in \mathbb{C}^n$ such that $z$ is an eigenvalue of $H_j$ for $j = 1, \ldots, k$. 
Lakatos and his ‘Monsters’

Imre Lakatos
1922–1974

Philosopher and historian of mathematics

His *Proofs and Refutations* begins with a dialog on Euler’s formula $V - E + F = 2$ relating the number of vertices, edges, and faces of a polyhedron.

The students in the dialog propose and debate *monster* polyhedra that violate the formula.
First Example

Consider the matrix

\[ A = \begin{pmatrix} 1 & 0 \\ 0 & J_{100} \end{pmatrix} \]

where

\[ J_n = \begin{pmatrix} 0 & 3/2 \\ 0 & \ddots \\ \vdots & \ddots & 3/2 \\ 0 & \ddots & \ddots & 0 \end{pmatrix} \in \mathbb{C}^{n \times n}. \]

We wish to compute the eigenvalue \( \lambda = 1 \) of \( A \).
First Example

Consider the matrix

\[ A = \begin{pmatrix} 1 & 0 \\ 0 & J_{100} \end{pmatrix} \]

where

\[ J_n = \begin{pmatrix} 0 & 3/2 \\ 0 & \ddots & \ddots \\ \cdot & \ddots & 3/2 \\ \cdot & \cdot & 0 \end{pmatrix} \in \mathbb{C}^{n \times n}. \]

We wish to compute the eigenvalue \( \lambda = 1 \) of \( A \).

Problem: \( \sigma(J_n) = \{0\} \) but as \( n \to \infty \),

\[ \sigma_\varepsilon(J_n) \to \{z \in \mathbb{C} : |z| < 3/2 + \varepsilon\}. \]

The desired eigenvalue is surrounded by \( \varepsilon \)-pseudoeigenvalues for \( J_n \) for \( \varepsilon \ll 1 \):

\[ \text{sep}(1, J_{100}) = \| (I - J_{100})^{-1} \| \approx 4.87 \times 10^{17}. \]
First Example

EigTool [Wright 2002] can be used to visualize ARPACK execution.
First Example

EigTool [Wright 2002] can be used to visualize ARPACK execution.
First Example

$\sigma_\varepsilon(A)$ for $\varepsilon = 10^{-1}, 10^{-3}, \ldots, 10^{-15}$. 
Summary of the First Example

This example is not especially troubling, as there are many points $z$ far to the right of the desired eigenvalue, $|z - \lambda| \gg 0$ for which there exist unit vectors $v$ such that

$$\|Av - zv\| \ll 1.$$

Consequence: transient behavior can differ significantly from the asymptotic behavior dictated by the true eigenvalue.
A Clean Example Showing the Failure of Exact Shifts

Find the largest-magnitude eigenvalue $\lambda = 1$ of $A$, which has eigenvector $u_1$:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & -2 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad u_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$  

Use the restarted Arnoldi algorithm with exact shifts, looking for $m = 1$ eigenvalue with $p = 1$ shift per iteration.
A Clean Example Showing the Failure of Exact Shifts

Find the largest-magnitude eigenvalue \( \lambda = 1 \) of \( \mathbf{A} \), which has eigenvector \( \mathbf{u}_1 \):

\[
\mathbf{A} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 6 & -2 \\
0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0
\end{pmatrix}, \quad \mathbf{u}_1 = \begin{pmatrix}
1 \\
0 \\
0 \\
0
\end{pmatrix}.
\]

Use the restarted Arnoldi algorithm with exact shifts, looking for \( m = 1 \) eigenvalue with \( p = 1 \) shift per iteration.

If the starting vector \( \mathbf{v}_1 \) has equal components in each entry, we obtain:

\[
\mathbf{v}_1 = \frac{1}{2} \begin{pmatrix}
1 \\
1 \\
1 \\
1
\end{pmatrix}, \quad \mathbf{v}_2 = \frac{1}{2\sqrt{35}} \begin{pmatrix}
-3 \\
9 \\
1 \\
-7
\end{pmatrix},
\]

so that \( \mathcal{K}_2(\mathbf{A}, \mathbf{v}_1) = \text{span}\{\mathbf{v}_1, \mathbf{A}\mathbf{v}_1\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \).

Now compute Ritz values for \( \mathcal{K}_2(\mathbf{A}, \mathbf{v}_1) \), and choose the smallest-magnitude Ritz value as the shift.
A Clean Example Showing the Failure of Exact Shifts

- Though nondiagonalizable, this matrix is not very ‘far from normal’:

\[
\text{sep}\left(1, \begin{pmatrix} 0 & 6 & -2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}\right) = 0.0837 \ldots
\]

The eigenvalue $\lambda_1 = 1$ would still likely be of interest for applications.
A Clean Example Showing the Failure of Exact Shifts

- Though nondiagonalizable, this matrix is not very ‘far from normal’:

\[ \text{sep}(1, \begin{pmatrix} 0 & 6 & -2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}) = 0.0837 \ldots. \]

The eigenvalue \( \lambda_1 = 1 \) would still likely be of interest for applications.
Transient Behavior

Since the spectral radius of this matrix is one, we should not observe exponential growth or decay in matrix powers.

The plot below exhibits mild transient growth consistent with the modest departure of $\mathbf{A}$ from normality.
A Clean Example Showing the Failure of Exact Shifts

Compute Arnoldi-Ritz values. Let $V_2 = [v_1 \ v_2]$, so that

$$H_2 = V_2^* A V_2 = \begin{pmatrix} 7/4 & 3/(4\sqrt{35}) \\ \sqrt{35}/4 & 5/4 \end{pmatrix}.$$ 

Note that

$$\det(\lambda - H_2) = \lambda^2 - 3\lambda + 2 = (\lambda - 1)(\lambda - 2)$$

and hence the Ritz values are $\theta_1 = 1, \theta_2 = 2$. 
A Clean Example Showing the Failure of Exact Shifts

Compute Arnoldi-Ritz values. Let $V_2 = [v_1 \ v_2]$, so that

$$H_2 = V_2^* A V_2 = \begin{pmatrix} \frac{7}{4} & \frac{3}{4\sqrt{35}} \\ \sqrt{35}/4 & \frac{5}{4} \end{pmatrix}.$$ 

Note that

$$\det(\lambda - H_2) = \lambda^2 - 3\lambda + 2 = (\lambda - 1)(\lambda - 2)$$

and hence the Ritz values are $\theta_1 = 1$, $\theta_2 = 2$.

The exact shift strategy selects $\theta_1 = 1$, giving the new starting vector

$$v_+ = (A - \theta_1 I)v_1 = \begin{pmatrix} 0 \\ 3 \\ 1 \\ -1 \end{pmatrix}.$$ 

This vector has no component in $u_1$, the eigenvector corresponding to $\lambda = 1$: convergence is now impossible!
A Clean Example Showing the Failure of Exact Shifts

- The first Arnoldi step gives $H_1 = (7/4)$ and $h_{2,1} = \sqrt{35}/4$, so the eigenvalue estimate after one iteration is poor.
A Clean Example Showing the Failure of Exact Shifts

- The first Arnoldi step gives \( H_1 = (7/4) \) and \( h_{2,1} = \sqrt{35}/4 \), so the eigenvalue estimate after one iteration is poor.
- If we adjust the component of the starting vector in the desired eigenvector

\[
\mathbf{v}_1 = \frac{1}{\sqrt{\alpha^2 + 3}} \begin{pmatrix} \alpha \\ 1 \\ 1 \\ 1 \end{pmatrix},
\]

then naturally \( H_2 \) depends on \( \alpha \):

\[
H_2 = \mathbf{V}_2^* \mathbf{A} \mathbf{V}_2 = \begin{pmatrix}
\frac{6 + \alpha^2}{3 + \alpha^2} & \frac{3\alpha^3}{(3 + \alpha^2)(24 + 11\alpha^2)^{1/2}} \\
\frac{(24 + 11\alpha^2)^{1/2}}{3 + \alpha^2} & \frac{3 + 2\alpha^2}{3 + \alpha^2}
\end{pmatrix}.
\]
A Clean Example Showing the Failure of Exact Shifts

- The first Arnoldi step gives $H_1 = \left( \frac{7}{4} \right)$ and $h_{2,1} = \sqrt{35}/4$, so the eigenvalue estimate after one iteration is poor.
- If we adjust the component of the starting vector in the desired eigenvector $v_1 = \frac{1}{\sqrt{\alpha^2+3}} \begin{pmatrix} \alpha \\ 1 \\ 1 \\ 1 \end{pmatrix}$, then naturally $H_2$ depends on $\alpha$:

$$H_2 = V_2^* A V_2 = \begin{pmatrix} \frac{6 + \alpha^2}{3 + \alpha^2} & \frac{3\alpha^3}{(3 + \alpha^2)(24 + 11\alpha^2)^{1/2}} \\ \frac{(24 + 11\alpha^2)^{1/2}}{3 + \alpha^2} & \frac{3 + 2\alpha^2}{3 + \alpha^2} \end{pmatrix}.$$  

However, $H_2$ always has the characteristic polynomial

$$\text{det}(\lambda - H_2) = \lambda^2 - 3\lambda + 2 = (\lambda - 1)(\lambda - 2)$$

and hence the Ritz values $\theta_1 = 1$, $\theta_2 = 2$. 
Conclusions to Draw from This Example

This example illustrates that the restarted Arnoldi algorithm with exact shifts may fail to converge:

▶ regardless of $\angle(v_1, u_1)$;
▶ even if $u_1$ is orthogonal to the complementary invariant subspace, i.e., the desired eigenvalue is *perfectly conditioned*;
▶ the desired eigenvalue is ‘well separated’.

How widespread is this sort of behavior?

Firstly, this example relies on the fact that $\lambda$ is contained well in the interior of the numerical range of $D$. 
Example: IFISS/Backward Facing Step

Dimension $N = 3287$

Matrices generated by IFISS package of Silvester, Elman, and Ramage.
Numerical range computed with `fvbis.m` of Braconnier and Higham.
Example: IFISS/Backward Facing Step

Dimension $N = 3287$

Matrices generated by IFISS package of Silvester, Elman, and Ramage.
Numerical range computed with `fvbis.m` of Braconnier and Higham.
Example: Tolosa Matrix

Dimension $N = 1090$

Godet-Thobie (CERFACS) and Bès (Aerospatiale)
Matrix Market/Non-Hermitian Eigenvalue Problem collection
Numerical range computed with $fv.m$ of Braconnier and Higham.
Example: Tolosa Matrix

Dimension $N = 1090$

![Graph of numerical range, $W(A)$](image)

Godet-Thobie (CERFACS) and Bès (Aerospatiale)
Matrix Market/Non-Hermitian Eigenvalue Problem collection
Numerical range computed with `fv.m` of Braconnier and Higham.
A Generalization

The clean $4 \times 4$ example can be generalized significantly. (Despite this fact, we believe such examples to be rare.)

**Theorem.** Suppose that

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & D \end{pmatrix} \in \mathbb{C}^{(n+1)\times(n+1)}$$

and suppose $v_1 \in \mathbb{C}^n$ is a vector such that $\{\lambda, \theta_1, \ldots, \theta_{k-1}\}$ are the $k \leq n$ Arnoldi Ritz values of $D$ drawn from the space $\mathcal{K}_k(D, v_1)$. (Assume no ‘lucky breakdown’ occurs.)

Then for any value of $\alpha \in \mathbb{C}$, the Arnoldi Ritz values from the space $\mathcal{K}_k(A, \hat{v}_1)$ are also $\{\lambda, \theta_1, \ldots, \theta_{k-1}\}$, where

$$\hat{v}_1 = \begin{pmatrix} \alpha \\ v_1 \end{pmatrix}$$

for any $\alpha \in \mathbb{C}$.
Implications

- This result implies that if there is some starting vector $v_1$ that gives $\lambda$ as a subordinate Ritz value of $D$, then there exist starting vectors for which the restarted Arnoldi algorithm with exact shifts applied to $A$ will annihilate the eigenvector associated with $\lambda$, regardless of the magnitude of that component in the starting vector.

- This result can be extended to systems where the desired eigenvector is orthogonal to the complementary invariant subspace via unitary similarity transformation.
A Generalization

Proof ($k = 2$ case for simplicity).

By the translation invariance of the Ritz values, it suffices to prove the result for

\[ A = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix} \]

with $D$ having the Ritz values $\{0, \theta_1, \ldots, \theta_{k-1}\}$. 
A Generalization

**Proof** \((k = 2\) case for simplicity). By the translation invariance of the Ritz values, it suffices to prove the result for

\[
A = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}
\]

with \(D\) having the Ritz values \(\{0, \theta_1, \ldots, \theta_{k-1}\}\).

We can write the following Krylov decomposition for \(\mathcal{K}_2(D, v_1)\):

\[
D[v_1 \ Dv_1] = [v_1 \ Dv_1] \begin{bmatrix} 0 & \gamma_0 \\ 1 & \gamma_1 \end{bmatrix} + [0 \ (D^2 - \gamma_1 D - \gamma_0)v_1],
\]

which we abbreviate \(DK = KC + be_k^*\).
A Generalization

Proof ($k = 2$ case for simplicity).

By the translation invariance of the Ritz values, it suffices to prove the result for

$$A = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}$$

with $D$ having the Ritz values $\{0, \theta_1, \ldots, \theta_{k-1}\}$.

We can write the following Krylov decomposition for $\mathcal{K}_2(D, v_1)$:

$$D[v_1 \ Dv_1] = [v_1 \ Dv_1] \begin{bmatrix} 0 & \gamma_0 \\ 1 & \gamma_1 \end{bmatrix} + [0 \ (D^2 - \gamma_1 D - \gamma_0)v_1],$$

which we abbreviate $DK = KC + be_k^*$.

We choose $\gamma_0$ and $\gamma_1$ so that $b = (D^2 - \gamma_1 D - \gamma_0)v_1$ is orthogonal to $\mathcal{K}_2(D, v_1) = \text{span}\{v_1, Dv_1\}$, which implies that the eigenvalues of $C$ will be the Arnoldi Ritz values; see (Stewart 2001).

In this case, $\det(\lambda - C) = \lambda^2 - \lambda \gamma_1 - \gamma_0 = \lambda(\lambda - \theta_1)$; i.e., $\gamma_0 = 0$. 


A Generalization

Since $b = (D^2 - \gamma_1 D - \gamma_0)v_1$ is orthogonal to $K_2(D, v_1) = \text{span}\{v_1, Dv_1\}$,

$$v_1^*(D^2 - \gamma_1 D - \gamma_0)v_1 = 0;$$

$$(Dv_1)^*(D^2 - \gamma_1 D - \gamma_0)v_1 = 0.$$
A Generalization

Since \( b = (D^2 - \gamma_1 D - \gamma_0)v_1 \) is orthogonal to \( K_2(D, v_1) = \text{span}\{v_1, Dv_1\} \),

\[
\begin{align*}
v_1^*(D^2 - \gamma_1 D - \gamma_0)v_1 &= 0; \\
(Dv_1)^* (D^2 - \gamma_1 D - \gamma_0)v_1 &= 0.
\end{align*}
\]

Now consider the Arnoldi Ritz values for \( A \) generated by \( \hat{v}_1 = \begin{pmatrix} \alpha \\ v_1 \end{pmatrix} \).

We have a similar Krylov decomposition,

\[
A[\hat{v}_1 \ A\hat{v}_1] = [\hat{v}_1 \ A\hat{v}_1] \begin{bmatrix} 0 & \hat{\gamma}_0 \\ 1 & \hat{\gamma}_1 \end{bmatrix} + [0 \ (A^2 - \gamma_1 A - \gamma_0)\hat{v}_1].
\]
A Generalization

Since \( b = (D^2 - \gamma_1 D - \gamma_0)v_1 \) is orthogonal to \( \mathcal{K}_2(D, v_1) = \text{span}\{v_1, Dv_1\} \),

\[
\begin{align*}
v_1^*(D^2 - \gamma_1 D - \gamma_0)v_1 &= 0; \\
(Dv_1)^*(D^2 - \gamma_1 D - \gamma_0)v_1 &= 0.
\end{align*}
\]

Now consider the Arnoldi Ritz values for \( A \) generated by \( \hat{v}_1 = \begin{pmatrix} \alpha \\ v_1 \end{pmatrix} \).

We have a similar Krylov decomposition,

\[
A[\hat{v}_1 \ A \hat{v}_1] = [\hat{v}_1 \ A \hat{v}_1] \begin{bmatrix} 0 & \hat{\gamma}_0 \\ 1 & \hat{\gamma}_1 \end{bmatrix} + [0 \ (A^2 - \hat{\gamma}_1 A - \hat{\gamma}_0) \hat{v}_1].
\]

To show that the Arnoldi Ritz values for \( (A, \hat{v}_1) \) agree with those for \( (D, v_1) \), we must show that the choice \( \hat{\gamma}_0 = \gamma_0 \) and \( \hat{\gamma}_1 = \gamma_1 \) will produce a vector \( \hat{b} = (A^2 - \hat{\gamma}_1 A - \hat{\gamma}_0)\hat{v}_1 \) that is orthogonal to \( \mathcal{K}_2(A, \hat{v}_1) \), i.e.,

\[
\begin{align*}
\hat{v}_1^*(A^2 - \gamma_1 A - \gamma_0)\hat{v}_1 &= 0; \\
(A\hat{v}_1)^*(A^2 - \gamma_1 A - \gamma_0)\hat{v}_1 &= 0.
\end{align*}
\]
A Generalization

Due to the structure of $A$ we have

$$A\hat{v}_1 = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} \alpha \\ v_1 \end{pmatrix} = \begin{pmatrix} 0 \\ Dv_1 \end{pmatrix}.$$ 

This fact, together with $\gamma_0 = 0$, implies that

$$\hat{v}_1^* (A^2 - \gamma_1 A - \gamma_0)\hat{v}_1 = v_1^* (D^2 - \gamma_1 D - \gamma_0)v_1 = 0;$$

$$(A\hat{v}_1)^* (A^2 - \gamma_1 A - \gamma_0)\hat{v}_1 = (Dv_1)^* (D^2 - \gamma_1 D - \gamma_0)v_1 = 0.$$
A Generalization

Due to the structure of $A$ we have

$$A\hat{v}_1 = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} \alpha \\ v_1 \end{pmatrix} = \begin{pmatrix} 0 \\ Dv_1 \end{pmatrix}.$$ 

This fact, together with $\gamma_0 = 0$, implies that

$$\hat{v}_1^*(A^2 - \gamma_1 A - \gamma_0)\hat{v}_1 = v_1^*(D^2 - \gamma_1 D - \gamma_0)v_1 = 0;$$

$$(A\hat{v}_1)^*(A^2 - \gamma_1 A - \gamma_0)\hat{v}_1 = (Dv_1)^*(D^2 - \gamma_1 D - \gamma_0)v_1 = 0.$$ 

This argument can be readily generalized to $k > 2$. 
A Final Example

The theorem suggests a method for constructing larger examples. Let

\[ A = \begin{pmatrix} \Lambda & 0 \\ 0 & D \end{pmatrix}, \]

where \( \Lambda \) is a normal \( 2 \times 2 \) matrix containing two desired eigenvalues: \( \sigma(\Lambda) \approx 2.94 \pm 1.76i \). These are the rightmost and largest magnitude eigenvalues of \( A \).

We take

\[ D = \begin{pmatrix} -0.5 & 5 \\ & -0.357 \ldots & \ddots \\ & & \ddots & 5 \\ & & & 0.5 \end{pmatrix} \in \mathbb{C}^{8 \times 8}. \]

\( D \) is diagonalizable but nonnormal, with eigenvalues uniformly spaced over \([-1/2, 1/2]\).
A Final Example: Eigenvalues

Compute $m = 2$ eigenvalues with $p = 2$ shifts per iteration.

- 2 sought-after eigenvalues of $\Lambda$
- 8 unwanted eigenvalues of $D$
A Final Example: Ritz Values

Compute $m = 2$ eigenvalues with $p = 2$ shifts per iteration.

- 2 sought-after eigenvalues of $\Lambda$
- 4 Ritz values
- 8 unwanted eigenvalues of $D$
A Final Example: Eigenvalue estimates and shifts

Compute $m = 2$ eigenvalues with $p = 2$ shifts per iteration.

- 2 sought-after eigenvalues of $\Lambda$
- 8 unwanted eigenvalues of $D$
- Eigenvalue approximations
- Exact shifts
A Final Example: Pseudospectra and Numerical Range

Compute $m = 2$ eigenvalues with $p = 2$ shifts per iteration.

- 2 sought-after eigenvalues of $\Lambda$
- 8 unwanted eigenvalues of $D$
- eigenvalue approximations
- exact shifts
Conclusions

- The Implicitly Restarted Arnoldi iteration with exact shifts, as implemented in ARPACK/eigs, is extremely useful algorithm that justly has a broad base of satisfied users.

- There exist non-Hermitian matrices and starting vectors for which where the exact shift algorithm will deflate desired eigenvalues.

- Such examples needn’t be pathologically nonnormal: these are examples where the sought-after eigenvalues are likely to have ‘physical’ significance.

- To better diagnose such problems and construct countermeasures, we need to understand the distribution of Ritz values for non-Hermitian matrices. Progress on this difficult problem will have broad impact.