Krylov methods for eigenvalue problems: theory and algorithms

Karl Meerbergen

KU LEUVEN

October 4-6th, 2017

Two lectures

1. Krylov methods for eigenvalue problems: theory and algorithms
   - Concepts of spectral approximation
   - Convergence/approximation theory
   - Algorithms
2. Krylov methods for eigenvalue problems: applications
   - Computing eigenvalues with largest real part (stability analysis)
   - Nonlinear eigenvalue problems

Outline of lecture 1

Eigenvalue problems

An eigenvalue problem is every mathematical problem that can be translated to

What is an eigenvalue problem

For matrix $A \in \mathbb{C}^{n \times n}$, find a particular $\lambda \in \mathbb{C}$ and $x \in \mathbb{C}^n$, $x \neq 0$ such that

$$Ax = \lambda x$$

or find a particular $\lambda \in \mathbb{C}$ and $x, y \in \mathbb{C}^n$, $x, y \neq 0$ such that

$$\begin{aligned} Ax &= \lambda x \\ y^* A &= \lambda y^* \end{aligned}$$

We call (λ, x) and eigenpair of A. (Later more on properties of eigenvalue problems)

Power method

Initial vector $x^{(0)}$ with $\|x^{(0)}\|_2 = 1$
for $k = 1, 2, \ldots$ **do**
$y^{(k)} = A \cdot x^{(k-1)}$
$x^{(k)} = y^{(k)} / \|y^{(k)}\|_2$
end for

Theorem (Convergence)

Let the n eigenvalues $\lambda_1, \ldots, \lambda_n$ of A be ordered as follows:

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$$

(λ_1 is the dominant eigenvalue; λ_1 is simple.)
$x^{(k)}$ converges to the dominant eigenvector.

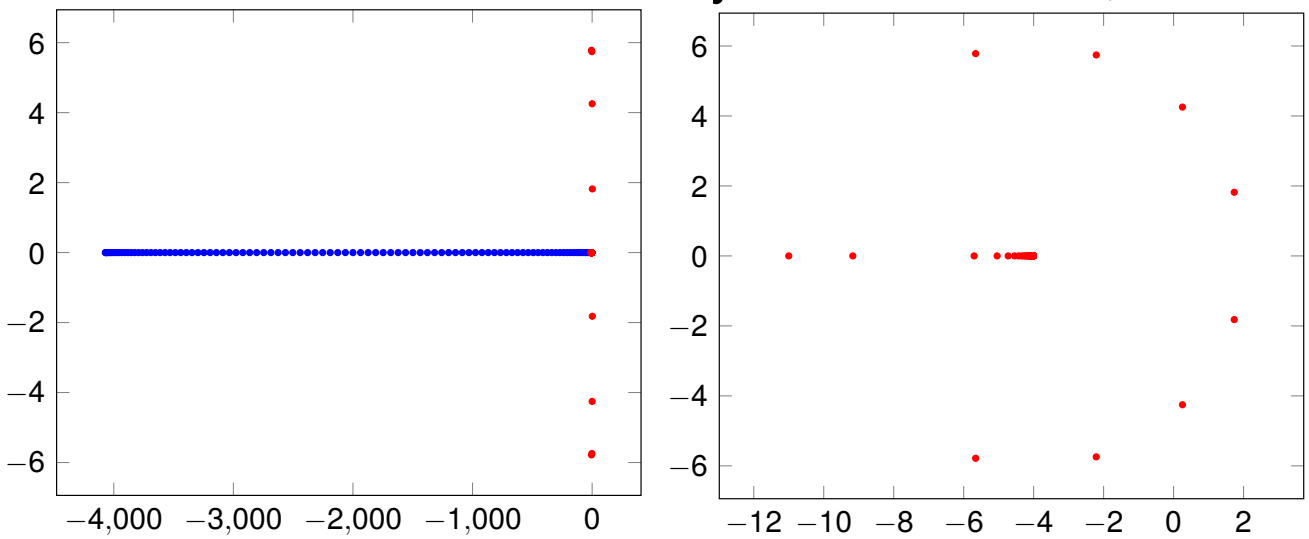
Theorem (Convergence speed)

The convergence rate of the dominant eigenvalue is $|\lambda_2|/|\lambda_1|$.

Olmstead benchmark problem

$$\begin{aligned} \frac{\partial u}{\partial t} &= (1-C)\frac{\partial^2 v}{\partial X^2} + C\frac{\partial^2 u}{\partial X^2} + Ru - u^3 \\ B\frac{\partial v}{\partial t} &= u - v \end{aligned}$$

u represents the speed of the fluid
v is related to viscoelastic forces.
The boundary conditions are $u(0) = u(1) = 0$ and $v(0) = v(1) = 0$.
Discretization with central differences
A is the Jacobian at the steady state for $B = 2$, $C = 0.1$ and $R = 4.5$.

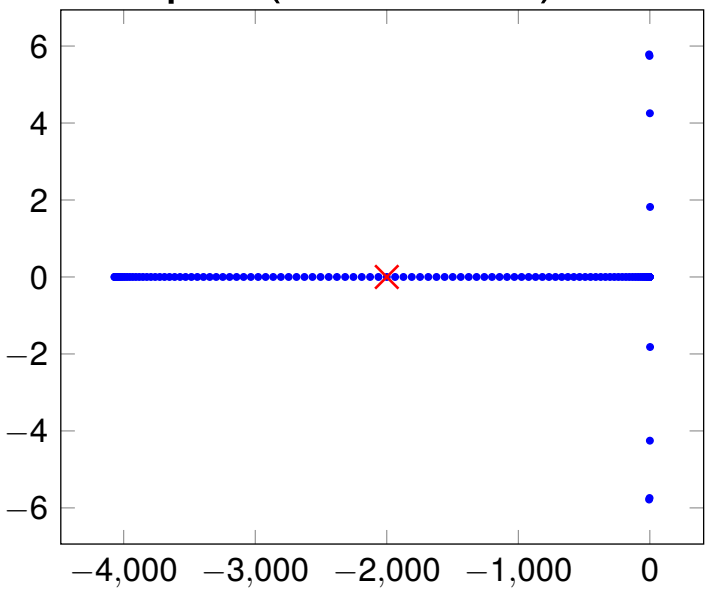


Shifted power method

Power on $A - \sigma I$.
Convergence to dominant eigenvalue of $A - \sigma I$ with rate of convergence $|\lambda_1 - \sigma|/|\lambda_2 - \sigma|$.
Implementation:

$$(A - \alpha I)x^{(k-1)} = Ax^{(k-1)} - \alpha x^{(k-1)}$$

Example (Olmstead):



- Shift $\sigma = -2000.0$
- Convergence rate: 0.99857 (slow!)

Two friends

- Subspace iteration
- Arnoldi method (Krylov)

Subspace iteration

Is the power method on a subspace, or a block of vectors:

Initial vectors $V^{(0)} = [v_1^{(0)}, \ldots, v_m^{(0)}] \in \mathbb{C}^{n \times m}$ with $\|V^{(0)}\|_2 = 1$
for $k = 1, 2, \ldots$ **do**
  $W^{(k)} = A \cdot V^{(k-1)}$
  Orthonormalize the columns of $W^{(k)} \rightarrow V^{(k)}$.
end for

Theorem (Convergence)

Let the n eigenvalues $\lambda_1, \ldots, \lambda_n$ of A be ordered as follows:

$$|\lambda_1| \geq \cdots \geq |\lambda_m| > |\lambda_{m+1}| \geq \cdots \geq |\lambda_n|$$

($\lambda_1, \ldots, \lambda_m$ are the m dominant eigenvalues.)
Range($V^{(k)}$) converges to the dominant invariant subspace.

Theorem (Convergence speed)

The rate of convergence is $|\lambda_{m+1}|/|\lambda_1|$.

Subspace iteration

Extracting eigenvalues by Galerkin projection

- $x = Vz$ with $z \in \mathbb{C}^m$.
- Galerkin:

$$Ax - \lambda x \perp \text{Range}(V)$$

becomes:

$$\begin{aligned} V^*(Ax - \lambda x) &= = 0 \\ V^*(AVz - \lambda Vz) &= 0 \\ (V^*AV)z &= \lambda z \end{aligned}$$

$$(V^*AV)z = \lambda z$$

Subspace iteration

Gram-Schmidt orthogonalization

- It is very (very) important that the columns of V are orthogonal
- Modified Gram-Schmidt is numerically unstable, i.e., columns of V are not orthogonal to machine precision
- Iterative Gram-Schmidt is the solution
  Gram-Schmidt for orthogonalization of vector w against columns of V:

  $$\tilde{w} = w - V(V^* w)$$
  $$v = \tilde{w}/\|\tilde{w}\|_2$$

Subspace iteration

Gram-Schmidt orthogonalization

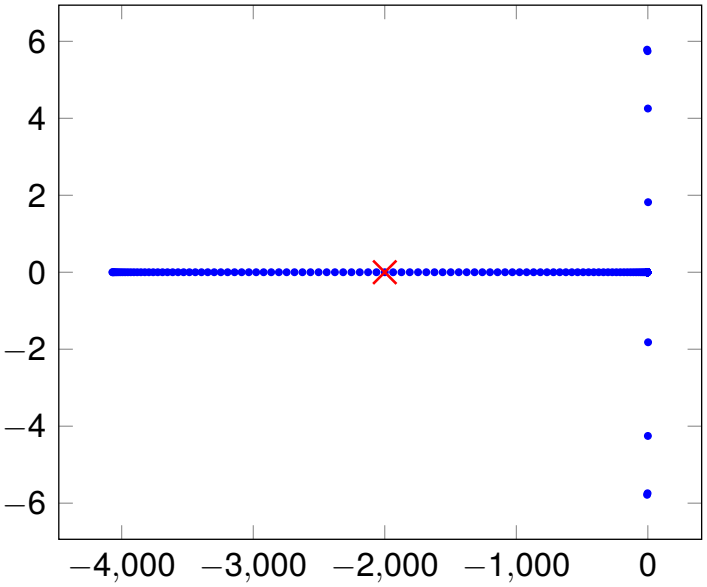
- It is very (very) important that the columns of V are orthogonal
- Modified Gram-Schmidt is numerically unstable, i.e., columns of V are not orthogonal to machine precision
- Iterative Gram-Schmidt is the solution
  Iterative Gram-Schmidt for orthogonalization of vector w against columns of V:

  $$\tilde{w} = w - V(V^* w)$$
  $$\tilde{\tilde{w}} = \tilde{w} - V(V^* \tilde{w})$$
  $$v = \tilde{\tilde{w}} / \|\tilde{\tilde{w}}\|_2$$

  Iterative Gram-Schmidt is backward stable: $[V, v]$ has orthonormal columns and w is spanned by the columns of $[V, v]$.

Subspace iteration

Example (Olmstead)



- Subspace iteration for $A + 2000I$.
- Convergence rate for the power method: 0.99857 (slow!)
- For subspace iteration with $m = 10$: 0.96587 (still slow)

Subspace (polynomial) iteration

Convergence of subspace iteration can be controlled by using shifts: (matrix iteration k)

$$\text{Range}(V^{(k)}) = \text{Range}((A - \sigma I)^k V^{(0)})$$

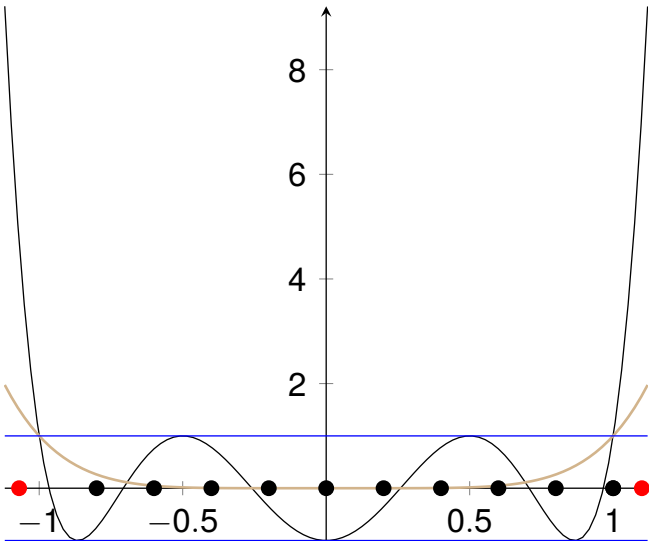
A better choice (often) is to choose a filter polynomial that removes the unwanted eigenvalues.

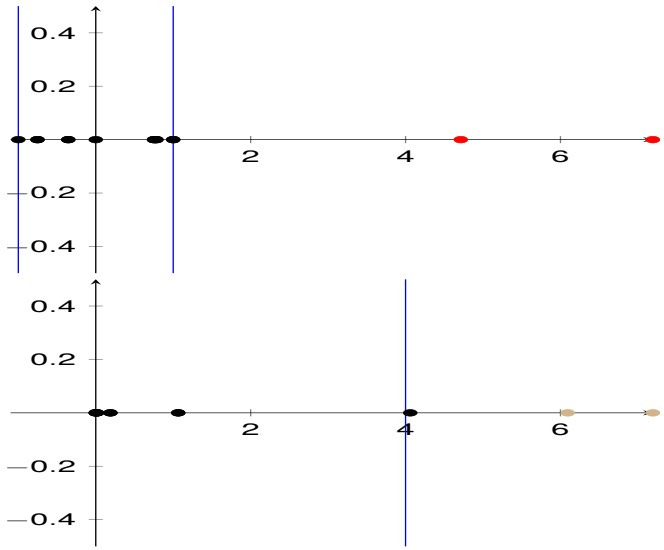
Initial vectors $V^{(0)} = [v_1^{(0)}, \ldots, v_m^{(0)}] \in \mathbb{C}^{n \times m}$ with $\|V^{(0)}\|_2 = 1$
for $k = 1, 2, \ldots, k$ **do**
  $W^{(k)} = (A - \sigma_k I) \cdot V^{(k-1)}$
  Orthonormalize the columns of $W^{(k)} \rightarrow V^{(k)}$.
end for

Chebyshev iteration

Chebyshev iteration:

Initial vectors $V^{(0)} = [v_1^{(0)}, \ldots, v_m^{(0)}] \in \mathbb{C}^{n \times m}$ with $\|V^{(0)}\|_2 = 1$
$W^{(-1)} = 0$
$W^{(0)} = V^{(0)}$
for $j = 1, 2, \ldots, k$ **do**
$W^{(j)} = 2AW^{(j-1)} - W^{(j-2)}$
end for
Orthonormalize the columns of $W^{(k)} \to V^{(k)}$.





Chebyshev iteration

Theorem (Convergence)

Let the n eigenvalues $\lambda_1, \ldots, \lambda_n$ of A be ordered as follows:

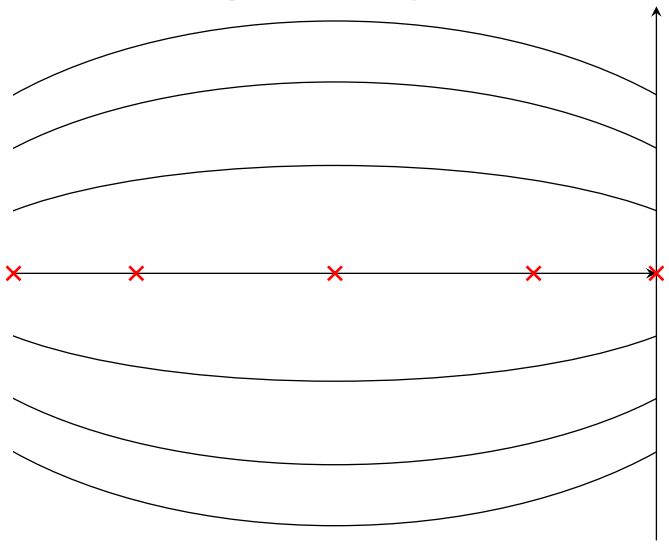
$$|T_k(\lambda_1)| \geq \cdots \geq |T_k(\lambda_m)| > |T_k(\lambda_{m+1})| \geq \cdots \geq |T_k(\lambda_n)|$$

($\lambda_1, \ldots, \lambda_m$ are the m dominant eigenvalues of $T_k(A)$.
Range($V^{(k)}$) converges to the m eigenvalues of A outside $[-1, 1]$.
The rate of convergence is $|T_k(\lambda_1)^{-1}|$.

For computing eigenvalues outside $[\alpha, \beta]$, shift and scale the Chebyshev polynomial:

$$T_k\left(\frac{A - \frac{\alpha+\beta}{2}}{\frac{\beta-\alpha}{2}}\right)$$

For complex eigenvalues:



Inverse iteration

Subspace iteration applied to $(A - \sigma I)^{-1}$

Initial vectors $V^{(0)} = [v_1^{(0)}, \ldots, v_m^{(0)}] \in \mathbb{C}^{n \times m}$ with $\|V^{(0)}\|_2 = 1$
for $k = 1, 2, \ldots, k$ **do**
    Solve $(A - \sigma I)W^{(k)} = .V^{(k-1)}$
    Orthonormalize the columns of $W^{(k)} \rightarrow V^{(k)}$.
end for

Theorem (Convergence)

Let the n eigenvalues $\lambda_1, \ldots, \lambda_n$ of A be ordered as follows:

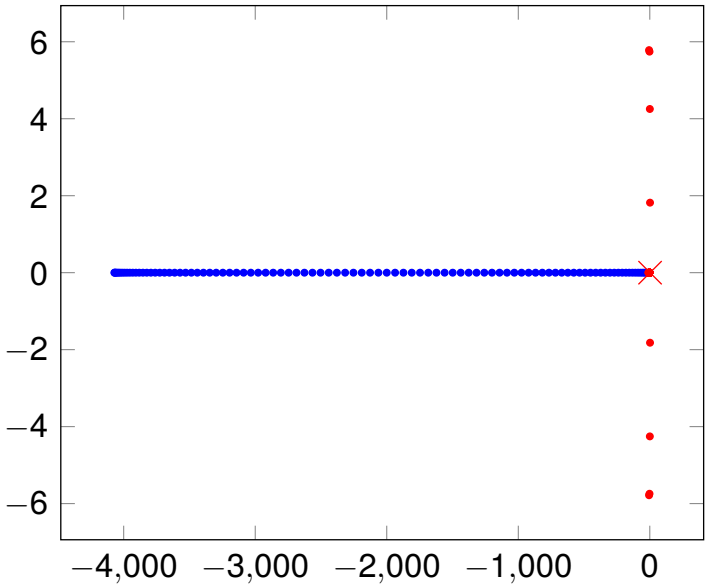
$$|\lambda_1 - \sigma| \leq \cdots \geq |\lambda_m - \sigma| < |\lambda_{m+1} - \sigma| \geq \cdots \geq |\lambda - n - \sigma|$$

$\lambda_1, \ldots, \lambda_m$ are the m eigenvalues of A nearest σ.
Range($V^{(k)}$) converges to the m eigenvalues of A o nearest σ.
The rate of convergence is $|\lambda_1 - \sigma|/|\lambda_{m+1} - \sigma|$.

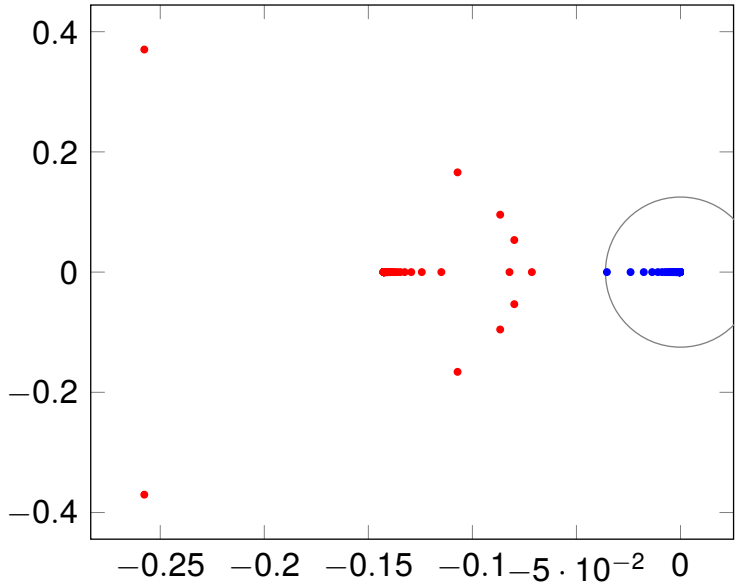
Inverse iteration

Example: Olmstead problem

Eigenvalues of A and shift σ



Eigenvalues of $(A - \sigma I)^{-1}$



Arnoldi method

Build a Krylov space (from a power sequence):

$$\{v_1, Av_2, A^2v_1, \ldots A^{k-1}v_1\}$$

Arnoldi algorithm produces orthonormal basis $V_k = [v_1, \ldots, v_k]$:

Given v_1 with $\|v_1\|_2 = 1$
for $j = 1, \ldots, k$ **do**
  $w_j = A \cdot v_j$
  Block Gram-Schmidt
    $h_{i,j} = v_i^* w_j$ for $i = 1, \ldots, j$
    $f_j = w_j - \sum_{i=1}^{j} v_i h_{i,j}$
    $h_{j+1,j} = \|f_j\|_2$
    $v_{j+1} = f_j / h_{j+1,j}$
  end block
end for

Eliminate w_j and f_j:

$$Av_j = \sum_{i=1}^{j} v_i h_{i,j} + f_j = \sum_{i=1}^{j+1} v_i h_{i,j}$$

Arnoldi method

Recurrence relations

$$Av_j = \sum_{i=1}^{j} v_i h_{i,j} + f_j = \sum_{i=1}^{j+1} v_i h_{i,j}$$

Define

$$H_k = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,k} \\ h_{2,1} & \ddots & & \vdots \\ 0 & \ddots & & \vdots \\ 0 & & h_{k,k-1} & h_{k,k} \end{bmatrix} \in \mathbb{C}^{k \times k}$$

Arnoldi factorization:

$$AV_k - V_k H_k = f_k e_k^T$$

From $V_k^* V_k = I$ and $V_k^* f_k = 0$, we find: $H_k = V_k^* A V_k$

Arnoldi method

Computation of Ritz values using Galerkin projection $x = V_k z$:

$$\begin{aligned} Ax - \lambda x &\perp \text{Range}(V_k) \\ V_k^*(AV_k z - \lambda V_k z) &= 0 \\ (V_k^* A V_k)z &= \lambda (V_k^* V_k) z \\ H_k z &= z \end{aligned}$$

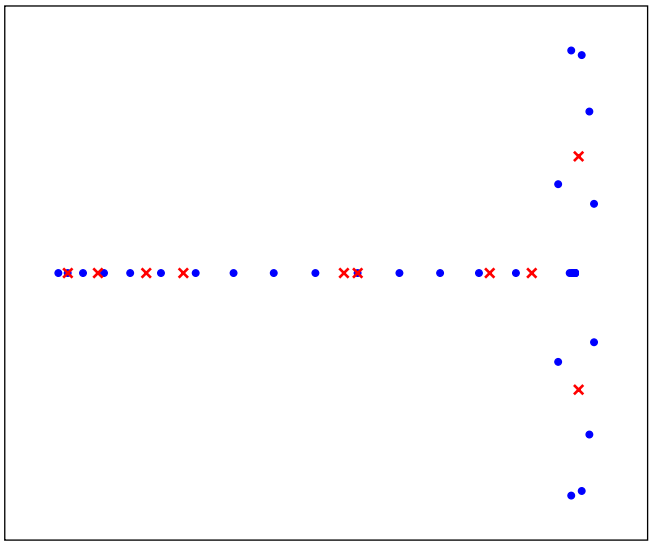
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 10$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

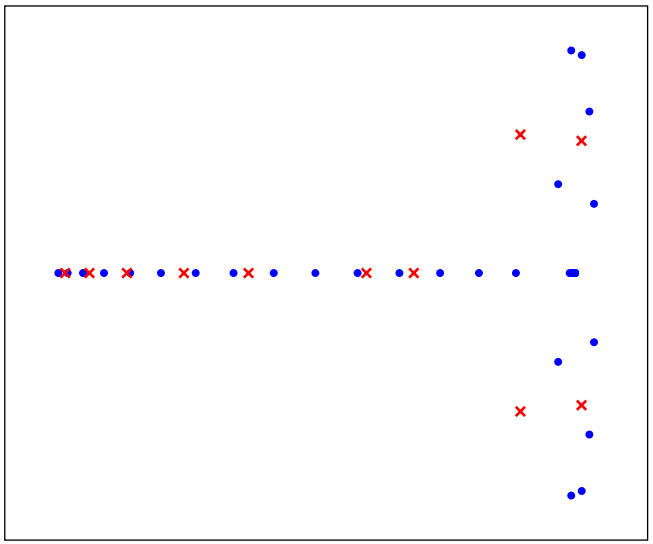
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 11$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

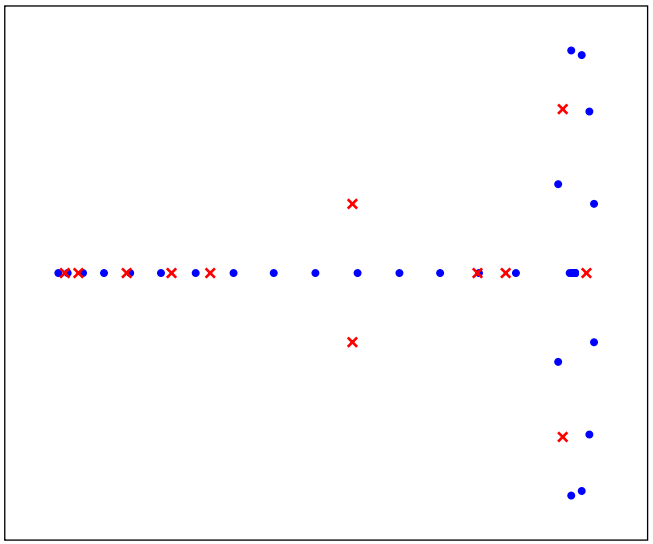
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 12$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

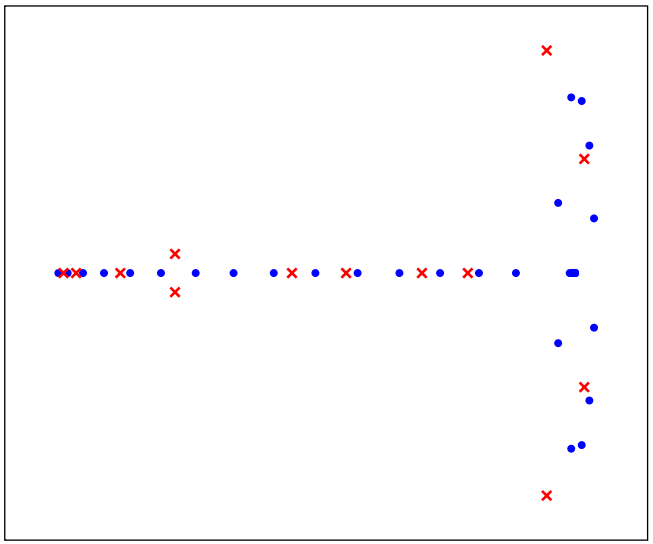
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 13$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

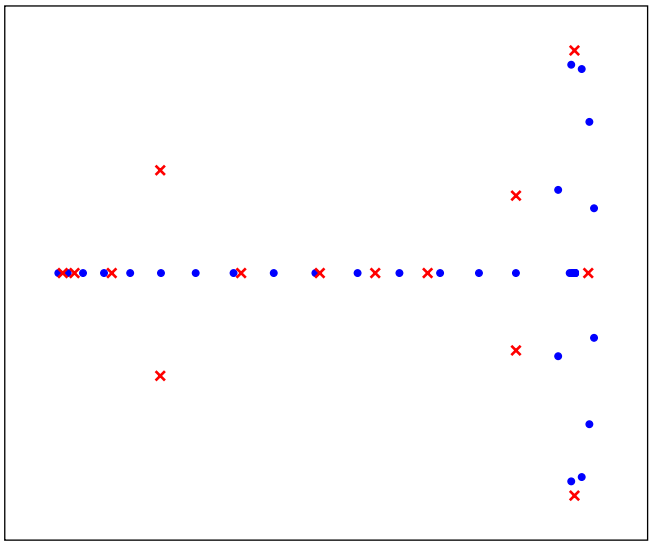
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 14$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

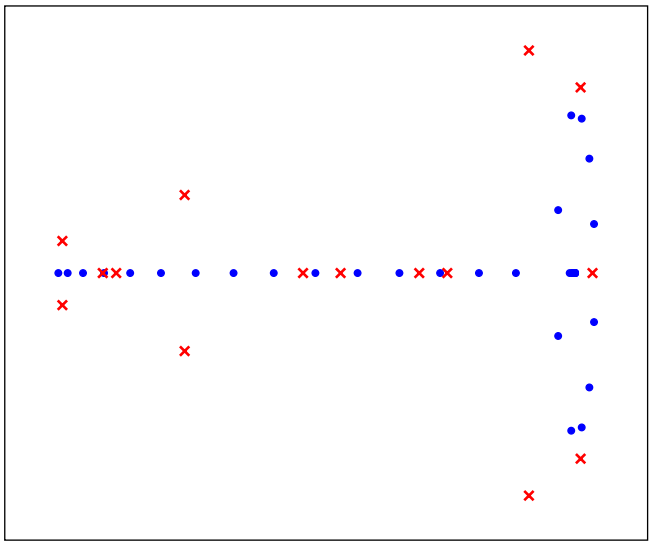
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 15$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

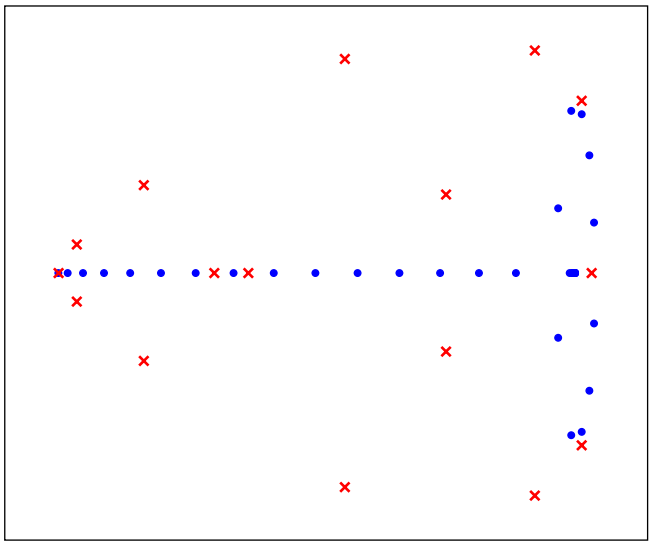
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 16$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

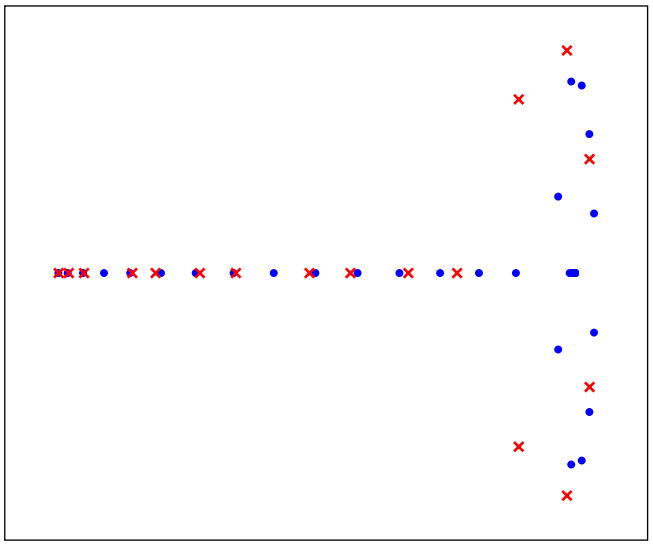
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 17$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

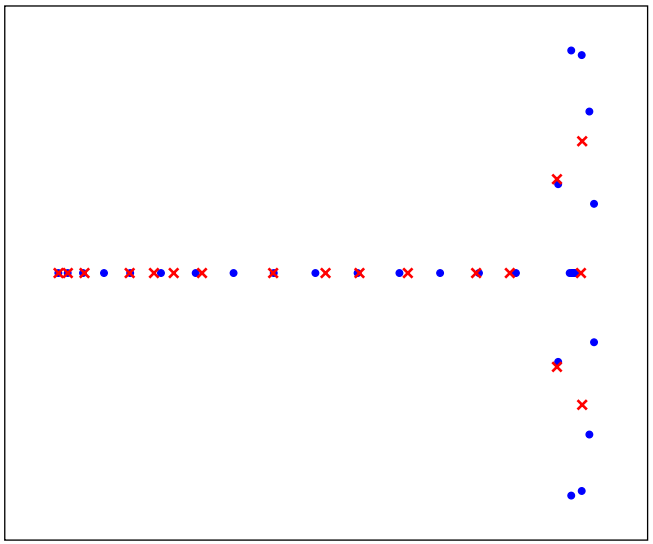
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 18$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

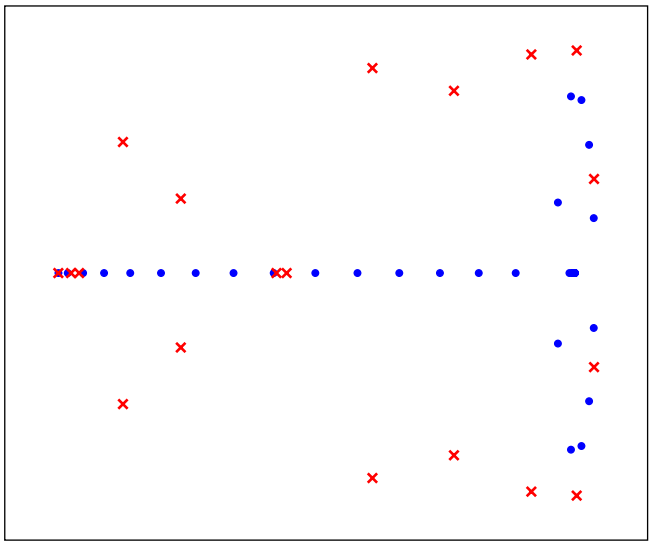
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 19$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

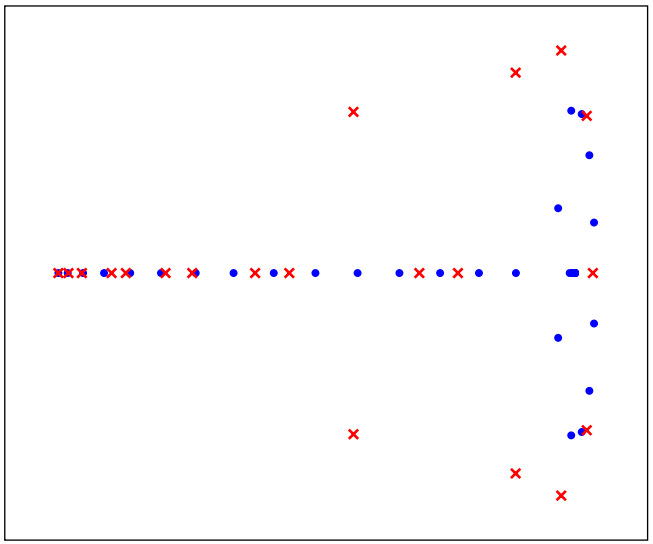
Arnoldi method

Convergence:

- Convergence rate for (λ_i, x_i):

$$\tan(v_1, y_i) \cdot \min_{p \in \mathbb{P}_k, p(\lambda_i)=1} \max_{\lambda \in \Lambda(A) \setminus \{\lambda_i\}} |p(\lambda)|$$

Example (Olmstead) – $k = 20$



- In words: well-separated extreme eigenvalues converge first.

- Krylov space is a space of polynomials. Fast convergence when there is a polynomial p, so that $p(A)v_1$ makes small angle with an eigenvector associated with λ_i.
- The Krylov spaces for A and $A - \sigma I$ are the same. Shifting does not change convergence.

Stopping criteria

- Residual: Given $\tilde{\lambda} \in \mathbb{C}$ and $\tilde{x} \in \mathbb{C}^n$, $\|\tilde{x}\|_2 = 1$

$$r = A\tilde{x} - \tilde{\lambda}\tilde{x}$$

- Backward error: $(\tilde{\lambda}, \tilde{x})$ is an eigenpair of

$$A - \frac{r\tilde{x}^*}{\tilde{x}^*\tilde{x}}$$

- In Arnoldi:

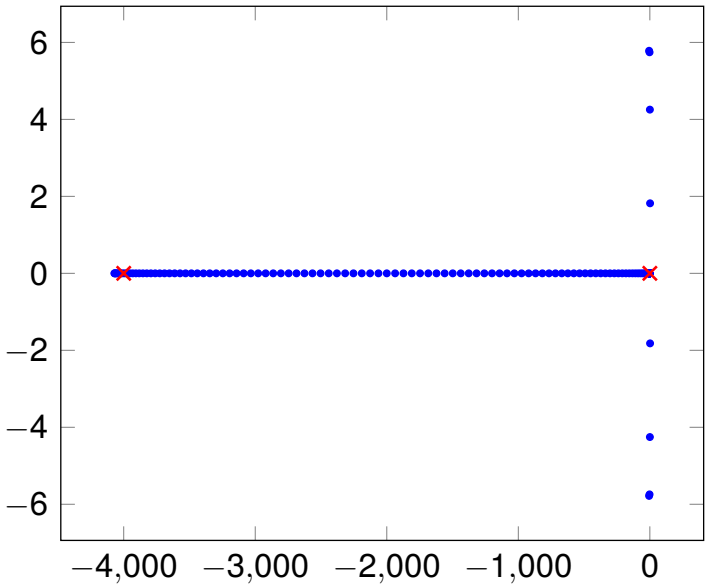
$$\begin{aligned} A\tilde{x} - \tilde{\lambda}\tilde{x} &= AV_k z - \tilde{\lambda} V_k z \\ &= V_k H_k z - \tilde{\lambda} V_k z + v_{k+1}\beta_k e_k^T z \\ &= f_k e_k^T z \\ \|A\tilde{x} - \tilde{\lambda}\tilde{x}\| &= h_{k+1,k}|e_k^T z| \end{aligned}$$

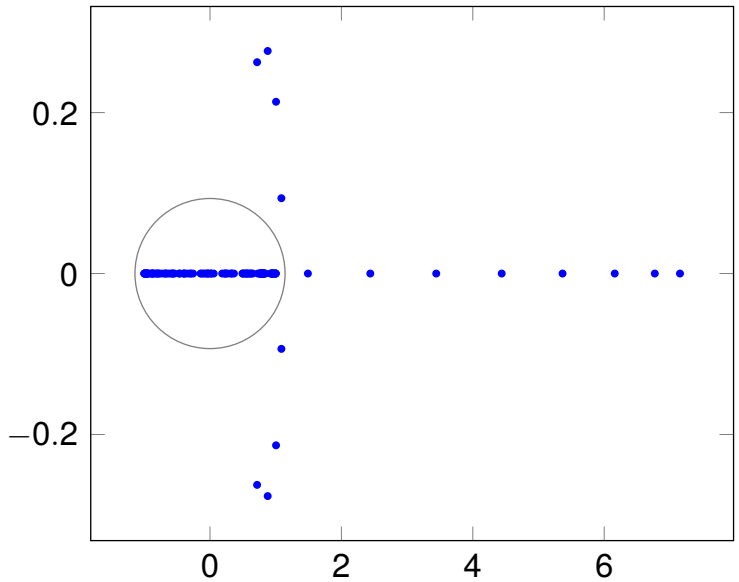
Matrix transformations

- The Arnoldi method is faster than subspace iteration (usually), but often still converges very slowly, in particular for large scale problems arising from PDEs for similar reasons as iterative methods for linear systems of equations.
- The main problem:
  - Memory cost: nk
  - Gram-Schmidt computational cost: nk^2
- Solutions:
  - Matrix transformation
  - Restart (as in restarted GMRES)

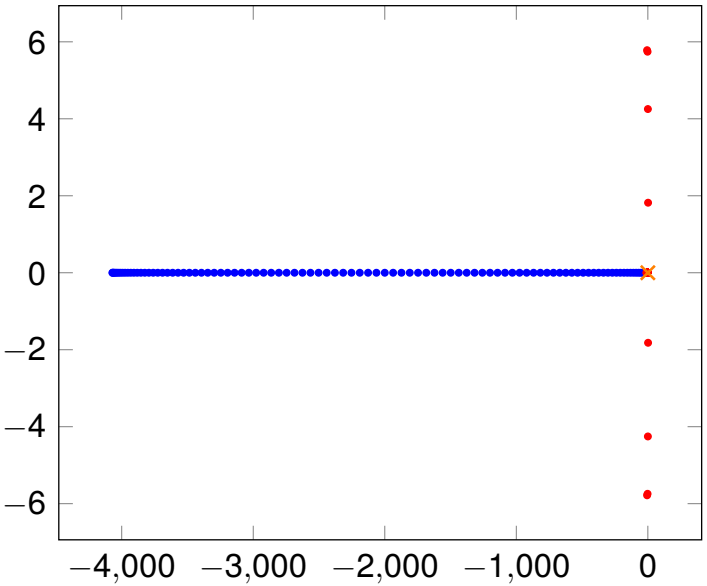
Chebyshev transformation

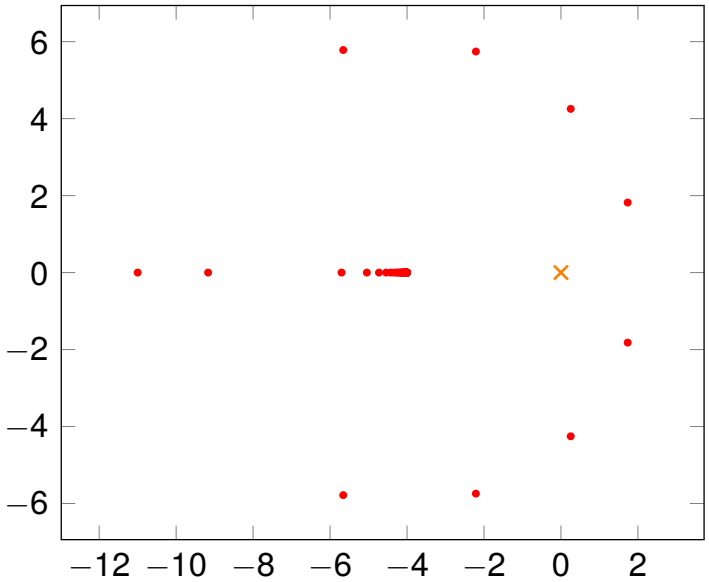
- As for subspace iteration, compute the eigenvalues of $T_k(A)$ instead of A.
- Eigenvalues of $T_k(A)$ are better separated, so, convergence is faster.





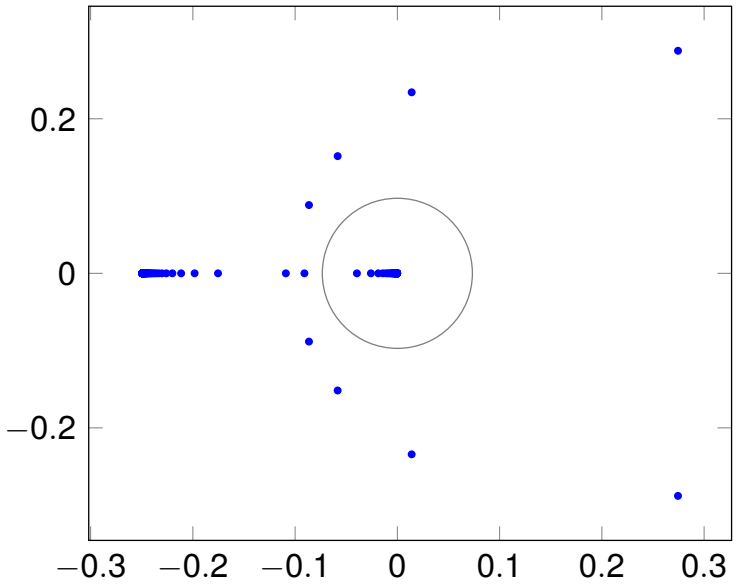
Shift-and-invert transformation





$$T = (A - \sigma I)^{-1}$$

- The most important transformation
- Based on inverse iteration



Shift-and-invert transformation

Convergence of the Shift-and-Invert Arnoldi method:

- Some say: convergence to the eigenvalues nearest σ: this is true for subspace iteration but not for Arnoldi.
- There is also convergence to well separated eigenvalues further away from σ.
- The eigenvalues nearest σ converge well, but other eigenvalues may also converge.
- Shift-and-invert is used in all kinds of situations, but keep this is in mind.

Implementation:

- One sparse *LU* factorization
- k backward solves
- This is what makes this method so popular

Implicit restarting

- Explicit restart
  - Improving one Ritz vector:

$$\{x_1, Ax_1, \ldots, A^{k-1}x_1\}$$

  - Improving two Ritz vectors:

$$\{(\alpha_1 x_1 + \alpha_2 x_2), A(\alpha_1 x_+ \alpha_2 x_2), \ldots, A^{k-1}(\alpha_1 x_+ \alpha_2 x_2)\}$$

    How do we choose α_1 and α_2?
  - Polynomial filter:

$$\{\phi_p(A)v_1, A\phi_p(A)v_1, \ldots, A^{k-1}\phi_p(A)v_1\}$$

    How do we choose the polynomial ϕ_p?
- Implicit restart

Implicit restart

- [Morgan, 1996] showed that the best linear combination is

$$v_1^+ = \rho_1 x_1 + \cdots, \rho_p x_p$$

with ρ_i the residual norms.

Implicit restart

- [Morgan, 1996] showed that the best linear combination is

$$v_1^+ = \rho_1 x_1 + \cdots, \rho_p x_p$$

  with ρ_i the residual norms.
- He proves that this is the same starting vector as

$$v_1^+ = \phi_p(A)v_1 \quad , \quad \phi_p(\lambda) = (\lambda - \lambda_1)\cdots(\lambda - \lambda_p)$$

Implicit restart

- [Morgan, 1996] showed that the best linear combination is

$$v_1^+ = \rho_1 x_1 + \cdots , \rho_p x_p$$

  with ρ_i the residual norms.
- He proves that this is the same starting vector as

$$v_1^+ = \phi_p(A)v_1 \quad , \quad \phi_p(\lambda) = (\lambda - \lambda_1)\cdots(\lambda - \lambda_p)$$

- [Sorensen 1992] showed that this is done by implicit restarting:
  1. QR factorization of $QR = \phi_p(H_k)$
  2. Keep the first p columns of Q
  3. Compute $V_p^+ = V_k Q$ and $H_p^+ = Q^* H_k Q$

Implicit restart

- [Morgan, 1996] showed that the best linear combination is

$$v_1^+ = \rho_1 x_1 + \cdots, \rho_p x_p$$

  with ρ_i the residual norms.
- He proves that this is the same starting vector as

$$v_1^+ = \phi_p(A)v_1 \quad , \quad \phi_p(\lambda) = (\lambda - \lambda_1)\cdots(\lambda - \lambda_p)$$

- [Sorensen 1992] showed that this is done by implicit restarting:
  1. QR factorization of $QR = \phi_p(H_k)$
  2. Keep the first p columns of Q
  3. Compute $V_p^+ = V_k Q$ and $H_p^+ = Q^* H_k Q$
- [M. & Spence 1997] and [Lehoucq, 1999] show that

$$\text{Range}(V_p^+) = \text{Range}(\phi_p(A)V_p)$$

  (= polynomial subspace iteration)

Implicit restart

Before 'implicit restart':

- Subspace dimension is k
- Ritz vectors $x_1, \ldots, x_p$.

After 'implicit restart':

- Subspace dimension is reduced from k to p.
- Ritz vectors $x_1, \ldots, x_p$.
- Recurrence relation: $AV_p^+ - V_p^+ H_p^+ = f_p^+ e_p^T$

$k - p$ additional Krylov steps:

- Subspace dimension increased from p to k.
- Recurrence relation: $AV_k^+ - V_k^+ H_k^+ = f_k^+ e_k^T$
- Is the same subspace as

$$\text{span}\{x_1, \ldots, x_p, Ax, \ldots, Ax^{k-p}\}$$

with x any from $\{x_1, \ldots, x_p\}$.

Implicit restart

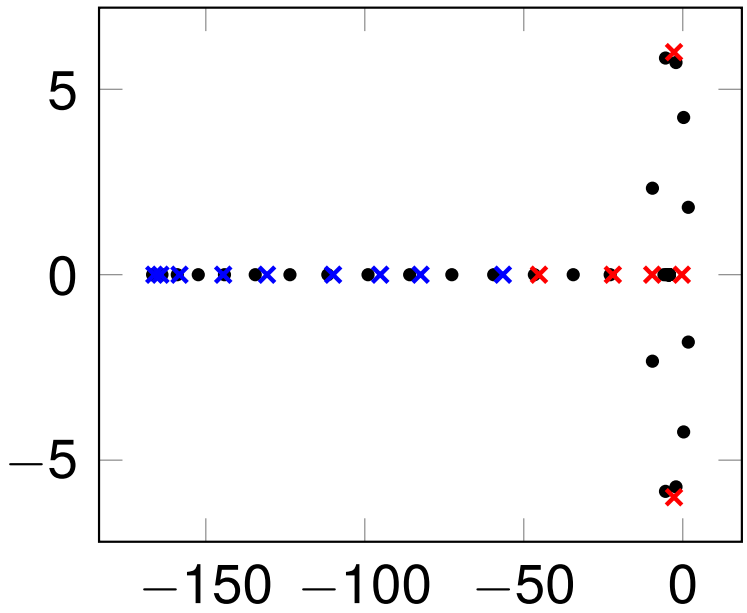
1: k Arnoldi steps with starting vector v_1: V_k, H_k
2: **while** Desired Ritz pairs have not converged **do**
3: Compute Ritz pairs and select ϕ_{k-p}.
4: QR on H_k with shifts: H_p^+, V_p^+.
5: $k-p$ Arnoldi steps to expand V_p^+ to V_k^+.
6: **end while**

- Convergence is similar to polynomial subspace iteration, but with cheaper iterations [Lehoucq, 1998]
- Convergence of full Arnoldi and restarted Arnoldi for selected eigenvalues is very much alike [Morgan 1995]
- In practice use the QR method's bulge chasing instead of Arnoldi on H_k.
- More numerically stable, but mathematically equivalent: Krylov Schur [Stewart 2001] (was known by Lehoucq in 1995).

Implicit restart

Polynomial filters $\phi_{k-p}(\lambda) = (\lambda - \sigma_1) \cdots (\lambda - \sigma_{k-p})$:

- Exact shifts: shifts are Ritz values (Sorensen & Morgan)
- Chebyshev shifts: select the parameters of the Chebyshev polynomial from the Ritz values and filter out the ellipse with unwanted eigenvalues
- Leja shifts (potential theory) [Calvetti, Reichel, Sorensen, 1994]
- Zero shifts (see further)



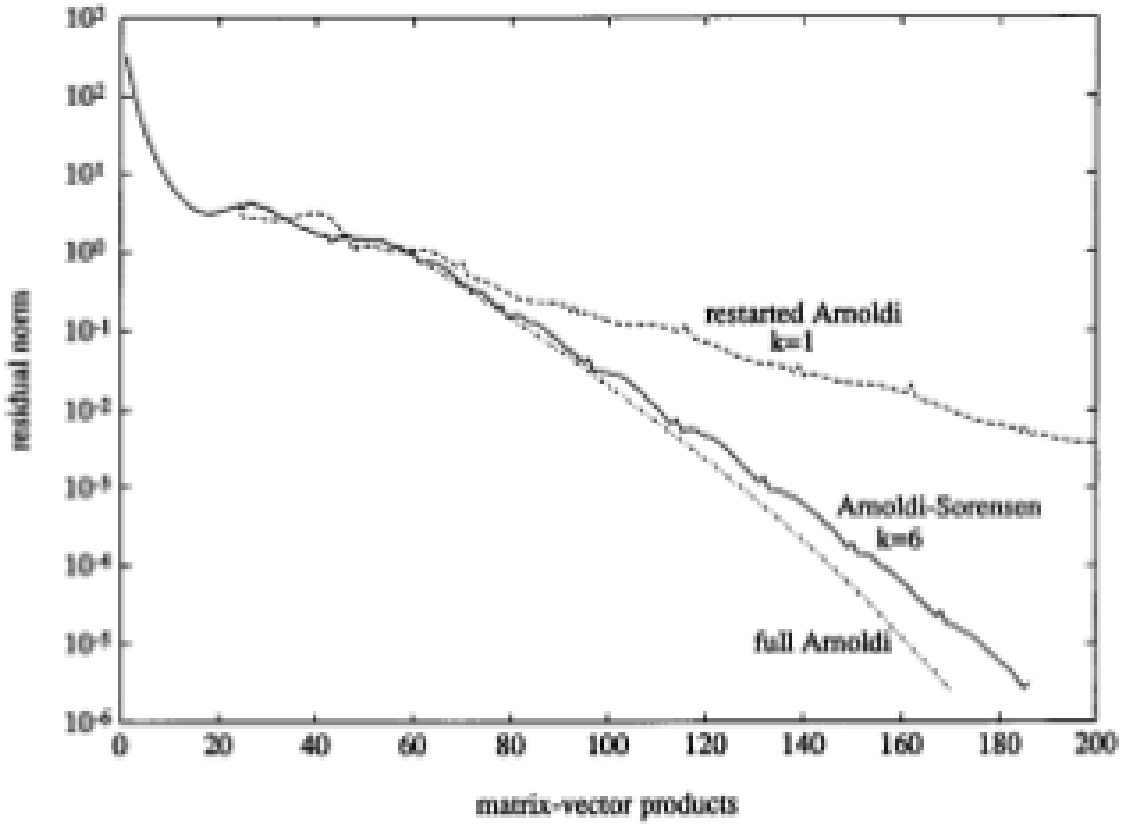
Implicit restart

Matrix is tridiagonal:

- Main diagonal: $1, \ldots, 1000$
- Superdiagonal: -0.1
- Subdiagonal: 0.1

Arnoldi:

- $k = 24, p = 6$
- 10 sweeps of implicit restarts are compared to full Arnoldi with $24 + 9 \cdot 18 = 186$ iterations



[Morgan, 1996]

Newton method

- Apply Newton (Raphson) to:

$$\begin{aligned} Ax - \lambda x &= 0 \\ x^* x &= 1 \end{aligned}$$

  for (λ, x)
- $\lambda^{(k+1)} = \lambda^{(k)} + \Delta\lambda$, $x^{(k+1} = x^{(k)} + \Delta x$:

$$\begin{bmatrix} A - \lambda^{(k)} I & -x^{(k)} \\ 2(x^{(k)})^* & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} Ax^{(k)} - \lambda^{(k)} x^{(k)} \\ \|x^{(k)}\|_2 - 1 \end{pmatrix}$$

- Explicitly normalize $x^{(k)}$ and on every iteration:

$$x^{(k+1)} = -\Delta\lambda (A - \lambda^{(k)} I)^{-1} x^{(k)}$$

Rational Krylov sequences

A generalization of the shift-and-invert Arnoldi method:

1: Given v_1 with $\|v_1\|_2 = 1$.
2: **for** $j = 1, \ldots, k$ **do**
3: Solve $(A - \sigma_j I)w_j = v_j$
4: **Block** Gram-Schmidt
5: $h_{i,j} = v_i^* w_j$ for $i = 1, \ldots, j$
6: $f_j = w_j - \sum_{i=1}^{j} v_i h_{i,j}$
7: $h_{j+1,j} = \|f_j\|_2$
8: $v_{j+1} = f_j / h_{j+1,j}$
9: **end block**
10: **end for**

Eliminate w_j and f_j:

$$(A - \sigma_j I)^{-1} v_j = \sum_{i=1}^{j} v_i h_{i,j} + f_j = \sum_{i=1}^{j+1} v_i h_{i,j}$$

Rational Krylov sequences

Recurrence relations

$$(A - \sigma_j I)^{-1} v_j = \sum_{i=1}^{j} v_i h_{i,j} + f_j = \sum_{i=1}^{j+1} v_i h_{i,j}$$

Rational Krylov factorization:

$$A V_k H_k - V_k K_k = (A - \sigma_k I) f_k e_k^T$$

with

$$K_k = I_k + H_k \Sigma_k \quad \text{with} \quad \Sigma_k = \text{diag}(\sigma_1, \ldots, \sigma_k)$$

Ritz values by Galerkin projection of $(A - \sigma_k I)^{-1}$ on Range(V_k):

$$K_k z = \lambda H_k z \quad , \quad x = V_k H_k z$$

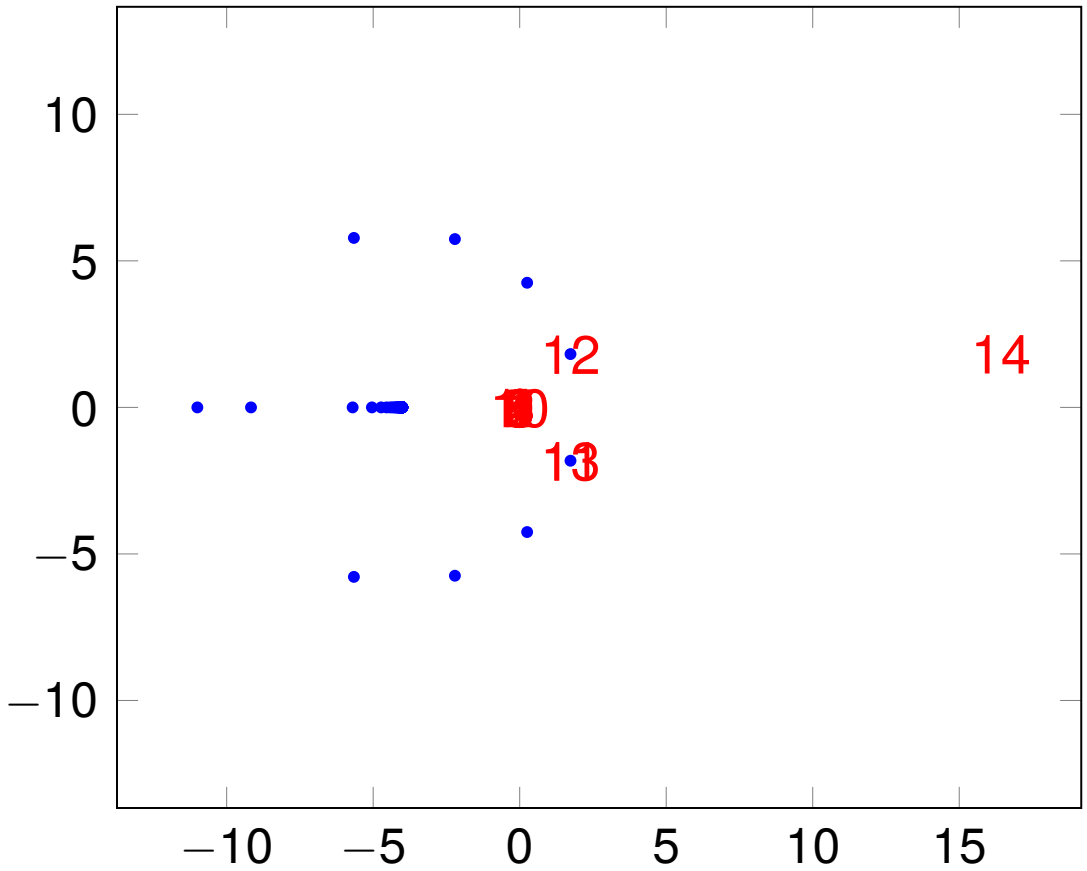
Residual

$$Ax - \lambda x = (A - \sigma_k I) f_k (e_k^T z)$$

Rational Krylov sequences

Example 1: rational Krylov as a Newton method for finding eigenvalue with largest real part:

1: Given v_1 with $\|v_1\|_2 = 1$.
2: Given $\sigma_1 = 0$.
3: **for** $j = 1, \ldots, k$ **do**
4: Solve $(A - \sigma_j)w_j = v_j$
5: Gram-Schmidt of w_j
6: Compute Ritz values
7: Select σ_{j+1} as the Ritz value with largest real part.
8: **end for**



Be careful: Newton steps may introduce large errors in the recurrence relation.

Rational Krylov sequences

Implementation issues:

- Linear systems have to be solved accurately

$$AV_kH_k - BV_kK_k = (A - \sigma_k I)f_k e_k^T + R_k$$

  with $\|R_k\| \approx \|(A - \sigma_j)w_j - v_j\|_2$.
- Linear solver is usually a direct method.
- Matrix factorization is often most expensive operation: reuse shifts σ_j.

Rational Krylov sequences

Slicing for symmetric eigenvalue problem (structural dynamics/acoustics)

Rational Krylov: Implicit restarting

- Restart with filtered starting vector/subspace:

$$v_1^+ = \phi_{k-p}(A)v_1 \quad \text{with} \quad \phi_{k-p}(z) = \frac{z-\mu_1}{z-\sigma_1}\cdots\frac{z-\mu_{k-p}}{z-\sigma_{k-p}}$$

$$\text{Range}(V_p^+) = \text{Range}(\phi_{k-p}(A)V_p)$$

- QZ step on H_k, K_k:

$$\begin{aligned} H_p^+ &= Q^* H_k Z \\ K_p^+ &= Q^* K_k Z \\ V_p^+ &= V_k Q \end{aligned}$$

- Recurrence relation:

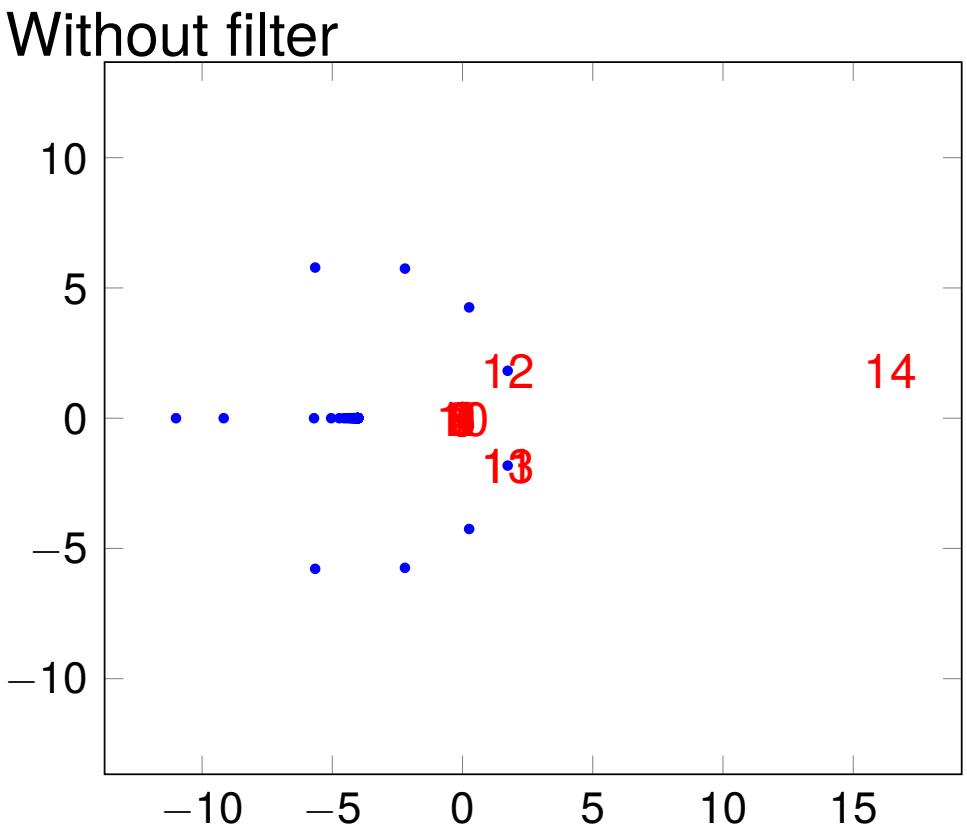
$$AV_p^+ H_p^+ - V_p^+ K_p^+ = (A - \sigma_k I) f_p^+ e_p^T$$

The first p poles are $\sigma_{k-p+1}, \ldots, \sigma_k$.

- Implementation:
  - Naive [De Samblanx, M. & Bultheel, 1997] (including Krylov-Schur)
  - Bulge chasing (QZ method) [Camps, M. & Vandebril, 2017]

Rational Krylov: Implicit restart with zero shift

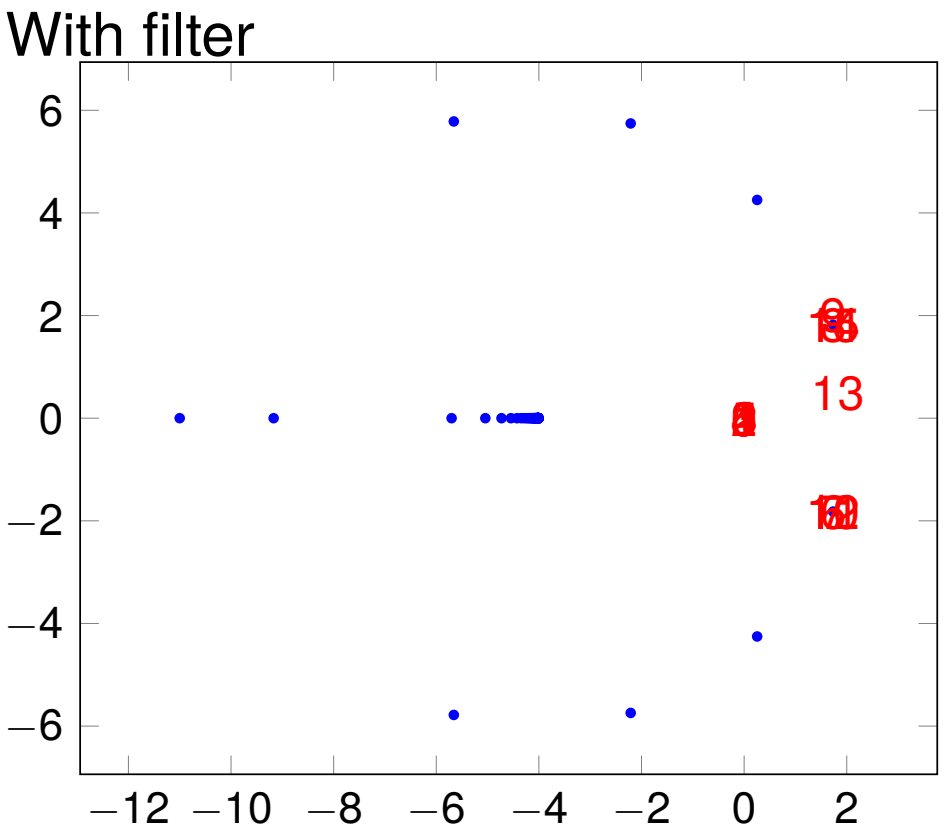
- Shift-and-invert transform has a cluster of eigenvalues near zero.
- This sometimes leads to spurious eigenvalues.
- The use harmonic Ritz values may help [Morgan, 1991] [Paige, Parlett, van der Vorst, 1995].
- Another way is to filter away the eigenvalues near zero [M. & Spence 1997]:



$$\text{Range}(V_{k-1}^{+}) = \text{Range}((A - \sigma_k I)^{-1} V_{k-1})$$

Rational Krylov: Implicit restart with zero shift

- Shift-and-invert transform has a cluster of eigenvalues near zero.
- This sometimes leads to spurious eigenvalues.
- The use harmonic Ritz values may help [Morgan, 1991] [Paige, Parlett, van der Vorst, 1995].
- Another way is to filter away the eigenvalues near zero [M. & Spence 1997]:



$$\text{Range}(V_{k-1}^{+}) = \text{Range}((A - \sigma_k I)^{-1} V_{k-1})$$

Generalized eigenvalue problems

Given matrices $A, B \in \mathbb{C}^{n \times n}$:

$$Ax = \lambda Bx$$

Regular eigenvalue problem: A and B do not have a common nullspace, i.e., there are α, β so that $\alpha A + \beta B$ is non-singular.

Shift-and-invert operator:

$$\begin{aligned} Ax &= \lambda Bx \\ (A - \sigma B)x &= (\lambda - \sigma)Bx \\ (\lambda - \sigma)^{-1}x &= (A - \sigma B)^{-1}Bx \end{aligned}$$

Apply Arnoldi on $(A - \sigma B)^{-1}B$ or use rational Krylov.

Generalized eigenvalue problems

Symmetric positive definite A and B.
[Grimes, Lewis, Simon, 1994]

- Choose shifts σ_j in between clusters of eigenvalues

- $(A - \sigma B)^{-1}B$ is nonsymmetric:
  - Use B inner product: $V_k^* B V_k = I$.
  - $(A - \sigma B)^{-1}B$ is self adjoint with the B-inner product:

$$y^* B\big((A - \sigma B)^{-1}Bx\big) = \big((A - \sigma B)^{-1}By\big)Bx$$

Generalized eigenvalue problems

Let B be singular and $A - \sigma B$ be nonsingular for some σ, then

1. for all $x : Bx = 0$, we have $(A - \sigma B)^{-1} Bx = 0$
2. which corresponds to $Ax = \infty Bx$.

Such problems arise from DAEs (differential algebraic equations).

The infinite eigenvalue is usually undesired, but it may hinder create spurious eigenvalues.
Tilted plane benchmark from rational Krylov [De Samblanx, M. & Bultheel, 1997]:

Iteration	Without filter	With filter
3	8.432	-8.4677
6	19.751	-9.4833
9	74.83	-9.48831

Implicit filtering: multiply the Krylov space with $(A - \sigma B)^{-1} B$

Other methods

- For both right and left eigenvectors:
  - One Krylov space with A for right eigenvectors
  - One Krylov space with A^* for left eigenvectors
  - Lanczos method, two-sided Arnoldi method
  - Higher risk for spurious eigenvalues $\rightarrow$ stabilize by an implicit restart with zero shift
- Linear systems with iterative solvers:
  - Jacobi-Davidson [Sleijpen & van der Vorst]
  - LOBPCG [Knyazev]
  - Tuned preconditioner [Spence & Freitag]
- Contour integral methods

Block Arnoldi method

Arnoldi's method applied to a block of b vectors $V_1 \in \mathbb{C}^{n \times b}$ with $b > 1$:

$$\{V_1, AV_1, A^2V_1, \ldots A^{k-1}V_1\}$$

Arnoldi algorithm produces orthonormal basis $V_k = [V_1, \ldots, V_k]$:

Given V_1 with $\|V_1\|_2 = 1$
for $j = 1, \ldots, k$ **do**
  $W_j = A \cdot V_j$
  Block Gram-Schmidt
    for $i = 1, \ldots, b$ **do**
      $V_{j+1} = []$
      Orthogonalize $w_{j,i} = W_j e_i$ against $[V_1, \ldots, V_{j+1}]$
      $V_{j+1} = [V_{j+1}\ w_{j,i}/\|w_{j,i}\|_2]$
    end for
  end block
end for

This is like subspace iteration, where the iterations are accumulated in a subspace.

Two-sided Krylov methods

For computing eigenvalues, right and left eigenvectors:

$$Ax = \lambda Bx \qquad y^*A = \lambda y^*B$$

Two Krylov spaces

- $(A - \sigma B)^{-1}B \Rightarrow V_k \Rightarrow x$
- $(A^* - \sigma B^*)^{-1}B^* \Rightarrow W_k \Rightarrow y$

Compute Ritz triples from reduced problem:

$$W^*AVz = \lambda W^*BVz$$

- Lanczos method: use B bi-othogonalization: $W^*BV = I$. [Bai & Ye 2001]
- Two-sided Arnoldi method: compute projection explicitly [Ruhe ...]
- Implicit restarting [De Samblanx & Bultheel, 1998] for Lanczos, [Hochstenbach & Zwaan, 2017] for Arnoldi

Statistical approaches

Jacobi-Davidson method

- Shift-and-invert with iterative solver
- In order to avoid the need for 'exact' solves, JD solves the *correction equation* iteratively:

$$(I - yx^*/(x^*y))(A - \lambda I)(I - xx^*/(x^*))z = -(A - \lambda I)x$$

$$\text{with} \quad \lambda = x^*Ax/(x^*x)$$

- Preconditioning possible, but hard
- I prefer to solve a shifted system inexactly:

$$(A - \sigma I)z = -(A - \lambda I)x$$

with σ the 'target'. The preconditioiner can be reused as long as σ is not changed.

Locally Optimal Block Preconditioned Conjugate Gradient (LOBPCG)

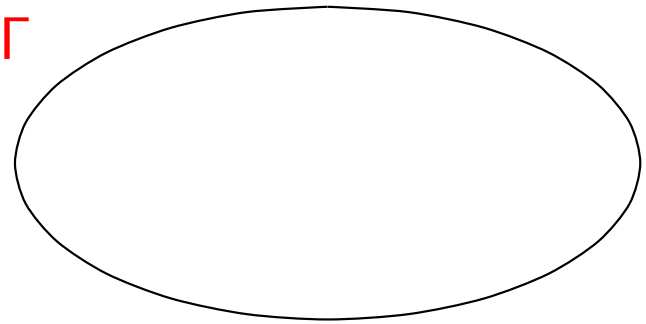
- For symmetric $Ax = \lambda Bx$ with positive definite B
- the Rayleigh quotient

$$\frac{x^T Ax}{x^T Bx}$$

  is maximum for the largest eigenvalue $\lambda_{\max}$ and minimum for the smallest eigenvalue $\lambda_{\min}$
- LOBPCG is the conjugate gradient method applied to this optimization problem

Contour integral methods

- Let $\Gamma \subset \mathbb{C}$ be a closed contour in the complex plane.
- Define



$$C_i = \int_\Gamma z^i (zI - A)^{-1} \mathrm{d}z \in \mathbb{C}^{n \times n}$$

- The rank of C_0 is the number of eigenvalues of A inside Γ.
- Eigenvalue problem:

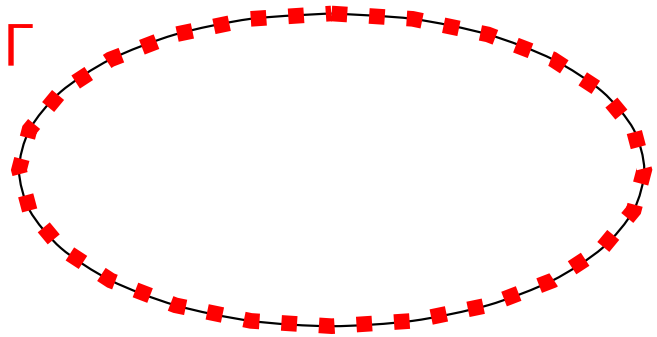
$$C_1 x = \lambda x \quad \text{with} \quad C_0 x = x$$

  if λ lies inside the contour.
- Two Methods:
  - Subspace iteration
  - Arnoldi

Contour integral methods

- Basis operation: let $V \in \mathbb{C}^{n \times m}$.

$$\begin{aligned} C_i V &= \int_\Gamma z^i (zI - A)^{-1} V \mathrm{d}z \\ &\approx \sum_{j=1}^{N} w_j z_j^i (z_j I - A)^{-1} V \end{aligned}$$



- Solution of N linear systems. Can be done in parallel.
- Subspace iteration.
  - Compute $V_0 = C_0 V$ and orthonormalize (rank revealing QR or SVD)
  - The rank of V_0 is the number of eigenvalues in the contour
  - Compute the eigenvalues of $V_0^* C_1 V_0$.
- (Block) Arnoldi:
  - Starting vector $v_1 = C_0 v$ and normalize
  - Perform Arnoldi with C_1 with starting vector v_1.
- Expensive methods!
- Contour integration leads to rational filter methods, which need significantly less points (smaller N)