Arnoldi Method

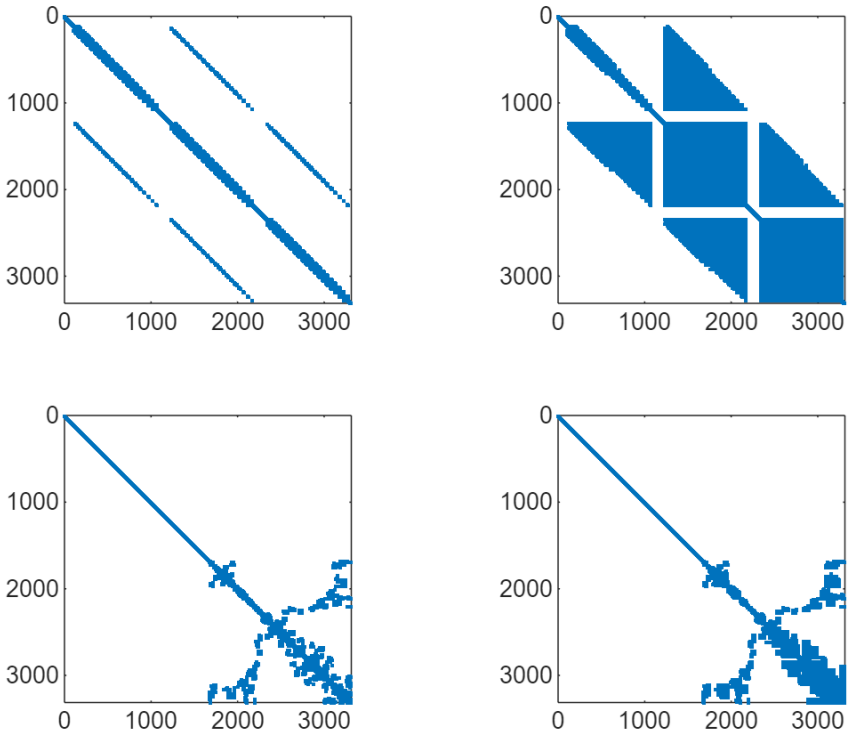
Professor Tim Moroney

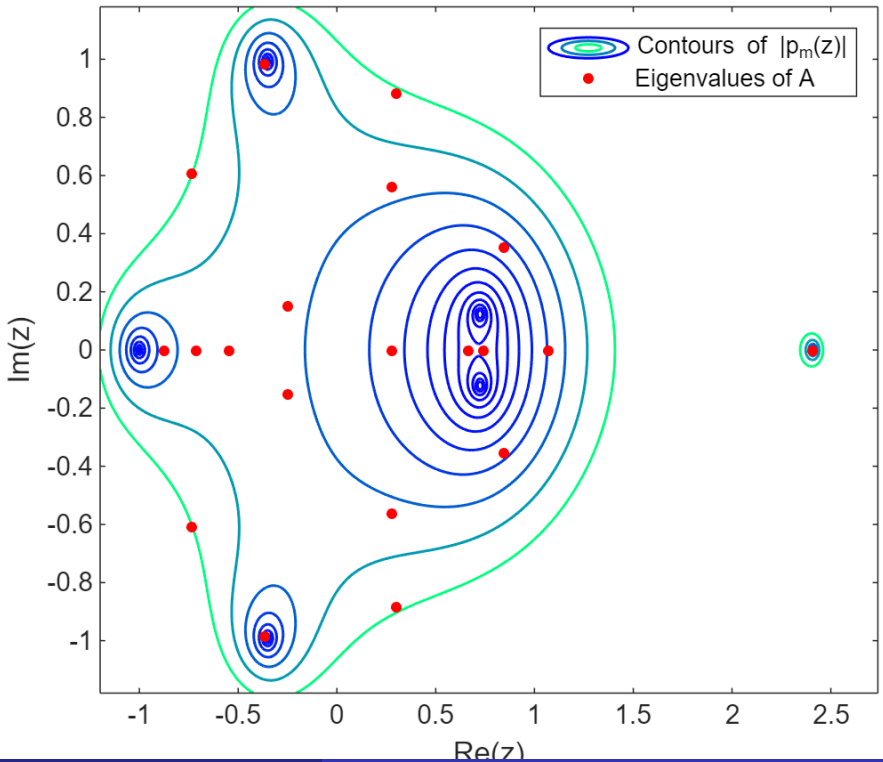
School of Mathematical Sciences, QUT

Outline

Iterative methods for sparse matrix problems

- Arnoldi method for eigenvalues (2 hr)
- GMRES for linear systems (2 hr)
- Extensions (1 hr)





Sparse Matrices

- Matrices with most entries zero: we store and operate only on the nonzeros
- Save memory and computation – essential for very large systems (e.g. PDE discretisations)
- Various storage formats (coordinate, CSR/CSC, etc.) – tradeoffs between memory usage, implementation complexity and operation speed
- Iterative methods (Arnoldi, GMRES, etc.) with only minimal requirements, e.g. matrix-vector product
- Applications to scientific computing, optimisation, simulation, machine learning, etc.

Eigenvalue problem

Today we consider the eigenvalue problem:

$$Ax = \lambda x$$

A is large, sparse, nonsymmetric, but, we assume, diagonalisable

Can be complex (everything works fine) but we present only real examples

Our goal is to compute some (not all) of the eigenvalues of A

In practice, we estimate the "extreme" eigenvalues of A (those near the edge of the spectrum – most often relevant in applications)

We'll consider two algorithms:

- Power method (stepping stone)
- Arnoldi method (main focus)

Power method

Suppose A has a dominant eigenvalue, $|\lambda_1| > |\lambda_2| \geq \cdots$ and has a complete eigenbasis $\{x_1, x_2, \ldots, x_n\}$.

Choose a starting vector b at random.

Power method

Suppose A has a dominant eigenvalue, $|\lambda_1| > |\lambda_2| \geq \cdots$ and has a complete eigenbasis $\{x_1, x_2, \ldots, x_n\}$.

Choose a starting vector b at random.

$$b = \sum_i c_i x_i = c_1 x_1 + \sum_{i \geq 2} c_i x_i$$

Power method

Suppose A has a dominant eigenvalue, $|\lambda_1| > |\lambda_2| \geq \cdots$ and has a complete eigenbasis $\{x_1, x_2, \ldots, x_n\}$.

Choose a starting vector b at random.

$$b = \sum_i c_i x_i = c_1 x_1 + \sum_{i \geq 2} c_i x_i$$

Then

$$Ab = c_1 A x_1 + \sum_{i \geq 2} c_i A x_i$$

Power method

Suppose A has a dominant eigenvalue, $|\lambda_1| > |\lambda_2| \geq \cdots$ and has a complete eigenbasis $\{x_1, x_2, \ldots, x_n\}$.

Choose a starting vector b at random.

$$b = \sum_i c_i x_i = c_1 x_1 + \sum_{i \geq 2} c_i x_i$$

Then

$$\begin{aligned} Ab &= c_1 A x_1 + \sum_{i \geq 2} c_i A x_i \\ &= c_1 \lambda_1 x_1 + \sum_{i \geq 2} c_i \lambda_i x_i \end{aligned}$$

Power method

Suppose A has a dominant eigenvalue, $|\lambda_1| > |\lambda_2| \geq \cdots$ and has a complete eigenbasis $\{x_1, x_2, \ldots, x_n\}$.

Choose a starting vector b at random.

$$b = \sum_i c_i x_i = c_1 x_1 + \sum_{i \geq 2} c_i x_i$$

Then

$$\begin{aligned} Ab &= c_1 A x_1 + \sum_{i \geq 2} c_i A x_i \\ &= c_1 \lambda_1 x_1 + \sum_{i \geq 2} c_i \lambda_i x_i \end{aligned}$$

and

$$A^k b = c_1 {\lambda_1}^k x_1 + \sum_{i \geq 2} c_i {\lambda_i}^k x_i$$

Power method

$$A^k b = c_1 {\lambda_1}^k x_1 + \sum_{i \geq 2} c_i {\lambda_i}^k x_i$$

Power method

$$
\begin{aligned}
A^k b &= c_1 {\lambda_1}^k x_1 + \sum_{i \geq 2} c_i {\lambda_i}^k x_i \\
&= {\lambda_1}^k \left(c_1 x_1 + \sum_{i \geq 2} c_i \left(\frac{\lambda_i}{\lambda_1} \right)^k x_i \right)
\end{aligned}
$$

Power method

$$A^k b = c_1 {\lambda_1}^k x_1 + \sum_{i \geq 2} c_i {\lambda_i}^k x_i$$

$$= {\lambda_1}^k \left(c_1 x_1 + \sum_{i \geq 2} c_i \left(\frac{\lambda_i}{\lambda_1} \right)^k x_i \right)$$

So as $k \to \infty$ the bracketed factor $\to c_1 x_1$ geometrically at rate $\left|\lambda_2 / \lambda_1\right|^k$.

Conclusion: the direction of $A^k b$ tends to the dominant eigenvector x_1.

Power method

Choose $b \neq 0$ at random
Set $v_1 = b/\|b\|$
For $j = 1, 2, \ldots m$

$$w_{j+1} = Av_j$$
$$v_{j+1} = w_{j+1}/\|w_{j+1}\| \qquad \text{(normalise)}$$
$$\mu_{j+1} = v_{j+1}^* A v_{j+1} \qquad \text{(Rayleigh quotient)}$$

End

v_{j+1} is normalised eigenvector estimate.

Rayleigh quotient yields eigenvalue estimate:

$$Av_{j+1} \approx \lambda_1 v_{j+1}$$
$$\mu_{j+1} := v_{j+1}^* A v_{j+1} \approx \lambda_1 v_{j+1}^* v_{j+1} = \lambda_1 \|v_{j+1}\| = \lambda_1$$

Power method

We start with b, and progressively calculate Ab, A^2b, $A^3b, \ldots$ (ignore scaling for now).

The final eigen-estimates use only the last vector in this sequence.

Could we do better by incorporating all the intermediate vectors too (somehow)?

Observation: By step m of the method we have built a basis for the subspace

$$\mathcal{K}_m(A, b) = \text{span}\{b, Ab, A^2b, \ldots, A^{m-1}b\}$$

which is known as the Krylov subspace of dimension m (for this particular A and b).

Krylov subspace

$$\mathcal{K}_m(A, b) = \text{span}\{b, Ab, A^2 b, \ldots, A^{m-1} b\}$$

This is a very natural subspace to consider: formed only by repeated matrix-vector products with A.

Clearly it captures something about how A acts on vectors. There should be some information about the eigenvectors lurking in there!

The power method however throws nearly all of it away, and only uses the final vector.

How can we make better use of the full subspace of information?

Projection methods

Start with a large-dimensional problem (e.g. our large, sparse eigenvalue problem).

Project the problem down onto a small-dimensional space (e.g. our Krylov subspace).

Solve it exactly (and cheaply!) in the small-dimensional space.

Lift the solution back up to the full scale to obtain an approximate solution to the large-dimensional problem.

So first question:

- What does it mean to "project" the eigenvalue problem onto a Krylov subspace?

Maybe we're getting ahead of ourselves

Here's the Krylov subspace:

$$\mathcal{K}_m(A, b) = \text{span}\{b, Ab, A^2b, \ldots, A^{m-1}b\}$$

The naive choice of basis for this space, literally formed by the vectors $b, Ab, A^2b, \ldots$, would be hopelessly ill-conditioned.

The whole point of the power method is that the later vectors all point in pretty much the same direction!

So second question (which we'll answer first):

- How do we build a well-conditioned basis for the Krylov subspace?

The Arnoldi process

$\beta = \|b\|, \qquad v_1 = b/\beta$

For $j = 1, 2, \ldots m$

$$h_{ij} = \langle Av_j, v_i \rangle, \qquad i = 1, 2, \ldots, j$$

$$w_j = Av_j - \sum_{i=1}^{j} h_{ij} v_i$$

$$h_{j+1,j} = \|w_j\| \qquad \text{if } h_{j+1,j} = 0 \text{ then stop}$$

$$v_{j+1} = w_j / h_{j+1,j}$$

End

At any step m the columns of $V_m = [v_1 \; v_2 \cdots v_m]$ form an orthonormal basis for the Krylov subspace $\mathcal{K}_m(A, b) = \text{span}\{b, Ab, A^2 b, ..., A^{m-1} b\}$.

We have also generated the matrix $\overline{H}_m$ whose $(i, j)th$ entry is h_{ij}.

For $j = 1, 2, \ldots m$

$$h_{ij} = \langle Av_j, v_i \rangle, \qquad i = 1, \ldots, j$$

$$w_j = Av_j - \sum_{i=1}^{j} h_{ij} v_i$$

$$h_{j+1,j} = \|w_j\| \qquad \text{if } h_{j+1,j} = 0 \text{ then stop}$$

$$v_{j+1} = w_j / h_{j+1,j}$$

End

For $j = 1, 2, \ldots m$

$$h_{ij} = \langle Av_j, v_i \rangle, \qquad i = 1, \ldots, j$$

$$w_j = Av_j - \sum_{i=1}^{j} h_{ij} v_i$$

$$h_{j+1,j} = \|w_j\| \qquad \text{if } h_{j+1,j} = 0 \text{ then stop}$$

$$v_{j+1} = w_j / h_{j+1,j} \implies w_j = h_{j+1,j}\, v_{j+1}$$

End

For $j = 1, 2, \ldots m$

$$h_{ij} = \langle A v_j, v_i \rangle, \qquad i = 1, \ldots, j$$

$$w_j = A v_j - \sum_{i=1}^{j} h_{ij} v_i \implies h_{j+1,j}\, v_{j+1} = A v_j - \sum_{i=1}^{j} h_{ij} v_i$$

$$h_{j+1,j} = \|w_j\| \qquad \text{if } h_{j+1,j} = 0 \text{ then stop}$$

$$v_{j+1} = w_j / h_{j+1,j} \implies w_j = h_{j+1,j}\, v_{j+1}$$

End

For $j = 1, 2, \ldots m$

$$h_{ij} = \langle Av_j, v_i \rangle, \qquad i = 1, \ldots, j$$

$$w_j = Av_j - \sum_{i=1}^{j} h_{ij} v_i \implies h_{j+1,j}\, v_{j+1} = Av_j - \sum_{i=1}^{j} h_{ij} v_i$$

$$h_{j+1,j} = \|w_j\| \qquad \text{if } h_{j+1,j} = 0 \text{ then stop}$$

$$v_{j+1} = w_j / h_{j+1,j} \implies w_j = h_{j+1,j}\, v_{j+1}$$

End

Rearranging the central formula:

$$Av_j = \sum_{i=1}^{j} h_{ij} v_i + h_{j+1,j}\, v_{j+1}, \qquad j = 1, \ldots m$$

For $j = 1, 2, \ldots m$

$$h_{ij} = \langle A v_j, v_i \rangle, \qquad i = 1, \ldots, j$$

$$w_j = A v_j - \sum_{i=1}^{j} h_{ij} v_i \implies h_{j+1,j}\, v_{j+1} = A v_j - \sum_{i=1}^{j} h_{ij} v_i$$

$$h_{j+1,j} = \|w_j\| \qquad \text{if } h_{j+1,j} = 0 \text{ then stop}$$

$$v_{j+1} = w_j / h_{j+1,j} \implies w_j = h_{j+1,j}\, v_{j+1}$$

End

Rearranging the central formula:

$$A v_j = \sum_{i=1}^{j} h_{ij} v_i + h_{j+1,j}\, v_{j+1}, \qquad j = 1, \ldots m$$

Combining:

$$\boxed{A v_j = \sum_{i=1}^{j+1} h_{ij} v_i, \qquad j = 1, \ldots m}$$

Arnoldi relation 1

$$Av_j = \sum_{i=1}^{j+1} h_{ij} v_i, \qquad j = 1, \ldots m$$

Arnoldi relation 1

$$Av_j = \sum_{i=1}^{j+1} h_{ij} v_i, \qquad j = 1, \ldots m$$

$$A \begin{bmatrix} \\ v_j \\ \\ \end{bmatrix} = \left[\begin{array}{c|c|c|c|c|c} & & & & & \\ v_1 & v_2 & \cdots & v_{j+1} & \cdots & v_{m+1} \\ & & & & & \end{array}\right] \begin{bmatrix} h_{1,j} \\ h_{2,j} \\ \vdots \\ h_{j+1,j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Arnoldi relation 1

$$Av_j = \sum_{i=1}^{j+1} h_{ij} v_i, \qquad j = 1, \ldots m$$

$$A \begin{bmatrix} \\ v_j \\ \\ \end{bmatrix} = \left[\begin{array}{c|c|c|c|c|c} & & & & & \\ v_1 & v_2 & \cdots & v_{j+1} & \cdots & v_{m+1} \\ & & & & & \end{array}\right] \begin{bmatrix} h_{1,j} \\ h_{2,j} \\ \vdots \\ h_{j+1,j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$AV_m(:, j) = V_{m+1} \overline{H}_m(:, j), \qquad j = 1, \ldots, m$$

Arnoldi relation 1

$$Av_j = \sum_{i=1}^{j+1} h_{ij} v_i, \qquad j = 1, \ldots m$$

$$A \begin{bmatrix} \\ v_j \\ \\ \end{bmatrix} = \left[\begin{array}{c|c|c|c|c|c} & & & & & \\ v_1 & v_2 & \cdots & v_{j+1} & \cdots & v_{m+1} \\ & & & & & \end{array}\right] \begin{bmatrix} h_{1,j} \\ h_{2,j} \\ \vdots \\ h_{j+1,j} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$AV_m(:, j) = V_{m+1} \overline{H}_m(:, j), \qquad j = 1, \ldots, m$$

Altogether,

$$\boxed{AV_m = V_{m+1} \overline{H}_m}$$

Arnoldi relation 2

$$AV_m = V_{m+1}\overline{H}_m$$

Arnoldi relation 2

$$AV_m = V_{m+1}\overline{H}_m$$

$$AV_m = \left[\begin{array}{c|c} V_m & v_{m+1} \end{array}\right] \left[\begin{array}{c} H_m \\ \hline 0 \quad 0 \quad \cdots \quad 0 \quad h_{m+1,m} \end{array}\right]$$

Arnoldi relation 2

$$AV_m = V_{m+1}\overline{H}_m$$

$$AV_m = \left[\begin{array}{c|c} V_m & v_{m+1} \end{array}\right] \left[\begin{array}{c} H_m \\ \hline 0 \quad 0 \quad \cdots \quad 0 \quad h_{m+1,m} \end{array}\right]$$

$$AV_m = \left[\begin{array}{c|c} V_m & v_{m+1} \end{array}\right] \left[\begin{array}{c} H_m \\ \hline h_{m+1,m}\, e_m{}^T \end{array}\right]$$

Arnoldi relation 2

$$AV_m = V_{m+1}\overline{H}_m$$

$$AV_m = \left[\begin{array}{c|c} V_m & v_{m+1} \end{array}\right] \left[\begin{array}{c} H_m \\ \hline 0 \quad 0 \quad \cdots \quad 0 \quad h_{m+1,m} \end{array}\right]$$

$$AV_m = \left[\begin{array}{c|c} V_m & v_{m+1} \end{array}\right] \left[\begin{array}{c} H_m \\ \hline h_{m+1,m}\, e_m{}^T \end{array}\right]$$

$$\boxed{AV_m = V_m H_m + h_{m+1,m}\, v_{m+1}\, e_m{}^T}$$

Arnoldi relation 3

$$AV_m = V_m H_m + h_{m+1,m}\, v_{m+1}\, {e_m}^T$$

Arnoldi relation 3

$$AV_m = V_m H_m + h_{m+1,m}\, v_{m+1}\, {e_m}^T$$

$$\begin{aligned} V_m^* A V_m &= V_m^* V_m H_m + h_{m+1,m}\, V_m^*\, v_{m+1}\, {e_m}^T \\ &= I_m H_m + h_{m+1,m}\, 0\, {e_m}^T \\ &= H_m \end{aligned}$$

Arnoldi relation 3

$$AV_m = V_m H_m + h_{m+1,m}\, v_{m+1}\, {e_m}^T$$

$$\begin{aligned} V_m^* A V_m &= V_m^* V_m H_m + h_{m+1,m}\, V_m^*\, v_{m+1}\, {e_m}^T \\ &= I_m H_m + h_{m+1,m}\, 0\, {e_m}^T \\ &= H_m \end{aligned}$$

$$\boxed{V_m^* A V_m = H_m}$$

The Arnoldi relations

$$
\begin{aligned}
AV_m &= V_{m+1}\overline{H}_m \\
AV_m &= V_m H_m + h_{m+1,m}\, v_{m+1}\, {e_m}^T \\
V_m^* A V_m &= H_m
\end{aligned}
$$

- A is $n \times n$, where n is typically large
- V_m is $n \times m$: m columns of n-dimensional vectors
- V_{m+1} is $n \times (m+1)$: it's just V_m but one iteration further along
- $\overline{H}_m$ is $(m+1) \times m$ (rectangular) upper Hessenberg
- H_m is $m \times m$ (square) upper Hessenberg: it's just $\overline{H}_m$ without the bottom row

For eigenvalue problems H_m is the relevant matrix.

Returning to our two questions

We have answered our second question from earlier:

- How do we build a well-conditioned basis for the Krylov subspace?

Answer: the Arnoldi process generates an orthonormal basis for $\mathcal{K}_m$ in the form of the columns of V_m.

Furthermore the Arnoldi relation reveals that $V_m^* A V_m = H_m$ where H_m is upper Hessenberg.

Returning to our two questions

We have answered our second question from earlier:

- How do we build a well-conditioned basis for the Krylov subspace?

Answer: the Arnoldi process generates an orthonormal basis for $\mathcal{K}_m$ in the form of the columns of V_m.

Furthermore the Arnoldi relation reveals that $V_m^* A V_m = H_m$ where H_m is upper Hessenberg.

We now return to the first question:

- What does it mean to "project" the eigenvalue problem onto a Krylov subspace?

Projecting A onto the Krylov subspace

- Consider an arbitrary vector $u \in \mathcal{K}_m$:
  $u = V_m\, y$ for some coordinate vector y.

Projecting A onto the Krylov subspace

- Consider an arbitrary vector $u \in \mathcal{K}_m$:
  $u = V_m\, y$ for some coordinate vector y.
- Now form $Au = AV_m y$
  (most likely this new vector does not belong to $\mathcal{K}_m$).

Projecting A onto the Krylov subspace

- Consider an arbitrary vector $u \in \mathcal{K}_m$:
  $u = V_m\, y$ for some coordinate vector y.
- Now form $Au = AV_m y$
  (most likely this new vector does not belong to $\mathcal{K}_m$).
- Orthogonally project this vector back onto $\mathcal{K}_m$:
  $\text{proj}_{\mathcal{K}_m}(A\,u) = V_m\, V_m^*\, A\, u = V_m\, V_m^*\, A\, V_m\, y$

Projecting A onto the Krylov subspace

- Consider an arbitrary vector $u \in \mathcal{K}_m$:
  $u = V_m\, y$ for some coordinate vector y.
- Now form $Au = AV_m y$
  (most likely this new vector does not belong to $\mathcal{K}_m$).
- Orthogonally project this vector back onto $\mathcal{K}_m$:
  $\text{proj}_{\mathcal{K}_m}(A\,u) = V_m\, V_m^*\, A\, u = V_m\, V_m^*\, A\, V_m\, y$
- This is of the form $V_m A_{\mathcal{K}_m}\, y$ so we identify the projected matrix

$$A_{\mathcal{K}_m} = V_m^*\, A\, V_m$$

Projecting A onto the Krylov subspace

- Consider an arbitrary vector $u \in \mathcal{K}_m$:
  $u = V_m\, y$ for some coordinate vector y.
- Now form $Au = AV_m y$
  (most likely this new vector does not belong to $\mathcal{K}_m$).
- Orthogonally project this vector back onto $\mathcal{K}_m$:
  $\text{proj}_{\mathcal{K}_m}(A\,u) = V_m\, V_m^*\, A\, u = V_m\, V_m^*\, A\, V_m\, y$
- This is of the form $V_m A_{\mathcal{K}_m}\, y$ so we identify the projected matrix

$$A_{\mathcal{K}_m} = V_m^*\, A\, V_m$$

But from the Arnoldi relation

$$V_m^* A V_m = H_m$$

Projecting A onto the Krylov subspace

- Consider an arbitrary vector $u \in \mathcal{K}_m$:
  $u = V_m\, y$ for some coordinate vector y.
- Now form $Au = AV_m y$
  (most likely this new vector does not belong to $\mathcal{K}_m$).
- Orthogonally project this vector back onto $\mathcal{K}_m$:
  $\text{proj}_{\mathcal{K}_m}(A\,u) = V_m\, V_m^*\, A\, u = V_m\, V_m^*\, A\, V_m\, y$
- This is of the form $V_m A_{\mathcal{K}_m}\, y$ so we identify the projected matrix

$$A_{\mathcal{K}_m} = V_m^*\, A\, V_m$$

But from the Arnoldi relation

$$V_m^* A V_m = H_m$$

So the projection of A onto $\mathcal{K}_m$ is simply

$$\boxed{A_{\mathcal{K}_m} = H_m}$$

Arnoldi method

We compute eigenvalues $\theta_i^{(m)}$ and eigenvectors $y_i^{(m)}$ of H_m

$$H_m \, y_i^{(m)} = \theta_i^{(m)} y_i^{(m)}$$

This is a small, m-dimensional eigenproblem, easy to solve.

Arnoldi method

We compute eigenvalues $\theta_i^{(m)}$ and eigenvectors $y_i^{(m)}$ of H_m

$$H_m y_i^{(m)} = \theta_i^{(m)} y_i^{(m)}$$

This is a small, m-dimensional eigenproblem, easy to solve.

Then map back to n-dimensions with

$$u_i^{(m)} = V_m y_i^{(m)}$$

The pair $(\theta_i^{(m)}, u_i^{(m)})$ are the Ritz values and Ritz vectors, respectively.

Ideally they are good approximations of true eigenvalue/eigenvector pairs of A.

Arnoldi residual

We can derive the eigenvalue residual by comparing $Au^{(m)}$ with $\theta^{(m)} u^{(m)}$:

$$A\, u^{(m)} = A\, V_m\, y^{(m)}$$

Arnoldi residual

We can derive the eigenvalue residual by comparing $Au^{(m)}$ with $\theta^{(m)} u^{(m)}$:

$$\begin{aligned} A\,u^{(m)} &= A\,V_m\,y^{(m)} \\ &= (V_m\,H_m + h_{m+1,m}\,v_{m+1}\,{e_m}^T)\,y^{(m)} \end{aligned}$$

Arnoldi residual

We can derive the eigenvalue residual by comparing $Au^{(m)}$ with $\theta^{(m)} u^{(m)}$:

$$\begin{aligned} A\,u^{(m)} &= A\,V_m\,y^{(m)} \\ &= (V_m\,H_m + h_{m+1,m}\,v_{m+1}\,{e_m}^T)\,y^{(m)} \\ &= V_m\,H_m\,y^{(m)} + h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \end{aligned}$$

Arnoldi residual

We can derive the eigenvalue residual by comparing $Au^{(m)}$ with $\theta^{(m)} u^{(m)}$:

$$\begin{aligned} A\,u^{(m)} &= A\,V_m\,y^{(m)} \\ &= (V_m\,H_m + h_{m+1,m}\,v_{m+1}\,{e_m}^T)\,y^{(m)} \\ &= V_m\,H_m\,y^{(m)} + h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \\ &= \theta^{(m)}\,V_m\,y^{(m)} + h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \end{aligned}$$

Arnoldi residual

We can derive the eigenvalue residual by comparing $Au^{(m)}$ with $\theta^{(m)} u^{(m)}$:

$$
\begin{aligned}
A\,u^{(m)} &= A\,V_m\,y^{(m)} \\
&= (V_m\,H_m + h_{m+1,m}\,v_{m+1}\,{e_m}^T)\,y^{(m)} \\
&= V_m\,H_m\,y^{(m)} + h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \\
&= \theta^{(m)}\,V_m\,y^{(m)} + h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \\
&= \theta^{(m)}\,u^{(m)} + h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)}
\end{aligned}
$$

Arnoldi residual

We can derive the eigenvalue residual by comparing $Au^{(m)}$ with $\theta^{(m)} u^{(m)}$:

$$\begin{aligned} A\,u^{(m)} &= A\,V_m\,y^{(m)} \\ &= (V_m\,H_m + h_{m+1,m}\,v_{m+1}\,{e_m}^T)\,y^{(m)} \\ &= V_m\,H_m\,y^{(m)} + h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \\ &= \theta^{(m)}\,V_m\,y^{(m)} + h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \\ &= \theta^{(m)}\,u^{(m)} + h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \end{aligned}$$

Eigenvalue residual

$$A\,u^{(m)} - \theta^{(m)}\,u^{(m)} = h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)}$$

Arnoldi residual

$$
\begin{aligned}
A\,u^{(m)} - \theta^{(m)}\,u^{(m)} &= h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)} \\
\|A\,u^{(m)} - \theta^{(m)}\,u^{(m)}\| &= \|h_{m+1,m}\,v_{m+1}\,{e_m}^T\,y^{(m)}\| \\
&= |h_{m+1,m}|\,\|v_{m+1}\|\,|{e_m}^T\,y^{(m)}| \\
&= |h_{m+1,m}|\,|{e_m}^T\,y^{(m)}| \\
&= |h_{m+1,m}|\,|{y^{(m)}}_m|
\end{aligned}
$$

Note the "lucky breakdown" in the Arnoldi process occurs when $h_{m+1,m} = 0$, so the Ritz values and vectors would in fact be true eigenvalues and eigenvectors of A in that case.