CS 450: Numerical Analysis
Eigenvalue Problems

University of Illinois at Urbana-Champaign

---

These slides have been drafted by Edgar Solomonik as lecture templates and supplementary material for the book “Scientific Computing: An Introductory Survey” by Michael T. Heath (slides).
Eigenvalues and Eigenvectors

- A matrix $A$ has eigenvector-eigenvalue pair (eigenpair) $(\lambda, x)$ if
  \[ Ax = \lambda x \]

- For any scalar $\alpha$, $\alpha x$ is also an eigenvector of $A$ with eigenvalue $\lambda$

- Generally, an eigenvalue $\lambda$ is associated with an eigenspace $\mathcal{X} \subseteq \mathbb{C}^n$ such that each $x \in \mathcal{X}$ is an eigenvector of $A$ with eigenvalue $\lambda$.

- The dimensionality of an eigenspace is at most the multiplicity of an eigenvalue (when less, matrix is defective, otherwise matrix is diagonalizable).

- Each $n \times n$ matrix has up to $n$ eigenvalues, which are either real or complex
  - The conjugate of any complex eigenvalue of a real matrix is also an eigenvalue.
  - The dimensionalities of all the eigenspaces (multiplicity associated with each eigenvalue) sum up to $n$ for a diagonalizable matrix.
  - If the matrix is real, real eigenvalues are associated with real eigenvectors, but complex eigenvalues may not be.
Eigenvalue Decomposition

- If a matrix $A$ is diagonalizable, it has an *eigenvalue decomposition*

  $\quad A = XDX^{-1}$

  where $X$ are the right eigenvectors, $X^{-1}$ are the left eigenvectors and $D$ are eigenvalues

  $\quad AX = [Ax_1 \ldots Ax_n] = XD = [d_{11}x_1 \ldots d_{nn}x_n].$

- If $A$ is Hermitian, its right and left singular vectors are the same by symmetry, hence in this case $X^{-1} = X^H$.

- More generally, any *normal* matrix, $A^H A = AA^H$, has unitary eigenvectors.

- $A$ and $B$ are *similar*, if there exist $Z$ such that $A = ZBZ^{-1}$

  - Normal matrices are *unitarily similar* ($Z^{-1} = Z^H$) to diagonal matrices
  - Symmetric real matrices are *orthogonally similar* ($Z^{-1} = Z^T$) to real diagonal matrices
  - Hermitian matrices are unitarily similar to real diagonal matrices
### Similarity of Matrices

#### Invertible similarity transformations $Y = XAX^{-1}$

<table>
<thead>
<tr>
<th>matrix ($A$)</th>
<th>reduced form ($Y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>arbitrary</td>
<td>bidiagonal</td>
</tr>
<tr>
<td>diagonalizable</td>
<td>diagonal</td>
</tr>
</tbody>
</table>

#### Unitary similarity transformations $Y = UAU^H$

<table>
<thead>
<tr>
<th>matrix ($A$)</th>
<th>reduced form ($Y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>arbitrary</td>
<td>triangular (Schur)</td>
</tr>
<tr>
<td>normal</td>
<td>diagonal</td>
</tr>
<tr>
<td>Hermitian</td>
<td>real diagonal</td>
</tr>
</tbody>
</table>

#### Orthogonal similarity transformations $Y = QAQ^T$

<table>
<thead>
<tr>
<th>matrix ($A$)</th>
<th>reduced form ($Y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>real</td>
<td>Hessenberg</td>
</tr>
<tr>
<td>real symmetric</td>
<td>real diagonal</td>
</tr>
<tr>
<td>SPD</td>
<td>real positive diagonal</td>
</tr>
</tbody>
</table>
Canonical Forms

- Any matrix is *similar* to a bidiagonal matrix, giving its *Jordan form*:

  \[ A = X \begin{bmatrix} J_1 & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \vdots & J_k \end{bmatrix} X^{-1}, \quad \forall i, \quad J_i = \begin{bmatrix} \lambda_i & 1 \\ \vdots & \ddots \\ & \ddots & \ddots \\ & & 1 \\ & & & \lambda_i \end{bmatrix} \]

  the Jordan form is unique modulo ordering of the diagonal Jordan blocks.

- Any diagonalizable matrix is *unitarily similar* to a triangular matrix, giving its *Schur form*:  
  \[ A = QTQ^H \]  
  where \( T \) is upper-triangular, so the eigenvalues of \( A \) is the diagonal of \( T \). Columns of \( Q \) are the *Schur vectors*.

- Real matrices are *orthogonally similar* to a block-triangular real matrix with 1 × 1 or 2 × 2 blocks (real Schur form)
Eigenvectors from Schur Form

Given the eigenvectors of one matrix, we seek those of a similar matrix: Suppose that $A = SBS^{-1}$ and $B = XDX^{-1}$ where $D$ is diagonal,

1. The eigenvalues of $A$ are $\{d_{11}, \ldots, d_{nn}\}$
2. $A = SBS^{-1} = SXDX^{-1}S^{-1}$ so $SX$ are the eigenvectors of $A$

Its easy to obtain eigenvectors of triangular matrix $T$:

1. One eigenvector is simply the first elementary vector.
2. The eigenvector associated with any diagonal entry (eigenvalue $\lambda$) may be obtaining by observing that

$$0 = (T - \lambda I)x = \begin{bmatrix} U_{11} & u & T_{13} \\ 0 & v^T & 1 \\ U_{33} & 0 & 0 \end{bmatrix} \begin{bmatrix} -U_{11}^{-1}u \\ 1 \\ 0 \end{bmatrix},$$

so it suffices to solve $U_{11}y = -u$ to obtain eigenvector $x$. 
For any vector $x$, the \textit{Rayleigh quotient} provides an estimate for some eigenvalue of $A$:

$$\rho_A(x) = \frac{x^H Ax}{x^H x}.$$ 

If $x$ is an eigenvector of $A$, then $\rho_A(x)$ is the associated eigenvalue.

Moreover, for $y = Ax$, the Rayleigh quotient is the best possible eigenvalue estimate given $x$ and $y$, as it is the solution to $x\alpha \approx y$.

The normal equations for this scalar-output least squares problem are

$$x^T x\alpha = x^T y \quad \Rightarrow \quad \alpha = \frac{x^T y}{x^T x} = \frac{x^T Ax}{x^T x}.$$
Perturbation Analysis of Eigenvalue Problems

- For non-defective \( A = XDX^{-1} \), the eigenvalues of \( A + \delta A = \hat{X}(D + \delta D)X^{-1} \) satisfy \( \|\delta D\| \leq \kappa(X)\|\delta A\| \):

  Note that the eigenvalues of \( X^{-1}(A + \delta A)X = D + X^{-1}\delta AX \) are also \( D + \delta D \). So if we have perturbation to the matrix \( \|\delta A\|_F \), its effect on the eigenvalues corresponds to a (non-diagonal/arbitrary) perturbation \( \delta \hat{A} = X^{-1}\delta AX \) of a diagonal matrix of eigenvalues \( D \), with norm

  \[
  \|\delta \hat{A}\|_F \leq \|X^{-1}\|_2\|\delta A\|_F\|X\|_2 = \kappa(X)\|\delta A\|_F.
  \]

- Gershgorin’s theorem allows us to bound the effect of the perturbation on the eigenvalues of a (diagonal) matrix:

  Given a matrix \( A \in \mathbb{R}^{n \times n} \), let \( r_i = \sum_{j \neq i} |a_{ij}| \), define the Gershgorin disks as

  \[
  D_i = \{ z \in \mathbb{C} : |z - a_{ii}| \leq r_i \}.
  \]

  The eigenvalues \( \lambda_1, \ldots, \lambda_n \) of any matrix \( A \in \mathbb{R}^{n \times n} \) are contained in the union of the Gershgorin disks, \( \forall i \in \{1, \ldots, n\} \), \( \lambda_i \in \bigcup_{j=1}^{n} D_j \).
Gershgorin Theorem Perturbation Visualization

- Top corresponds to Gershgorin disks on complex plane of 4-by-4 real matrix.
- Bottom part corresponds to bounds on Gershgorin disks of $X^{-1}(A + \delta A)X$, which contain the eigenvalues $D$ of $A$ and the perturbed eigenvalues $D + \delta D$ of $A + \delta A$ provided that $||\delta A||$ is sufficiently small.
Conditioning of Particular Eigenpairs

Consider the effect of a matrix perturbation on an eigenvalue $\lambda$ associated with a right eigenvector $x$ and a left eigenvector $y$, $\lambda = y^H Ax / y^H x$

For a sufficiently small perturbation $\delta A$, the eigenvalue $\lambda$ is perturbed to an eigenvalue $\hat{\lambda}$ of $\hat{A} = A + \delta A$. The eigenvalue perturbation, ignoring error due to the change in eigenvectors, is

$$|\hat{\lambda} - \lambda| \approx |y^H \delta Ax / y^H x| \leq \frac{||\delta A||}{|y^H x|}.$$

This is small if $x$ is near-parallel to $y$ and large if they are near-perpendicular.

A more accurate eigenvalue approximation than Rayleigh quotient for a normalized perturbed eigenvector (e.g., iterative guess) $\hat{x} = x + \delta x$, can be obtained with an estimate of both eigenvectors (also $\hat{y} = y + \delta y$),

$$|\hat{\lambda}_{xAx} - \lambda| \approx |\delta x^H Ax + x^H A \delta x| \leq |\lambda||\delta x| + \left( |\lambda||y^H x| + |1 - y^H x \cdot ||A|| \right)||\delta x||$$

$$|\hat{\lambda}_{yAx} - \lambda| \approx \frac{|\delta y^H Ax + y^H A \delta x|}{y^H x} \leq |\lambda||\delta x|| \frac{1}{|y^H x|} + ||\delta y||.$$
Google's PageRank

A well-known application of eigenproblems is the problem of ranking $n$ web-pages

- Based on web-data, we compute transition probability from webpage $i$ to webpage $j$ at $a_{ij}$, so $\sum_j a_{ij} = 1$

- We seek a measure of webpage popularity, which we can take to be the probability $x_i$ of a web-surfer being on webpage $i$ as opposed to the other $n - 1$ webpages

- This vector of probabilities, $x$, is given by the stationary probability vector, which satisfies

  \[ A^T x = x \]

- For such a transition probability (stochastic) matrix $A$, all eigenvalues are at most 1 in absolute value
Power Iteration

- **Power iteration** can be used to compute the largest eigenvalue of a real symmetric matrix $A$:

$$x^{(i)} = Ax^{(i-1)} \quad \text{(typically with normalization of } x^{(i)} \text{ at each step).}$$

For a random $x^{(0)}$, power iteration converges eigenvalue of $A$ with largest modulus, $\lim_{i \to \infty} \rho_A(x^{(i)}) = \lambda_{\max}(A)$. If this eigenvalue has multiplicity one, power iteration converges to the dominant eigenvector.

- The error of power iteration decreases at each step by the ratio of the largest eigenvalues:

Assuming $A$ is diagonalizable with eigenvectors $U$ and $V^H = U^{-1}$,

$$x^{(k)} = A^k x^{(0)} = (UDV^H)^k x^{(0)} = UD^k V^H x^{(0)} = \sum_{i=1}^{n} u_i \lambda_i^k v_i^H x^{(0)}.$$

The coefficient $\alpha^{(i,k)}$ associated with the maximum $\lambda_i$ and dominant eigenvector $u_i$ grows relatively, since $|\alpha^{(i,k)}/\alpha^{(j,k)}| = (|\lambda_i|/|\lambda_j|)^k \left[\alpha^{(i,0)}/\alpha^{(j,0)}\right]$. 

**Demo:** Power iteration and its Variants
If the error at the $k$th step with respect to the desired solution is $e_k$, $r$th order convergence implies that $\lim_{k \to \infty} \|e_k\|/\|e_{k-1}\|^r \leq C$.

- Power iteration achieves linear convergence with $C = |\lambda_2|/|\lambda_1|$ assuming decreasing order, $|\lambda_i| \geq |\lambda_{i+1}|$ and that $|\lambda_1| > |\lambda_2|$.
- Convergence of order $r > 1$ (superlinear) implies that the number of digits of correctness increases by a factor of $r$ at each step.
- For $r > 1$, error $e_k \leq \epsilon$ is achieved after $O(\log_r(\log(1/\epsilon)))$ steps.
- Having achieved superlinear convergence ($r > 1$), methods differ only by constant factors in complexity.
Inverse and Rayleigh Quotient Iteration

- **Inverse iteration** uses LU/QR/SVD of $A$ to run power iteration on $A^{-1}$
  - For a randomly chosen $x^{(0)}$, solving
    
    $$Ax^{(i)} = x^{(i-1)} \text{ (typically with normalization of } x^{(i)} \text{ at each step).}$$

    converges to $\lim_{i \to \infty} \rho_A(x^{(i)}) = \lambda_{\text{min}}(A)$ provided there is a unique eigenvalue with minimum magnitude.

- Inverse iteration on $A - \sigma I$ converges to the eigenvalue closest to $\sigma$, since all eigenvalues are shifted by $\sigma$.

- **Rayleigh quotient iteration** provides rapid convergence to an eigenpair

    $$\left(A - \rho_A(x^{(i-1)})I\right)x^{(i)} = x^{(i-1)},$$

    since at each step the relative magnitude largest eigenvalue of $(A - \rho_A(x^{(i-1)})I)^{-1}$ grows. Formally, it achieves cubic convergence, but requires matrix refactorization at each step.
Deflation

- Power, inverse, and Rayleigh-quotient iteration compute a single eigenpair, to obtain further eigenpairs, can perform deflation.

- Given eigenvalue $\lambda_1$ and right eigenvector $x_1$, seek $v$ so that $B = A - \lambda_1 uv^T$ has eigenvalues $0, \lambda_2, \ldots, \lambda_n$, where

  $$A = XD Y^T = \sum_{i=1}^{n} \lambda_i x_i y_i^T.$$

- Ideal choice would be $v = y_1^T$, i.e., the left eigenvector associated with $\lambda_1$, as then the $n - 1$ other eigenvectors of $B$ would be the same as those of $A$.

- For symmetric matrices $y_1 = x_1$, but for nonsymmetric, obtaining $y_1$ may require more work.

- Good alternative choice for nonsymmetric matrices is to deflate with Schur vectors, which preserves the Schur decomposition, e.g.,

  $$B = QTQ^T - \lambda_1 q_1 q_1^T = Q(T - \lambda_1 Q^T q_1 q_1^T Q)Q^T = Q(T - \lambda_1 e_1 e_1^T)Q^T.$$
We can always compute an orthogonal similarity transformation to reduce a general matrix to **upper-Hessenberg** (upper-triangular plus the first subdiagonal) matrix $H$, i.e. $A = QHQ^T$:

*We can perform successive two-sided application of Householder reflectors*

$$
A = \begin{bmatrix}
    h_{11} & a_{12} & \cdots \\
    a_{21} & a_{22} & \\
    \vdots & \ddots & \\
\end{bmatrix} = Q_1 \begin{bmatrix}
    h_{11} & a_{12} & \cdots \\
    h_{21} & t_{22} & \\
    0 & \ddots & \\
\end{bmatrix} = Q_1 \begin{bmatrix}
    h_{11} & h_{12} & \cdots \\
    h_{21} & h_{22} & \\
    0 & \ddots & \\
\end{bmatrix} Q_1^T = \cdots
$$

Subsequent columns can be reduced by induction, so we can always stably reduce to upper-Hessenberg with roughly double the cost of QR.

*In the symmetric case, Hessenberg form implies tridiagonal:*

*If $A = A^T$ then $H = QAQ^T = H^T$, and a symmetric upper-Hessenberg matrix must be tridiagonal.*
Simultaneous and Orthogonal Iteration

- **Simultaneous iteration** provides the main idea for computing many eigenvectors at once:
  - Initialize $X_0 \in \mathbb{R}^{n \times k}$ to be random and perform
    $$X_{i+1} = AX_i.$$  
  - Observe that $\lim_{i \to \infty} \text{span}(X_i) = S$ where $S$ is the subspace spanned by the $k$ eigenvectors of $A$ with the largest eigenvalues in magnitude.

- Orthogonal iteration performs QR at each step to ensure stability
  $$Q_{i+1}R_{i+1} = AQ_i$$
  - $Q_i$ has the same span as $X_i$ in orthogonal iteration.
  - QR has cost $O(nk^2)$ while product has cost $O(n^2k)$ per iteration.
  - Can use this to compute the right singular vectors of matrix $M$ by using $A = M^T M$ (no need to form $A$, just multiply $Q_i$ by $M^T$ then $M$).
  - Small number of iterations suffice to obtain reasonable low-rank approximation of $M$, and ultimately $Q$ converges to singular vectors in its truncated SVD.
If $A$ has distinct eigenvalues and $R_i$ has positive decreasing diagonal, the $j$th column of $Q_i$ converges to the $j$th Schur vector of $A$ linearly with rate $\max(|\lambda_{j+1}/\lambda_j|, |\lambda_j/\lambda_{j-1}|)$.

- Convergence of the first column of $Q_i$ follows by analogy to power iteration
- Span of first $j$ columns of $Q_i$ converges to the span of the first $j$ Schur vectors with rate $|\lambda_{j+1}/\lambda_j|$
- Hence orthogonal iteration converges similarly to $k$ instances of inverse iteration with shifts chosen near the $k$ largest magnitude eigenvalues
QR Iteration

- QR iteration reformulates orthogonal iteration for \( n = k \) to reduce cost/step,
  - Orthogonal iteration computes \( \hat{Q}_{i+1} \hat{R}_{i+1} = A \hat{Q}_i \)
  - QR iteration computes \( A_{i+1} = R_i Q_i \) where \( A_i = Q_i R_i \) at iteration \( i \)
  - Hence \( A_{i+1} \) is similar to \( A \), as \( A_{i+1} = Q_i^T A_i Q_i = \left( \prod_{j=0}^{i-1} Q_j \right)^T A \left( \prod_{j=0}^{i-1} Q_j \right) \)

- If orthogonal iteration starts with \( \hat{Q}_1 = Q_0 \), then \( \hat{Q}_i = \prod_{j=0}^{i-1} Q_j \),
  - By induction, \( \hat{Q}_i^T A \hat{Q}_i = A_i = Q_i R_i \)
  - Hence, the QR factorization of \( A \hat{Q}_i = \hat{Q}_i Q_i R_i = \hat{Q}_{i+1} \hat{R}_{i+1} \)

- QR iteration converges to triangular \( A_i \) if the eigenvalues are distinct in modulus, and in general converges to block-triangular form with a block for each set of eigenvalues of equal modulus.
QR Iteration with Shift

QR iteration can be accelerated using shifting:

\[ Q_i R_i = A_i - \sigma_i I, \quad A_{i+1} = R_i Q_i + \sigma_i I \]

*note that \( A_{i+1} \) is similar to \( A_i \), since we can reorganize the above as*

\[ R_i Q_i = Q_i^T (A_i - \sigma_i I) Q_i, \]

\[ Q_i (A_{i+1} - \sigma_i I) Q_i^T = Q_i R_i, \]

*and observe that \( R_i Q_i \) is similar to \( Q_i R_i \).*

*The shift is selected to accelerate convergence to an eigenvalue (pair):*

*We can select the shift as the bottom right element of \( A_i \) or last diagonal entry adjacent to nonzero subdiagonal entry (Wilkinson shift)*

*Wilkinson shift accelerates convergence (follows from analogy of orthogonal iteration to inverse iteration)*

*Complex eigenvalues require more sophisticated shifts, ”implicit double shift” converges to real Schur form while avoiding complex arithmetic*
QR Iteration Complexity

- QR iteration is accelerated by first reducing to upper-Hessenberg or tridiagonal form:

  Reduction to upper-Hessenberg or tridiagonal in the symmetric case, costs $O(n^3)$ operations and can be done in a similar style to Householder QR.

Given an upper-Hessenberg matrix, $H_i = A_i$

- reduction to upper-triangular requires $n - 1$ Givens rotations, if $G_i$ rotates the $(i + 1)$th row into the $i$th to eliminate the $i$th element on the first subdiagonal, $R_i = G_1^T \cdots G_{n-1}^T H_i$

- computation of $H_{i+1} = RQ$ can be done by application of the $n - 1$ Givens rotations to $R$ from the right $H_{i+1} = R_i G_{n-1} \cdots G_1$.

  Both cost $O(n^2)$, for $O(n^3)$ overall if QR iteration converges in $O(n)$ steps.

Given a tridiagonal matrix, the same two general steps are required, but now each step costs $O(n)$, so overall the eigenvalues and eigenvectors of a tridiagonal matrix can be computed with $O(n^2)$ work.
A variety of methods exists for the tridiagonal eigenproblem:

- **QR iteration** requires $O(1)$ QR factorizations per eigenvalue, $O(n^2)$ cost to get eigenvalues, $O(n^3)$ for eigenvectors. The last cost is not optimal.

- **Divide and conquer**
  - partition tridiagonal matrix as
    \[ T = \begin{bmatrix} \hat{T}_1 & \hat{T}_2 \\ \end{bmatrix} + t_{n/2+1,n/2} \begin{bmatrix} e_{n/2} \\ e_1 \end{bmatrix} \begin{bmatrix} e_{n/2}^T \\ e_1^T \end{bmatrix} \]
  - solve two independent eigenvalue problems recursively, recover eigenvalues of $T$ via solving "secular equation"

- **Major alternatives to divide and conquer** include spectral bisection and the MRRR algorithm.
Krylov subspace methods work with information contained in the $n \times k$ matrix

$$K_k = \begin{bmatrix} x_0 & Ax_0 & \cdots & A^{k-1}x_0 \end{bmatrix}$$

We seek to best use the information from the matrix vector product results (columns of $K_k$) to solve eigenvalue problems.

Assuming $K_n$ is invertible, the matrix $K_n^{-1}AK_n$ is a companion matrix $C$:

Letting $k_n^{(i)} = A^{i-1}x$, we observe that

$$AK_n = \begin{bmatrix} Ak_n^{(1)} & \cdots & Ak_n^{(n-1)} & Ak_n^{(n)} \end{bmatrix} = \begin{bmatrix} k_n^{(2)} & \cdots & k_n^{(n)} & Ak_n^{(n)} \end{bmatrix},$$

therefore premultiplying by $K_n^{-1}$ transforms the first $n - 1$ columns of $AK_n$ into the last $n - 1$ columns of $I$,

$$K_n^{-1}AK_n = \begin{bmatrix} K_n^{-1}k_n^{(2)} & \cdots & K_n^{-1}k_n^{(n)} & K_n^{-1}Ak_n^{(n)} \end{bmatrix} = \begin{bmatrix} e_2 & \cdots & e_n & K_n^{-1}Ak_n^{(n)} \end{bmatrix}$$
Krylov Subspaces

Given $Q_k R_k = K_k$, we obtain an orthonormal basis for the Krylov subspace,

$$K_k(A, x_0) = \text{span}(Q_k) = \{p(A)x_0 : \text{deg}(p) < k\},$$

where $p$ is any polynomial of degree less than $k$.

The Krylov subspace includes the $k - 1$ approximate dominant eigenvectors generated by $k - 1$ steps of power iteration:

- The approximation obtained from $k - 1$ steps of power iteration starting from $x_0$ is given by the Rayleigh-quotient of $y = A^k x_0$.
- This vector is within the Krylov subspace, $y \in K_k(A, x_0)$.
- Consequently, Krylov subspace methods will generally obtain strictly better approximations of the dominant eigenpair than power iteration.
The $k \times k$ matrix $H_k = Q_k^T A Q_k$ minimizes $\|A Q_k - Q_k H_k\|_2$:

Let $M = A Q_k - Q_k H_k$, then

$$\|M\|_2 \geq \max(\| (I - Q_k Q_k^T) M \|_2, \| Q_k Q_k^T M \|_2)$$

Since $(I - Q_k Q_k^T) M = (I - Q_k Q_k^T) A Q_k$, the first term is independent of $H_k$. The second term is minimized (zero) with the choice $H_k = Q_k^T A Q_k$.

$H_k$ is Hessenberg, because the companion matrix $C_k$ is Hessenberg:

$$H_k = Q_k^T A Q_k = R_k K_k^{-1} A K_k R_k^{-1} = R_k C_k R_k^{-1}$$

is a product of three matrices: upper-triangular $R_k$, upper-Hessenberg $C_k$, and upper-triangular $R_k^{-1}$, which results in upper-Hessenberg $H_k$. 
Rayleigh-Ritz Procedure

The eigenvalues/eigenvectors of $H_k$ are the Ritz values/vectors:

$$H_k = XDX^{-1}$$

eigenvalue approximations based on Ritz vectors $X$ are given by $Q_kX$.

The Ritz vectors and values are the ideal approximations of the actual eigenvalues and eigenvectors based on only $H_k$ and $Q_k$:

Assuming $A$ is a symmetric matrix with positive eigenvalues, the largest Ritz value $\lambda_{\text{max}}(H_k)$ will be the maximum Rayleigh quotient of any vector in $K_k = \text{span}(Q_k)$,

$$\max_{x \in \text{span}(Q_k)} \frac{x^T A x}{x^T x} = \max_{y \neq 0} \frac{y^T Q_k^T A Q_k y}{y^T y} = \max_{y \neq 0} \frac{y^T H_k y}{y^T y} = \lambda_{\text{max}}(H_k),$$

which is the best approximation to $\lambda_{\text{max}}(A) = \max_{x \neq 0} \frac{x^T A x}{x^T x}$ available in $K_k$. The quality of the approximation can also be shown to be optimal for other eigenvalues/eigenvectors.

Demo: Arnoldi vs Power Iteration
Arnoldi iteration computes the $i$th column of $H_n$, $h_i$ and the $i$th column of $Q_n$ directly using the recurrence $Aq_i = Q_nh_i = \sum_{j=1}^{i+1} h_{ji}q_j$

Note that

$$q_i^TAq_j = q_i^T(Q_nH_nQ_n^T)q_j = e_i^TH_ne_j = h_{ij}.$$

The Arnoldi algorithm computes $q_{i+1}$ from $q_1, \ldots, q_i$ by first computing $u_i = Aq_i$ then orthogonalizing,

$$q_{i+1}h_{i+1,i} = u_i - \sum_{j=1}^{i} q_jh_{ji}, \quad h_{ji} = q_j^Tu_i$$

then computing the norm of the vector to obtain $h_{i+1,i}$, yielding the $i$th column of $H_n$. 

**Demo:** Arnoldi Iteration

**Demo:** Arnoldi Iteration with Complex Eigenvalues
Lanczos Iteration

- Lanczos iteration provides a method to reduce a symmetric matrix to a tridiagonal matrix:

  *Arnoldi iteration on a symmetric matrix will result in an upper-Hessenberg matrix $H_n$ as before, except that it must also be symmetric, since*

  $$H_n^T = (Q_n^T A Q_n)^T = Q_n^T A^T Q_n = Q_n^T A Q_n = H_n,$$

  *which implies that $H_n$ must be tridiagonal.*

- After each matrix-vector product, it suffices to orthogonalize with respect to two previous vectors:

  *Since $h_{ji} = 0$ if $j - i > 1$, given $u_i = A q_i$, it suffices to compute only $h_{ii} = q_i^T u_i$ and $h_{i+1,i} = h_{i,i+1} = \| u_i - q_i h_{ii} \|_2$.*
The cost of matrix-vector multiplication when the matrix has \( m \) nonzeros is \( m \) multiplications and at most \( m \) additions, so roughly \( 2m \) in total.

The cost of orthogonalization at the \( k \)th iteration of a Krylov subspace method is

- \( O(nk) \) for \( k \) inner products in Arnoldi,
- \( O(n) \) in Lanczos, since only 2 orthogonalizations needed.
- For Arnoldi with \( k \)-dimensional subspace, in total, orthogonalization costs \( O(nk^2) \), matrix-vector products cost \( O(mk) \), so generally desire \( nk < m \).
Restarting Krylov Subspace Methods

- In finite precision, Lanczos generally loses orthogonality, while orthogonalization in Arnoldi can become prohibitively expensive:
  - Arnoldi cost of orthogonalization dominates if $k > m/n$.
  - In Lanczos, reorthogonalizing iterate to previous guesses can ensure orthogonality in the presence of round-off error.
  - Selective orthogonalization strategies control when and with respect to what previous columns of $Q$, each new iterate $u_j = Aq_j$ should be orthogonalized.

- Consequently, in practice, low-dimensional Krylov subspace methods are constructed repeatedly using carefully selected new starting vectors:
  \[\text{If we wish to find a particular eigenvector isolate some eigenspaces, restarting is beneficial}\]
  - can orthogonalize to previous eigenvector estimates to perform deflation,
  - can pick starting vector as Ritz vector estimate associated with desired eigenpair,
  - given new starting vector, can discard previous Krylov subspace, which helps make storing the needed parts of $Q$ possible.
A generalized eigenvalue problem has the form $Ax = \lambda Bx$, 

$$AX = BXD$$

$$B^{-1}A = XDX^{-1}$$

Generalized eigenvalue problems arise frequently, especially in solving partial differential equations.

When $A$ and $B$ are symmetric and $B$ is SPD, we can perform Cholesky on $B$, multiply $A$ by the inverted factors, and diagonalize it:

$$AX = LL^TXD$$

$$L^{-1}AL^{-T}L^TX = \hat{L}^T\hat{X}D$$

Specialized canonical forms and methods exist for the generalized eigenproblem with fewer constraints on $B$ and better cost/stability.