

CS 450: Numerical Analysis – Lecture 12

Eigenvalue Problems: Definitions, Applications, and Conditioning

Edgar Solomonik

October 1, 2026

Today

- Eigenvalues, eigenvectors, eigenvalue decomposition
- Applications
- Conditioning

Eigenvalues and eigenvectors

Given $A \in \mathbb{C}^{n \times n}$, an *eigenpair* is $\lambda \in \mathbb{C}$, $x \in \mathbb{C}^n$ with

$$Ax = \lambda x \quad (\text{a left eigenpair satisfies } y^H A = \lambda y^H).$$

Note:

- $\lambda = 0$ can be an eigenvalue; x is then in the null space of A .
- If (λ, x) is an eigenpair, so is $(\lambda, \alpha x)$ for any $\alpha \neq 0$; if (λ, z) is also an eigenpair, so is $(\lambda, \alpha x + \beta z)$ for any α, β with $\alpha x + \beta z \neq 0$.
- In general, each eigenvalue λ is associated with an *eigenspace* $V = \{x : Ax = \lambda x\}$; the *geometric multiplicity* of λ is $\dim(V)$.
- Eigenvalues are the roots of the *characteristic polynomial*,

$$p(\lambda) = \det(\lambda I - A) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n); \quad (12.1)$$

the degree of $(\lambda - \lambda_i)$ in $p(\lambda)$ is the *algebraic multiplicity* of λ_i .

- $A \in \mathbb{R}^{n \times n}$ may have complex eigenvalues ($p(\lambda)$'s roots are complex), in which case they are paired: $\lambda_1 = \alpha + \beta i$, $\lambda_2 = \alpha - \beta i$.

Eigenvalue decomposition

- $A \in \mathbb{C}^{n \times n}$ is *diagonalizable* if there exists an invertible $X \in \mathbb{C}^{n \times n}$ such that

$$D = X^{-1}AX = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \text{ is diagonal.}$$

- (d_{ii}, x_i) are eigenpairs, with x_i the columns of X .
- A is diagonalizable iff the algebraic multiplicity of λ equals the geometric multiplicity of λ , for each eigenvalue λ .
- Any $A \in \mathbb{C}^{n \times n}$ has a *Jordan form*

$$A = XJX^{-1}, \quad J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{bmatrix}, \quad J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}.$$

Applications of eigenvalue problems

(1) Computing $\|A\|_2$.

$$\max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}, \quad \min_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}, \quad \min_{x \neq 0} \frac{x^T Ax}{x^T x}, \quad \max_{x \neq 0} \frac{x^T Ax}{x^T x}.$$

The last two are quadratic programs when A is SPD. The first two are obtained via $A = USV^T$:

$$A^T A = VS^2V^T, \quad AA^T = US^2U^T$$

are both eigenvalue decompositions. If A is symmetric:

$$\|A\|_2 = \max_{\lambda} |\lambda|, \quad \|A^{-1}\|_2 = \frac{1}{\min_{\lambda} |\lambda|}.$$

(2) Time evolution and steady-state behavior.

$$y_{k+1} = Ay_k \quad \implies \quad y_k = A^k y_0.$$

If $A = XDX^{-1}$, then

$$A^2 = XDX^{-1}XDX^{-1} = XD^2X^{-1}, \quad A^k = XD^kX^{-1}.$$

Expanding $y_0 = \sum_{i=1}^n c_i x_i$ in the eigenbasis (the columns of X),

$$A^k y_0 = \lambda_1^k c_1 x_1 + \cdots + \lambda_n^k c_n x_n.$$

For typical y_0 (i.e. $c_1 \neq 0$), if λ_1 is the unique eigenvalue of largest modulus, with eigenvector x_1 ,

$$\lim_{k \rightarrow \infty} \frac{A^k y_0}{\|A^k y_0\|_2} = x_1 \quad (\text{up to a sign/phase}). \quad (12.2)$$

The recurrence $y_{k+1} = Ay_k$ describes

- linearized/discretized ODEs $\dot{y} = f(y)$;
- dynamical systems.

Graph and network analysis. Given $G = (V, E)$ with edge weights $w : E \rightarrow \mathbb{R}$ and $|V| = n$, define the *adjacency matrix* $A \in \mathbb{R}^{n \times n}$,

$$a_{ij} = \begin{cases} w(e) & \text{if } e = (v_i, v_j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Stochastic matrices. $A \in \mathbb{R}^{n \times n}$ is a *stochastic matrix* if

$$A\mathbf{1} = \mathbf{1} \quad (\text{rows sum to 1}), \quad A \geq 0 \quad (\text{nonnegative}).$$

The entry a_{ij} describes a Markov-chain transition probability,

$$a_{ij} = p(i \rightarrow j), \quad w(e) = p(i \rightarrow j) \text{ for } e = (v_i, v_j).$$

If G is connected, $\lambda_{\max}(A) = 1$, and the left eigenvector for $\lambda_{\max}$ is x such that

$$x^T A = x^T, \quad x \geq 0 \quad (\text{Perron–Frobenius theorem}).$$

Normalized so that $\sum_i x_i = 1$, $\{x_1, \dots, x_n\}$ is the probability distribution of the Markov chain as $t \rightarrow \infty$.

Google's PageRank. Given websites $V = \{v_1, \dots, v_n\}$ with links E , assign the weights $w(e)$ uniformly for each website (row of A), or based on observed transition rates. Rank according to the left eigenvector for the largest eigenvalue of A .

Nonlinear eigenvalue problems. In quantum physics, the time-independent Schrödinger equation asks for an electron wavefunction $\psi : \mathbb{R}^3 \rightarrow \mathbb{C}$ satisfying

$$H\psi = E\psi,$$

where $H\psi : \mathbb{R}^3 \rightarrow \mathbb{C}$ again, H is a functional (differential) operator that encodes the potential, and $E \in \mathbb{R}$. As stated this is a linear operator eigenproblem; it becomes a *nonlinear* eigenvalue problem when H depends on ψ , as in mean-field models (Hartree–Fock, density functional theory).

Similarity transformations

$A, B \in \mathbb{C}^{n \times n}$ are *similar* if there exists an invertible $X \in \mathbb{C}^{n \times n}$ with

$$XAX^{-1} = B \quad \Longleftrightarrow \quad X^{-1}BX = A.$$

If (λ, y) is an eigenpair of A , then (λ, Xy) is an eigenpair of B :

$$BXy = XAX^{-1}Xy = XAy = \lambda Xy.$$

The eigenvalue decomposition and the Jordan form are similarity transformations from A to D and to J . Algorithms for the eigenvalue decomposition compute a series of similarity transformations,

$$A \rightarrow X_1AX_1^{-1} \rightarrow X_2X_1AX_1^{-1}X_2^{-1} \rightarrow \cdots \rightarrow D.$$

Hermitian matrices

Given $A \in \mathbb{C}^{n \times n}$, suppose

$$A^H = A^{*T} = \begin{bmatrix} a_{11}^* & \cdots & a_{n1}^* \\ \vdots & & \vdots \\ a_{1n}^* & \cdots & a_{nn}^* \end{bmatrix} = A.$$

Then:

- A is diagonalizable;
- $A = QDQ^H$, with Q unitary (the eigenvectors are orthonormal) and D real (the eigenvalues are real).

If $A \in \mathbb{R}^{n \times n}$ with $A^T = A$:

- A is Hermitian;
- $A = QDQ^T$, with Q orthogonal and D real.

Normal matrices

If $A = QDQ^H$ with Q unitary and D diagonal (possibly complex), then A is *normal*, i.e.

$$AA^H = A^HA,$$

and vice versa. All Hermitian matrices are normal.

matrix type	similarity transformation	eigenvalue type / target
SPD	orthogonal	positive
symmetric positive semidefinite	orthogonal	nonnegative
symmetric real	orthogonal	real
Hermitian	unitary	real
normal	unitary	complex
diagonalizable	invertible	complex
arbitrary	invertible	Jordan form

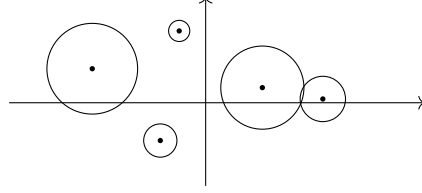
The Jordan form and the characteristic polynomial are sensitive to perturbations; we will exploit alternative routes in algorithms.

Conditioning of the eigenvalue decomposition

Suppose A is diagonalizable, $A = XDX^{-1}$. If $A \rightarrow A + \delta A$, how much does D (the eigenvalues) change? We have

$$X^{-1}(A + \delta A)X = D + \underbrace{X^{-1}\delta AX}_{\text{small}}.$$

Gershgorin disk theorem. Every eigenvalue of A is contained in the union of the disks centered at a_{ii} with radius $\sum_{j \neq i} |a_{ij}|$:



Hence the eigenvalues of $D + X^{-1}\delta AX$ are perturbed (relative to those of D) by at most

$$O(\|X^{-1}\delta AX\|) = O(\|X\|_2 \|X^{-1}\|_2 \|\delta A\|_2) = O(\kappa(X) \|\delta A\|_2). \quad (12.3)$$

The (absolute) conditioning of the eigenvalue problem is $\kappa(X)$ (the Bauer–Fike theorem). If A is normal, $\kappa(X) = 1$.

CS 450: Numerical Analysis – Lecture 13

Power Iteration and the Rayleigh Quotient

Edgar Solomonik

October 6, 2026

Today

- Power iteration and variants
- Convergence rates
- Rayleigh quotient

Power iteration

Last time we saw that $\lambda(A)$ describes the behavior of systems where A is applied repeatedly—Markov chains via stochastic matrices, dynamical systems and ODEs. The same principle is useful algorithmically:

$$y = Ax^{(k)}, \quad x^{(k+1)} = y / \|y\|_2, \quad x^{(0)} \text{ random.}$$

Normalization keeps $\|x^{(k)}\|_2$ and $\|y\|_2$ bounded and preserves direction:

$$x^{(k)} = \frac{A^k x^{(0)}}{\|A^k x^{(0)}\|_2}.$$

Suppose $A = QDQ^T$ with $Q^T Q = I$. Then there exists $z \in \mathbb{R}^n$ such that

$$x^{(0)} = z_1 q_1 + \cdots + z_n q_n, \quad A^k x^{(0)} = \lambda_1^k z_1 q_1 + \cdots + \lambda_n^k z_n q_n.$$

If $|\lambda_1| > |\lambda_2| \geq \cdots$ and $z_1 \neq 0$, then the direction of $x^{(k)}$ converges to the direction of q_1 .

Generally, for nonsymmetric A , $x^{(k)}$ will converge in direction to the leading eigenvector if

- the largest eigenvalue in modulus ($\lambda = a + bi$, $|\lambda| = \sqrt{a^2 + b^2}$) is distinct and has algebraic multiplicity 1;
- $x^{(0)}$ has a nonzero coefficient for the leading eigenvector in a Jordan-basis expansion—true with probability 1 for random $x^{(0)}$.

How fast do we converge?

Convergence rate. Consider a sequence $\alpha_1, \alpha_2, \dots$ of positive errors, $\alpha_k \rightarrow 0$. If

$$\lim_{k \rightarrow \infty} \frac{\alpha_{k+1}}{\alpha_k^r} \leq C \quad \text{for some } C > 0, \ r \geq 1, \quad (13.1)$$

then α converges at least at r th order:

- $r = 1$: linear (requires $C < 1$);
- $r = 2$: quadratic;
- $r = 3$: cubic.

For $r > 1$: $\alpha_{k+1} = O(\alpha_k^r)$; for example, if $\alpha_{k+1} = \alpha_k^3$ and $\alpha_k = 10^{-3}$, then $\alpha_{k+1} = 10^{-9}$. Asymptotically, the number of accurate digits improves by at least approximately a factor of r per iteration. The constant C is important for $r = 1$: it bounds the asymptotic error reduction factor.

Given a target accuracy ε :

$$k = O(\log_{1/C}(1/\varepsilon)) \text{ iterations needed if } r = 1, \quad k = O(\log_r \log(1/\varepsilon)) \text{ iterations needed for } r > 1.$$

Power iteration achieves linear convergence (under the stated assumptions) with

$$C = |\lambda_2| / |\lambda_1|.$$

Inverse and shifted inverse iteration

Assume the required inverse exists and the power-iteration assumptions hold for the transformed matrix.

Inverse iteration.

$$y = A^{-1}x^{(k)} \quad (\text{solve } Ay = x^{(k)}), \quad x^{(k+1)} = y / \|y\|_2.$$

Inverse iteration is power iteration with A^{-1} : it converges in direction to the eigenvector of the smallest-modulus eigenvalue, with $C = |\lambda_n| / |\lambda_{n-1}|$. Cost of k iterations: $O(n^3)$ for LU of A , $O(n^2)$ per iteration for the triangular solves—total $O(n^3 + kn^2)$.

Shifted inverse iteration.

$$y = (A - \sigma I)^{-1}x^{(k)}, \quad x^{(k+1)} = y / \|y\|_2.$$

Shifting maps the spectrum:

$$\lambda(A - \sigma I) = \lambda(A) - \sigma, \quad \lambda((A - \sigma I)^{-1}) = \frac{1}{\lambda(A) - \sigma},$$

so the largest-modulus eigenvalue of $(A - \sigma I)^{-1}$ is

$$\frac{1}{\lambda^* - \sigma}, \quad \text{where } \lambda^* \text{ is the eigenvalue of } A \text{ closest to } \sigma.$$

The shift σ controls which eigenvalue we converge to. The cost is the same as inverse iteration, and the convergence is analogous (contraction factor $|\lambda^* - \sigma| / |\lambda^{**} - \sigma|$, with λ^{**} the next-closest eigenvalue to σ).

Rayleigh quotient and Rayleigh quotient iteration

For $A \in \mathbb{R}^{n \times n}$ and $0 \neq x \in \mathbb{R}^n$, define the *Rayleigh quotient*

$$\rho(A, x) = \frac{x^T A x}{x^T x}.$$

- $\{\rho(A, x) : 0 \neq x \in \mathbb{R}^n\}$ is the real “field of values” of A .
- If (λ, x) is an eigenpair of A , then $\rho(A, x) = \lambda$.
- Given $y \neq 0$ near an eigenvector, estimate the eigenvalue by solving

$$\min_{\lambda} \|Ay - \lambda y\|_2,$$

a linear least-squares problem $y\lambda \cong Ay$;

$$\lambda = \frac{y^T A y}{y^T y}$$

from the normal equations.

- $\rho(A, x)$ is a good estimator of eigenvalues in the iterates of power iteration and (shifted) inverse iteration.

Rayleigh quotient iteration.

$$y = (A - \rho(A, x^{(k)})I)^{-1}x^{(k)}, \quad x^{(k+1)} = y / \|y\|_2. \quad (13.2)$$

For symmetric A , the method achieves a *cubic* convergence rate locally near a simple eigenpair, assuming the shifted solves are well-defined. The eigenpair reached depends on the starting vector. However, the shift changes every step, so the LU factorization must be redone at each step: $O(n^3k)$ cost for k iterations.

Numerical stability. As the shift approaches an eigenvalue, the shifted system becomes ill-conditioned. This need not spoil the normalized eigenvector: for symmetric A near a simple eigenpair, the largest amplified errors can lie along the desired eigendirection. Use LU with pivoting and monitor the eigenpair residual $\|Ax^{(k)} - \rho(A, x^{(k)})x^{(k)}\|_2$. Stop when it is sufficiently small, before another solve: the shifted matrix is singular at an exact eigenpair, and roundoff limits attainable accuracy.

CS 450: Numerical Analysis – Lecture 14

Deflation, Orthogonal Iteration, and QR Iteration

Edgar Solomonik

October 15, 2026

Today

Goal: develop algorithms for the full eigendecomposition.

- Deflation
- Schur form
- Orthogonal iteration
- QR iteration

Deflation

Power, inverse, and Rayleigh-quotient iterations compute a single eigenvector. *Deflation* is a technique to reduce A with eigenvalues $\lambda_1, \dots, \lambda_n$ to $\tilde{A}$ with eigenvalues $0, \lambda_2, \dots, \lambda_n$.

Given $A \in \mathbb{R}^{n \times n}$ with $A^T = A$ and an eigenpair (λ_1, x_1) with $\|x_1\|_2 = 1$,

$$\tilde{A} = A - \lambda_1 x_1 x_1^T = \lambda_2 x_2 x_2^T + \dots + \lambda_n x_n x_n^T. \quad (14.1)$$

For general (diagonalizable) A , we want left and right eigenpairs for λ_1 ,

$$Ax_1 = \lambda_1 x_1, \quad y_1^H A = \lambda_1 y_1^H,$$

as in

$$A = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^H, \quad \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^H = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^{-1}$$

(so $y_i^H x_j = \delta_{ij}$), and then

$$\tilde{A} = A - \lambda_1 x_1 y_1^H = \lambda_2 x_2 y_2^H + \dots + \lambda_n x_n y_n^H.$$

Deflation plus power/inverse iteration is suitable for computing a few extremal eigenvalues—too expensive for the full eigenvalue decomposition.

Direct eigendecomposition algorithms

Suppose $A \in \mathbb{R}^{n \times n}$ with $|\lambda_1| > \dots > |\lambda_n|$ (eigenvalues distinct in modulus; for real A this forces all eigenvalues to be real, since complex eigenvalues come in conjugate pairs of equal modulus).

Orthogonal iteration. Computes the k leading eigenvectors at once. Pick a random initial guess $X^{(0)} \in \mathbb{R}^{n \times k}$; for $i = 0, 1, \dots$ until convergence to tolerance:

$$Y = AX^{(i)}, \quad \text{factor } Y = QR, \quad X^{(i+1)} = Q.$$

The QR factorization normalizes the leading column of Y to get $X^{(i+1)}$ and orthogonalizes the other columns against the prior ones. Consequently, column 1 converges to the eigenvector with eigenvalue λ_1 , and column i converges to the i th Schur vector (for symmetric A , the eigenvector with eigenvalue λ_i). The rate of convergence for column i is

$$\max \left\{ \frac{|\lambda_i|}{|\lambda_{i-1}|}, \frac{|\lambda_{i+1}|}{|\lambda_i|} \right\}.$$

Cost per iteration: $AX^{(i)}$ is $O(n^2k)$; $Y = QR$ is $O(nk^2)$.

Note: A is used only in matrix multiplication, hence we need only a way to multiply by A and not its explicit entries. E.g., we can compute the leading singular vectors of B by orthogonal iteration on $A = B^T B$ without forming A , by applying $B^T(BX^{(i)})$ —a good idea for sparse B .

Convergence to the Schur form. When the eigenvalues are distinct in magnitude and $k = n$, orthogonal iteration converges to the *Schur form*

$$A = QTQ^T \quad (T \text{ triangular}), \quad \text{and for } A \in \mathbb{C}^{n \times n}, \quad A = UTU^H.$$

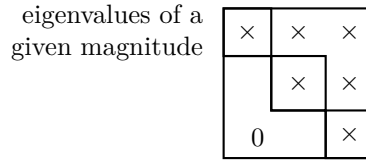
The Schur form may be derived from the Jordan form: $A = XJX^{-1}$ with $X = QR$ gives

$$A = Q \underbrace{RJR^{-1}}_{\text{triangular}} Q^T.$$

Indeed, the Schur basis is a fixed point of orthogonal iteration: if $X^{(i)} = Q$, then

$$AX^{(i)} = QT = Y \quad (\text{already a QR factorization}), \quad X^{(i+1)} = Q.$$

If $|\lambda_j| = |\lambda_{j+1}|$ for some j , orthogonal iteration converges to block-triangular form, with the eigenvalues of a given magnitude grouped in a diagonal block:



E.g., any orthogonal matrix is a stationary point of orthogonal iteration for $A = I$; oscillation is also possible within the diagonal blocks.

Note: the Schur form provides the eigenvalues of A : T is similar to A , and the eigenvalues of a triangular matrix are its diagonal entries, since

$$\det(T - t_{ii}I) = \prod_j (t_{jj} - t_{ii}) = 0$$

—the shifted matrix is triangular, so its determinant is the product of its diagonal entries, and the i th factor is zero.

QR iteration

When $n = k$, we can accelerate orthogonal iteration via a reformulation:

$$A = Q^{(0)}R^{(0)}; \quad \text{for } i = 0, 1, \dots \text{ until convergence to tolerance: } Q^{(i+1)}R^{(i+1)} = R^{(i)}Q^{(i)}.$$

This is equivalent to orthogonal iteration with $X^{(0)} = Q^{(0)}$.

Lemma. QR iteration = orthogonal iteration; in particular,

$$Q^{(i)}R^{(i)} = Q^{(i-1)T} \dots Q^{(0)T} A Q^{(0)} \dots Q^{(i-1)}, \quad X^{(i)} = Q^{(0)} \dots Q^{(i-1)}. \quad (14.2)$$

Proof by induction. Base case $i = 1$: from $A = Q^{(0)}R^{(0)}$,

$$Q^{(1)}R^{(1)} = R^{(0)}Q^{(0)} = Q^{(0)T} A Q^{(0)},$$

and

$$AX^{(0)} = A Q^{(0)} = Q^{(0)} Q^{(1)} R^{(1)} \implies X^{(1)} = Q^{(0)} Q^{(1)}.$$

Inductive step:

$$Q^{(i)}R^{(i)} = R^{(i-1)}Q^{(i-1)} = Q^{(i-1)T} (Q^{(i-1)}R^{(i-1)}) Q^{(i-1)} \stackrel{\text{ind. hyp.}}{=} Q^{(i-1)T} \dots Q^{(0)T} A Q^{(0)} \dots Q^{(i-1)},$$

and

$$AX^{(i-1)} \stackrel{\text{ind. hyp.}}{=} AQ^{(0)} \dots Q^{(i-1)} = \underbrace{Q^{(0)} \dots Q^{(i)}}_{X^{(i)}} R^{(i)} \quad (\text{a QR factorization of } AX^{(i-1)}). \quad \square$$

Cost: $O(n^3)$ per iteration; the main benefit is that $R^{(i)}Q^{(i)}$ may already be near triangular, unlike A .

Reduction to upper-Hessenberg form

Accelerate QR iteration by preprocessing. Goal:

$$Q A Q^T = H \quad (\text{upper Hessenberg: one nonzero subdiagonal}); \quad \text{if } A^T = A, \text{ } H \text{ is tridiagonal.}$$

Take $Q = Q_1 \dots Q_{n-2}$, a product of Householder reflectors, where Q_i is chosen from the entries below the subdiagonal of column i (so that it zeroes them while leaving rows $1, \dots, i$ alone). The first step:

$$Q_1^T A = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}, \quad Q_1^T A Q_1 = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix},$$

since the right-multiplication by Q_1 combines columns $2, \dots, n$ only, preserving the new zeros.

Upper-Hessenberg reduction.

- $H \leftarrow A$;
- for $i = 1$ to $n - 2$: pick the reflector Q_i to annihilate $H[i+2:n, i]$, and update $H \leftarrow Q_i^T H Q_i$;
- output H .

Cost: n iterations $\times O(n^2)$ (the cost of applying Q_i to H) $= O(n^3)$ total.

After upper-Hessenberg reduction, QR iteration has cost $O(n^2)$ /iteration via Givens rotations. When $A^H = A$, H is tridiagonal, and QR iteration has cost $O(n)$ /iteration.

Shifted QR iteration

Convergence in $O(1)$ iterations per eigenvalue is possible for QR iteration via *shifting*: with $A^{(0)} = A$,

$$Q^{(i+1)} R^{(i+1)} = A^{(i)} - \sigma^{(i)} I, \quad A^{(i+1)} = R^{(i+1)} Q^{(i+1)} + \sigma^{(i)} I, \quad (14.3)$$

so that

$$A^{(i+1)} = Q^{(i+1)T} A^{(i)} Q^{(i+1)}$$

remains similar to A —the diagonal shift preserves the eigenvectors, and adding it back after the swap preserves the eigenvalues. The shift is chosen to annihilate the last remaining subdiagonal nonzero of $A^{(i)}$ (deflation: once it is negligible, the problem splits).

Wilkinson shift. The simplest shift is the Rayleigh-quotient shift $\sigma^{(i)} = a_{nn}^{(i)}$. The *Wilkinson shift* is more robust: take $\sigma^{(i)}$ to be the eigenvalue of the trailing 2×2 block

$$\begin{bmatrix} a_{n-1,n-1}^{(i)} & a_{n-1,n}^{(i)} \\ a_{n,n-1}^{(i)} & a_{n,n}^{(i)} \end{bmatrix}$$

closest to $a_{nn}^{(i)}$; for symmetric tridiagonal matrices this makes shifted QR iteration provably convergent, with cubic local convergence. The classical algorithm uses this shift; in practice, the standard implementation converges in $O(1)$ iterations per eigenvalue.

Recent development. “Pseudospectral shattering” gives provable convergence of QR iteration with bounded backward error: push the eigenvalues apart (in magnitude) by a random perturbation; a more advanced shifting strategy is required.

CS 450: Numerical Analysis – Lecture 15

QR Iteration in Practice and Krylov Eigensolvers

Edgar Solomonik

October 20, 2026

Today

Goal: survey state-of-the-art eigensolvers.

- QR iteration acceleration
- Krylov solvers: Arnoldi and Lanczos

QR iteration in practice

Initial reduction. $(n - 2)$ Householder similarity transformations, at $O(n^3)$ cost (not iterative), reduce A to upper-Hessenberg form—tridiagonal when $A^H = A$:

$$QAQ^H = \begin{bmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix} \quad (\text{non-symmetric}), \quad QAQ^H = \begin{bmatrix} \times & \times & & & \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix} \quad (\text{Hermitian}).$$

QR iteration on an upper-Hessenberg matrix. Factoring $A = A^{(0)} = QR$ takes $(n - 1)$ Givens rotations, at $O(n^2)$ cost ($O(n)$ for tridiagonal); then

$$A^{(1)} = RQ, \quad Q = \underbrace{Q_1 \cdots Q_{n-1}}_{\text{Givens rotations}},$$

again at $O(n^2)$ cost to apply the rotations to R .

Shifted QR iteration.

$$A^{(0)} - \sigma_0 I = QR, \quad A^{(1)} = RQ + \sigma_0 I, \quad A^{(1)} - \sigma_1 I = QR, \quad \dots$$

Each step is a similarity transformation:

$$A^{(1)} = Q^T (A^{(0)} - \sigma I) Q + \sigma I = Q^T A^{(0)} Q.$$

We can pick σ to accelerate convergence:

- the Wilkinson shift uses the eigenvalue of the trailing 2×2 block closest to the last diagonal entry (Lecture 14);
- more sophisticated shifts guarantee convergence in $O(n)$ total iterations.

Can we forgo an iterative solver altogether. No! The Abel–Ruffini theorem (1824): there is no formula in radicals for the roots of general polynomials of degree $n \geq 5$ —the characteristic polynomial included (and every polynomial is the characteristic polynomial of its companion matrix, which appears below). Moreover, extracting all eigenvalues or the full eigenvalue decomposition from even a tridiagonal matrix remains complicated; several viable alternatives to QR iteration are available.

Krylov subspace methods

Returning to extreme eigenvalue computation: Krylov solvers. The $O(n^3)$ cost is expensive, especially if A is sparse—e.g., if A is a finite-difference operator, it has $O(1)$ nonzeros per row and $O(n)$ cost to apply (matrix–vector product).

Krylov subspace methods seek the optimal approximation given the power-iteration iterates:

$$\text{Kr}(A, x_0, s) = \text{span}\{x_0, Ax_0, \dots, A^{s-1}x_0\}, \quad K_s = \begin{bmatrix} x_0 & \cdots & A^{s-1}x_0 \end{bmatrix}, \quad (15.1)$$

seeking the largest/smallest eigenvector of A within $\text{Kr}(A, x_0, s)$.

Suppose K_n is full rank (otherwise we can often stop early—a lucky breakdown). Then

$$AK_n = \begin{bmatrix} Ax_0 & \cdots & A^n x_0 \end{bmatrix},$$

and collecting $A^n x_0$ in the basis K_n , $v = K_n^{-1} A^n x_0$,

$$K_n^{-1} AK_n = \begin{bmatrix} 0 & \cdots & 0 & v_1 \\ 1 & & & v_2 \\ & \ddots & & \vdots \\ & & 1 & v_n \end{bmatrix} = C_n, \quad (15.2)$$

the *companion matrix*—upper Hessenberg. Factoring $K_n = Q_n R_n$,

$$A = K_n C_n K_n^{-1} = Q_n \underbrace{R_n C_n R_n^{-1}}_{\text{upper-Hessenberg, } H_n} Q_n^T; \quad \text{if } A^H = A, \ H_n \text{ is tridiagonal.} \quad (15.3)$$

The leading part

$$H_s = Q_s^T A Q_s$$

(with Q_s the first s columns of Q_n) will be our general target. Note that K_n itself is ill-conditioned—its columns tend to the largest eigenvector—so we work with the orthonormal basis Q_s , never with K_n directly.

Ritz values and vectors. The eigenvalues of H_s are the best approximations to eigenvalues of A available from $\text{Kr}_s(A, x_0)$. E.g. (for symmetric A),

$$\min_{\substack{x \in \text{Kr}_s(A, x_0) \\ \|x\|_2=1}} x^T A x = \min_{\substack{y \in \mathbb{R}^s \\ \|y\|_2=1}} y^T Q_s^T A Q_s y = \lambda_{\min}(H_s) \quad (x = Q_s y). \quad (15.4)$$

The corresponding eigenvector approximation is $Q_s y$, for y an eigenvector of H_s .

Lanczos and Arnoldi

Our goal becomes to compute Q_s and H_s .

Arnoldi iteration. $\text{ARNOLDI}(A, x_0, s)$: seek $Q_s = [q_1 \ \cdots \ q_s]$ with $H_s = Q_s^T A Q_s$.

- $q_1 = x_0 / \|x_0\|_2$
- for $i = 1$ to s :
 - $y = Aq_i$
 - $H_{j,i} = \langle q_j, y \rangle$ for $j = 1, \dots, i$ (column i of H_s , from the existing q_i)
 - if $i < s$:
 - * $w = y - q_1 H_{1,i} - \cdots - q_i H_{i,i}$
 - * $H_{i+1,i} = \|w\|_2$
 - * $q_{i+1} = w / H_{i+1,i}$

Each pass first completes column i of H_s from the basis vector q_i already in hand; while iterations remain, the just-computed column is then used to orthogonalize y , giving the subdiagonal entry $H_{i+1,i}$ and the next basis vector q_{i+1} . The final pass ($i = s$) only fills in the last column.

Lanczos iteration. For $A^H = A$, use that H_s is tridiagonal: $\langle q_j, y \rangle = 0$ for $j < i - 1$, so each step orthogonalizes against q_{i-1} and q_i only (a three-term recurrence).

Finite precision. In floating-point arithmetic, Lanczos vectors can lose orthogonality, producing repeated approximations to already-converged eigenvalues. Reorthogonalization may be needed, increasing the work beyond the ideal three-term recurrence. The $O(n)$ orthogonalization cost below excludes this additional work.

Cost. Both algorithms take one matrix–vector product per iteration— $O(n^2)$ for dense A —plus the orthogonalization at step k : $O(nk)$ for Arnoldi, $O(n)$ for Lanczos, which can dominate when the matrix–vector product is cheap.

Arnoldi/Lanczos achieve a reduction similar to the direct one for $s = n$, and are preferable for $s \ll n$: good estimates of the extremal eigenvalues.

CS 450: Numerical Analysis – Lecture 16

Eigensolver Review, Generalized Eigenproblems, and the SVD

Edgar Solomonik

October 22, 2026

Today

Goal: review (and demo) eigenvalue computation algorithms and their properties.

- Power iteration and variants
- Orthogonal/QR iteration
- Krylov solvers

Iterative solvers

Power iteration. $\text{POWERITERATION}(A, x^{(0)}, \tau)$: normalize the nonzero starting vector $x^{(0)}$ and use the Rayleigh quotient $\rho(A, x) = x^H A x / x^H x$ for $x \neq 0$. For $i = 0, 1, \dots$, compute $\rho_i = \rho(A, x^{(i)})$ and stop when

$$\|Ax^{(i)} - \rho_i x^{(i)}\|_2 \leq \tau.$$

Otherwise, update

$$x^{(i+1)} = Ax^{(i)} / \|Ax^{(i)}\|_2.$$

Here $\tau > 0$ is a chosen absolute tolerance. For a relative criterion, scale the tolerance by an estimate of the matrix norm. Small changes in the Rayleigh quotient alone do not ensure an accurate eigenpair. Converges in direction to an eigenvector for the eigenvalue of A of largest magnitude if

- this eigenvalue is simple and strictly larger in magnitude than the others;
- $x^{(0)}$ has a nonzero coefficient along its eigenvector in an eigenvector or Jordan basis (true with probability one for a continuous random starting vector).

Shifting and inversion. For $\sigma \notin \lambda(A)$,

$$\lambda(A) = \{\lambda_1, \dots, \lambda_n\} \implies \lambda((A - \sigma I)^{-1}) = \left\{ \frac{1}{\lambda_1 - \sigma}, \dots, \frac{1}{\lambda_n - \sigma} \right\}.$$

Power iteration on $(A - \sigma I)^{-1}$ can be done by LU of $A - \sigma I$; it converges to the eigenpair with eigenvalue closest to σ (under analogous conditions).

Rayleigh quotient iteration. Pick $\sigma_i = \rho(A, x^{(i)})$, i.e. a new shift at each step. For Hermitian A , convergence is locally cubic near a simple eigenpair (Lecture 13), assuming the shifted solves are well-defined. As the shift approaches that eigenvalue, the inverse-iteration factor

$$\frac{|\lambda_2((A - \sigma_i I)^{-1})|}{|\lambda_1((A - \sigma_i I)^{-1})|}$$

tends to zero. Downsides: a new factorization is generally needed at every step, and the increasingly ill-conditioned shifted systems can cause numerical difficulties. Check the eigenpair residual before attempting another shifted solve.

Krylov solvers

Let's use all power iteration iterates instead of the last. The inverse-power expression below assumes A is nonsingular:

$$S = \{x^{(0)}, Ax^{(0)}, A^2x^{(0)}, \dots, \underbrace{A^{s-1}x^{(0)}}_y\} = \{A^{-(s-1)}y, \dots, A^{-1}y, y\} : \quad (16.1)$$

the same set then contains the inverse-iteration iterates of y . Check which directions grow *and* shrink fastest $\rightarrow$ extremal eigenvalues. Maintain orthonormal basis vectors $q_1, \dots, q_s$ in $Q = [q_1 \cdots q_s]$, assuming the Krylov subspace has dimension s :

$$\text{Kr}(A, x^{(0)}, s) = \text{span}(\{q_1, \dots, q_s\}) = \text{span } S.$$

Then (for Hermitian A)

$$\begin{aligned} \max_{\substack{x \in \text{Kr}(A, x^{(0)}, s) \\ x \neq 0}} |\rho(A, x)| &= \max_{0 \neq y \in \mathbb{C}^s} |\rho(A, Qy)| = \max_{y \neq 0} \left| \frac{y^H Q^H A Q y}{y^H Q^H Q y} \right| \\ &= \max_{y \neq 0} \left| \frac{y^H H y}{y^H y} \right| = \max_j |\lambda_j(H)|, \end{aligned} \quad (16.2)$$

where $H = Q^H A Q$. The power iterates in S , and the inverse iterates represented there when A is nonsingular, are particular choices of $x = Qy$. Thus optimizing over the whole subspace is at least as good in the $|\rho(A, x)|$ objective; equality can occur.

Lanczos and Arnoldi = power iteration + orthogonalization against the prior directions; for Hermitian A , use that

$$Q^H A Q = \begin{bmatrix} \times & \times & & & \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \times & \times & \times \\ & & & \times & \times \end{bmatrix},$$

which implies orthogonality to all except the last two directions in exact arithmetic.

Orthogonal and QR iteration

Compute many eigenvectors, or the full eigendecomposition, at once.

Orthogonal iteration. $\text{OI}(A, X^{(0)})$, with $X^{(0)} \in \mathbb{C}^{n \times k}$ having orthonormal columns and $k \leq n$ a parameter: at each iteration,

$$Y = AX^{(i)}, \quad Y = QR \text{ (reduced QR)}, \quad X^{(i+1)} = Q.$$

For $k < n$, changes in the subspace can be monitored by the deviation of the singular values of $(X^{(i)})^H X^{(i-1)}$ from 1, for $i \geq 1$. If $|\lambda_1| \geq \cdots \geq |\lambda_n|$ and $|\lambda_k| > |\lambda_{k+1}|$, a suitable starting subspace converges to the leading k -dimensional invariant subspace of A (the span of its leading k eigenvectors when A is diagonalizable). For nonsymmetric A , the individual columns need not converge to eigenvectors.

For $k = n$, the span is always $\mathbb{C}^n$, so this subspace-change test is not useful. Instead, monitor $(X^{(i)})^H A X^{(i)}$ approaching triangular form (or real Schur form in real arithmetic). With distinct-magnitude eigenvalues and a suitable starting basis, the entries below the diagonal tend to zero, yielding a Schur factorization

$$A = QTQ^H, \quad T \text{ upper triangular.}$$

QR iteration. A simplification of orthogonal iteration for $k = n$, with shifts and $A^{(0)} = A$:

$$\begin{aligned} A - \sigma^{(0)}I &= Q^{(0)}R^{(0)}, \\ A^{(1)} &= R^{(0)}Q^{(0)} + \sigma^{(0)}I, \\ A^{(1)} - \sigma^{(1)}I &= Q^{(1)}R^{(1)}, \\ A^{(2)} &= R^{(1)}Q^{(1)} + \sigma^{(1)}I, \\ &\vdots \end{aligned} \tag{16.3}$$

Each step satisfies $A^{(i+1)} = (Q^{(i)})^H A^{(i)} Q^{(i)}$, so the eigenvalues are preserved. In practice: reduce directly to upper-Hessenberg form first, then iterate with shifting. At fixed accuracy, $O(n)$ shifted steps is a typical practical estimate, giving $O(n^3)$ total work including the initial reduction; this is not an unconditional iteration bound. In the complex Schur form $A = QTQ^H$, the Schur vectors are eigenvectors if $AA^H = A^H A$; otherwise, solve $(T - \lambda I)y = 0$ for a nonzero y using the triangular structure, and return $x = Qy$.

Demo: Orthogonal and QR iteration, on the [course website](#) under Chapter 4.

Further topics in eigenvalue problems

Generalized eigenvalue problems. Given real matrices $A, B \in \mathbb{R}^{n \times n}$, the pair (λ, x) with $x \neq 0$ is a generalized eigenpair if

$$Ax = \lambda Bx.$$

- These arise naturally in applications (“matrix pencils”).
- If B is nonsingular, premultiplication gives $B^{-1}Ax = \lambda x$. However, $B^{-1}A$ may be dense and need not be symmetric even when A and B are symmetric. Applying $B^{-1}A$ implicitly requires a solve with B per product; an ill-conditioned B can amplify errors.
- If B is SPD, factor $B = LL^T$:

$$L^{-1}AL^{-T} \underbrace{L^T x}_y = \lambda \underbrace{L^T x}_y, \tag{16.4}$$

an ordinary eigenvalue problem for $L^{-1}AL^{-T}$, which is symmetric when A is symmetric.

- This Cholesky reduction requires B to be SPD. Even then, explicitly forming $L^{-1}AL^{-T}$ may destroy sparsity.

Application to the SVD. We often want the SVD for low-rank approximation. For real $A \in \mathbb{R}^{n \times n}$, recall the Eckart–Young–Mirsky theorem (Lecture 11):

$$\min_{B: \text{rank}(B) \leq r} \|A - B\|_2 \quad \text{is attained by} \quad B = \sum_{i=1}^r \sigma_i u_i v_i^T, \quad \text{where} \quad A = \sum_{i=1}^n \sigma_i u_i v_i^T, \quad \sigma_i \geq \sigma_{i+1}$$

(and ditto for $\|A - B\|_F$). Storing B in factored form as $U_r \Sigma_r V_r^T$ requires $O(nr)$ storage and $O(nr)$ work to apply to a vector; B itself may be dense. Alternative approaches for computing this approximation include:

- orthogonal iteration on AA^T ;
- eigenvalue decomposition of $A^T A$; or
- “bidiagonalization”: the stable general SVD algorithm, similar to tridiagonal reduction of $A^T A$.