AM 205: lecture 21

- Today: eigenvalue sensitivity
- Reminder: midterm starts today
  - Posted online at 5 PM on Thursday 9th
  - Deadline at 5 PM on Saturday 11th
  - Covers material up to and including lecture 19
  - Open book. No collaboration allowed.
  - Send questions as private messages on Piazza. Please don’t post any public follow-up discussions until Nov 17.
In some cases, the eigenvectors of $A$ can be chosen such that they are orthonormal

$$v_i^* v_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

In such a case, the matrix of eigenvectors, $Q$, is unitary, and hence $A$ can be unitarily diagonalized

$$A = QDQ^*$$
Eigenvalue Decomposition

**Theorem:** A hermitian matrix is unitarily diagonalizable, and its eigenvalues are real.

But hermitian matrices are not the only matrices that can be unitarily diagonalized... $A \in \mathbb{C}^{n \times n}$ is **normal** if

$$A^* A = AA^*$$

**Theorem:** A matrix is unitarily diagonalizable if and only if it is normal.
Gershgorin’s Theorem

Due to the link between eigenvalues and polynomial roots, in general one has to use iterative methods to compute eigenvalues.

However, it is possible to gain some information about eigenvalue locations more easily from Gershgorin’s Theorem.

Let $D(c, r) \equiv \{ x \in \mathbb{C} : |x - c| \leq r \}$ denote a disk in the complex plane centered at $c$ with radius $r$.

For a matrix $A \in \mathbb{C}^{n \times n}$, $D(a_{ii}, R_i)$ is called a Gershgorin disk, where

$$R_i = \sum_{j=1}^{n} |a_{ij}|, \quad \text{for } j \neq i.$$
Gershgorin’s Theorem

**Theorem**: All eigenvalues of $A \in \mathbb{C}^{n \times n}$ are contained within the union of the $n$ Gershgorin disks of $A$

**Proof**: See lecture
Gershgorin’s Theorem

Note that a matrix is **diagonally dominant** if

\[ |a_{ii}| > \sum_{j=1, j \neq i}^{n} |a_{ij}|, \quad \text{for } i = 1, 2, \ldots, n \]

It follows from Gershgorin’s Theorem that a diagonally dominant matrix cannot have a zero eigenvalue, hence **must be invertible**

For example, the finite difference discretization matrix of the differential operator \(-\Delta + I\) is diagonally dominant

In 2-dimensions, \((-\Delta + I)u = -u_{xx} - u_{yy} + u\)

Each row of the corresponding discretization matrix contains diagonal entry \(4/h + 1\), and four off-diagonal entries of \(-1/h\)
Sensitivity of Eigenvalue Problems

We shall now consider the sensitivity of the eigenvalues to perturbations in the matrix $A$

Suppose $A$ is nondefective, and hence $A = VDV^{-1}$

Let $\delta A$ denote a perturbation of $A$, and let $E \equiv V^{-1}\delta AV$, then

$$V^{-1}(A + \delta A)V = V^{-1}AV + V^{-1}\delta AV = D + E$$
Sensitivity of Eigenvalue Problems

For a nonsingular matrix $X$, the map $A \rightarrow X^{-1}AX$ is called a similarity transformation of $A$

**Theorem:** A similarity transformation preserves eigenvalues

**Proof:** We can equate the characteristic polynomials of $A$ and $X^{-1}AX$ (denoted $p_A(z)$ and $p_{X^{-1}AX}(z)$, respectively) as follows:

$$p_{X^{-1}AX}(z) = \det(zI - X^{-1}AX)$$
$$= \det(X^{-1}(zI - A)X)$$
$$= \det(X^{-1}) \det(zI - A) \det(X)$$
$$= \det(zI - A)$$
$$= p_A(z),$$

where we have used the identities $\det(AB) = \det(A) \det(B)$, and $\det(X^{-1}) = 1/\det(X) \square$
Sensitivity of Eigenvalue Problems

The identity \( V^{-1}(A + \delta A)V = D + E \) is a similarity transformation

Therefore \( A + \delta A \) and \( D + E \) have the same eigenvalues

Let \( \lambda_k, k = 1, 2, \ldots, n \) denote the eigenvalues of \( A \), and \( \tilde{\lambda} \) denote an eigenvalue of \( A + \delta A \)

Then for some \( w \in \mathbb{C}^n \), \((\tilde{\lambda}, w)\) is an eigenpair of \((D + E), i.e.\)

\[
(D + E)w = \tilde{\lambda}w
\]
Sensitivity of Eigenvalue Problems

This can be rewritten as

\[ w = (\tilde{\lambda}I - D)^{-1}Ew \]

This is a promising start because:

- we want to bound \( |\tilde{\lambda} - \lambda_k| \) for some \( k \)
- \((\tilde{\lambda}I - D)^{-1}\) is a diagonal matrix with entries \( 1/(\tilde{\lambda} - \lambda_k) \) on the diagonal
Sensitivity of Eigenvalue Problems

Taking norms yields

\[ \| \mathbf{w} \|_2 \leq \| (\tilde{\lambda} \mathbf{I} - \mathbf{D})^{-1} \|_2 \| \mathbf{E} \|_2 \| \mathbf{w} \|_2, \]

or

\[ \| (\tilde{\lambda} \mathbf{I} - \mathbf{D})^{-1} \|_2^{-1} \leq \| \mathbf{E} \|_2 \]

Note that the norm of a diagonal matrix is given by its largest entry (in abs. val.)\(^1\)

\[ \max_{\mathbf{v} \neq \mathbf{0}} \frac{\| \mathbf{Dv} \|}{\| \mathbf{v} \|} = \max_{\mathbf{v} \neq \mathbf{0}} \frac{\| (\mathbf{D}_{11} \mathbf{v}_1, \mathbf{D}_{22} \mathbf{v}_2, \ldots, \mathbf{D}_{nn} \mathbf{v}_n) \|}{\| \mathbf{v} \|} \]

\[ \leq \left\{ \max_{i=1,2,\ldots,n} |D_{ii}| \right\} \max_{\mathbf{v} \neq \mathbf{0}} \frac{\| \mathbf{v} \|}{\| \mathbf{v} \|} \]

\[ = \max_{i=1,2,\ldots,n} |D_{ii}| \]

---

\(^1\)This holds for any induced matrix norm, not just the 2-norm
Hence \( \| (\tilde{\lambda} I - D)^{-1} \|_2 = 1/|\tilde{\lambda} - \lambda_k^*| \), where \( \lambda_k^* \) is the eigenvalue of \( A \) closest to \( \tilde{\lambda} \).

Therefore it follows from \( \| (\tilde{\lambda} I - D)^{-1} \|_2^{-1} \leq \| E \|_2 \) that

\[
|\tilde{\lambda} - \lambda_k^*| = \| (\tilde{\lambda} I - D)^{-1} \|_2^{-1} \\
\leq \| E \|_2 \\
= \| V^{-1} \delta A V \|_2 \\
\leq \| V^{-1} \|_2 \| \delta A \|_2 \| V \|_2 \\
= \text{cond}(V) \| \delta A \|_2
\]

This result is known as the Bauer–Fike Theorem.
Sensitivity of Eigenvalue Problems

Hence suppose we compute the eigenvalues, $\tilde{\lambda}_i$, of the perturbed matrix $A + \delta A$

Then Bauer–Fike tells us that each $\tilde{\lambda}_i$ must reside in a disk of radius $\text{cond}(V)\|\delta A\|_2$ centered on some eigenvalue of $A$

If $V$ is poorly conditioned, then even for small perturbations $\delta A$, the disks can be large: sensitivity to perturbations

If $A$ is normal then $\text{cond}(V) = 1$, in which case the Bauer–Fike disk radius is just $\|\delta A\|_2$
Note that a limitation of Bauer–Fike is that it does not tell us which disk \( \tilde{\lambda}_i \) will reside in.

Therefore, this doesn’t rule out the possibility of, say, all \( \tilde{\lambda}_i \) clustering in just one Bauer–Fike disk.

In the case that \( A \) and \( A + \delta A \) are hermitian, we have a stronger result.
Sensitivity of Eigenvalue Problems

**Weyl’s Theorem:** Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq \cdots \leq \tilde{\lambda}_n$ be the eigenvalues of hermitian matrices $A$ and $A + \delta A$, respectively. Then $\max_{i=1, \ldots, n} |\lambda_i - \tilde{\lambda}_i| \leq \|\delta A\|_2$.

Hence in the hermitian case, each perturbed eigenvalue must be in the disk\(^2\) of its corresponding unperturbed eigenvalue!

---

\(^2\)In fact, eigenvalues of a hermitian matrix are real, so disk here is actually an interval in $\mathbb{R}$
The Bauer–Fike Theorem relates to perturbations of the whole spectrum.

We can also consider perturbations of individual eigenvalues.

Suppose, for simplicity, that $A \in \mathbb{C}^{n \times n}$ is symmetric, and consider the perturbed eigenvalue problem:

$$(A + E)(v + \Delta v) = (\lambda + \Delta \lambda)(v + \Delta v)$$

Expanding this equation, dropping second order terms, and using $Av = \lambda v$ gives:

$$A\Delta v + Ev \approx \Delta \lambda v + \lambda \Delta v$$
Sensitivity of Eigenvalue Problems

Premultiply $A\Delta v + Ev \approx \Delta \lambda v + \lambda \Delta v$ by $v^*$ to obtain

$$v^* A\Delta v + v^* Ev \approx \Delta \lambda v^* v + \lambda v^* \Delta v$$

Noting that

$$v^* A\Delta v = (v^* A\Delta v)^* = \Delta v^* Av = \lambda \Delta v^* v = \lambda v^* \Delta v$$

leads to

$$v^* Ev \approx \Delta \lambda v^* v,$$

or

$$\Delta \lambda = \frac{v^* Ev}{v^* v}$$
Finally, we obtain

\[ |\Delta \lambda| \approx \frac{|v^* E v|}{\|v\|^2_2} \leq \frac{\|v\|_2 \|E v\|^2_2}{\|v\|^2_2} = \|E\|_2, \]

so that \( |\Delta \lambda| \lesssim \|E\|_2 \)

We observe that

- perturbation bound does not depend on \( \text{cond}(V) \) when we consider only an individual eigenvalue
- this individual eigenvalue perturbation bound is asymptotic; it is rigorous only in the limit that the perturbations \( \to 0 \)
Algorithms for Eigenvalue Problems
Power Method
Power Method

The **power method** is perhaps the simplest eigenvalue algorithm. It finds the eigenvalue of \( A \in \mathbb{C}^{n \times n} \) with largest modulus.

1. Choose \( x_0 \in \mathbb{C}^n \) arbitrarily.
2. **for** \( k = 1, 2, \ldots \) **do**
3. \( x_k = Ax_{k-1} \)
4. **end for**

**Question:** How does this algorithm work?
**Power Method**

Assuming $A$ is nondefective, then the eigenvectors $v_1, v_2, \ldots, v_n$ provide a basis for $\mathbb{C}^n$

Therefore there exist coefficients $\alpha_i$ such that $x_0 = \sum_{j=1}^n \alpha_j v_j$

Then, we have

$$x_k = Ax_{k-1} = A^2 x_{k-2} = \cdots = A^k x_0$$

$$= A^k \left( \sum_{j=1}^n \alpha_j v_j \right) = \sum_{j=1}^n \alpha_j A^k v_j$$

$$= \sum_{j=1}^n \alpha_j \lambda_j^k v_j$$

$$= \lambda_n^k \left( \alpha_n v_n + \sum_{j=1}^{n-1} \alpha_j \left[ \frac{\lambda_j}{\lambda_n} \right]^k v_j \right)$$
Then if $|\lambda_n| > |\lambda_j|$, $1 \leq j < n$, we see that $x_k \to \lambda_n^k \alpha_n v_n$ as $k \to \infty$.

This algorithm converges linearly: the error terms are scaled by a factor at most $|\lambda_{n-1}|/|\lambda_n|$ at each iteration.

Also, we see that the method converges faster if $\lambda_n$ is well-separated from the rest of the spectrum.
Power Method

However, in practice the exponential factor $\lambda^k_n$ could cause overflow or underflow after relatively few iterations.

Therefore the standard form of the power method is actually the normalized power method:

1: choose $x_0 \in \mathbb{C}^n$ arbitrarily
2: for $k = 1, 2, \ldots$ do
3: $y_k = Ax_{k-1}$
4: $x_k = y_k / \|y_k\|$  
5: end for
Convergence analysis of the normalized power method is essentially the same as the un-normalized case.

Only difference is we now get an extra scaling factor, \( c_k \in \mathbb{R} \), due to the normalization at each step.

\[
x_k = c_k \lambda_n^k \left( \alpha_n v_n + \sum_{j=1}^{n-1} \alpha_j \left[ \frac{\lambda_j}{\lambda_n} \right]^k v_j \right)
\]
Power Method

This algorithm directly produces the eigenvector $v_n$

One way to recover $\lambda_n$ is to note that

$$y_k = Ax_{k-1} \approx \lambda_n x_{k-1}$$

Hence we can compare an entry of $y_k$ and $x_{k-1}$ to approximate $\lambda_n$

We also note two potential issues:

1. We require $x_0$ to have a nonzero component of $v_n$
2. There may be more than one eigenvalue with maximum modulus
Issue 1:

- In practice, very unlikely that \( x_0 \) will be orthogonal to \( v_n \)
- Even if \( x_0^* v_n = 0 \), rounding error will introduce a component of \( v_n \) during the power iterations

Issue 2:

- We cannot ignore the possibility that there is more than one “max. eigenvalue”
- In this case \( x_k \) would converge to a member of the corresponding eigenspace
An important idea in eigenvalue computations is to consider the “shifted” matrix \( A - \sigma I \), for \( \sigma \in \mathbb{R} \)

We see that

\[
(A - \sigma I)v_i = (\lambda_i - \sigma)v_i
\]

and hence the spectrum of \( A - \sigma I \) is shifted by \(-\sigma\), and the eigenvectors are the same.

For example, if all the eigenvalues are real, a shift can be used with the power method to converge to \( \lambda_1 \) instead of \( \lambda_n \).
Power Method

**Python example:** Consider power method and shifted power method for

\[ A = \begin{bmatrix} 4 & 1 \\ 1 & -2 \end{bmatrix}, \]

which has eigenvalues \( \lambda_1 = -2.1623, \lambda_2 = 4.1623 \)
Inverse Iteration
The eigenvalues of $A^{-1}$ are the reciprocals of the eigenvalues of $A$, since

$$Av = \lambda v \iff A^{-1}v = \frac{1}{\lambda}v$$

**Question:** What happens if we apply the power method to $A^{-1}$?
Inverse Iteration

**Answer:** We converge to the largest (in modulus) eigenvalue of $A^{-1}$, which is $1/\lambda_1$ (recall that $\lambda_1$ is the smallest eigenvalue of $A$).

This is called **inverse iteration**

1: choose $x_0 \in \mathbb{C}^n$ arbitrarily
2: **for** $k = 1, 2, \ldots$ **do**
3: solve $Ay_k = x_{k-1}$ for $y_k$
4: $x_k = y_k / \|y_k\|$
5: **end for**
Inverse Iteration

Hence inverse iteration gives $\lambda_1$ without requiring a shift.

This is helpful since it may be difficult to determine what shift is required to get $\lambda_1$ in the power method.

**Question:** What happens if we apply inverse iteration to the shifted matrix $A - \sigma I$?
Inverse Iteration

The smallest eigenvalue of $A - \sigma I$ is $(\lambda_{i^*} - \sigma)$, where

$$i^* = \arg \min_{i=1,2,\ldots,n} |\lambda_i - \sigma|,$$

and hence...

**Answer:** We converge to $\tilde{\lambda} = 1/(\lambda_{i^*} - \sigma)$, then recover $\lambda_{i^*}$ via

$$\lambda_{i^*} = \frac{1}{\tilde{\lambda}} + \sigma$$

Inverse iteration with shift allows us to find the eigenvalue closest to $\sigma$
Rayleigh Quotient Iteration
Rayleigh Quotient

For the remainder of this section (Rayleigh Quotient Iteration, QR Algorithm) we will assume that $A \in \mathbb{R}^{n \times n}$ is real and symmetric\(^3\).

The **Rayleigh quotient** is defined as

$$r(x) \equiv \frac{x^T A x}{x^T x}$$

If $(\lambda, v) \in \mathbb{R} \times \mathbb{R}^n$ is an eigenpair, then

$$r(v) = \frac{v^T A v}{v^T v} = \frac{\lambda v^T v}{v^T v} = \lambda$$

---

\(^3\)Much of the material generalizes to complex non-hermitian matrices, but symmetric case is simpler
Rayleigh Quotient

**Theorem:** Suppose $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix, then for any $x \in \mathbb{R}^n$ we have

$$\lambda_1 \leq r(x) \leq \lambda_n$$

**Proof:** We write $x$ as a linear combination of (orthogonal) eigenvectors $x = \sum_{j=1}^{n} \alpha_j v_j$, and the lower bound follows from

$$r(x) = \frac{x^T A x}{x^T x} = \frac{\sum_{j=1}^{n} \lambda_j \alpha_j^2}{\sum_{j=1}^{n} \alpha_j^2} \geq \lambda_1 \frac{\sum_{j=1}^{n} \alpha_j^2}{\sum_{j=1}^{n} \alpha_j^2} = \lambda_1$$

The proof of the upper bound $r(x) \leq \lambda_n$ is analogous \(\square\)
Corollary: A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive definite if and only if all of its eigenvalues are positive.

Proof: ($\Rightarrow$) Suppose $A$ is symmetric positive definite (SPD), then for any nonzero $x \in \mathbb{R}^n$, we have $x^T A x > 0$ and hence

$$\lambda_1 = r(v_1) = \frac{v_1^T A v_1}{v_1^T v_1} > 0$$

($\Leftarrow$) Suppose $A$ has positive eigenvalues, then for any nonzero $x \in \mathbb{R}^n$

$$x^T A x = r(x)(x^T x) \geq \lambda_1 \|x\|_2^2 > 0$$

$\square$
Rayleigh Quotient

But also, if \( x \) is an approximate eigenvector, then \( r(x) \) gives us a good approximation to the eigenvalue

This is because estimation of an eigenvalue from an approximate eigenvector is an \( n \times 1 \) linear least squares problem: \( x\lambda \approx Ax \)

\( x \in \mathbb{R}^n \) is our “tall thin matrix” and \( Ax \in \mathbb{R}^n \) is our right-hand side

Hence the normal equation for \( x\lambda \approx Ax \) yields the Rayleigh quotient, \( i.e. \)

\[ x^T x\lambda = x^T Ax \]
Rayleigh Quotient

**Question**: How accurate is the Rayleigh quotient approximation to an eigenvalue?

Let's consider $r$ as a function of $x$, so $r : \mathbb{R}^n \rightarrow \mathbb{R}$

\[
\frac{\partial r(x)}{\partial x_j} = \frac{\partial}{\partial x_j} (x^T Ax) \frac{x^T x}{x^T x} - \frac{(x^T Ax) \partial}{\partial x_j} (x^T x) \frac{x^T x}{(x^T x)^2}
\]

\[
= \frac{2(Ax)_j}{x^T x} - \frac{(x^T Ax) 2x_j}{(x^T x)^2}
\]

\[
= \frac{2}{x^T x} (Ax - r(x)x)_j
\]

(Note that the second equation relies on the symmetry of $A$)
Therefore

\[ \nabla r(x) = \frac{2}{x^T x} (Ax - r(x)x) \]

For an eigenpair \((\lambda, v)\) we have \(r(v) = \lambda\) and hence

\[ \nabla r(v) = \frac{2}{v^T v} (Av - \lambda v) = 0 \]

This shows that eigenvectors of \(A\) are stationary points of \(r\)
Suppose \((\lambda, v)\) is an eigenpair of \(A\), and let us consider a Taylor expansion of \(r(x)\) about \(v\):

\[
    r(x) = r(v) + \nabla r(v)^T (x - v) + \frac{1}{2} (x - v)^T H_r(v)(x - v) + H.O.T.
\]

Hence as \(x \to v\) the error in a Rayleigh quotient approximation is

\[
    |r(x) - \lambda| = O(\|x - v\|^2_2)
\]

That is, the Rayleigh quotient approx. to an eigenvalue squares the error in a corresponding eigenvector approx.