Applied Math II Notes

MTH5361

Spring 2026

Lecture by Professor Ronald Morgan, Notes by Chris F Lin

Wed 1/21

Chapters 7.1+7.2: eigenvalues and eigenvectors. starting with eigenvalues:

Definition 1. λ is an **eigenvalue** if it satisfies for square matrix A and vector $z \neq \vec{0}$,

$$Az = \lambda z$$

So what are some uses of eigenvalues? Spectral theory. In general, can tell us a few things

- natural frequency of vibrations (can result in resonance, increased amplification)
- energy levels
- understanding iterative methods

Theorem 1. Invertible matrix theorem: The following for square matrix $A \in \mathbb{R}^{n \text{ x } n}$

- A invertible (nonsingular)
- A has linearly independent columns
- $\det(A) \neq 0$
- A has rank 0
- $Ax = \vec{0}$ has only the trivial solution
- coumns of A span $\mathbb{R}^n$

We will now derive the characteristic equation:

$$Az = \lambda z, z \neq \vec{0} \rightarrow Az - \lambda z = \vec{0} \tag{1}$$

$$(A - \lambda I)z = \vec{0} \tag{2}$$

and since $z \neq \vec{0}$, then $A-\lambda I$ is singular by the invertible matrix theorem, and therefore $\det(A-\lambda I) = 0$.

Definition 2. Characteristic equation $\rho_A(\lambda)$ of matrix A is:

$$\rho_A(\lambda) = \det(A - \lambda I) = 0$$

where λ are eigenvalues of A. $\rho(\lambda)$ also known as characteristic polynomial.

The characteristic polynomial gives our eigenvalues via its roots, so using $A - \lambda I$, we can get associated eigenvectors. Typically, we don't use the characteristic equation to solve for eigenvalues:

- hard to find characteristic equation
- for ill-conditioned problems, hard to get actual roots

Example 1. $A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Our characteristic equation is therefore:

$$\rho(\lambda) = \det(A - \lambda I) = \det \begin{pmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & 1-\lambda \end{pmatrix} = (1-\lambda)\big[(-\lambda)^2 - (1)(1)\big] = (1-\lambda)(\lambda^2 - 1) = -(\lambda-1)^2(\lambda+1).$$

Therefore, $\lambda = 1$ (multiplicity 2) and $\lambda = -1$.

For $\lambda = 1$:

$$N(A - \lambda I) = N\left(\begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right) \to \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \vec{x} = \vec{0} \tag{3}$$

$$\to \begin{cases} -x_1 + x_2 = 0, \\ x_1 - x_2 = 0 \end{cases} \tag{4}$$

$$\to x_1 = x_2,\ x_3 \text{ free} \tag{5}$$

$$\to N(A - \lambda I) = \text{span}\{(1,1,0)^T, (0,0,1)^T\}. \tag{6}$$

Therefore, for $\lambda = 1$, we have two associated eigenvectors. For $\lambda = -1$, we get $(1, -1, 0)^T$.

We can actually infer this because we know it must be orthogonal to the other two vectos. Note that although we have 3 eigenvectors, we technically have an infinite amount. We will be using eigenpairs to diagonalize our matrices A. Note *that not all matrices all diagonalizable.* We will also be looking at norms and other relevant information.

Friday 01/23

Today, norms of vectors and matrices

Definition 3. 2-norm: for $v \in \mathbb{R}^n$:

$$\|v\|_2 = \sqrt{v_1^2 + \cdots + v_n^2}$$

If v has imaginary components: $\|v\|_2 = \sqrt{\langle v, v \rangle}$

Definition 4. inner/dot product of two vectors $v, w \in \mathbb{R}^n$:

$$\langle v, w \rangle = v \cdot w = v_1 w_1 + \cdots + v_n w_n = v^T w$$

For complex, need to do conjugate: $v \cdot w = v^* w$ (complex conjugate + transpose, $*$ or H operation for hermitian)

That last equality is technically matrix multiplication (becomes a 1 by 1 matrix).

Definition 5. 1-norm:

$$\|v\|_1 = \sum |v_i|$$

Definition 6. infinite-norm:

$$\|v\|_\infty = \max_i \{v_i\}$$

Note, if we have an n-norm, then taking $n \to \infty$ we get the infinite norm. We can also define matrix norms:

Definition 7. matrix 2-norm, assume $A \in \mathbb{R}^{n \times n}$:

$$\|A\|_2 = \max_{x \neq \vec{0}} \frac{\|Ax\|_2}{\|x\|_2}$$

In other words, "how does matrix expand the length of a vector?" Similar to vectors, also have other norms for matrices.

Definition 8. matrix 1-norm:

$$\|\dot{\|}_1 = \max_x \frac{\|Ax\|_1}{\|x\|_1}$$

Definition 9. matrix ∞-norm:

$$\|\dot{\|}_\infty = \max_x \frac{\|Ax\|_\infty}{\|x\|_\infty}$$

Example 2. $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$: not easy to compute!

The 2-norm is important, but hard to compute by hand. 1 and ∞ norms are easier.

Example 3. $\|A\|_1 =?$ Let $A = \begin{pmatrix} 1 & 2 \\ -3 & 0 \end{pmatrix}$ $\|Ax\|_1 = \| \begin{pmatrix} x_1 + 2x_2 \\ -3x_1 \end{pmatrix}$ Therefore, we have:

$$\frac{4\|x_1\| + 2\|x_2\|}{\|x_1\| + \|x_2\|}$$

Let's guess and check. if $x = (1, 0)^T$, $Ax = (1, -3)^T$, and so $\|Ax\|_1 = 4$. Similarly, if $x = (0, 1)^T$, then norm is 2. Remember that instead of doing dot products, we can think of it as linear combination of the columns of A. Through trial and error, we see that:

$$\|A\|_1 = \text{max column sum of absolute values}$$

Now for *orthogonal vectors*:
Note, we will only use the standard inner product (and so only standard orthogonality).

Definition 10. vectors $v, w \in \mathbb{R}^n$ are **orthogonal** if $v \cdot w = 0$.

Example 4. Orthogonal set of vectors: $\{(1,1,1)^T, (1,-1,0)^T\}$. If we normalize the vectors, they're considered *orthonormal vectors*.

We also have similar definition for matrices:

Definition 11. Orthonormal matrix A has orthonormal columns.
Example: $A = \begin{pmatrix} 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & 0 \end{pmatrix}$

Definition 12. Orthogonal matrix are square orthonormal matrices. Columns are norm 1 and orthogonal.
(use these because they tend to reduce round-off error)

A special class of orthogonal matrices are *rotation matrices*: $\begin{pmatrix} \cos & \sin \\ -\sin & \cos \end{pmatrix}$

Definition 13. square matrices $A, B = FAF^{-1}$ are **similar** if $A, B \in \mathbb{R}^{n \times n}$, and the operation $P^{-1}AP$ is called a **similarity transformation** on A.

An important characteristic is that **similar matrices share eigenvalues**. this means that to find eigenvalues of A, we can instead find the eigenvalues of similar matrices; so we want to apply *similar transformations* until we're able to find a matrix that we can easily compute the eigenvalues for.

Theorem 2. Similar matrices have the same eigenvalues.

Proof. by definition:

$$Az = \lambda z \to FAz = \lambda Fz \to (FAF^{-1})(Fz) = \lambda Fz = \lambda(Fz)$$

□

Note: this not only tells us about the eigenvalues, but also tells us about the eigenvectors: our new eigenvector is Fz. Therefore, to reverse, we need to track all the transformations F. For numerical properties, we typically want F to be orthogonal.

Mon 1-26

Last time, discussed orthogonal matrices. What do you call it if complex? *unitary matrix*.

Definition 14. Unitary matrix U is a square matrix with orthonormal columns and satisfies (* the complex conj, transpose opeeration):

$$U^*U = I$$

Definition 15. square matrix A is **symmetric** if $A^T = A$.
A is **hermitian** if: $A^* = A$

many properties for symmetric matrices; these extend to hermitian matrices too. There is also a middleground: can have complex and symmetric (only). Now we discuss diagonalizing matrices. We have three different ways:

$$\begin{aligned} AP = PD, D \text{ diagonal} \\ P^{-1}DP = D \\ A = PDP^{-1} \end{aligned}$$

Each of these three has the same eigenvalues because P is a similarity transformation matrix.

Theorem 3. A is diagonalizable if and only if A has a full set of linearly independent eigenvectors

Proof. ($\leftarrow$) let $z_1, \dots z_n$ be linearly independent, and matrix $P = [z_1, \dots, z_n]$. now:

$$AP = [Az_1, \dots, Az_n]$$

this is because each column is a linear combination of A where the ith column of P is the coefficients. We also know that $Az_1 = \lambda_1 z_1$. Therefore:

$$AP = \cdots = [\lambda_1 z_1 \dots \lambda_n z_n] = PD$$

where D is diagonal matrix with λ_i along the diagonals. We also need to know that P is invertible. This is true because the columns are linearly independent, so P is invertible by invertible matrix theorem.
$\rightarrow$ A is diagonal, we want to show full set of eigenvectors.
A diagonalizable $\rightarrow$ $AP = PD$, P invertible. Looking at first column:

$$A\vec{p_1} = d_1\vec{p_1} \rightarrow p_1 \text{ is an eigenvector}$$

therefore, all p_i are eigenvectors and since P is invertible, teh columns P_i are linearly independent. □

What about if we can't diagonalize?

Example 5. $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ Only has 1 eigenvector. So what can we do? jsut ignore it. why? In real computation, we generally have diagonalizable matrices. Even if not, we won't run into this problem when we performming the computations.

Theorem 4. If A has distinct eigenvalues, it is *diagonalizable*.

Proof. if $\{z_1, \ldots, z_n\}$ eigenvectors corresponding to eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$, we wts linearly independent, then can use last theorem. Assume linearly *dependent*. By definition, an arbitrary:

$$\lambda_n = \sum_1^{n-1} c_i z_i$$

Additionally, we have:

$$\lambda_n z_n = \sum_1^{n-1} c_i \lambda_i z_i$$

Therefore, if we multiply the first equation by λ_j and subtract from the second, we get:

$$\vec{0} = c_1(\lambda_1 - \lambda_n)z_1 + \cdots + c_{n-1}(\lambda_{n-1} - \lambda_n)z_{n-1}$$

We know that c_i nonzero, λ_i distinct, so we can say not all coefficients are 0. Therefore, by definition, $\{z_i\}$ are linearly dependent. We can therefore keep repeating this until our set reduces to only $\{z_1\}$. But our sete has to be linearly independent, so we have a contradiction. □

Next time, discuss left and right eigenvalues. So what other forms can we put matrices into?

- If *diagonalizable*: $A = PDP^{-1}$
- if *not*: (will show next time)

What if it's too tedious to diagonalize or its not practical? There are intermediate forms, ie "sure form' or upper triangular.

Definition 16. for an eigenpair (λ, x) of A: $Ax = \lambda x$. So if $y^*A = \lambda y^*$, then y^* is a **left eigenvector**. another way of writing:

$$(y^*A)^* = (\lambda y^*)^* \leftrightarrow A^*y = \bar{\lambda}y \rightarrow \lambda^* = \bar{\lambda}$$

Theorem 5. Left and right eigenvectors corresponding to different eigenvalues are orthogonal

How about if matrices are symmetric?

Proof. $i \neq j$

$$y_j^* A x_i = y_j^*(\lambda_i x_i) = \lambda_i y_j^* x_i$$

on LHS: $= (A^*y)^* x_i = (\bar{\lambda_j} y)^* x_i = \lambda_j y_i^* x_i$ Therefore, $y_j^* x_i = 0$ since $\lambda_i, \lambda_j \neq 0$ □

wed 1/28

So what if A can't be diagonalized or it's too expensive? A is similar to: $\begin{pmatrix} J_1 & \ldots & \\ \ldots & J_2 & \ldots \\ \ldots & \ldots & J_p \end{pmatrix}$ where $J = \begin{pmatrix} \lambda_i & 1 & \ldots & \ldots \\ \ldots & \lambda_i & 1 & \ldots \\ \ldots & \ldots & \ldots & \lambda_i \end{pmatrix}$ for k repeats, $k-1$ 1's in off diagonal in block.

Example 6. eigenvalues 2, 3, 7, 9: $\begin{pmatrix} 2 & & & & & & \\ & 3 & & & & & \\ & & 7 & 1 & & & \\ & & & 7 & & & \\ & & & & 9 & 1 & \\ & & & & & 9 & 1 \\ & & & & & & 9 \end{pmatrix}$ 2, 3 are unique eigenvalues. 7 is a double eigenvalue, 9 is a triple. In this form, the off diagonals/blocks don't impact the eigenvalue, but they do impact the eigenvector.

Theorem 6. Cayley-Hamilton theorem states that for square matrix A, A satisfies its own characteristic equation.

Example 7. $A = UTU^*$, where $T = \begin{pmatrix} 1 & 1 & * \\ & 1 & * \\ & & 2 \end{pmatrix}$ We know we can get this form (schur form) using unitary transformations ($U^*AU = T$). We want to show that T satisfies $\rho(T) = 0$ therefore we can also apply to matrix A.

$$\begin{aligned} \rho(\lambda) &= (\lambda-1)(\lambda-1)(\lambda-2) = 2 \rightarrow \\ \rho(T) &= (T-I)(T-I)(T-2I) \\ &= \begin{pmatrix} 0 & 1 & * \\ & 0 & * \\ & & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & * \\ & 0 & * \\ & & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & * \\ & -1 & * \\ & & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 10 & * \\ 0 & 0 & * \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & * \\ 0 & -1 & * \\ 0 & 0 & 0 \end{pmatrix} = 0 \end{aligned}$$

left and right evectors have different uses:

- how to "deflate" from matrix
- spectral decomposition (using L/R eigenvector)

rehash: λ is an eigenvalue and x, y right/left eigenvectors. showed that left/right with diff eigenvalues are orthogonal. So what if A is symmetric?

$$\begin{aligned} \text{hermitian} \rightarrow A^* = A = A^T, \lambda_i \text{ all real} \\ \rightarrow Ay = \bar{\lambda} y = \lambda y \\ \rightarrow \text{if y is left evector, then right is also an evector} \end{aligned}$$

Therefore, for hermitian matrices, we dont need to distinguish between L/R eigenvectors.
Deflation can remove "effects" of eigenvalues. We can remove eigenvalues to help make solving a system of linear equations easier!

Assume $y^*x = 1$, with x, y right and left eigenvectors. our deflation operation is thus $A + \sigma xy^*$:

$$\begin{aligned}(A + \sigma xy^*)x &= Ax + \sigma xy^y x \\ &= Ax + \sigma x \\ &= Ax + \sigma x \\ &= (\lambda + \sigma)x\end{aligned}$$

Therefore, for the same eigenvector x, we have different eigenvalue $\lambda + \sigma$, hence a shift of λ.
Sometimes, we dont have both left and right vectors, so how to deflate if we dont have both? doesnt work as good, can still deflate but then eigenvectors change.
Something we don't like about $U^*AU = T$, if we start with real matrix A, can still get complex eigenvalues. we prefer to do only real arithmetic: so have quasi-schur form for real matrices. Can go to *quasi-schur* form with real orthogonal-similarity transformations. One way of getting complex eigenvalues os nonsymmetric matrices.

Definition 17. quasi-schur form: uppertriangular except for 2 by 2 blocks.

The key thing here is that it keeps things "real".

Example 8. $A = \begin{pmatrix} 1 & 1 & 7 \\ -1 & 1 & -2 \\ 0 & 0 & 3 \end{pmatrix}$ transforming $A \to A'$: eigenvalues are 3 and $1 \pm i$. Note the blocks: for the $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, this corresponds with $1 \pm i$ and 3 corresponds with 3 eigenvalue.

Can we also have multiple 1 by 1 matrices of duplicates? yes. Apply similarity matrices to get J-C form

Theorem 7. Sohur Triangularization: Every square matrix A is *unitarily similar* to an *upper triangular* matrix. So: $U^*AT = T$ and diagon of T are e-value of A

In practice, can just use orthogonal instead of unitary. From upper triangular, easier to compute eigenpairs.

Fri-01/30

Note for problem 7.17, this is deflation without both left and right eigenvectors. Now, spectral decomposition. This is for diagonalizable matrices with distinct eigenvalues.

$$A = \lambda_1 \frac{x_1 y_1^*}{y_1^* x_1} + \cdots = \sum_{i=1}^{n} \lambda_i \frac{x_i y_i^*}{y_i^* x_i}$$

Note these are all rank 1 matrices, so can also write as outer products ($x_i y_i^* \to$ matrix rank 1, $y_i^* x_i \to$ single value). Remember that it's equivalent to PDP^{-1}. So what if not distinct?

$$A = \lambda_1 G_1 + \cdots + \lambda_k G_k$$

- For λ_i multiplicity 2, then G_i rank 2
- instead of $x_1^* x_1 = G_1$, $\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix}$
- can more for distinct eigenvalue case

let $y_i^* x_i = 1 \to A = \lambda_1 x_1 y_1^* + \dots$

Proof. want to show Av, for arbitrary v equal to $(\lambda_1 x_1 y_1^* + \dots)v$
Therefore: $A = \lambda_1 x_1 y_1^* + \dots$
begin with $v = \beta_1 x_1 + \dots + \beta_n x_n$. Expand by basis of eigenvectors (assumption due to diagonalizable assumption):

$$Av = A(\beta_1 x_1 + \dots + \beta_n x_n) = \beta_1 \lambda_1 x_1 + \dots + \beta_n \lambda_n x_n$$

$$(\lambda_1 x_1 y_1^* + \dots)v = (\lambda_1 x_1 y_1^* + \dots + \lambda_n x_n y_n^*)(\beta_1 x_1 + \dots + \beta_n x_n)$$

get equality because $y_i^* x_j = 0$ by orthogonality and $y_1^* x_1 = 1$ □

(beginning 7.3)
Functions of diagonalizable matrices:
What if we want to compute e^A, matrix A? what is e^x? can use taylor expansion:

$$e^x = 1 + x + \frac{x^2}{2!} + \dots$$

We can therefore calculate the same way:

$$e^A = I + A + A^2/2! + \dots$$

$$A^2 = PDP^{-1}PDP^{-1}$$
$$PD^2P^{-1}$$

extends to: $A^n = PD^nP^{-1}$, so therefore:

$$e^A = P(I + D + D^2 + \dots)P^{-1} \tag{7}$$

$$= P \begin{pmatrix} 1 + \lambda_1 + \dots + \lambda_1^n/n! & 0 & 0 \\ \dots & & \\ 0 & \dots & 1 + \lambda_n + \dots + \lambda_n^j/j! \end{pmatrix} P^{-1} \tag{8}$$

Looks easy, but requires diagonalizability.

Mon 2/1

Definition 18. triangle property of norms (A matrix, x vector):

$$\|Ax\| \leq \|A\|\|x\|$$

Same property for all norms, and similar for matrix mutiplication:

Definition 19. Triangle proeprty for norms of matrix multiplication:

$$\|AB\| \leq \|A\|\|B\|$$

Definition 20. condition number $K(A) = \|A\|\|A^{-1}\|$

For iterative methods, condition number tells us how fast we get convergence.
What's the norm of a diagonal matrix? $\|D\| = \max_i |d_i|$. Can show through trial and error or 2 norm. In general, if we have A diagonalizable, we can write:

$$f(A) = Pf(D)P^{-1}$$

where we apply function $f(x)$ to each of the diagonal values (ie eigenvalues).

Definition 21. Newmann series (or alternatively, geometric series of matrices)

$$(I - A)^{-1} = I + A + A^2 + \dots$$

We will use this to define $(I - A)^{-1}$. Want to eventually work back to get form $Pf(A)P^{-1}$ if $f(A) = (I - A)^{-}1$. In general:

$$\begin{aligned}
(I - A)^{-1} &= PP^{-1} + PDP^{-1} + \dots \\
&= P(I + D + D^2 + \dots)P^{-1} \\
&= P \begin{pmatrix} 1 + \lambda_1 + \lambda_1^2 + \dots & 0 & \dots \\ \dots & & \\ 0 & \dots & 1 + \lambda_n + \lambda_n^2 + \dots \end{pmatrix} P^{-1} \\
&= \begin{pmatrix} \frac{1}{1-\lambda_1} & \dots & 0 \\ 0 & \dots & \frac{1}{1-\lambda_n} \end{pmatrix} \\
&= P(I - D)^{-}1P^{-1} \\
&= Pf(D)P^{-1}
\end{aligned}$$

For this to be possible, we require all $\lambda_i < 1$.

Theorem 8.

$$\|\lambda_i\| \leq \|A\|, \forall i$$

Proof.

$$\begin{aligned}
&Az = \lambda z, z \neq \vec{0} \\
&\quad \rightarrow \|Az\| = |\lambda|\|z\| \rightarrow |\lambda| = \frac{\|Az\|}{\|z\|} = \|A\|
\end{aligned}$$

or:

$$\|A\|\|z\| \geq |\lambda|\|z\| \rightarrow \|A\| \geq |\lambda|$$

□

We're now heading towards theorem that bounds how much λ_i can change when matrix is perturbed. This is important due to potential noise, round-off error.

Theorem 9. A diagonalizable, then: If $\|A\| < 1$, then $(I - A)^{-1}$ exists.

In other words, if we're changing identity by small amount, then we still have invertibility.

Proof.

$$\sum_{j=1}^{k} A^j = \sum_{j=1}^{k} PD^jP^{-1} = P(\sum D^j)P^{-1}$$
$$= P \begin{pmatrix} \lambda_1 + \cdots + \lambda_1^k & \dots & 0 \\ \dots & & \\ 0 & \dots & \lambda_n + \cdots + \lambda_n^k \end{pmatrix} P^{-1}$$

Taking limit on both sides $k \to \infty$, we get $\frac{1}{1-\lambda_i}$, but only true if $|\lambda_i| < 1$ for convergence. Therefore, since $\|A\| < 1$, then by previous theorem, $|\lambda_i| < 1$. therefore, series converges and inverse exists. □

Wed 2/4

Theorem 10. Bauer-Fike theorem Assume A diagonaliable, $B = A + E$. λ_i eigenvalues of A, β an eigenvalue of B, and $K(P) = \|P\|\|P^{-1}\|$. Then:

$$\min_{\lambda_i} |\beta - \lambda_i| \leq K(P)\|E\|$$

- E can be thought of as a perturbation.
- This shows how how much eigenvalues change when a matrix is perturbed

Tools to help prove:

- IF A, B invertible, then so is AB
- IF B or C is not invertible, then neither is BC.

BF-Proof:

Proof. Assume β not eigenvalue of A (otherwise trivial). So $(\beta I - A)^{-1}$ exists and can be represetned with neumann series. Therefore:

$$(\beta I - A)^{-1}(\beta I - B) = (\beta I - A)^{-1}(\beta I + A - A - B) = I + (\beta I - A)^{-1}(A - B) = I - (\beta I - A)^{-1}E$$

Since $\beta I - B$ is not invertible then the right hand side of the above equality is also not invertible. Since we also said: If $\|A\| < 1$ implies $(I - A)$ is invertible, then the contrapositve can be applied here and we get:

$$\|(\beta I - A)^{-1}E\| \geq 1$$

$$\begin{aligned}
\rightarrow 1 &\leq \|(\beta I - A)^{-1}\| \cdot \|E\| \\
&= \|(\beta I - PDP^{-1})^{-1}\|\|E\| \\
&= \|(\beta PP^{-1} - PDP^{-1})^{-1}\|\|E\| \\
&= \|(P(\beta I - D)P^{-1})^{-1}\|\|E\| \\
&\leq \|P\|\|P^{-1}\|\|E\|\|(\beta I - D)^{-1}\| \\
&= K(P)\|E\| \min_i |\beta - \lambda_i|^{-1}
\end{aligned}$$

$$\rightarrow \min_i |\beta - \lambda_i| \leq K(P)\|E\|$$

□

Example 9. $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}, E = \begin{pmatrix} -0.66 & -.073 & .0026 \\ -.051 & .034 & .015 \\ -.017 & .006 & .013 \end{pmatrix}$
$\lambda = (-.5157, .1509, 11.345)^T, \|E\| = .1, B = A + E \rightarrow \beta = (-.5711, .2142, 11.338)^T$
Theorem states change in eigenvalue limited by $K(P)\|E\| = 1 \times .1 = 0.1$.

Note that $K(P) = 1$ for all symmetric matrices because P can be orthonormal: $P^{-1} = P^T$.

Example 10.

$$A = E = \begin{pmatrix} 1 & 100 & 0 \\ 0 & 4 & 100 \\ 0 & 0 & 6 \end{pmatrix}$$

$\lambda = 1, 4, 6$. So $A+E \rightarrow \beta = (-1.91, 6.45 \pm 4.66i)^T$. Our eigenvalue is now complex. Our eigenvalues changed a lot because our condition number changed drastically: $K(A) = 3561$ (ie eigenvectors not close to being orthogonoal).

Now to discuss first method for solving eigenvalue: not good but leads to better methods with this simple idea. Choose starting vector S, can be random or chosen heuristically. Now continously multiply by A: (note can preserve previous calculation so net step is still only multiplying by A):

$$\begin{aligned}
&As \\
&A(As) = A^2 s \\
&\cdots = A^3 s \\
&\cdots
\end{aligned}$$

What does this give us if continued? Assume that A is diagonalizable. Therefore, assume λ_i eigenvalues and wlog, assume $\lambda_i \leq \lambda_{i+1}$ so λ_n largest eigenvalue. Respective eigenvectors z_i. Therefore, because diagonalizable, our eigenvectors z_i form a basis of $\mathbb{R}^n$, so we can expand our random vector s:

$$s = \sum \beta_i z_i$$

Now, multiply both sides by A:

$$\begin{aligned} As &= \beta_1 Az_1 + \cdots + \beta_n Az_n \\ &= \beta_1 \lambda_1 z_1 + \cdots + \beta_n \lambda_n z_n \end{aligned}$$

Can repeat this n times to get:

$$A^j s = \beta_1 \lambda_1^j z_1 + \cdots + \beta_n \lambda_n^j z_n$$

Let's "normalize" by dividing both sides by $\beta_n \lambda_n^j$:

$$\frac{A^j s}{\beta_n \lambda_n^j} = (\frac{\beta_1}{\beta_n})(\frac{\lambda_1}{\lambda_n})^j z_1 + \cdots + z_n$$

Therefore, we've been able to isolate and our eigenvectors z_n, corresponding to largest eigenvalue. Couple questions:

- How does this fail?
- How fast is convergence?

Fri 2/6

Last time, we used power method to isolate z_n largest eigenvalue's eigenvector.

- What's rate of convergence?

  observe $(\frac{\lambda_{n-1}}{\lambda_n})^j$ is largest term, so this is our determining factor: $|\frac{\lambda_{n-1}}{\lambda_n}|$
- How good is this?
- *Complexity*: one matrix-vector product per iteration
- *rate of convergence*: in real life, eigenvaalues are close to one another, so $\frac{\lambda_{n-1}}{\lambda_n}$ is close to 1 so close convergence
- Only finds largest eigenvalue

What can we do to improve this in the event our largest two eigenvalues are very close to eachother? two ways:

1. Using subspaces (or Krylov subspaces)
2. Shifts and inverting to retain eigenvector: $z = (\lambda - \sigma)(A - \sigma I)^{-1} z$

   this is practical for smaller matrices because we can easily calculate the inverse. Method 1 better for large matrices

Comments on power method:

- typically normalize after each step
- suppose x approximates an eigenvector: we can estimate the eigenvalue using *Rayleigh Quotient*

Definition 22. Rayleigh Quotient (ρ): given vector x with norm 1:

$$\rho = \frac{x^*Ax}{x^*x} = x^*Ax$$

residual vector (r):

$$r = Ax = \rho x$$

Residual norm:

$$\|r\| = \frac{\|Ax - \rho x\|}{\|x\|}$$

So power method improvements: shift, shift + invert

Definition 23. Shifted Power Method:

$$A - \sigma I$$

Example 11. if eigenvalues are: $\{-4, -2, -1, 3, 7, 9, 10\}$: If we use power method to find evector corresponding with 10, we have a rate of $\frac{9}{10}$. If we shift by instead $\sigma = 2$, we now are finding for the new eigenvalues:

$$\{-6, -4, -3, 1, 6, 7, 8\}$$

where 8 corresponds to the original 10. Here, we get a rate of $\frac{7}{8}$, which is an improvement. Note that shifting too much will result in lack of convergence.

- If $\sigma = 3$, our first and last no longer converge as they're equal, so will bounce between two vectors that are a linear combination of z_1, z_n in our example
- shifting more will cause -6 to be our new largest eigenvalue, so we will instead be solving for a different eigenvalue.

Mon 2-9

Example 12. Let A have eigenvalues $\lambda \in \{1, 3, 5, 7, 9, 10, 11\}$

- What eigenvector will power method find and what rate?

  $\lambda = 11, @\frac{10}{11}$
- What if use $\sigma = -6, A - 6I$

  Technically no convergence

- $\sigma = 7$?

  5.5, $\frac{4.5}{5.5}$

- A^{-1}?

  this is inverse power method: need to solve A^{-1}, or solve $Aw = s$. If we use standard guassian elimination, $o(n^2)$. So while finding the inverse is expensive, we can get a much better convergence rate.

  $\{1/11, 1/5, \ldots, 1\}$; solving $\frac{1}{\lambda_i}$ corresponds with λ_i: rate is now $\frac{1}{3}$, so very fast

- $\lambda = 4.5$: $(A - 4.5I)^{-1}$. This is inverse+shift method

From our last example, we see that we can get infinitely small, so infinitely fast rate of convergence. However, we do need to know eigenvalues to know how much to shift, and inverse is still expensive. So what's next? It's still impractical, and sometimes we may not be able to factor.
If we have a good eigenvector, we can instead use *Rayleigh Quotient*

Wed 2-11

Last will discuss Rayleigh quotient iteration:

Algorithm 1. Rayleigh Quotient Iteration: begin with random y, starting shift σ. We start by solving:

$$(A - 4I)w = y^{(0)}$$

We now have a new $y^{(1)}$, and let's normalize: $y^{(1)} = \frac{w}{\|w\|}$. The next step is going to be the rayleigh-quotient:

$$\rho = y^{(1)} A y$$

and so next iteration: $w = (A - \rho I)^{-1} y^{(1)}$

This gives very fast convergence, but have a new problem: we no longer know which eigenvalue/eigenvector we're solving for due to the changing shifts. A different initialization can result ina different solution every time.
Now we discuss different ways of finding orthogonal matrices (heading towards *QR iterative method*).

Definition 24. Q is an **orthogonal matrix** if it is square and has orthonormal columns.

Example 13. $Q = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}$

What's great about orthogonal matrices? There is less roundoff error.

Theorem 11. $\forall x \in \mathbb{R}^n$, given orthogonal matrix $Q \in \mathbb{R}^{n \times n}$, then:

$$\|Qx\| = \|x\|$$

Proof. $Q^TQ = I \leftrightarrow Q^{-1} = Q^T \leftrightarrow QQ^T = I$ because Q is a square matrix. We also have: $\|v\|_2 = \sqrt{v_1^2 + \cdots + v_n^2} = \sqrt{v^Tv}$, or $v^Tv = \|v\|_2^2$. Now:

$$\|Qx\|^2 = (Qx)^T(Qx) = x^TQ^TQx = x^Tx = \|x\|^2$$

□

Example 14. A rotation matrix is an example of an orthogonal matrix.

$$Q = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

Easy to show that $Q^TQ = I$.

So why do we care for the rotation matrix?
Goal: apply similarity transforms until $A \to R^A$ is upper triangular. Unfortunately, simply applying rotation matrix is a similarity transform. We need QAQ^T. So we no longer get an upper triangular matrix, but we can still get close.

Friday 2-13-26

Definition 25. QR factorization $A = QR$, Q orthonormal, R uppertriangular.

Note that all matrices are QR factorizable. If A is complex, then we instead require Q to be unitary. Some uses of QR factorization:

- can solve $Ax = b$ with higher accuracy than Gaussian-Elimination
- can solve least squares
- reveals rank of matrix (SVD is better but more expensive)
- most importantly, we can use it to find eigenvalues

We ofen use householder transformations (reflections, but here we will be using rotations).

Example 15. $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, Q is our rotation matrix. We want to apply our transformation so that our bottom left value is 0; therefore we need:

$$-\sin\theta + 3\cos\theta = 0 \leftrightarrow \tan\theta = 3 \to \theta = \arctan(3) \to \sin\theta = 3/\sqrt{10}, \cos\theta = 1/\sqrt{10}$$

(In general for 2×2, for matrix $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, $\tan\theta = C/A$.) So from above:

$$\to AQ_1 = \begin{pmatrix} \sqrt{10} & * \\ 0 & * \end{pmatrix}$$

So we have $Q_1A = R$, and therefore $A = Q_1^TR$, or since Q is orthonormal, $A = QR$.

Example 16. For any general 3×3 matrix, need to apply our rotation so that we can achieve a 0 in three different places in the matrix. Begin with the following:

$$Q_1 = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

This fixes the third row, while changing the top two and resulting in a zero:

$$Q_1 A = \begin{pmatrix} ** & ** & ** \\ 0 & ** & ** \\ * & * & * \end{pmatrix}$$

Now,we apply a different rotation matrix that fixes the second row, but introduces a zero in the third row:

$$Q_2 = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \to Q_2(Q_1 A) = \begin{pmatrix} * * * & * * * & * * * \\ 0 & ** & ** \\ 0 & * * * & * * * \end{pmatrix}$$

Finally, we need to introduce one more zero while maintaining the zeros already introduced:

$$Q_3 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} (Q_2(Q_1 A)) = \begin{pmatrix} * * * & * * * & * * * \\ 0 & * * ** & * * ** \\ 0 & 0 & * * ** \end{pmatrix}$$

Ideally we want to transform $A \to R$, an upper triangular matrix. $QR = A$ does not preserve eigenvalues, so we need to use similarity transformations.
If we perform AQ^T, this undos the uppertriangular form we perviously achieved. Unfortunately, we cannot find eigenvalues in a finite number of steps, and we cannot factor a degree 5+ polynomial since we cannot solve in finite number of steps.
HOWEVER, we can get closer to the upper traingular form than our original A:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} = \begin{pmatrix} * & * & * \\ ** & ** & ** \\ 0 & ** & ** \end{pmatrix}$$

$$\to \begin{pmatrix} * & * & * \\ ** & ** & ** \\ 0 & ** & ** \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}$$

From the above, the first multiplication maintains values from the first row, and second following multiplication maintains the values in the first column. Overall, we end up with what is called an *upper-hessenberg form*

Definition 26. Matrix A is in **Upper-Hessenberg Form** if it is an upper triangular matrix with an additional off-diagonal.:

$$\begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{pmatrix}$$

This leads into the QR iteration:

Definition 27. QR **Iteration:**

1. $A = QR$; perform QR factorization
2. $\hat{A} = RQ$: perform multiplication
3. Repeat where next iteration, $\hat{A}$ is our new A to perform QR factorization on.

We have a few observations (9 total, only three shown this lecture):

1. This is a similarity transformation!

   Proof. If $\hat{A} = RQ$, then let $A = QR$, $Q^T A = R$ implies $\hat{A} = Q^T AQ$ □

2. (skip for now)
3. The last row converges to the form: $[0, \ldots, 0, *]$ at a rate similar to the inverse power method. Similar to the inverse power method, we also need to perform a factorization. So how can we improve? by using shifts. Instead of $A = QR$, apply $A - \sigma I = QR$.

1 Monday 2-16-2026

Last time, QR iteration. We use this for small and medium size matrices.

- $O(n^3)$ (compared to gaussian elimination, which is $O(2/3n^3)$)
- what about methods for large matrices? this leads to methods for those.

Continuation of properties previously mentioned:

1. This is a similarity transformation
2. Preserves symmetry

   Proof.
   $$(\hat{A})^T = (Q^T AQ)^T = Q^T A^T Q = Q^T AQ = \hat{A}$$
   □

3. Last row converges to $[0, \ldots, 0, *]$ where $*$ becomes an eigenvalue at rate of inverse power-method

4. can add shifts: $A - \sigma I = QR \rightarrow \hat{A} - \sigma I = RQ$, $\hat{A} = RQ + \sigma I$

   Instead, this converges like shift + inverse power method

   Using shift converges even better than Rayleigh Quotient (Use eval of 2×2 lowest submatrix)

5. QR factorization expensive

   Instead, first transform into something that is easier to QR factorize (*upper hessenberg* H)

   start QR with H; this reduces from $O(n^2)$ to $O(n)$ now

6. We can deflate to smaller problems along the way

   Once reduce to the last row with only a value in the last nth column, we can instead now solve for only the $n-1$ subset of the matrix, ignoring the last row

7. For real matrices with complex eigenvalues, use a double shift to preserve real component

8. implicit shift (chase the bulge)

   don't technically need to do the explicit QR factorization

We use all these to help understand what is needed to write a good QR algorithm!

Friday 2-20-2026

For large matrices, QR iteration struggles (because QR factorization is too expensive). What are ways to deal with large matrices? Power method is one; this is only simple matrix-vector multiplication, especially good when using sparse matrices. Like the QR method, we will use power method as a basis for next method.

Definition 28. Krylov Subspace K defined for a vector $s \in \mathbb{R}^n$, square matrix $A \in \mathbb{R}^{n \times n}$, is:

$$K := \text{span}\{S, As, A^2 s, \ldots, A^{m-1} s\}$$

Note the similarity to the power method, which continues for an endless number of powers. So if we have a krylov subspace K, we can define any $y \in K$ as:

$$y = c_1 s + c_2 As + \cdots + s_m A^{m-1} s$$

Why is this better than power method?

- power only needs $A^j s$, the "best" one
- similar to power method, if eigenvalues are close, takes longer to converge for them
- in contrast, this solves for the smallest eigenvalues first.

Krylove Convergence Analysis
Assume A is diagonalizable with eigenpairs (λ_i, z_i) where $|\lambda_1 \leq \ldots |\lambda_n|$. For vector s, we can therefore expand it as:

$$s := \beta_1 z_1 = \cdots + \beta_n z_n$$

if we have a $y \in K(A)$, then $y = c_1 s + c_2 As + \cdots + \ldots c_m A^{m-1} s$ (note like a polynomial so can be thought of like a polynomial method).
therefore:

$$\rightarrow p(\alpha) = c_1 + c_2\alpha + c_3\alpha^2 + \cdots + c_m\alpha^{m-1} \tag{9}$$
$$p(\alpha) = c_1 I + c_2 A + \cdots + c_m A^{m-1} \tag{10}$$
$$\rightarrow p(A)s = c_1 Is + c_2 As + \cdots + c_m A^{m-1} s \tag{11}$$
$$\rightarrow y = p(A)s \tag{12}$$

So what happens if we apply the polynomial to an eigenvector:

$$y = p(A)s = \beta p(A)z_1 + \cdots + \beta_n p(A)z_n$$

Will demonstrate $p(A)z_z$ as an example, but not a proof:

Example 17. if $p(A) = A^3 - 2A$, then:

$$p(A)z_1 = A^3 z_1 - 2A_2 \tag{13}$$
$$A^2(Az_1) - 2Az_1 \tag{14}$$
$$= A^2\lambda_1 z_1 - 2\lambda_1 z_1 \tag{15}$$
$$= A(Az_1)\lambda_1 - 2\lambda_1 z_1 \tag{16}$$
$$= A\lambda_1^2 z_1 - 2\lambda_1 z_1 \tag{17}$$
$$= \cdots = (\lambda_1^3 - 2\lambda_1)z_1 = p(\lambda_1)z_1 \tag{18}$$

Therefore in general, we have $p(A)z_1 = p(\lambda_1)z_1$, so:

$$y = \beta_1 p(\lambda_1)z_1 + \cdots + \beta_n p(\lambda_n)z_n$$

Therefore, if we want to compute λ_1, we need to find z_1, so simply need to find $p(\lambda_i)$ to be small at all eigenvalues *except* λ_i

2 Mon Feb 23

Review the gram-schmidt orthogonalization
Now continuing: we want to find the eigenpairs of a large sparse matrix.

$$K = \text{span}\{s, As, A^2 s, \ldots, A^{m-1}s\}$$

then $y = p(A)s = \beta_1 p(\lambda_1)z_1 + \cdots + \beta_n p(\lambda_n)z_n$; if K is of degree m, then polynomial p is of degree $m-1$ (but can also be less).
In general, we want $p()$ to be small at all λ_i except λ_j, the eigenvalue we're looking for. In general, this takes a high degree polynomial to do this.

$$p(\lambda_1) >>> p(\lambda_2), \ldots, p(\lambda_n) \approx \epsilon$$

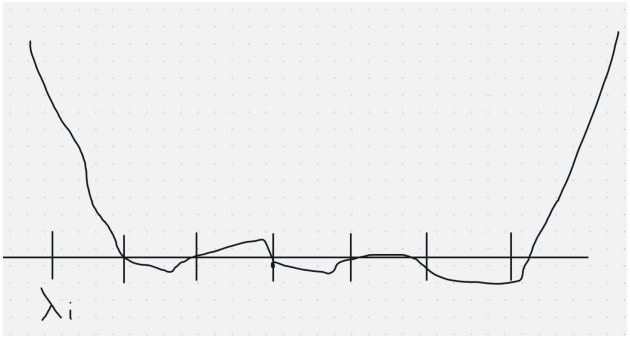

Figure 1: Fast Convergence Example

So what's the criteria for convergence? When do we get fast convergence? Ideally, λ_i is at the *exterior* of the spectrum (figure 1). More difficult if λ_i is not *separated* form the rest of the eigenvalues (figure 2). How about if "near" exterior? Starting vector affects convergence a little

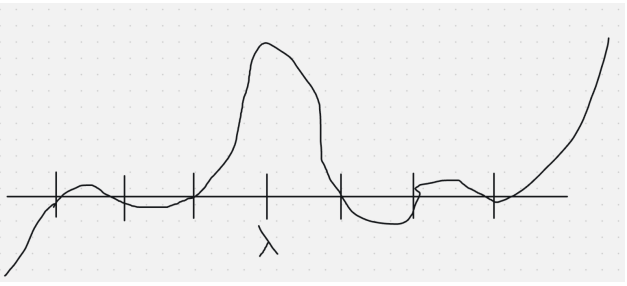

Figure 2: Slow Convergence Example

(only the beginning). Gaps in spextrum also has an effect of making convergence bettter; lower degree (figure 3) Now for K kryolov subspace, let $y \in K$, we want a useful y so how do we do this?

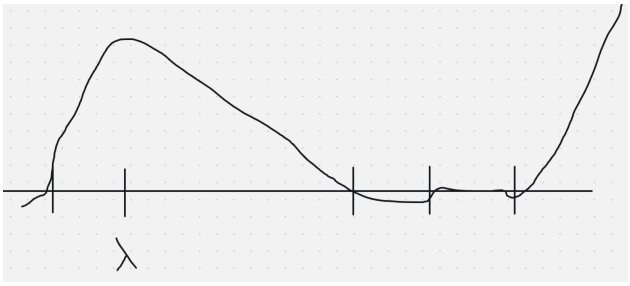

Figure 3: Near Exterior Convergence Example

Use **rayleigh-ritz procedure** (almost a generalization of rayleigh quotient).

Definition 29. Rayleigh-ritz Procedure: given any subspace S:

1. find othonormal basis for S (ie using gram-schmidth), call this $\{v_1, \ldots, v_m\}$, let $V = [v_1 | \ldots | v_m]$ (not a square matrix but will be very tall and skinny)
2. $H = V^T AV$

3. find epair of H. Note, H is easier to find pairs because smaller

   eigenvalues θ_i called R itz values, eigenvectors g_i ritz vectors

4. make g_i longer: find $y_i : VG_i$

   y_i is an approximation of eigenvectors, call them ritz vectors.

This procedure approximates the eigenpairs of A.
Now beginning Arnoldi Recurrence:

Definition 30. Arnoldie Recurrence: This is a combination of Krylov + gram-schmidt
Let s starting vector; we're trying to find $V = [v_1, \dots, v_n]$

1. $v_1 = s/\|s\|$
2. $w = Av_1$ (this makes it belong to the Krylov subspace)
3. do: $w = Av_1 - \alpha v_1$, α chosen such that $w \perp v_1$

   we do this via:

$$v_1^T w = v_1^T(Av_1 - \alpha v_1) \tag{19}$$
$$= v_1^T AV_1 - \alpha v_1^T v_1 = 0 \tag{20}$$
$$\to \alpha = v_1^T AV_1 \tag{21}$$

4. now have $v_2 = w/\|w\|$ (note v_2 orth to v_1)
5. next do $w = Av_2 - \alpha v_2 - \beta v_1$ (and orthogonal to both v_2, v_1)
6. solve β to be perpendicular to v_1: $\beta = v_1^T AV_2$
7. ...

Next time, will discuss methods to make it cheaper.

3 wed 2-25-2026

Krylov convergence for eigenvalues:

Example 18. A evals $1, \dots, 100$. Will find 1 faster than 3. Mentioned two properties:

1. faster for exterior evals

   Look up Yousef Saad's books on finding eigenvalues, also solving linear equations

2. well-separated (relative to the whole spectrum) eigenvalues are found faster, take lower degree polynomials.

3. range of eigenvalues has impact:

   If A range $1, \dots, 100$ and B range $1, \dots, 1000$:

$$\frac{2-1}{100-1} \text{ compared to } \frac{2-1}{1000-1}$$

Back to Arnoldi Recurrence: this develops orthonormal basis for our Krylov Subspace. We want $v = [v_1, \dots, v_j]$ with $\text{span}\{v_1, \dots, v_j\} = \text{span}\{s, As, \dots, A^{j-1}s\}$. We don't directly use the Krylov subspace because it results in very similar vectors as $j \to \infty$, so can result in numerical errors/ roundof errors.

$$v_1 = s/\|s\|, \hat{v_2} = Av_1 - \alpha_1 v_1$$

not surprising we're multiplying by A because want krylove subspace. α chosen to make $\hat{v_2} \perp v_1$, then also normalize $v_2 = \hat{v_2}/\|v_2\|$.
We now have: $\text{span}\{v_1, v_2\}$ an orthonormal set, and $\text{span}\{v_1, v_2\} = \text{span}\{s, As\}$.
Similarly, we continue to do:

$$\hat{v_3} = Av_2 - \alpha_2 v_2 - \beta_2 v_1, v_3 = \hat{v_3}/\|v_3\|$$

From here, how do we get H? (instead of just $H = V^*AV$)
In fact:

$$\alpha_1 = h_{11}, \tag{22}$$
$$\alpha_2 = h_{22}, \|v_2\| = h_{21}, \beta_2 = h_{12}, \dots \tag{23}$$

From here, we will show $H = V^TAV$ is upper hessenberg.

Proof. want to show: $h_{ij} = 0$ if $i > j + 1$, otherwise $h_{ij} = v_i^*AV_j$.
claim: if $\forall j \in \text{span}\{v_1, \dots, v_j\}$, then $Av_j \notin \text{span}\{v_1, \dots, v_i\}$
Instead (use v_1 as an example):

$$Av_1 = \hat{v_2} + \alpha v_1 = \|\hat{v_2}\|v_2 + \alpha v_1 \tag{24}$$
$$= \text{linear combination of } v_1, v_2 \tag{25}$$

So: $Av_j \in \text{span}\{v_1, \dots, v_{j+1}\}$
Does this increase to a higher or lower Kryolov subspace?

$$Av_j = c_1v_1 + \cdots + c_{j+1}v_{j+1} \tag{26}$$
$$\to h_{ij} = v_j^*(c_1v_1 + \cdots + c_{j+1}v_{j+1}) = 0 \text{by orthogonality to the vectors} \tag{27}$$

□

Definition 31. An alteration is the **Lanczos 3-term Recurrence** (1950, arnoldi 1951)
Have $\{v_1, \dots, v_j\}$, want v_{j+1} and suppose A is symmetric.

$$\hat{v_{j+1}} = Av_j - \alpha_j v_j - \beta_j v_{j-1}$$

$$\hat{v_{j+1}} = \hat{v_{j+1}}/\|v\|$$

We only need orthogonality to previous 2 terms; automatically have orthogonality to the rest

Will show proof or orthogonality to everything else:

Proof. want to show $\hat{v_{j+1}} = Av_j - \alpha v_j - \beta v_{j-1}$ perpendicular to $v_1, \ldots, v_{j-2}$
will show for $i = 1, \ldots, j-2$:

$$v_i^*(\hat{v_{j+1}}) = v_i^*(Av_j - \alpha v_j - \beta v_{j-1}) \tag{28}$$
$$= v_i^* AV_j - \alpha v_i^* v_j - \beta v_j^* v_{j-1} \tag{29}$$
$$= v_i^* AV_j \tag{30}$$
$$= v_i^* A^* v_j = (Av_i)^* v_j \tag{31}$$

However, $Av_i \in \text{span}\{v_1, \ldots, v_{i+1}\}$ so we have the above is 0, and so perpendicular for all i up to $j-1$ □

Friday 2-27

(7.5: Symmetric and Hermitian matrices)

Definition 32. Normal matrices: $A * A = AA*$
Symmetric (real) Matrices: $A^T = A$
Hermitian (complex) Matrices: $A^* = A$

Very briefly, going back to Arnoldi recurrence and how to show that why the components of H are simply the constants $\alpha_,\beta_j$ generated by each step:

Proof. Let S be starting vector. Then by defn:

$$v_1 = s/\|s\|, \hat{v_2} = Av_1 - \alpha_1 v_1, \alpha = \ldots, h_{11} = \alpha_1 \tag{32}$$
$$v_2 = \hat{v_2}/\|\hat{v_2}\|, \hat{v_3} = Av_1 - \alpha_2 v_1 - \beta_2 v_2, h_{22} = \alpha_2, h_{21} = \|\hat{v_2}\|, h_{12} = \beta_2 \tag{33}$$

For α_1, we need $\hat{v_2} \perp v_1$, so $v_1^* \hat{v_2} = v_1^*(Av_1 - \alpha v_1) \to \alpha = \frac{v_1^* \hat{v_2} - v_1^* A v_1}{\|v_1^* v_1\|}$
So why is $h_{11} = \alpha$?

$$H = V^*AV = \begin{pmatrix} v_1^* \\ \dots \\ v_k^* \end{pmatrix} A \begin{pmatrix} v_1 & \dots & v_k \end{pmatrix} \tag{34}$$

$$\rightarrow \text{ to get } h_{11} := e_1^* H e_1 \qquad e_1 = \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} \tag{35}$$

$$= e_1^*(\vec{h_1}1 + \vec{h_2} * 0 + \dots + \vec{h_k} * 0) \qquad e_i = \begin{pmatrix} 0 \\ \dots \\ 1 \\ \dots \\ 0 \end{pmatrix} \tag{36}$$

$$= e_1^* \vec{h_1} = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \vec{h_1} \tag{37}$$

$$= e_1^* V^* A V e_1 \tag{38}$$

Note that:

$$Ve_1 = \vec{v_1} \tag{39}$$

$$e_1^* V^* = (Ve_1)^* = \vec{v_1} \tag{40}$$

$$\rightarrow = \vec{v_1}^* A \vec{v_1} = \alpha \tag{41}$$

□

Now discussing convergence for finding eigenvalues of Krylov subspace.

- in code shown, "restarted arnoldi" restarts with a single vector
- builds out 50 vectors then restarts to 1 for a total of 35 cycles
- if target is $.1, .2, \dots, 10, 11, \dots, 4910$, takes 10x longer than integers 1 to 5000
- instead, try restarting using a ritz vector (restarting with multiple is harder)
- how about, instead we find 15 evals but restart with 20 around 200 iterations, buffer of extra 5 makes it faster
- lowering restart to 15, 15th takes significantly longer time (due to loss of buffer)
- new example: positive and negative eigenvalues: makes our eigenvalues interior values.

(Now back to 7.5 in text)

Theorem 12. Eigenvalues of symmetric (or hermitian) matrices are real. Also, eigenvectors of distinct eigenvalues are orthogonal

Proof. (pt 1) given (λ, z), then $z^*Az = z^*(\lambda z) = \lambda z^* z$ because symmetric, $A = A^*$

$$\rightarrow = z^*A^*z = (Az)^*z \tag{42}$$
$$= (\lambda z)^* z^* \tag{43}$$
$$= \bar{\lambda} z^* z \tag{44}$$

Therefore, $(\lambda - \bar{\lambda})z^*z = 0$
Since z not zero vector, we get $\lambda = \bar{\lambda}$, and therefore λ is real. □

Monday 3/2

Theorem 13. U orthogonal/unitary or orthogonal, then $\|Ux\| = \|x\|$

Proof. $\|Ux\|^2 = (Ux)^*(Ux) = x^*U^*Ux = x^*x = \|x\|^2$ □

Let's assume A symmetric with real eigenvalues. Second part of last theorem was that eigenvectors orthogonal if corresponding to different eigenvalues:

Proof. assume $\lambda_i \neq \lambda_j \in \mathbb{R}$. Then:

$$z_j^* A z_i = z_j^*(\lambda_i z_i) = \lambda_i (z_j^* z_i) \tag{45}$$
$$\text{(or)} = (A^* z_j)^* z_i = (\lambda_j z_j)^* z_i = \lambda_j z_j^* z_i \tag{46}$$
$$\rightarrow (\lambda_j - \lambda_i)(z_j^* z_i) = 0 \tag{47}$$

Since $\lambda_i \neq \lambda_j$, $z_i \perp z_j$ □

Note that this proof is similar to the proof for left and right eigenvalues. Now, we know that if A symmetric then always diagonalizable: $A = PDP^{-1}$. We have P is orthogonal, but what if we ahve mutiple eigenvalues? then not always orthonormal in this case, but we can still *choose* them to be so.
So: In general, we choose P to be orthonormal, therefore

$$P^{-1} = P^* \rightarrow A = PDP^*, \text{ diagonal } D$$

Spectral decomposition; another formulation (with these additonal assumptions):

$$A = \sum \lambda_i z_i z_i^*$$

before, we needed distinct λ_i and needed both L/R evectors. Here, we need only the eigenvector and no longer need the distinct eigenvalue assumption!

Theorem 14. order $\lambda_1 \geq \cdots \geq \lambda_n$, A symmetric. Then:

$$\max_{\|x\|=1} x^* A x = \lambda_1$$

In other words, max rayleigh quotient is the largest eigenvalue

Proof. x^*x, expand x in terms of $z_1, \dots, z_m$, which are orthonormal:

$$x = \alpha_1 z_1 + \cdots + \alpha_n z_n$$

Therefore,

$$x^*x = (\alpha_1 z_1 + \cdots + \alpha_n z_n)^*(\dots)$$

We now get n^2 terms, but most cancel out due to orthogonality. So equal to:

$$= \|\alpha_1\|^2 + \cdots + \|\alpha_n\|^2 = 1$$

Now look at $x^*Ax = x^*(A(\alpha_1 z_1 + \cdots + \alpha_n z_n))$:

$$= (\alpha_1 z_1 + \cdots + \alpha_n z_n)^* A(\alpha_1 z_1 + \cdots + \alpha_n z_n) \tag{48}$$

$$= \alpha_1^2 \lambda_! + \alpha_n^2 \lambda_n = \sum \|\alpha_i\|^2 \lambda_i \tag{49}$$

$$\leq \lambda \sum \|\alpha_i\|^2 = \lambda_1 \tag{50}$$

therefore, $\forall x$, $x^*Ax \leq \lambda_1$.
Now choose $x = z_1$

$$^*Ax = z_1^* A z_1 = \lambda_1 z_1^* z_1 = \lambda_1 \tag{51}$$

$$\to \max_{\|x\|} x^*Ax \geq \lambda_1 \tag{52}$$

$$\to \max_{\|x\|} x^*Ax = \lambda_1 \tag{53}$$

□

wed march 3

remember if Q has orthonormal columns then $\|Qv\| = \|v\|$. We previously had the theorem describing the maximum rayleighquotient. We will now present a different proof (as shown in text):

Proof. $A = UDU^*$, U orthogonal, D diagonal, let $y = U^*x \to x = Uy$

$$\max_{\|x\|=1} x^*UDU^*x \tag{54}$$

$$= \max_{\|x\|=1} y^*Dy = \max_{\|y\|=1} y^*Dy \tag{55}$$

$$= \sum_1^n \lambda_i \bar{y_i} y_i \qquad \text{if } y = (y_1, \dots, y_n)^T \tag{56}$$

$$\leq \lambda_1 \sum \bar{y_i} y_i = \lambda_1 \|y\| = \lambda_1 \tag{57}$$

Next, pick a specific vector: let $x = U^*(1, 0, \ldots, 0)^T \leftrightarrow y = (1, 0, \ldots 0)^T$

$$\rightarrow \max_{\|y\|=1} y^* Dy \geq y^* Dy = (1, 0, \ldots, 0) D (1, 0, \ldots, 0)^T = \lambda_1 \tag{58}$$

$$\rightarrow \max_{\|x\|=1} x^* Ax = \lambda_1 \tag{59}$$

□

Will now discuss three big theorems:

Theorem 15. Courant-Fischer Theorem (minimax theorem) If eigenvalues are: $\lambda_1 \geq \cdots \geq \lambda_n$:

$$\lambda_i = \min_{\dim V = n-i+1} \max_{x \in V, \|x\|=1} x^* Ax$$

In other wrods, all eigenvalues can be characterized as the min/max of rayleigh quotients.

Proof. Let $y = U^*x$, $U^*AU = D$
We need to show:

$$\lambda_i = \min_{\dim V = n-i+1} \max_{y \in V} y^* Dy$$

(proof similar to previous proof). Let V be dimension $n - i + 1$, $S_v := \{y \in V, \|y\| = 1\}$
let:

$$S_V^1 = \{y \in V \cap F, y\| = 1\} \text{ where } F = \text{span}\{e_1, \ldots, e_i\}$$

We first note that $V \cap F \neq \{\vec{0}\}$ because otherwise, $\dim(V + f) = \dim(V) + \dim(F)$, but $\dim(V) = n - i + 1$ and $\dim(F) = i$, so sums not equal.
From here, we say thatS_V^1 has vectors of the form: $y = (y_1, \ldots, y_i, 0, \ldots, 0)^T$ and $\|y\| = 1 \rightarrow \sum_{j=1}^{i} \|y\|^2 = 1$. Therefore:

$$\rightarrow y^* Dy = \sum_{j=1}^{i} \lambda_j (y_j)^2 \geq \lambda_i \sum_{j=1}^{i} \|y_j\|^2 = \lambda_i \tag{60}$$

$$\rightarrow \max_{y \in V, \|y\|=1} y^* Dy = \max_{S_v} y^* Dy \geq \max_{S_V} y^* Dy \geq \lambda_i \tag{61}$$

Since we stated this $\forall V$, we therefore have:

$$\min_{\dim V = n-i+1} \max_{y \in V, \|y\|=1} y^* Dy \; geq \lambda_i$$

now: choose $\tilde{V} = \{e_1, \ldots, d_{i-1}\}^\perp$ So if $y \in \tilde{V}$, y has 0's in position 1 through $i - 1$. This dimension is of $n - (i - 1) = n - i + 1$. Therefore:

$$y^* Dy = \sum_{i}^{n} \lambda_j \|y_j\|^2 =\leq \lambda_i \sum \|y_j\|^2 = \lambda_i$$

so:

$$\max_{\{} y \in \tilde{V}\} y^* Dy \leq \lambda_i$$

We chose a specific $\tilde{V}$, therefore:

$$\min_{y \in V} \max_{y \in V} y^* Dy \leq \lambda_i$$

□

Fri March 5

(more on courant-fischer)

$$\lambda_i = \min_{\dim V = n-i+1} \max_{x \in V} x^* A x$$

Example 19. $A = diag(1, 2, 3, 4, 5)$ so $\lambda_1 = 5, \ldots, \lambda_5 = 1$. So:

$$\lambda_4 = 2 = \min_{\dim V = 2} = \max_{x \in V} x^* A x$$

So what is V here that gives min? we only want to use smaller eigenvalues.

$$V = \text{span}\{(1, 0, \ldots, 0)^T), (0, 1, 0, \ldots 0)^T\}$$

This negates effects of evals $3, 4, 5$. Therefore:

$$\max x^* A x = 2, \text{if we let } x = (0, 1, 0, 0)^T$$

If we do this and use a random initilaization, we dont get the min that we want.

Theorem 16. Assume A symmetric. Let $B = A + E$ with all hermitian. Let:

$$A \text{ with eigenvalues } \lambda_1 \geq \cdots \geq \lambda_n$$
$$E \text{ with eigenvalues } \epsilon_1 \geq \cdots \geq \epsilon_n$$
$$B \text{ with eigenvalues } \beta_1 \geq \cdots \geq \beta_n$$

Then:

$$\lambda_i \epsilon_1 \geq \beta_i \geq \lambda_i + \epsilon_n$$

This tells us how much an eigenvalue can move if the matrix is perturbed. We will use courant-fischer in proof. Remember from BF theorem:

$$\min_{\lambda_i} \|\beta - \lambda_i\| \leq K(\rho)\|E\|$$

becomes equal to 1. Therefore, P can be chosen to be orthogonal and thus:

$$K(P) = \|P\|\|P^{-1}\| = \|P\|\|P^*\| = \max \|Px\| = \max \|x\| = 1$$

Last step due to P being orthonormal and x being a unit vector. Then:

$$\to \cdots \leq K(\rho)\|E\| = 1 * \|E\| = \max \epsilon_i \tag{62}$$
$$\to \lambda_i - \epsilon_1 \leq \beta_i \leq \lambda_i + \epsilon_1 \tag{63}$$

Final theorem is also similar.

Theorem 17. Cauchy-Interlace Theorem: Let A hermitian with evalues $\lambda_1 \geq \cdots \geq \lambda_n$ and B of form:

$$B = \begin{pmatrix} A & c \\ c^* & \alpha \end{pmatrix}$$

for any $\vec{c}, \alpha$. Then:

$$\beta_1 \geq \lambda_1 \geq \beta_2 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq \beta_{n-1}$$

Note that B maintains $Hermitian$ property that A has. So what happens if we choose an outrageous c? what does theorem tell us? Our $\beta_1, \ldots, \beta_{n+1}$ likely makes the range very large, but we still maintain the interweaving pattern.
Claim: For symmetric matrix, then the approximate evalues are "extra" accurate assuming our vector is fairy accurate.
note: (B-O result): O means bounded by something, we can control size:

$$\rho = \lambda_1 + O(\epsilon) \to \|\rho - \lambda_1\| \leq \epsilon \text{ for small} \epsilon$$

Result: $y = z + O(\epsilon)$, then $\rho =]\lambda_1 + O(\epsilon^2)$

Proof. (pseudoproof) $y = z_1 + \alpha_2 z_2 + \cdots + \alpha_n z_n$, z_i orthonormal. then:

$$\begin{aligned} Ay =& \lambda_1 z_1 + \alpha_2 \lambda_2 z_2 + \cdots + \alpha_n \lambda_n z_n \\ & \to y^* Ay = \lambda_1 + \alpha_2^2 \lambda_2 + \cdots + \alpha_n^2 \lambda_n \\ & = \lambda_1 + O(\epsilon^2) \qquad\qquad (\text{bc } \alpha_i = O(\epsilon)) \end{aligned}$$

This is only possible due to symmetry. Also, $y^* y = 1 + \alpha_2^2 + \cdots + \alpha_n^2 = 1 + O(\epsilon^2)$

$$\begin{aligned} \to \rho = \frac{y^* Ay}{y^* y} &= \frac{\lambda_1 + O(\epsilon^2)}{1 + O(\epsilon^2)} && (64) \\ &= (\lambda_1 + O(\epsilon^2))(1 - O(\epsilon^2)) && (65) \\ &= \lambda_1 + O(\epsilon^2) + \ldots && (66) \\ &= \lambda_1 + O(\epsilon^2) && (67) \end{aligned}$$

Note second equality is due to geometric series: $1/(1 + O(\epsilon^2)) = 1 - O(\epsilon^2) + O(\epsilon^4) + \ldots$ □

Monday 3-16

We are now moving back to chapter 7.4 and discussing differential equations:

- derivatives
- finite different methods (bounded value problems)
- finite element method (bounded value problems)
- (...) (initial value problems)

Systems of Differential Equations

- If you have several diffeq, can write as matrix:

$$u' = Au, u = (u_1, u_2, \ldots)^T, u(0) = \vec{c} \in \mathbb{R}^n (\text{ this is our initial condition})$$

- time derivative: du/dt

Assume A diagonalizable. Claim: solution is of form $u = e^{At}c$

Proof.

$$\begin{aligned}
&A = PDP^{-1}, D = diag(\lambda_1, \dots) \\
&\to e^{At} = Pe^{Dt}P^{-1} = Pdiag(e^{\lambda_1 t}, \dots, e^{\lambda_n t})P^{-1} \\
&\to \frac{d}{dt}(e^{At}) = Pdiag(\lambda_1 e^{\lambda_1 t}, \dots, \lambda_n e^{\lambda_n t})P^{-1} \\
&= PDe^{Dt}P^{-1} = PDP^{-1}Pe^{Dt}P^{-1} = Ae^{At} \\
&\to \frac{d}{dt}(e^{At}) = Ae^{At} \\
&\to \text{ if } u = e^{At}c, u' = Ae^{At}c = Au
\end{aligned}$$

Also, $u(0) = e^0 c = c$. (look in book for uniqueness) □

stability: some diffeq have more stable solutions:
Systems are stable if all eigenvalues are in left-hand side of complex plane

- If solution decays, error from numerical methods decays
- inversely, grows if solution grows

Trefethen (1990s) said just because all eigenvalues in LH, not guaranteed to always be stable.

- could have transitional instability
- how? consider pseudoeigenvalues: eigenvalues of slightly perturned matrix. These slightly extend into RH plane. This can happen when evectors are not perfectly orthogonal. rbations of matrix mean small perturbations of evals from our theorems, our perturbations mean it is controlled.
- so when A is nonsymmetric, we get instability

Example 20. solve:

$$u' = Au \qquad A = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \tag{68}$$

$$u(0) = c \qquad c = (10, 10)^T \tag{69}$$

remember solution is of form $u = e^{At}c$.

- first diagonalize A

- second find evals/evects

(skipping ahead), get evals of $\lambda_1 = 0, \lambda_2 = 5$ with corresponding evectors $(2/\sqrt{5}, 1/\sqrt{5}), (1/\sqrt{5}, -2/\sqrt{5})$. Now write spectral decomposition: $A = \sum \lambda_i z_i z_i^T$:

$$\begin{aligned}\vec{u} = e^{At}c &= \sum e^{\lambda_i t} z_i z_i^T c \\ &\dots \\ &= (12, 6)^T + e^{5t}(-2, 4)^T\end{aligned}$$

Next, how do we calculate a derivative on the computer? suppose we want $f'(a)$:

Definition 33. forward difference formula:

$$f'(a) \approx \frac{f(a+h) - f(a)}{h}$$

We can use taylor expansion to find a better approximation.

- can see our approximation is much worse than the actual
- how can we get better approximation by using more points?
- taylor expansion tells us how good our approx is

Definition 34. Taylor Expansion:

$$f(a+h) = f(a) + f'(a)h + f''(a)h^2/2! + \dots$$

Wed 3/18

Using forward difference, our (leading) error term is $o(h)$; we want to get a better formula. Can do this by using more points @ $a - h$

Definition 35. backward diff: $f'(a) \approx \frac{f(a)-f(a-h)}{h}$ **central diff**: $f'(a) \approx \frac{f(a+h)-f(a-h)}{2h}$

Note the central is simply the average of forward and backward. Central diff is generally better to estimate actual $f'(a)$ Derive:

Proof. $f(a-h) = f(a) - f'(a)h + f''(a)h^2/2 - f'''(a)h^3/3! + \dots$
therfore:

$$\begin{aligned}f(a+h) - f(a) &= \text{this eliminates h2 leading term, so h3 becomes new leading error term} \\ &= (f(a) + f'(a)h + f''(a)h^2/2 + \dots) - (f(a) - f'(a)h + f''(a)h^2/2 - \dots) \\ &= f'(a)2h + f''(a)2h^3/3!\end{aligned}$$

Therefore by rearranging:

$$f'(a) = 1/2h(f(a+h) - f(a-h)) - f'''(a)h^2/3!$$

Leading error term now $o(h^2)$ □

Example 21. $f(x) = e^x \sin x$, approximate $f'(1) \approx 3.7560492$. Let $h = 0.01$.

- forward: 3.7707
- central: 3.7560219

Now for boundary value problems: will do 1D and 2D

Example 22.

$$\begin{aligned} -u'' &= 4 \\ u'(0) &= 3 \\ u(1) &= 0 \end{aligned}$$

- only care between 0 and 1
- also need h to determine how to break up interval, choose $h = 1/2$ so have 3 points
- solving for u @ $u_0 = 0, u_1 = 1/2, u_1 = 1$
- can interpolate between to find values inbetween
- already know $u_2 = 0$ due to b.c.
- to solve, use diff @ all interior points, and use b.c. at all end points

at $x = 1/2, -u'' = 4$ so:

$$\text{(use diff formula)}, -(\frac{u_2 - 2u_1 + u_0}{(1/2)^2}) = 4$$

equivalently, $2u_1 - u_0 = 1$
Using bc for $x = 0$, $u'(0) = 3 \rightarrow \frac{u_1 - u_0}{1/2} = 3$. Use forward diff to stay within boundary. Can then take both above results and place them in matrix:

$$\begin{pmatrix} -1 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 3/2 \end{pmatrix}$$

Solve to get $u_1 = -1/2$ and $u_0 = -2$

4 Fri 03/20

review of finite different method:

Example 23. steady state heat equation:

$$\begin{aligned} &u_{yy} + u_{xx} = 0 \\ &u(x,0) = u(0,y) = u(1,y) = 0 \\ &u(x,1) = 1000 \end{aligned}$$

Let $h = 1/3$, unknowns are interior points:

$$\begin{aligned} u_1 &= u(1/3, 1/3) \\ u_2 &= u(2/3, 1/3) \\ u_3 &= u(1/3, 2/3) \\ u_4 &= u(2/3, 2/3) \end{aligned}$$

at each point, we can apply central different: $f''(a) \approx \frac{f(a+h)-2f(a)+f(a-h)}{h^2}$
Around point at $(1/3, 1/3)$:

$$\begin{aligned} u_{xx} + u_{yy} &= 0 \\ &\to \frac{u_2 - 2u_1 + (0)}{h^2} + \frac{u_3 - 2u_1 + (0)}{h^2} = 0 \\ &\to -4u_1 + u_2 + u_3 = 0 \end{aligned}$$

Around point at $(2/3, 1/3)$:

$$\begin{aligned} &\to \frac{0 - 2u_2 + u_1}{h^2} + \frac{u_4 - 2u_2 + (0)}{h^2} = 0 \\ &\to u_1 - 4u_2 + u_4 = 0 \end{aligned}$$

Around point $(1/3, 2/3)$:

$$\begin{aligned} &\to \frac{1000 - 2u_3 + u_1}{h^2} + \frac{u_4 - 2u_3 + (0)}{h^2} = 0 \\ &\to u_1 - 4u_3 + u_4 = -1000 \end{aligned}$$

Around point $(2/3, 2/3)$:

$$\begin{aligned} &\to \frac{1000 - 2u_4 + u_2}{h^2} + \frac{-2u_4 + u_3}{h^2} = 0 \\ &\to u_2 + u_3 - 4u_4 = 1000 \end{aligned}$$

This system of equations is:

$$\begin{pmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1000 \\ 1000 \end{pmatrix} \tag{70}$$

Resulting in:

$$u_1 = u_2 = 125$$
$$u_3 = u_4 = 375$$

Comment: problem gets more complex if not steady state: differential equation can become: $u_t = c(u_{xx} + u_{yy})$

Now finite element method:

- very different from finite difference method
- will only deal with 1-D to understand idea

Example 24.

$$-u'' = 4$$
$$u'(0) = 3$$
$$u(1) = 0$$
$$h = 1, x \in [0, 1]$$

Will not need to use the derivative formulas. Instead:

$$-u'' = 4 \rightarrow -u''v = 4v$$
$$\Longrightarrow -\int_0^1 u''v = 4\int_0^1 v$$

Will do integration by parts but first we discuss our "test function" $v()$; we will use piece-wise linear. The ϕ_i (figure 4) will be used as test functions, pick as form of u to be of form:

$$u \approx \tilde{u} = \alpha_0\phi_0 + \alpha_1\phi_1 + \alpha_2\phi_2$$

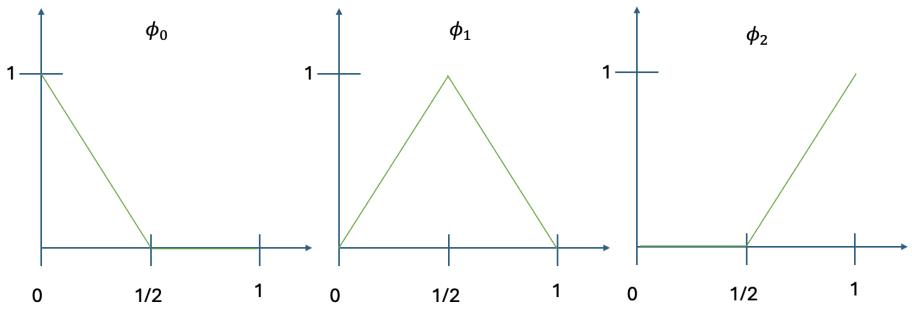

Figure 4: Test Functions

We do integration by parts to get rid of the 2nd derivative (due to discontinuities): let $g = u'$, $f = v$ then:

$$\int_0^1 vu'' = [vu']_0^1 - \int_0^1 v'u'$$
$$\updownarrow$$
$$-\int u''v = -[vu']_0^1 + \int_0^1 v'u' = 4\int_0^1 v$$

Can now substitute and use boundary conditions:

$$0 = u(1) = \alpha_0\phi_0(1) + \alpha_1\phi(1) + \alpha_2\phi(1) = \alpha_2$$

Also:

$$-u'(1)v(1) + u'(0)v(0) + \int_0^1 v'u' = 4\int_0^1 v$$

Now we know $\alpha_2 = 0$, so only need ϕ_0, ϕ_1:

$$u = \alpha_0\phi_0 + \alpha_1\phi_1$$
$$v = \phi_0, \phi_1$$

Plug in 1 by 1 and complete unknowns: Let $v = \phi_0$:

$$-u'(1)\phi_0(1) + u'(0)\phi_1(0) + \int_0^1 (\alpha_0\phi_0 + \alpha_1\phi_1)'\phi_0' = 4\int_0^1 \phi_0$$

dont know $u'(1)$ but not important since $\phi_0(1) = 0$, and use boundary condition $u'(0) = 3$, so:

$$3(1) + \alpha_0\int_0^{1/2}(-2)(-2) + \alpha_1\int_0^{1/2}(2)(-2) = 4(1/4)$$

This gives us $\alpha_0 - \alpha_1 = -1$, and repeat for $v = \phi_1$.

Comments

- we can increase complexity by instead, raising ϕ_i to quadratic functions (but still piecewise)
- how about 2D? triangles with tetrahedrals

5 Monday Mar 23-2026

last time, we did an example. Continuing for $v = \phi_1$, we get:

$$-u'(1)\phi_1(1) + u'(0)\phi_1(0) + \int_0^1 (\alpha_0\phi_0 + \alpha_1\phi_1)'\phi_1' = 4\int_0^1 \phi_1$$
$$\leftrightarrow \alpha_0\int_0^1 \alpha_0'\phi_1' + \alpha_1\int_0^1 \phi_1'\phi_1' = 4\int_0^1 \phi_1$$
$$= \alpha_0\int_0^{1/2}(-2)(2) + \alpha_1(\int_0^{1/2}(2)(2) + \int_{1/2}^1(-2)(-2)) = 4(1/2)$$
$$= -2\alpha_0 + 4\alpha_1 = 2 \leftrightarrow -\alpha_0 + 2\alpha_1 = 1$$

We then have system of equations:

$$\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \tag{71}$$

So we get $\alpha_1 = 0, \alpha_0 = 1$ and a final estimate of:

$$\tilde{u} = -\alpha_0$$

comparing to real solution of $u(x) = -2x^2 + 3x - 1$:

$$\text{at } 0 : u(0) = -1$$
$$\text{at } 1/2 : u(1/2) = 0$$

We get an exact match at our three node points here. Now we look at dirac delta functions and usage in FEM functions.

Definition 36. dirac delta function $\int \delta(x) = 1$, $\delta(x) = 0, x \neq 0$

Backing up: we want to find the derivative of the step function:

$$f(x) = \begin{cases} 1 & x \geq 1/2 \\ 0 & x < 1/2 \end{cases} \tag{72}$$

traditionally, cant take derivative due to discontinuity, but if we get really close (within d of $1/2$):

$$f(x) \approx g_d(x) = \begin{cases} 0 & x < 1/2 - d \\ 1 & x > 1/2 + d \\ \frac{1}{2d}(x - 1/2) + 1/2 & x \in [1/2 - d, 1/2 + d] \end{cases} \tag{73}$$

then:

$$\frac{d}{dx} g_d(x) = \begin{cases} 0 & x < 1/2 - d \\ \frac{1}{2d} & x \in [1/2 - d, 1/2 + d] \\ 0 & x > 1/2 + d \end{cases}$$

If we take $d \to \infty$m then $g_d(x) \to f(x)$, so we define:

$$f'(x) = \lim_{d \to 0} g'_d(x)$$

then two properties:

- $f(1/2) = \infty$
- $\int f(x) = 1$

We can generalize this:

$$\delta(x - a) = \begin{cases} \infty & x = a \\ 0 & \text{else} \end{cases}$$
$$\int \delta(x - a) = 1$$

Also have **sifting property**:

$$\int_{-\infty}^{\infty} h(x)\delta(x-a)dx = h(a)$$

We now do FEM where delta function is used:

Example 25.

$$\begin{aligned} -u'' &= 2\delta(x-1/4) \\ u'(0) &= 3 \\ u(1) &= 0 \end{aligned}$$

will need to assume u is continuous. What if discontinuous? then u' gives a $\delta()$, but we dont have $u'' = \delta'(x)$. therefore, we need to assume u is continuous so that u'' will exist. Let's look at the exact solution:

$$u'' = -2\delta(x-1/4) \rightarrow u' = \begin{cases} c+0 & x < 1/4 \\ c-2 & x > 1/4 \end{cases}$$

$$\rightarrow u = \begin{cases} 3x + c_2 & x < 1/4 \\ x + c_3 & x > 1/4 \end{cases} \rightarrow u(1) = 0, c_3 = -1$$

Since u continuous we have $3/4 + c_2 = 1/4 - 1 \rightarrow c_2 = -3/2$ and thus:

$$u = \begin{cases} 3x - 3/2 & x < 1/4 \\ x - 1 & x > 1/4 \end{cases} \tag{74}$$

6 wed 3/25

last time, did example with delta functions. (see handout) Now, looking into initial value problems (IVT) or time-dependent problems. have two types: ODE and PDE. We have already done some PDE with time-dependence. General form:

$$\frac{du}{dt} = f(u,t) \tag{75}$$

$$u(0) = v_0 \tag{76}$$

$$t \in [0,T], k = \text{ some step size} \tag{77}$$

Definition 37. Euler's Method:

1. initial slope is: $f(v_0, 0_)$
2. follow slope to next value: (k, v_1)
3. repeat and continue for $v_2, v_3, \ldots$

our slope is: $(v_1 - v_0)/k = f(u_0, 0)$. therefore, $v_1 = v_0 + kf(u_0, 0)$. In general, we have:

$$v_{j+1} = v_j + kf(u_j, t_k) \tag{78}$$

Example 26.

$$\frac{du}{dt} = u \tag{79}$$
$$u(0) = 1 \tag{80}$$
$$t \in [0, 2], k = 0.2 \tag{81}$$

note the diffeq means we have exponential growth. Walking through a few steps:

$$\begin{aligned} v_0 &= 1 \\ v_1 &= v_0 + kf(v_0, t_0) \\ &= 1 + 1/5(1) = 6/5 \\ v_2 &= v_1 + kf(v_1, t_1) \\ &= 6/5 + 1/5(6/5) = 1.44 = (1.2)^2 \\ \rightarrow v_j &= (1.2)^j \end{aligned}$$

Note that after 10 steps, $v_10 \approx 6.1917$, but true solution is $e^2 = 7.39$; these are not close. can either use a smaller k but ideally get find a better method.

Definition 38. Implicit method/backward euler:

$$v_{j+1} = v_j + k \cdot f(v_{j+1}, t_{j+1}) \tag{82}$$

This is "implicit" becuase v_{j+1} used before we know it; use it to solve itself. For PDEs, we can solve because we have both time and postiion. The following is an implementable version:

Definition 39. Predictor corrector version of Backward Euler's

$$\text{Predictor} : \tilde{v}_{j+1} = v_j + kf(v_j, t_j) \tag{83}$$
$$\text{Corrector} : v_{j+1} = v_j + kf(\tilde{v}_{j+1}, t_{j+1}) \tag{84}$$

This turns out to have the same accuracy. It instead overshoots the solution, so to improve we try a midpoint rule:

Definition 40. Midpoint Rule:

$$v_{j+1/2} = v_j + h/2f(v_j, t_j) \tag{85}$$
$$v_{j+1} = v_j + hf(v_{j+1/2}, t_{j+1/2}) \tag{86}$$

7 Fri 26/03/27

Last time we discussed midpoint rule. we first use slope at initial timestep to get a midpoint, then recalculate slope.

Example 27.

$$\begin{aligned} \frac{du}{dt} &= u = f(u_j, t_j) \\ u(0) = 1, k &= 0.2 \\ \rightarrow v_{1/2} &= v_0 + .2/2 f(1,0) \\ &= 1 + 1/10(1) = 1.1 \\ \rightarrow v_1 &= v_0 + .2 f(1.1, 0.1) \\ &= 1 + 1/5(1.1) = 1.22 \end{aligned}$$

Compared with true solution: $e^{0.2} = 1.2214$. With eulers, 1.2 and backwards eulers 1.24.

For trapezoid, we will use points at both beginning and end and average their slopes. compared to midpoint, which only uses one slope at middle. The leading errors are about the same. The trapezoid rule is *implicit* method:

Definition 41. Trapezoid Rule:

$$v_{j+1} = v_j + \frac{k}{2}(f(v_j, t_j) + f(v_{j+1}, t_{j+1})) \tag{87}$$

Here is a hybrid of eulers and trapezoid: another predictor-corrector method:

Definition 42. Heun's Method:

$$\tilde{v}_{j+1} = v_j + k f(v_j, t_j) \tag{88}$$

$$v_{j+1} = v_j + \frac{k}{2}\left(f(v_j, t_j) + f(\tilde{v}_{j+1}, t_{j+1})\right) \tag{89}$$

Now discuss accuracy and stability. We have two choices: primarily runge-kutta methods

Definition 43. Runge-Kutta Methods: Three slope estimates

$$s_1 = f(v_j, t_j) \tag{90}$$

$$s_2 = f(v_j + k/2 s_1, t_j + k/2) \tag{91}$$

$$s_3 = f(v_j + 3/4 k s_1, t_j + 3/4 k) \tag{92}$$

$$v_{j+1} = v_j + k(3/9 s_1 + 1/3 s_2 + 4/9 s_3) \tag{93}$$

For stability, let's first look at eulers for example: let $\frac{du}{dt} = \lambda u$, $\lambda < 0$, so when does $v_j \rightarrow 0$? Eulers becomes:

$$v_{j+1} = v_j + k f(v_j, t_j) \tag{94}$$

$$= v_j + k\lambda v_j \tag{95}$$

$$v_{j+1} = (1 + k\lambda) v_j \tag{96}$$

So we need $(1 + k\lambda)^n \rightarrow 0$, or in otherwords $|1 + k\lambda| < 1 \leftrightarrow -1 < 1 + k\lambda < 1$, so for stability, we need:

$$k\lambda > -1 \rightarrow k < -2/\lambda \tag{97}$$

Example 28. $\frac{du}{dt} = -10u$, $u(0) = 1$, $k < \frac{-2}{-10} = \frac{1}{5}$

Stability for backward eulers as implicit:

$$v_{j+1} = v_j + kf(v_{j+1}, t_{j+1}) \tag{98}$$
$$v_{j+1} = v_j + k\lambda v_{j+1} \tag{99}$$
$$(1 - k\lambda)v_{j+1} = v_j \tag{100}$$
$$v_{j+1} = \frac{1}{1 - k\lambda} v_j \tag{101}$$
$$\text{need } |\frac{1}{1 - k\lambda}| < 1 \tag{102}$$
$$\to -1 < \frac{1}{1 - k\lambda} < 1 \tag{103}$$

$k\lambda < 0$ since $\lambda < 0$, so always true/convegent, therefore backward euler is always convergent

8 Wed 2026/03/30

last time, showed stability analysis for backward eulers, now we show for backward euler as a predictor corrector:

$$\tilde{v}_{j+1} = v_j + kf(v_j, t_j) \tag{104}$$
$$v_{j+1} = v_j + kf(\tilde{v}_{j+1}, t_{j+1}) \tag{105}$$

Assume $du/dt = \lambda u$, $\lambda < 0$:

$$\tilde{v}_{j+1} = v_j + k\lambda v_j = (1 + k\lambda)v_j \tag{106}$$
$$v_{j+1} = v_j + k\lambda \tilde{v}_{j+1} = v_j + k\lambda(1_k\lambda)v_j \tag{107}$$
$$= (1 + k\lambda + (k\lambda)^2)v_j \tag{108}$$
$$\text{want } |1 + k\lambda + (k\lambda)^2| < 1 \text{ for convergence} \tag{109}$$
$$\tag{110}$$

Let $k\lambda = x$, so want $g(x) = 1 + x + x^2 < 1$. this is between $[-1, 0]$. For $-1 < x$, we therefore need $k < -1/\lambda$ for stability condition. Note there is a separation from accuracy; stability is a **minimal** condition.

9 Wed 26/04/01

PDEs: three general types

- parabolic
- elliptical

- hyperbolic

Of these three, parabolic and hyperbolic are time dependent. Think of these classification as the same with funcitons. With functions, they have the form:

- parabolic: $y = A(x - x_0)^2 + b$
- elliptic: $(x - a)^2/A + (y - b)^2/B = c$
- hyperbolic: $(x - a)^2/A - (y - b)^2/B = c$

When it comes to classifying PDEs, the formulas have a similar shape, except instead of exponents, these represent the level of the derivative.

- Parabolic: $\frac{du}{dt} = \alpha^2 \frac{d^2u}{dx^2}$
- Elliptic: $\frac{d^2u}{dx^2} + \frac{d^2u}{dy^2} = f(x)$
- Hyperbolic: $\frac{d^2u}{dx^2} = \frac{d^2u}{dt^2} + C$

Of these, parabolic and hyperbolic both contain time dependencies. Elliptical PDEs do not. Parabolic PDEs are also classified as *heat equations* while hyperbolic are *wave equations*. We first go into the heat (or diffusion equations). We will only be discussing the computational solution.

Example 29.

$$\frac{du}{dt} = \alpha^2 \frac{d^2u}{dx^2} \tag{111}$$
$$u(x, 0) = f(x) \tag{112}$$
$$u(0, t) = 0 \tag{113}$$
$$u(l, t) = 0 \tag{114}$$

The last two equations are the initial boundary conditions. In this case, since the ends are insulated, we know that as time goes to infinite, we would expect our solution to go to 0. Usually, α^2 is dependent on the material. To solve, we'll use finite difference: need to define intervals in both x (h) and t (k) dimensions.
Given values at $f(0)$, $f(h)$, $f(2h)$, we're looking for values at k, $2k$, $3k$, ...

- So why is du/dt function of spatial second derivative?... something to think about...

Let $w_{i,j}$ be approximate value of u at position i, time j, then at (i, j), we have:

$$\frac{du}{dt} = \alpha^2 \frac{d^2u}{dx^2} = \alpha^2(\dots) \tag{115}$$

Replacing using forward difference and $w_{i,j}$ notation, we get:

$$\frac{w_{i,j+1} - w_{i,j}}{k} = \alpha^2 \frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{2h} \tag{116}$$

note that the right hand side we acutally use central difference. This can be rearranged to get:

$$w_{i,j+1} = w_{i,j} + k\alpha^2/h^2(w_{i+1,j} - 2w_{i,j} + w_{i-1,j}) \tag{117}$$

Let $\lambda = \frac{k\alpha^2}{h^2}$, then $wi, j+1 = (1-2\lambda)w_{i,j} + \lambda w_{i+1,j} + \lambda w + i - 1, j$ This can be better visualized in matrix form. If:

$$\vec{w^{(0)}} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \dots \\ f(x_{m-1}) \end{pmatrix} \tag{118}$$

Subsequent time steps can be written as:

$$\left(\vec{w^{(j)}} = \begin{pmatrix} w_{1,j} \\ \dots \\ w_{(m-1,j)} \end{pmatrix} \right) \tag{119}$$

Then:

$$\vec{w^{(j+1)}} = Aw^{(j)} \tag{120}$$

where matrix A is of form:

$$A = \begin{pmatrix} 1-2\lambda & \lambda & 0 & \dots & 0 \\ \lambda & 1-2\lambda & \lambda & \dots & 0 \\ \dots & & & & \end{pmatrix} \tag{121}$$

This is a tridiagonal matrix

We now discuss the stability of this solution. We know we want our solution to eventually decay to 0, so if we get an error it goes away too. Let the initial error be:

$$e^{(0)} = \begin{pmatrix} e_1^{(0)} \\ \dots \\ e_m^{(0)} \end{pmatrix} \tag{122}$$

For one time step:

$$w^{(1)} = A(w^{(0)} + e^{(0)}) = Aw^{(0)} + Ae^{(0)} \tag{123}$$

We want:

$$(A)^n e^{(0)} \to 0 \tag{124}$$

or alternatively, all eigenvalues must have magnitudes less than 1: $|\mu_i| < 1$ For tridiagonal matrices of the form in the above example, the eigenvalues are:

$$\mu_i = 1 - 4\lambda(\sin(i\pi/2m))^2 \tag{125}$$

So all we need is:

$$0 \leq \lambda(\sin(i\pi/2m))^2 \leq 1/2 \tag{126}$$

In other words, we need:

$$k \leq \frac{h^2}{2\alpha^2} \tag{127}$$

This means that we need very small time steps; this is why we generally dont want to use this method, very computationally expensive.

10 Wed 4/8/2020

List of previously discussed PDE classes:

- Elliptic

  example is 2D steady state heat equation:

$$\frac{d^2u}{dx^2} + \frac{d^2u}{dy^2} = 0$$

- Parabolic

  ex is 2D heat/diffusion equation

$$\frac{du}{dt} = k(\frac{d^2u}{dx^2} + \frac{d^2u}{dy^2})$$

- Hyperbolic

  ex 1D wave equation

$$\frac{d^2u}{dt^2} = k\frac{d^2u}{dx^2}$$

For 1D heat equation, we can solve using forward difference method:

$$\frac{w_{i,j+1} - w_{i,j}}{k} = \alpha^2(\frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2}) \tag{128}$$

We showed earlier that this can be rewritten similar to power method:

$$w^{j+1} = Aw^j$$

We showed that for stability/convergence, we need all eigenvalues of A to be less than 1 in magnitude, or $k \leq \frac{h^2}{2\alpha}$. this is however not accurate and not always stable.
Instead, we solve using backward difference equation; this makes it more stable (but not more accurate):

$$A = \begin{pmatrix} 1+2\lambda & -\lambda & \dots & 0 \\ -\lambda & 1+2\lambda & -\lambda * \dots & 0 \\ \dots & & & \end{pmatrix} \tag{129}$$

The eigenvalues for this matrix have been shown to be:

$$u_i = 1 + 4\lambda(\sin(\frac{i\pi}{2m})^2) \tag{130}$$

In this case, we have all eigenvalues $u_i > 1$. this is called **implicit method**; so how is this more stable? $Aw^{(j)} = w^{(j-1)}$ instead of $w^{(j+1)} = Aw^{(j)}$. For backwards, we have to solve system of linear equations. Since we know eigenvalues always greater than 1, then eval of $A^{-1} < 1$ and positive, so always stable. The time derivative component technically makes it less accurate, so instead of backward difference, we can just use central difference. However, this will result in less stability.

11 Sat April 11

Had problem with forward difference being unstable and inaccurate. We used backward diff to make it more stable but still inaccurate in time. Now we implemented central difference in time:

Definition 44. Richardson's Method: for heat equation $\frac{du}{dt} = \alpha^2 \frac{d^2u}{dx^2}$

$$\frac{w_{i,j+1} - w_{i,j-1}}{2k} = \text{(same as previously shown) } \alpha^2(\frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2}) \tag{131}$$

This gives us in matrix form:

$$w^{j+1} = \begin{pmatrix} -4\lambda & 2\lambda & 0 & \dots \\ 2\lambda & -4\lambda & 2\lambda & \dots \\ \dots & & & \end{pmatrix} w^j + w^{j-1} \tag{132}$$

where $\lambda = \frac{\alpha^2 k}{h^2}$. for the above: $\mu_i > 1$ if $\lambda > 1/8$; before, we have stability for $\lambda > 1/2$, so this is more unstable.

We now introduce crank-nicolson, the golden standard. this uses 6 total points:

Definition 45. Crank-Nicolson:

$$\frac{w_{i,j+1} - w_{i,j}}{k} = \alpha^2/2(\frac{2_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2} + \frac{w_{i+1,j+1} - 2w_{i,j+1} + w_{i-1,j+1}}{h^2}) \tag{133}$$

This gives matrix form:

$$Aw^{(j+1)} = Bw^{(j)} \to w^{(j+1)} = A^{-1}Bw^{(j)} \tag{134}$$

where A is:

$$\begin{pmatrix} 1+\lambda & \lambda/2 & \dots & 0 \\ \lambda/2 & 1+\lambda & \lambda/2 & \dots \\ \dots & & & \end{pmatrix}$$

For this A definition since it is once again a tridiagonal matrix, we know that A is always stable.

Note that this is similar to doing an average of two central differences. Now we move onto wave equation:

Definition 46. Wave equation

$$\frac{d^2u}{dt^2} = c^2 \frac{d^2u}{dx^2}$$

Think of a vibrating string. If we have a higher d^2u/dx^2, then we have more concavity. We have three classes of wave equations:

- $\frac{d^2u}{dt^2} = c^2 \frac{d^2u}{dx^2}$
- $\frac{du}{dt} = a\frac{du}{dx}$
- (kDV equation, nonlinear): $\frac{du}{dt} + \frac{d^3u}{dx^3} - 6u\frac{du}{dx} = 0$

Will now introduce analytical solution for first problem:

Definition 47. 1D wave equation:

$$\frac{du}{dt} + a\frac{du}{dx} = 0$$

initial equation:

$$u(x, 0) = q(x)$$

unbounded $x \in (-\infty, \infty)$ and $t \geq 0$

We will use *change of variables* method: Let $\tau = t$, $\xi = x - at$. This makes it look like we're "following teh wave". Therefore:

$$u(\xi(x,t), \tau(x,t)) = u(x,t) \tag{135}$$

We use chain rule to get:

$$\frac{du}{dt} = \frac{du}{d\xi}\frac{d\xi}{dt} + \frac{du}{d\tau}\frac{d\tau}{dt} \tag{136}$$

$$= \frac{du}{d\xi}(-a) + \frac{du}{d\tau}(1) \tag{137}$$

$$\frac{du}{dx} = \frac{du}{d\xi}\frac{d\xi}{dx} + \frac{du}{d\tau}\frac{d\tau}{dx} \tag{138}$$

$$= \frac{du}{d\xi}(1) + \frac{du}{d\tau}(0) \tag{139}$$

Plugging these both these in to our original PDE, we get:

$$\frac{du}{d\tau} = 0 \rightarrow u = f(\xi) = f(x - at)$$

Using our inial conditions:

$$u(x, 0) = q(x) = f(x), t = 0$$

Therefore, our final solution is:

$$u(x,t) = q(x - at) \tag{140}$$

So what happens with an initial $q(x)$? In this sitation, it maintains the shape and simply moves as time progresses with velocity a.

12 monday april 13

Remember, last time we discussed three classes of wave equations and introduced analytical solution to the simple equation. it's not always possible to solve a wave equation in general because we need specific b.c. and conditions.

Definition 48. d'Alembert Solution for standard wave equation:

$$\frac{d^2u}{dt^2} = c^2\frac{d^2u}{dx^2} \tag{141}$$

$$-\infty < x < \infty \tag{142}$$

$$u(x,0) = g(x) \tag{143}$$

$$u_t(x,0) = h(x) \tag{144}$$

Let $\mu = x+ct$ and $\xi = x-ct$. This is motivated by the fact that it has been observed that after the initial time, two waves form and move in opposite directions. Substituting and using chain rule, we get:

$$\frac{du}{dt} = \frac{du}{d\mu}\frac{d\mu}{dt} + \frac{du}{d\xi}\frac{d\xi}{dt} = c\frac{du}{d\mu} - c\frac{du}{d\xi} \tag{145}$$

$$\frac{d^2u}{dt^2} = c\frac{d^2u}{d\mu^2}\frac{d\mu}{dt} + c\frac{d^2u}{d\mu d\xi}d\xi dt - c[\frac{d^2u}{d\xi d\mu}\frac{du}{dt} + \frac{d^2u}{d\xi^2}\frac{d\xi}{dt}] \tag{146}$$

We assum that: $\frac{d^2u}{d\mu d\xi} = \frac{d^2u}{d\xi d\mu}$, so the second derivative can be simplified to:

$$\frac{d^2u}{dt^2} = c^2\frac{d^2u}{d\mu^2} - 2c^2\frac{d^2u}{d\mu d\xi} + c^2\frac{d^2u}{d\xi^2}$$

We plug these two into our PDE ; this reduces down to:

$$\frac{d^2u}{d\xi d\mu} = 0$$

Inegrating first by xi then mu gives:

$$\frac{du}{d\mu} = f(\mu) \to u = F(\mu) + G(\xi) = F(x+ct) + G(x-ct)$$

Now we introduce our initial conditions:

$$g(x) = u(x,0) = F(x) + G(x) \tag{147}$$

differentiating:

$$\frac{du}{dt} = \frac{d}{dt}(F+G) = cF'(x+ct) - cG'(x-ct) \tag{148}$$

$$h(x) = u_t(x,0)- = cF'(x) - cG'(x) \tag{149}$$

We integrate again to get:

$$cF(x) - cG(x) = \int_{-\infty}^{x} h(\xi)d\xi + C_1 \tag{150}$$

Combining equations 147 (multiplied by C) and 150:

$$c(F(x)+G(x)) + cF(x) - cG(x) = 2cF(x) \tag{151}$$

$$\to F(x) = \frac{1}{2c}g(x) + \frac{1}{2c}\int_{-\infty}^{x} h(\xi) + C \tag{152}$$

$$\to G(x) = \frac{1}{2c}g(x) + \frac{1}{2c}\int_{-\infty}^{x} h(\xi) + C \tag{153}$$

$$\to u = \frac{1}{2c}g(x+ct) + \frac{1}{2c}\int_{-\infty}^{x+ct} h(\xi) + C \tag{154}$$

$$+ \frac{1}{2c}g(x-ct) + \frac{1}{2c}\int_{-\infty}^{x-ct} h(\xi) + C \tag{155}$$

So our final solution is of form:

$$u(x,t) = \frac{1}{2c}(g(x+ct) + g(x-ct)) + \frac{1}{2c}\int_{x-ct}^{x+ct} h(\xi)d\xi \tag{156}$$

So what happens if we have given an initial wave? how about as time passes?

13 Wed April 15

Last time, discussed d'alembert solution to the equation:

$$\frac{d^2u}{dt^2} = \alpha^2 \frac{d^2u}{dx^2}$$

This has initial conditions:

$$u(0,t) = u(l,t) = 0, \forall t > 0 \tag{157}$$
$$u(x,0) = f(x) \forall x \in [0,l] \tag{158}$$
$$\frac{du}{dt}(x,0) = g(x) \forall x \in [0,l] \tag{159}$$

Few comments:

- has inifinite spatial domain
- $u_t(x,0)$ is how height is changing

Assume initial wave has no veloxcity. Then, given an initial wave, it splits into two waves moving in the opposite direction with exactly half the height of the initial shape: $g(x+ct)$ and $g(x-ct)$. how about if we introduce an initial velocity? at $t=0$, we have $g(x) = u(x,0) = 0$, so nothing happens initially. Then, given an initial velocity $h(x)$ which is 1 within a range but 0 elsewhere, because $g(\cdot) + g(\cdot)$ all zero, so we're only concerned with the component: $\frac{1}{2c}\int_{x-ct}^{x+ct} h(\xi)d\xi$ If we choose a small t for $x-ct$, the string becomes slightly deformed. as t increases, it becomes a wider zone.
Now how to discretize to solve numerically? If we apply ifnite difference, both are 2nd derivatives so we can use central difference method. If we apply this for both sides, we should get good accuracy. Suppose j, want $j+1$ points. Then:

$$\frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{k^2} = \alpha^2 \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{k^2}$$

Like before, we end up with the form:

$$\rightarrow w_{i,j+1} = \dots \tag{160}$$
$$w^{(i+1)} = \begin{pmatrix} 2(1-\lambda^2) & \lambda^2 & \dots & 0 \\ \lambda^2 & \dots & & \\ \dots & & & \end{pmatrix} \vec{w}^{(j)} - \vec{w}^{(j-1)} \tag{161}$$

Here, $\lambda = \frac{\alpha k}{h}$.
We have two problems with this solution:

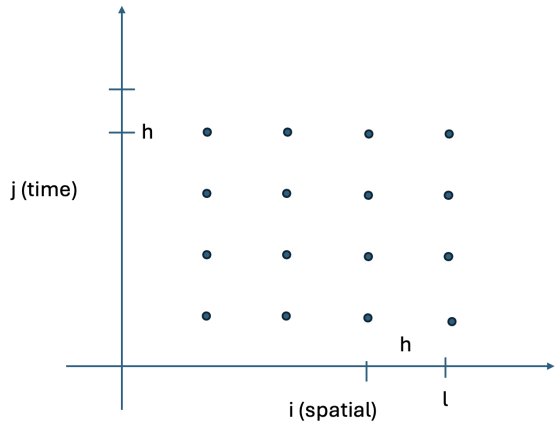

Figure 5: Wave Discretization

- Not always stable; need $\lambda < 1$ (so small time steps)
- How do we begin? need to use both $f(x)$ and $g(x)$ initial conditions. $g(x)$ tells us how it changes in time, so we can use eulers (or forward difference).

14 Friday 4/17 (missed lecture)

topic on SVD

15 Monday April 20

Theorem 18. For square nonsingular matrix with singular values $\sigma_1 \geq \cdots \geq \sigma_n > 0$ (always > 0 for nonsingular so last inequality redundant), then σ_n is distance to nearest singular matrix, where distance is the norm of the difference of the matrices.

Partial proof. By SVD, $A = U\Sigma V^T$, where U, V are orthogonal by svd and Σ is a diagonal matrix with the singlar values on its diagonal. We want to show there exists a singular matrix of distance σ_n away. Let's define one:

$$B = U \begin{pmatrix} \sigma_1 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix} V^T \tag{162}$$

Our first claim is that B is singular. To show this, we want to show there is a nonzero x such that

$Bx = 0$: let $x = v_n$, the last column of V. Then:

$$\rightarrow U(\dots)V^T v_n \tag{163}$$

$$= U(\dots)\begin{pmatrix} 0 \\ \dots \\ 1 \end{pmatrix}^T \tag{164}$$

$$= U\vec{0} \tag{165}$$

$$= \vec{0} \tag{166}$$

We have shown B is singular. now we caluclate the difference:

$$A - B = U\begin{pmatrix} 0 & \dots & 0 \\ \dots & & \\ 0 & \dots & \sigma_n \end{pmatrix} V^T \tag{167}$$

$$\rightarrow \|A - B\| = \max_{\|x\|=1} \|U(\cdot)V^T x\| \tag{168}$$

Remebmer that if Q orthonormal, then $\|Qx\| = \|x\|$. U is orthonormal, so the above is equal to:

$$= \max_{\|x\|=1} \|(\cdot)v^T x\| \tag{169}$$

We let $y = v^t x$, so our problem is then:

$$\max_{\|y\|=1} \| \begin{pmatrix} 0 & \dots & 0 \\ \dots & & \\ 0 & dots & \sigma_n \end{pmatrix} y = \max_{\|y\|=1} \| \begin{pmatrix} 0 \\ \dots \\ \sigma_n y_n \end{pmatrix} \| = \|\sigma_n y_n\| = \sigma_n \tag{170}$$

where the last equality occurs if we choose $y = (0,\ dots, 0, 1)^T$. □

Now we discuss applications of SVD in **least squares problems**. We wish to solve the problem:

$$\min_x \|b - Ax\| \leftrightarrow Ax = b \tag{171}$$

naturally, we want to use gaussian elimination, but this doesn't always work. if our matrix is singular, there are two possiblities: either no solution or infinite solutions.
If our matrix A is a rectangle, we likewise cannot diretly solve, but we want to instead just get as close as possible.

16 Wed April 22

Last time, discussed two formulations: least squares $Ax = b$, or alternatively the least squares problem of 'get as close as possible": $\min_x \|b - Ax\|$. We call $b - Ax$ the residual norm.

Theorem 19. Best approximation theorem:
$\|y - \tilde{y}\|$ minimal if and only if we use the orthogonal projection

Proof. (uses pythagorean theorem) □

We have three possible solutions:

1. derive normal equation: $A^TAx = A^Tb$; this is a good solution but there exist better ways
2. QR factorization
3. SVD

For 1: define subspace $S := col(A)$; projection b onto it is $\hat{b}$, but can rewrite $\hat{b}$ as columns of A: Since $\hat{b} \in S$ then $b = x_1\vec{a_1} + \cdots = A\vec{x}$ for some x then the residual is:

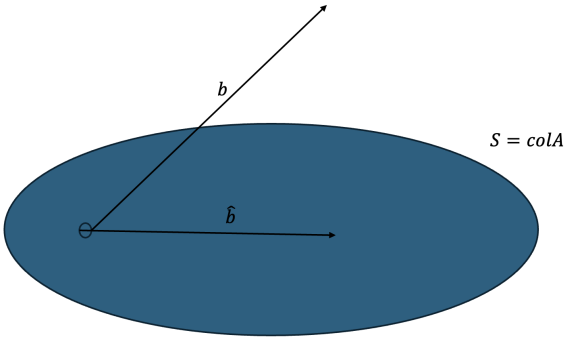

Figure 6: Projection Visualization

$$b - \vec{b} = b - A\vec{x} \tag{172}$$

to minimize the difference, use the best approximation theorem, therefore we need $b - Ax \perp colA$, or $A^T(b - Ax) = \vec{0}$, or $A^Tb = A^TAx$, which is the solution to the normal equation.
Nxt, SVD solution. We awnt to solve:

$$\min_x \|b - Ax\|$$

by SVD of A, we have $A = U\Sigma V^T$. We actually don't need all of U and Σ: let $\hat{U}$ be the first n columns and $\hat{\Sigma}$ be the first n rows that contains all n singular values so it is a diagonal matrix. then, our minimization problem becomes:

$$\begin{aligned}\min_x \|b - Ax\| &= \min_x \|b - U\Sigma V^T x\| \\ &= \min_x \|U^T(b - U\Sigma V^T x)\| \\ &= \min_x \|U^Tb - \Sigma V^T x\|\end{aligned}$$

we once again let $y = V^Tx$ so the above continues as:

$$\begin{aligned}&= \min_y \|U^Tb - \Sigma y\| \\ &= \min_y \|() - \begin{pmatrix}\sigma \\ 0\end{pmatrix} y\| = \min_y \| \begin{pmatrix}s1 \\ s2\end{pmatrix}\end{aligned}$$

but how do we minimize this? we can safely solve for: $\hat{U^T}b = \Sigma y$: only top n are computed. So let:

$$U^T b = \begin{pmatrix} \hat{U}b \\ U^u b \end{pmatrix} \tag{173}$$

We can solve for the top part to become zero, bottom unchanged.

Example 30.

$$f1 : x_2 = 1/3x_1 \tag{174}$$
$$f2 : x_2 = 1/3x_1 + 1 \tag{175}$$
$$f3 : x_2 = 1 - 1/3x_1 \tag{176}$$

If we format as matrix:

$$\begin{pmatrix} 1/3 & -1 \\ 1/3 & -1 \\ -1/3 & -1 \end{pmatrix}, b = (0, -1, -1)^T \tag{177}$$

to find solution, we have three ways:

1. graph + estimate (7) note for graph, $1/3x_1 = x_2$ same as $x_1 = 3x_2$ but it does chagne the

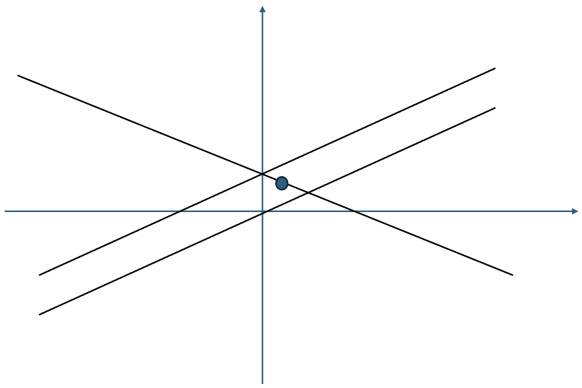

Figure 7: graph solution

least-squares solution.
Visually, we select the point that is closest to each of the lines

2. normal equation:

$$A^T A = \cdots = \begin{pmatrix} 3 & -3 \\ -3 & 27 \end{pmatrix}, A^T b = (0, 18)^T$$
$$\rightarrow x = (3/4, 3/4)^T$$

3. SVD (shown next time)

17 Fri April 24

Remember for SVD solution to least squares: to solve $\min_x \|b - Ax\|$, we can instead use SVD decomp and solve $\hat{\Sigma}y = \hat{U^T}b$, then transform $x = Vy$.

Example 31.

$$Ax = b \leftrightarrow \begin{pmatrix} 1 & -1 \\ 1 & -3 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ -3 \\ 3 \end{pmatrix} \tag{178}$$

By SVD comp, we know that:

$$U = \begin{pmatrix} -.59 & .39 \\ -.59 * .39 & \\ .55 & .84 \end{pmatrix}, \Sigma = \begin{pmatrix} 5.2 & 0 \\ 0 & 1.5 \end{pmatrix}, V = \begin{pmatrix} -.12 & .99 \\ .99 & .12 \end{pmatrix} \tag{179}$$

In this case, since we only have 2 singular values which is the size of hte matrix, our $U = \hat{U}$ and $\Sigma = \hat{\Sigma}$. So solving:

$$\hat{U}^T b = \begin{pmatrix} 2.41 \\ 1.36 \end{pmatrix} \to y = (\hat{\Sigma})^{-1}\hat{U}^T b = \begin{pmatrix} .653 \\ .836 \end{pmatrix}$$
$$\to x = Vy = (3/43/4)^T$$

This is the same solution as the normal equation solution as we expected.

Theorem 20. For $A \in \mathbb{R}^{m \times n}$:

$$Rank(A) = \#\text{ of nonzero singular values}$$

Proof. Recall **rank theorem**: $rank(A) + dim(null(A)) = \#$ of col of A
$A = U\Sigma V^T$ by svd. note that by defn of V, $V^T v_i = (0, \ldots, 1, \ldots, 0)^T$, 1 in the ith index. and v_i is the ith column of V. So:

$$\begin{aligned} Av_1 &= U\Sigma V^T v_1 \\ &= U\Sigma \begin{pmatrix} 1 \\ 0 \\ \ldots \\ 0 \end{pmatrix} = U \begin{pmatrix} \sigma_1 \\ 0 \\ \ldots \\ 0 \end{pmatrix} = \sigma_1 u_1 \end{aligned}$$

can generalize this for all columns; therefore, $u_i \in col(A)$ since $\sigma_i u_i = Av_i$. Also, we know that u_i are linearly independent. Therefore, we know that $dim(A) \geq r$, r the number of nonzero singular values of A. similarly, for $k = r + 1 ton$:

$$Av_k = 0$$

Therefore $v_k, k \in [r+1, n]$ $v_k \in Null(A)$. Therefore, since v_i also linearly independent, we know that $dim(Null(A)) \geq n - 1$.
We no apply rank theorem: $rank(A) = \#\text{ cols} - dim(Null(A)) \leq n - (n - r) = r$ Therefore, $rank(A) = r$ □

Theorem 21. $\|A\| = \sigma_1 =$ largest singular value, A square.

Proof.

$$\begin{aligned}\|A\| &= \max_x \|Ax\| \\ &= \max_x \|U\Sigma V^T x\| \\ &= \max_x \|\Sigma V^T x\| \\ &= \max_y \|\Sigma y\| \\ &= \|\Sigma\| = \sigma_1\end{aligned}$$

Another way to prove the last part:

$$\|(\sigma_1 y_1 \dots \sigma_n y_n)\| \geq \|(\sigma_1 \cdot 1, 0 \dots, 0)\| = \sigma_1$$

last line due to picking $y = (1, 0, \dots, 0)^T$. now to show $\leq \sigma_1$:

$$= \sqrt{(\sigma_1 y_1)^2 + \dots + (\sigma_n y_n)^2} \leq \sqrt{\sigma_1^2 (\sum y_i^2)|} = \sigma_1$$

□

Note that if σ_1 is singular value of A, then $1/\sigma_1$ is largest singular value of A^{-1}.

18 Monday april 27

Last topic covering krylov subspace methods for $Ax = b$ with A large matrix. Let's review krylov for eigenvalues:

$$K := Span\{s, As, \dots, A^{n-1}s\}$$

Here, s usually random. So if $y \in K$, then $y = p(A)s$. Assume A is diagonalizable. From convergence analysis, $s = \beta_1 z_1 + \dots + \beta_n z_n$, where z_i are eigenvectors. If we apply the polynomial, then:

$$p(A)s = \beta_1 p(\lambda_1) + \dots + \beta_n p(\lambda_n)$$

For y z_1, need $p(\lambda_1) >>> p(\lambda_2) > \dots$. We want a large $p(\lambda_1)$ while being small elsewhere. In general, we have convergence if:

- eigenvalues well separated from others
- eigenvalues on/near exterior

Now **krylov subspace methods** with linear equations: solving for $Ax = b$. If A is small, can just use gaussian elimination. In this case, our krylov subspace is:

$$K_b := span\{b, Ab, A^2 b, \dots, A^{n-1} b\}$$

By convergence analysis: $b = \sum \beta_i z_i$, with eigenvectors z_i. polynomial is then:

$$p(A)b = \sum \beta_i p(\lambda_i)$$

Let $\tilde{x} \in K$, want to define how good this solution is:

$$\tilde{x} = \sum A^i b = \sum \beta_i p(\lambda_i) z_i$$

Instead of requiring $\hat{x}$ to resemble an eigenvector, we consider the residual norm:

$$r = b - A\hat{x}.$$

We want $r \to \mathbf{0}$, or $\|r\| < r_{\text{tolerance}}$. Let $\tilde{x} = p(A)b$. Then

$$r = b - A\tilde{x} = b - Ap(A)b = (I - Ap(A))\, b.$$

Define

$$q(\alpha) = 1 - \alpha p(\alpha),$$

so that

$$q(A) = I - Ap(A), \quad \text{and hence} \quad r = q(A)b.$$

To analyze r, expand b in the eigenbasis of A:

$$b = \beta_1 z_1 + \cdots + \beta_n z_n,$$

where $Az_i = \lambda_i z_i$. Then

$$q(A)b = q(A)\left(\beta_1 z_1 + \cdots + \beta_n z_n\right) = \beta_1 q(\lambda_1) z_1 + \cdots + \beta_n q(\lambda_n) z_n.$$

Thus, to make the residual small, we want $q(\lambda_i)$ to be small for all eigenvalues λ_i.

Why not choose $q(\alpha) = 0$? Because we must satisfy the constraint

$$q(0) = 1,$$

which follows from the definition $q(\alpha) = 1 - \alpha p(\alpha)$. Note that q is a polynomial of degree m, while p has degree $m - 1$.

When is the problem easier to solve?

- When the spectrum is well-separated from 0
- When the spectrum is clustered

In class, we demonstrated this by comparing spectra $\{1, 20\}$ and $\{-10, -1\} \cup \{1, 10\}$. introduces 10 spikes initially for the 10 negative values, then 4 extra spikes later due to round-off error; should get 10 total if doesn't converge completely. If we comepare $\{1, 20\}$ and$\{-1, 2, 20\}$, this results in a similar number of iterations.

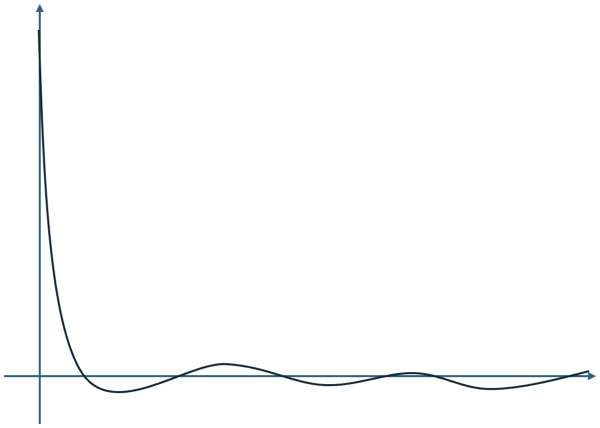

Figure 8: Krylov polynomial

19 Wed 4/29

four examples of how the singular values vary given $A = U\Sigma V^T$:

- A^{-1}

$$\begin{aligned} A &= U\Sigma V^T \\ A^{-1} &= (U\Sigma V^T)^{-1} \\ &= V^T\Sigma^{-1}U^{-1} \\ &= V\Sigma^{-1}U^T \end{aligned}$$

  note that in addition to the singular values being recipricols, the singular vectors are also flipped

- A^T

$$\begin{aligned} A &= U\Sigma V^T \\ A^T &= V\Sigma^T U^T \\ &= V\Sigma^T U^T \\ &= V\Sigma U^T \end{aligned}$$

  so same singular values

- A^2 no guaranteed relationship

- AA^TA

$$\begin{aligned} &= U\Sigma V^T V\Sigma U^T A \\ &= U\Sigma^2 U^T U\Sigma V^T \\ &= U\Sigma^3 V^T \end{aligned}$$

On a side note, let's show what the trace of the matrix inverse is:

$$\begin{aligned} Tr(A^{-1}) &= Tr(A^{-1} - \sum \frac{1}{\sigma_i} V u^*) + Tr(\sum \frac{1}{\sigma_i} V u^*) \\ &= Tr(A^{-1} - p(A)) + Tr(p(A)) \end{aligned}$$

Combine the above to get:

$$Tr(A^{-1} + p(A)) + Tr(p(A))$$

or:

$$Tr(A^{-1} - p_1(A)) + Tr(p_1(A) - p_2(A)) + Tr(p_2(A) - p_3(A)) + Tr(p_3(A))$$

20 Monday May 4

Krylov subspace: given problem $Ax = b$ with m number of iterations: $K = span\{b, Ab, A^2b, \ldots, A^{m-1}b\}$

- if $\hat{x} \in K$, then $\hat{x} = p(A)b$, degree p is $m - 1$ or less
- $r = q(A)b$, $q(x) = 1 - \alpha p(\alpha)$, so q is degree m or less with $q(0) = 1$

Remember we start with b and not a random vector. Need $q(\lambda_i) \approx 0$ for all eigenvalues λ_i.

Example 32. Let $\lambda_i \in [10, \ldots, 1000]$. This is "easy" because no small λ_i; smallest is $\lambda_1 = 10$
If instead: $\lambda_i \in [.1, .2, \ldots, 10, 11, \ldots, 100]$, convergence is relative to $\sqrt{\lambda_n/\lambda_1}$, so this new one is about 10x harder/slower.