MATH 2121 - Lecture #22

For webwork: use sqrt(x) instead of $x^{0.5}$

Today:

1. Review of Least-squares solutions
   (new: line of best fit)

2. Symmetrices and their eigenvalues

1 Last time: least-squares problems

Want to solve Ax=b

Definition. If A is an $m \times n$ matrix and $b \in \mathbb{R}^m$ then $\boxed{A^\top A x = A^\top b}$ is consistent.

A solution to $A^\top A x = A^\top b$ is called a *least-squares solution* to the equation $Ax = b$.

Let $\|v\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \geq 0$ for $v \in \mathbb{R}^n$. Recall $\|v\| = 0$ if and only if $v = 0$.

Fact. A vector $s \in \mathbb{R}^n$ is a least-squares solution to $Ax = b$ if and only if

$$\|b - As\| \leq \|b - Ax\| \quad \text{for all } x.$$

The linear system $Ax = b$ is consistent if and only if $\|b - Ax\| = 0$ for some $x \in \mathbb{R}^n$.

If $Ax = b$ is consistent then all least-squares solutions s have $\|b - As\| = 0$ so $As = b$.

If $Ax = b$ is inconsistent, still at least one least-squares solution s, but $\|b - As\| > 0$.

a LS solution to $Ax=b$ is a vector $s \in \mathbb{R}^n$ that minimizes the distance $\|b - As\|$

Picture of a least-squares solution to $Ax=b$ (inconsistent)

when system is

LS solutions to $Ax=b$

are the exact solutions to

$Ax = \text{proj}_{\text{Col}A}(b)$

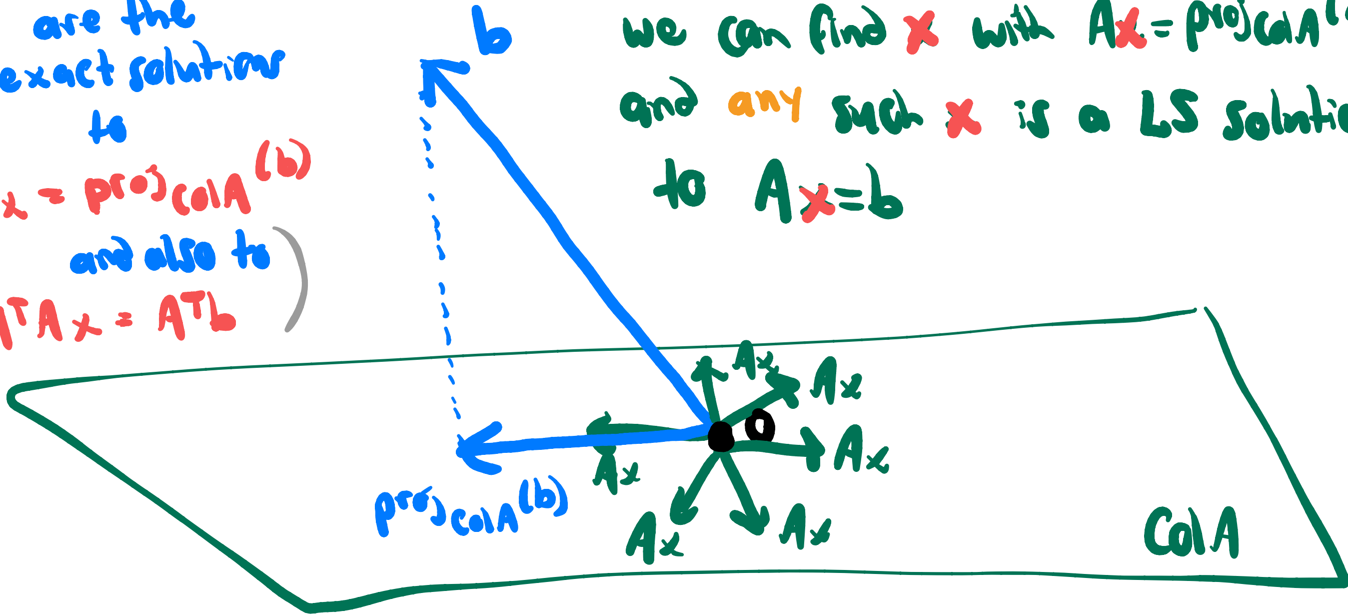
and also to

$\left(A^TAx = A^Tb \right)$

because $\text{proj}_{\text{Col}A}(b) \in \text{Col}A$,

we can find x with $Ax = \text{proj}_{\text{Col}A}(b)$

and any such x is a LS solution to $Ax=b$



no vector Ax is equal to b so best we can do is find x such that Ax is as close to b as possible

Theorem. Let A be an $m \times n$ matrix. The following properties are equivalent:

(a) $Ax = b$ has a unique least-squares solution for each $b \in \mathbb{R}^m$.

(b) The columns of A are linearly independent.

(c) $A^\top A$ is invertible.

$$A^T A x = A^T b$$

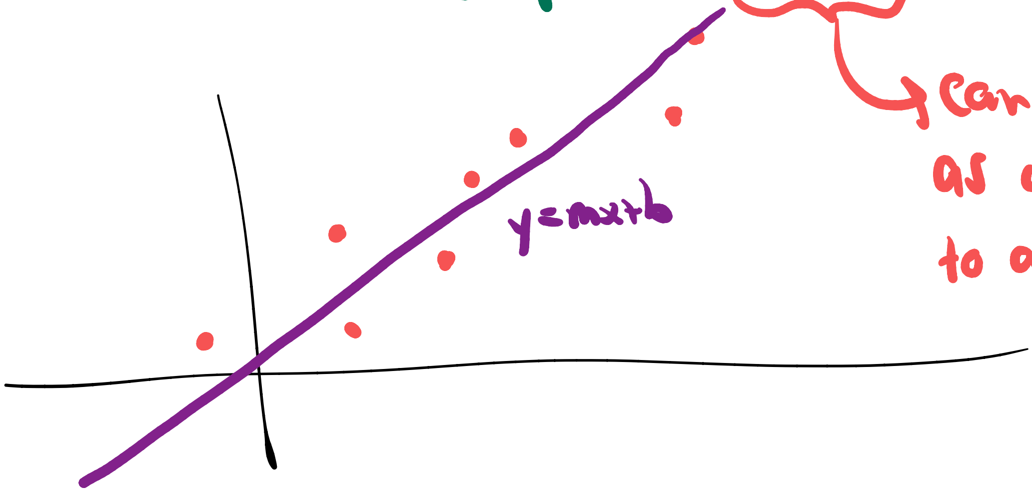
given by

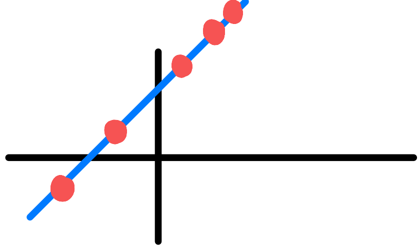
$$x = (A^T A)^{-1} A^T b$$

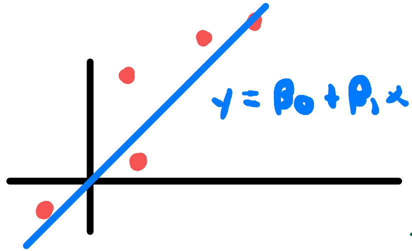
New application: computing line of best fit

if you have some datapoints and you want to fit a line $y = mx + b$ to approximate the data, how to compute m and b?

can find these as a LS solution to a linear system







Example (Lines of best fit)**.** Suppose we have n data points $(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$. Want to find $\boxed{\beta_0, \beta_1 \in \mathbb{R}}$ such that $\boxed{y = \beta_0 + \beta_1 x}$ is the *line of best fit* for this data.

If our points are all on the same line, then for some $\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \in \mathbb{R}^2$ we would have

Can solve $$\boxed{b_i = \beta_0 + \beta_1 a_i} \quad \text{for } i = 1, 2, \ldots, n,$$ (a linear system)

meaning that $x = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ is an exact solution to the linear system $\boxed{Ax = b}$ where

$$A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_n \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

If the given points are not on the same line, then no exact solution to $Ax = b$ exists, and **we should instead try to find a least-squares solution** to the system.

$$A \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 a_1 \\ \beta_0 + \beta_1 a_2 \\ \beta_0 + \beta_1 a_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \end{bmatrix}$$

want to find β_0, β_1 such that
$y = \beta_0 + \beta_1 x$ is close to these pts

To be concrete, suppose we have four points $\boxed{(2,1),(5,2),(7,3), \text{ and } (8,3)}$ so that

① Form:

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}. \qquad A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_k \end{bmatrix} \quad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}$$

The least-squares solutions to $Ax = b$ are the exact solutions to $A^\top A x = A^\top b$.

②

$$A^\top A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \quad \text{and} \quad A^\top b = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

2x2

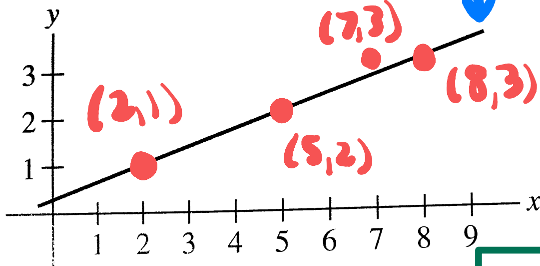
The matrix $A^\top A$ is invertible. So a least-squares solution is

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (A^\top A)^{-1} A^\top b = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix} = \begin{bmatrix} \text{y-intercept} \\ \text{slope} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

③ Solve $A^\top A x = A^\top b \rightsquigarrow x = (A^\top A)^{-1} A^\top b$

From prev slide:

Thus our line of best fit for the data is $\boxed{y = \frac{2}{7} + \frac{5}{14}x}$:



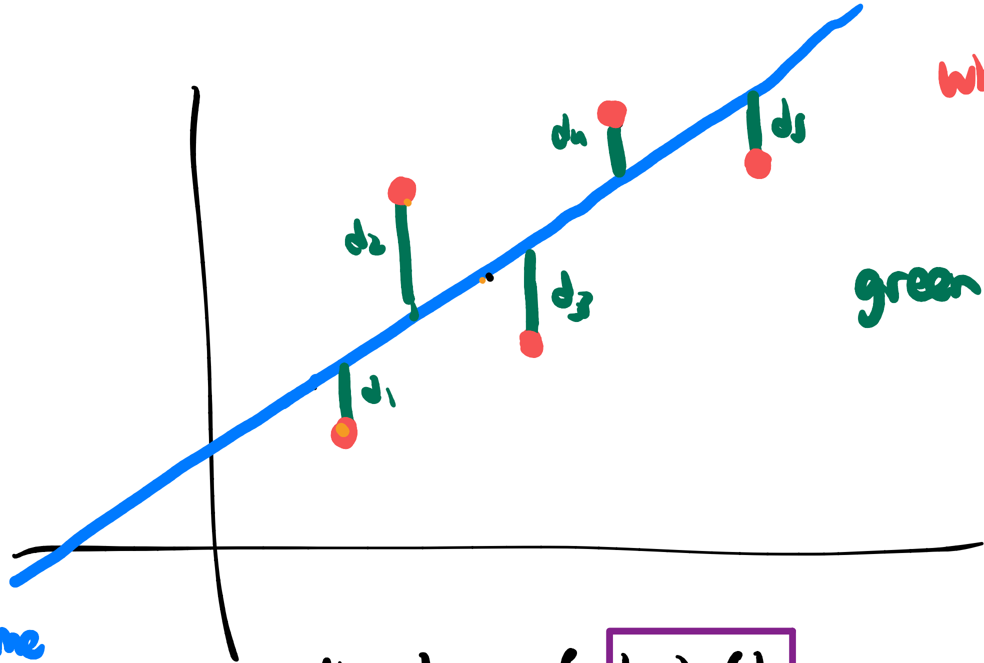
Fix $d \geq 1$

$$\boxed{\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_d x_i^d = y_i}$$

The process to find a "polynomial of best fit" for some datapoints $(x_1, y_1) \dots (x_r, y_r)$ is similar (just will involve more variables)

Question: what does our defn of line of best fit optimize?

what does "best" mean?



green lines = vertical distances from points to line

some line

the line of best fit minimizes this sum:

error measurement → $\sum_i (d_i)^2 \geq 0$

when this sum is small, the line is a good fit

2 Symmetric matrices

$\boxed{\text{A matrix } A \text{ is } \textit{symmetric} \text{ if } A^\top = A.}$

Goal today: these matrices have a different characterization in terms of diagonalization

This happens if A is square and $A_{ij} = A_{ji}$ for all i, j.

Example. $\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$ and $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}$ and $\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$ are symmetric.

$A = PDP^{-1}$

(D diag)

$\begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 6 & -6 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \end{bmatrix}$ are not symmetric.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Proposition. If A is symmetric and $k \in \{1, 2, 3, \dots\}$ then A^k is also symmetric.

Proof. If $A = A^\top$ then $(A^k)^\top = (AA \cdots A)^\top = A^\top \cdots A^\top A^\top = (A^\top)^k = A^k$. $\square$

Proposition. If A is an invertible symmetric matrix then A^{-1} is also symmetric.

Proof. This is because $(A^{-1})^\top = (A^\top)^{-1}$. $\square$

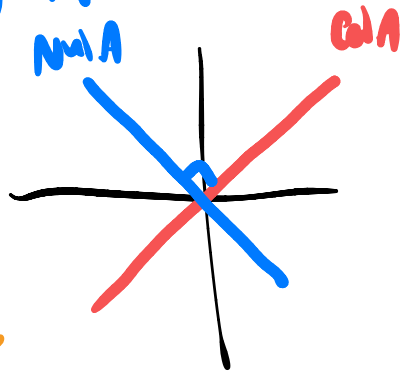
* powers and inverses preserve (transpose) symmetry

Fact If A is any matrix, not nec. square, then A^TA is symmetric: $(A^TA)^T = A^T(A^T)^T = A^TA$

What else do we know?

if $A = A^T$ then $\mathrm{Col}(A)^\perp \overset{\text{always true}}{=} \mathrm{Null}(A^T) \overset{\text{true when } A^T = A}{=} \mathrm{Null}(A)$

So the column space and nullspace of symmetric matrices are orthogonal



Ex If $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ has $\underbrace{\text{rank } 1}_{\text{so Col A is a line } (\dim \mathrm{Col}\, A = 1)}$ then

Illustrate eigenvector/value properties of $A = A^T$ by example:

Recall how we can diagonalize a matrix.

Example. Let $\boxed{A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}} = A^T$

all real eigenvalues

Then $\det(A - xI) = (8 - x)(6 - x)(3 - x)$ so the eigenvalues of A are $\boxed{8, 6, \text{ and } 3.}$

By constructing bases for the null spaces of $A - 8I$, $A - 5I$, and $A - 3I$, we find that the following are eigenvectors of A:

$$v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \text{ with eigenvalue 8.} \qquad v_2 = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \text{ with eigenvalue 6.}$$

$$v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ with eigenvalue 3.}$$

Observe v_1 v_2 v_3 are Orthogonal!

These eigenvectors are actually an orthogonal basis for $\mathbb{R}^3$.

Nontrivial to find, easy to check: $Av_2 = \begin{bmatrix} -6+2-2 \\ 2-6-2 \\ 1+1+10 \end{bmatrix} = \begin{bmatrix} -6 \\ -6 \\ 12 \end{bmatrix} = 6v_2$

Converting these vectors to unit vectors gives an $\boxed{\text{orthonormal basis of eigenvectors:}}$ *

$$u_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \qquad u_2 = \begin{bmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \qquad u_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}.$$

As usual we then have $\boxed{A = PDP^{-1}}$ where

$$\boxed{P = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}} \quad \text{and} \quad \boxed{D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}}.$$

Since the columns of P are orthonormal, we actually have $P^\top = P^{-1}$ so $A = PDP^\top$.

The special properties in this example turn out to hold for all symmetric matrices.

because u_1 u_2 u_3 are orthonormal

$$P^{-1} = [u_1 \; u_2 \; u_3]^{-1} = P^T = \begin{bmatrix} u_1^T \\ u_2^T \\ u_3^T \end{bmatrix}$$

P is an orthogonal matrix

In this example: $A = A^T$ has all real eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$

and we can find an orthogonal matrix P

(with $P^{-1} = P^T$)

such that $A = PDP^{-1} = PDP^T$

$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$

Will say: "$A = A^T$ is orthogonally diagonalizable" if $Av = \lambda v$ then $v \cdot w = 0$

$Aw = \mu w$

$\lambda \neq \mu$

$A = A^T$

Theorem. Suppose A is a symmetric matrix.

$\boxed{\text{Then any two eigenvectors from different eigenspaces of } A \text{ are orthogonal.}}$

So if $A = A^\top$ is $n \times n$ and $u, v \in \mathbb{R}^n$ are such that

$$Au = au \quad Av = bv \quad \text{for numbers } a \neq b,$$

symmetry is used here

then $u \bullet v = 0$.

(But if u, v are eigenvectors of A with **same eigenvalue** then could have $u \bullet v \neq 0$.)

Proof. Let u and v be eigenvectors of A with eigenvalues a and b, where $a \neq b$.

Then $au \bullet v = Au \bullet v = (Au)^\top v = u^\top A^\top v = u^\top Av = u \bullet Av = u \bullet bv$.

But $au \bullet v = a(u \bullet v)$ and $u \bullet bv = b(u \bullet v)$, so $a(u \bullet v) = b(u \bullet v)$, thus $(a-b)(u \bullet v) = 0$.

Since $a - b \neq 0$, it follows that $u \bullet v = 0$. ∎

<u>pf idea:</u> $u^T Av$ is both $\begin{cases} a(u \cdot v) \\ b(u \cdot v) \end{cases}$ if $a \neq b$, then these can only be equal if $u \cdot v = 0$

Recall that a matrix P is *orthogonal* if P is invertible and $\boxed{P^{-1} = P^{\top}.}$

Definition. A matrix A is *orthogonally diagonalizable* if there is an orthogonal matrix P and a diagonal matrix D such that $A = PDP^{-1} = PDP^{\top}$.

When A is orthogonally diagonalizable and $A = PDP^{-1} = PDP^{\top}$, the diagonal entries of D are the eigenvalues of A, and the columns of P are the corresponding eigenvectors; moreover, these eigenvectors form an orthonormal basis of $\mathbb{R}^n$.

In fact, it follows by the arguments in our earlier lectures about diagonalizable matrices that **an $n \times n$ matrix A is orthogonally diagonalizable if and only if there is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for A.**

Surprisingly, there is a much more direct characterization:

* **Theorem.** A square matrix $\boxed{A \text{ is orthogonally diagonalizable}}$ if and only if $\boxed{A = A^{\top}.}$

nontrivial

equivalent

in words: we can find an orthogonal basis of eigenvectors for A iff $A = A^{T}$

easy direction of <u>Thm</u> orthogonally diagonalizable $\Leftrightarrow$ Symmetric

$\Rightarrow$

The complete proof of previous theorem is in lecture notes.

We will just note some key parts of the argument.

Proposition. If A is orthogonally diagonalizable then A is symmetric.

Proof. If X, Y, Z are $n \times n$ matrices then $(XYZ)^\top = Z^\top (XY)^\top = Z^\top Y^\top X^\top$.

Suppose $\boxed{A = PDP^\top}$ where D is diagonal. Then $D = D^\top$ and $(P^\top)^\top = P$, so

$$A^\top = (PDP^\top)^\top = \underbrace{(P^\top)^\top}_{=P} \underbrace{D^\top}_{=D \text{ as } D \text{ is diagonal}} P^\top = PDP^\top = A.$$

$\square$

was a practice problem

$(v^T A \bar{v})^T = \bar{v}^T A^T (v^T)^T$

$= \bar{v}^T A v$ if $A = A^T$

Proposition. Suppose A is a symmetric matrix with all real entries.

Then all complex eigenvalues of A belong to $\mathbb{R}$.

Proof. Suppose A is a symmetric $n \times n$ matrix with real entries, so that $A = A^\top = \overline{A}$.

Let $\boxed{v \in \mathbb{C}^n.}$ Consider $\overline{v}^\top A v$. For example, if $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ and $v = \begin{bmatrix} 1+i \\ 1-i \end{bmatrix}$ then

$$\overline{v}^\top A v = \begin{bmatrix} 1-i & 1+i \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} = \begin{bmatrix} 3+i & 3-i \end{bmatrix} \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} = 4.$$

* $\boxed{\text{The number } \overline{v}^\top A v \text{ always belongs to } \mathbb{R}}$ since $\overline{\overline{v}^\top A v} = v^\top A \overline{v} = (\overline{v}^\top A v)^\top = \overline{v}^\top A v$.

(The last equality holds since both sides are 1×1 matrices, i.e., scalars.)

holds as $A = A^T$

Now suppose $\underline{v \in \mathbb{C}^n}$ is an eigenvector for A with eigenvalue $\underline{\lambda \in \mathbb{C}}$.

Then $\overline{v}^\top A v = \overline{v}^\top (\lambda v) = \lambda(\overline{v}^\top v) \in \mathbb{R}$. But $\underline{\overline{v}^\top v \in \mathbb{R}}$ so we must also have $\lambda \in \mathbb{R}$. $\square$

$\bar{v}^T v = \bar{v}^T B v$ for $B = I = B^T$

Thm: If $A = A^T$ then can find P with $P^{-1} = P^T$ and D diagonal with $A = PDP^{-1}$

How?!

To orthogonally diagonalize an $n \times n$ symmetric matrix A, we just need to find an orthogonal basis of eigenvectors $v_1, v_2, \ldots, v_n$ for $\mathbb{R}^n$.

Then $A = UDU^\top$ with $U = \begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix}$ where $u_i = \frac{1}{\|v_i\|} v_i$ and D is the diagonal matrix of the corresponding eigenvalues.

If all eigenspaces of A are 1-dimensional, then any basis of eigenvectors will be orthogonal. If A has an eigenspace of dimension greater than one, then after finding a basis for this eigenspace, **it may be necessary to apply the Gram-Schmidt process to convert this basis to one that is orthogonal.**

Can either: find basis for 1-eigspace → GS → take union for all eigspaces

or: find bases for each eigspace → take union → GS

First: ① find eigvals λ for $A = A^T$ ② compute for each Nul$(A - \lambda I)$

③ After finding basis for each Nul$(A - \lambda I)$

when $A = A^T$, can either do GS process on each basis, then collect

or

can collect the bases into one long list and do GS process

(order doesn't matter if $A = A^T$ as eigenspaces are orthogonal) an orthogonal diagonalization

$U = (U^{-1})^T = (U^T)^{-1}$ $(\Leftrightarrow U^{-1} = U^T)$

Corollary. If $A = UDU^\top$ where

$$U = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \text{ has orthonormal columns and } D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

then $\boxed{A = \lambda_1 u_1 u_1^\top + \lambda_2 u_2 u_2^\top + \cdots + \lambda_n u_n u_n^\top.}$ ← multiply out $[u_1 \dots u_n]\, D\, [u_1 \dots u_n]^T$

* Each product $u_i u_i^\top$ is an $n \times n$ matrix of **rank 1**.

One calls this expression a *spectral decomposition* of A.

Practice problem

rank $M = 1$

$\Updownarrow$

$M = vw^T$

$$A = U \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} U^T$$

Example. Let $A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$. A spectral decomposition of A is given by

$$A = \underbrace{\begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}}_{U} \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \underbrace{\begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}}_{U^T}$$

$$= 8 \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} + 3 \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

$$\begin{bmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{bmatrix} + \begin{bmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{bmatrix}.$$