Lecture #23 – MATH 2121

① Review : symmetric matrices
orthogonal matrices

② New : singular value decompositions

Orthog. matrices: $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ or any perm. matrix

$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ or any rotation matrix

1 Last time: symmetric matrices

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

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

* **Example.** $\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ is symmetric but $\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$ is not.

A matrix U is $\boxed{\textit{orthogonal} \text{ if } U \text{ is invertible and } U^{-1} = U^\top}$.

This happens precisely when U is square with orthonormal columns.

$v \cdot v = 1$
$v \cdot w = 0$
for distinct columns v, w

An $n \times n$ matrix A is $\boxed{\textit{orthogonally diagonalizable}}$ if there is an orthogonal matrix U and a diagonal matrix D such that $A = UDU^{-1} = UDU^\top$.

* In this case, columns of U are an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for A, and eigenvalues of these eigenvectors are the diagonal entries of D.

$A = UDU^T = UDU^{-1}$ w/ D diagonal

The following summarizes the main results from last time:

$A = UDU^T = UDU^{-1}$ $\qquad$ $A = A^T$

Theorem.

(1) A square matrix is $\boxed{\text{orthogonally diagonalizable}}$ if and only if it is $\boxed{\text{symmetric.}}$

(2) Eigenvectors with different eigenvalues for a symmetric matrix are orthogonal.

(3) All (complex) eigenvalues of a symmetric matrix A are real. The characteristic polynomial of A has all real roots and can be completely factored as

$$\det(A - xI) = (\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x)$$

for some (not necessarily distinct) **real numbers** $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$.

Appl. of (3) To show that a polynomial $p(x)$ has all real roots, just need to find $A = A^T$ with $p(x) = \det(A - xI)$

Fact $v^2 - w^2 = (v-w)(v+w)$ for any $v, w \in \mathbb{C}$

Example. Suppose $A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$ for some $a, b \in \mathbb{R}$. $A = A^T$

How does the preceding theorem apply to this generic 2-by-2 matrix? Since

$$\det(A - xI) = \det \begin{bmatrix} a-x & b \\ b & a-x \end{bmatrix} = (a-x)^2 - b^2 = (a-b-x)(a+b-x),$$

the eigenvalues of A are $a-b$ and $a+b$.

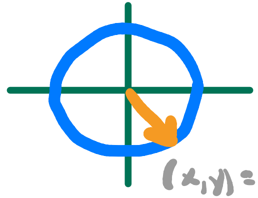
* The vector $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is an eigenvector for A with eigenvalue $a-b$. $\begin{bmatrix} a & b \\ b & a \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} a-b \\ b-a \end{bmatrix}$

* The vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector for A with eigenvalue $a+b$.

These eigenvectors are orthogonal, as predicted by the theorem. We can convert them to unit vectors by multiplying each vector by the reciprocal of its length.

replace $v \in \mathbb{R}^2$ by $\frac{1}{\|v\|} v$

Note: $U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ for $\theta = -\frac{\pi}{4}$



$(x,y) = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$
$= (\cos\theta, \sin\theta)$

This gives the orthonormal eigenvectors

unit vectors: $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

which form an orthonormal basis for $\mathbb{R}^2$. So

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} = A = UDU^{-1} = UDU^{\top}$$

where $\boxed{U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \text{ and } D = \begin{bmatrix} a-b & 0 \\ 0 & a+b \end{bmatrix}}$.

Check:

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} a-b & 0 \\ 0 & a+b \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ +\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$A = U \cdot D \cdot U^{T}$$

2 Singular value decomposition

SVD

"$A = U\Sigma V^T$"

Today, we'll prove the existence of *singular value decompositions*, which give a sort of approximate orthogonal diagonalization for any matrix, not just symmetric ones.

Let A be an $m \times n$ matrix.

* Then $\boxed{A^\top A \text{ is a symmetric } n \times n \text{ matrix,}}$ since $(A^\top A)^\top = A^\top (A^\top)^\top = A^\top A$.

It follows from our results last time that $A^\top A$ has all real eigenvalues.

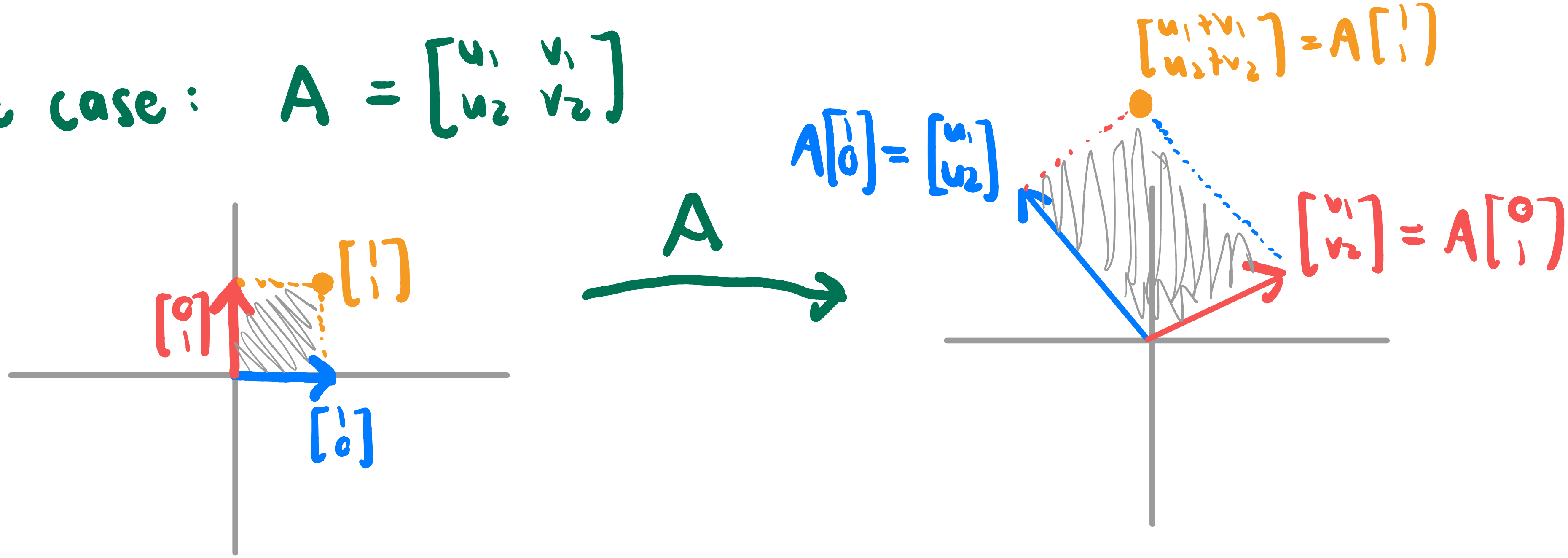
A stronger statement holds.

Will turn out that the numbers $\sigma_1 \geq \sigma_2$ (ellipse radii) on next slide are square roots of eigenvalues of $A^T A$

Idea of an SVD for a matrix A

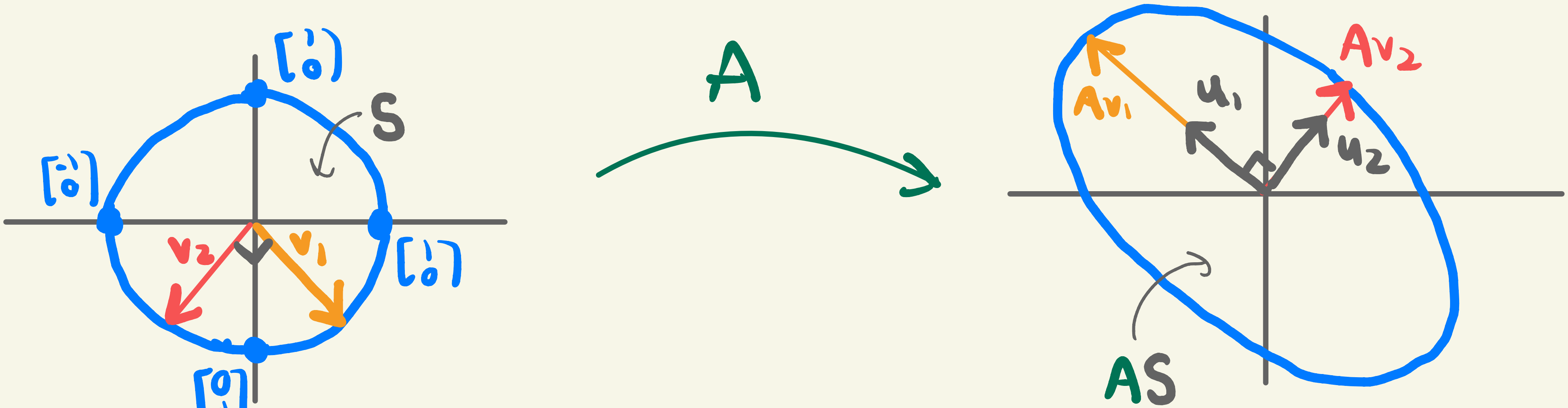
Usual way of writing a matrix encodes how it transforms unit square

2x2 case: $A = \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix}$



The SVD for a matrix instead encodes how the matrix transforms the unit circle (or n-dim sphere)

The SVD for a matrix instead encodes how the matrix transforms the unit circle (or n-dim sphere)



this is an ellipse centered at origin

v_1 v_2 are orthonormal basis for $\mathbb{R}^2$

u_1 u_2 are also orthonormal

$$u_i \stackrel{\text{def}}{=} \frac{1}{\|Av_i\|} Av_i$$

$$\sigma_i = \|Av_i\|$$

(not unique)

$$\sigma_1 \geq \sigma_2 \geq 0$$

nontrivial exercise

$$A = \underbrace{\begin{bmatrix} u_1 & u_2 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} v_1 & v_2 \end{bmatrix}^T}_{V^T}$$

call this decomp. an SVD for A

If $A^\top A v = \lambda v$ for $0 \neq v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$

then $\boxed{\lambda = \|Av\|^2 \geq 0}$

Lemma. All eigenvalues of $A^\top A$ are nonnegative real numbers. [In fact, if λ is an eigenvalue of $A^\top A$ and $v \in \mathbb{R}^n$ is a unit vector with $A^\top A v = \lambda v$, then $\lambda = \|Av\|^2$.]

Proof. If $v \in \mathbb{R}^n$ has $\|v\| = 1$ and $A^\top A v = \lambda v$ then

$$0 \leq \|Av\|^2 = (Av) \bullet (Av) = (Av)^\top (Av) = v^\top A^\top A v = v^\top (\lambda v) = \lambda (v^\top v) = \lambda \|v\|^2 = \lambda.$$

□

The preceding lemma allows us to make the following definition.

Definition. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of $A^\top A$ arranged in decreasing order. Define $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, 2, \ldots, n$.

The numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ are the *singular values* of A.

The singular values of a matrix A are **the squares roots of the eigenvalues of** $A^\top A$, which are guaranteed to be nonnegative (so have well-defined square roots).

$n = \#$ columns of A

$\det(A^\top A - xI) = (\lambda_1 - x) \cdots (\lambda_n - x)$

Ex Suppose $D = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix}$ is diagonal

Then $D^T D = \begin{bmatrix} d_1^2 & & & \\ & d_2^2 & & \\ & & \ddots & \\ & & & d_n^2 \end{bmatrix}$ is also diagonal

The eigvals of D are $d_1 \; d_2 \cdots d_n$

The eigvals of $D^T D$ are $d_1^2 \; d_2^2 \cdots d_n^2$

arranged in decreasing order

The singular values of D are $|d_1|, |d_2|, \ldots, |d_n|$

we have $\sigma_1 = \max\{|d_i|\}$ and $\sigma_n = \min\{|d_i|\}$

3 cols $\qquad\qquad\qquad\qquad$ 3×3

Example. Suppose $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$. Then $A^\top A = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$.

This matrix $A^\top A$ has characteristic polynomial

$$\det(A^\top A - xI) = (360 - x)(90 - x)x$$

so the eigenvalues of $A^\top A$ are $\boxed{\lambda_1 = 360,\ \lambda_2 = 90, \text{ and } \lambda_3 = 0.}$

The singular values of A are $\sigma_1 = \sqrt{360} = 6\sqrt{10} > \sigma_2 = \sqrt{90} = 3\sqrt{10} > \sigma_3 = 0$.

$$\begin{cases} \sigma_1 = \sqrt{360} \\ \sigma_2 = \sqrt{90} \\ \sigma_3 = 0 \end{cases}$$

<u>Note</u> 2 nonzero singular values

<u>Also</u> rank $A = 2$

Fact rank A = # of nonzero singular values

Setup: $\boxed{v_1\ v_2 \cdots v_r}$ is orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A^TA

A sequel to the lemma above, about the eigenvectors of $A^\top A$:

Suppose $v_1, v_2, \ldots, v_n$ is an orthonormal basis of $\mathbb{R}^n$ composed of eigenvectors of $A^\top A$. Assume that if $\lambda_i \in \mathbb{R}$ is eigenvalue of v_i then $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

Theorem. Suppose A has r nonzero singular values.

Then $Av_1, Av_2, \ldots, Av_r$ is an orthogonal basis for the column space of A.

Consequently rank $A = r$.

Proof. Choose indices $i \neq j$. Then $v_i \bullet v_j = 0$ so also $v_i \bullet \lambda_j v_j = 0$. Then

$$(Av_i)^\top Av_j = v_i^\top A^\top Av_j = v_i^\top(\lambda_j v_j) = v_i \bullet \lambda_j v_j = 0.$$

This shows that $Av_1, Av_2, \ldots, Av_r$ are orthogonal vectors in Col A.

Since $\|Av_i\| = \sqrt{\lambda_i} > 0$, these vectors are nonzero, so are linearly independent.

Intuitive reason: the vectors $v_1, v_2, \ldots, v_r$ (r = rank A) are orthog. radius vectors for unit n-dim sphere that became radius vectors of corresponding n-dim ellipse when multiplied by A

Fact from earlier: $\|Av_i\| = \sqrt{\lambda_i}$

To see that these vectors span the column space of A, suppose $y \in \mathrm{Col}\, A$. $\Rightarrow$ $y = Ax$ for some x

Then $y = Ax$ for some vector $x \in \mathbb{R}^n$, which we can write as

$$\boxed{x = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n}$$

as v's are basis for $\mathbb{R}^n$

for some $c_1, c_2, \ldots, c_n \in \mathbb{R}$.

$\boxed{\text{If } i > r \text{ then } Av_i = 0 \text{ since } \|Av_i\| = \sqrt{\lambda_i} = 0.}$ Therefore

$\lambda_i = 0$

$$y = Ax = c_1 Av_1 + c_2 Av_2 + \cdots + c_r Av_r + \underbrace{c_{r+1} Av_{r+1} + \cdots + c_n Av_n}_{=0} = c_1 Av_1 + c_2 Av_2 + \cdots + c_r Av_r.$$

We conclude that $\boxed{Av_1, Av_2, \ldots, Av_r}$ is a basis for $\mathrm{Col}\, A$. □

$\rightarrow$ span Col A (and are orthogonal)

* **Corollary.** For any matrix A: rank A = number of nonzero singular values of A.

orthogonal = (inverse = transpose) = (orthonormal columns + square)

Theorem (Existence of SVDs)**.** Let A be an $m \times n$ matrix with rank r.

Suppose $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ are the nonzero singular values of A.

Then we can write $A = U\Sigma V^\top$ where ← called an SVD for A

not unique

U is some $m \times m$ orthogonal matrix.

V is some $n \times n$ orthogonal matrix.

unique

Σ is the $m \times n$ matrix $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ where $D = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$.

Definition. A factorization $A = U\Sigma V^\top$ with U, V, Σ as above is a *singular value decomposition* of A. We sometimes abbreviate by writing *SVD* instead of *singular value decomposition*.

If A is $m \times n$ and $A = U\Sigma V^T$ is an SVD then

U is $m \times m$, Σ is $m \times n$, V is $n \times n$

U and V in an SVD $A = U\Sigma V^\top$ are not uniquely determined by A, but Σ is.

* The columns of U are called *left singular vectors* of A.

* The columns of V are called *right singular vectors* of A.

Proof that an SVD of A exists. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of $A^\top A$. ①

The singular values of A are $\sigma_i = \sqrt{\lambda_i}$ for each $i = 1, 2, \ldots, n$. ②

Let $v_1, v_2, \ldots, v_n$ be a list of corresponding orthonormal eigenvectors for $A^\top A$. ③

Then $\lambda_{r+1} = \cdots = \lambda_n = 0$ and $Av_1, Av_2, \ldots, Av_r$ is orthogonal basis for $\operatorname{Col} A$.

For each $i = 1, 2, \ldots, r$, define $\boxed{u_i = \frac{1}{\|Av_i\|}Av_i = \frac{1}{\sqrt{\lambda_i}}Av = \frac{1}{\sigma_i}Av_i.}$ ④

* If rank $A = m$, then we're done: $U = [u_1 \cdots u_m]$

$V = [v_1 \cdots v_n]$

$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \end{bmatrix}$

① Find eigvals of A^TA

② Take square roots

③ Find orthonormal eigvectors $\underbrace{v_1, v_2, \ldots, v_n}_{\in \mathbb{R}^n}$ for A^TA

④ Define $\underbrace{u_i}_{\in \mathbb{R}^m}$ = (unit vector in dir of Av_i) for $i = 1, 2, \ldots, \operatorname{rank} A = r$

Then $u_1, u_2, \ldots, u_r$ is an orthonormal basis for $\operatorname{Col} A$. If rank $A < m$:

*

We can choose vectors $u_{r+1}, u_{r+2}, \ldots, u_m \in \mathbb{R}^m$ such that the extended list of vectors $u_1, u_2, \ldots, u_m$ is an orthonormal basis for $\mathbb{R}^m$. Make any such choice, and define

$$U = \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}.$$

These matrices are orthogonal by construction, and

Check that it works

$$\begin{aligned} AV &= \begin{bmatrix} Av_1 & Av_2 & \ldots & Av_n \end{bmatrix} \\ &= \begin{bmatrix} Av_1 & Av_2 & \ldots & Av_r & 0 & \ldots & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \ldots & \sigma_r u_r & 0 & \ldots & 0 \end{bmatrix}. \end{aligned}$$

If Σ is the matrix given in the theorem, then we also have

$$U\Sigma = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \ldots & \sigma_r u_r & 0 & \ldots & 0 \end{bmatrix} = AV$$

so $U\Sigma V^\top = AVV^\top = AI = A$, which confirms the theorem statement. $\square$

$\text{rank}\, A = 2 = r$

$\sigma_1 = \sqrt{360} > \sigma_2 = \sqrt{90}$

Example. Again suppose $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$.

To find a singular value decomposition for A, there are three steps.

1. Find an orthogonal diagonalization of $A^\top A$.

   $A^\top A$ is a 3×3 matrix, and by the usual methods of row reducing $A - \lambda I$ to find a basis for $\text{Nul}(A - \lambda I)$ for each eigenvalue λ, you can check that

$$v_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \qquad \text{and} \qquad v_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

   is an orthonormal basis of $\mathbb{R}^3$ consisting of eigenvectors of $A^\top A$.

   The corresponding eigenvalues are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$.

2. Set up V and Σ. Following the proof of the theorem, we have

$$\boxed{V = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix}} \quad \text{and} \quad D = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

for $\sigma_1 = \sqrt{\lambda_1} = \sqrt{360}$ and $\sigma_2 = \sqrt{\lambda_2} = \sqrt{90}$.

Since Σ has the same size as A, we get $\boxed{\Sigma = \begin{bmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{bmatrix}}$.

3. Construct U. We have $U = \begin{bmatrix} u_1 & u_2 \end{bmatrix}$ where $u_i = \frac{1}{\sigma_i} A v_i$.

In this case you can compute that

$$\frac{1}{\sigma_1} A v_1 = u_1 = \frac{1}{\sqrt{360}}\begin{bmatrix} 18 \\ 6 \end{bmatrix} \quad \text{and} \quad u_2 = \frac{1}{\sqrt{90}}\begin{bmatrix} 3 \\ -9 \end{bmatrix} = \frac{1}{\sigma_2} A v_2$$

which means that we can write $\boxed{U = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & 1 \\ 1 & -3 \end{bmatrix}}$.

rank A = # rows (easy case)

Putting everything together produces the singular value decomposition.

Check

$$A = U\Sigma V^\top = \overset{U}{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}} \overset{\Sigma}{\begin{bmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{bmatrix}} \overset{V^T}{\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}}.$$

Be careful to note that the third matrix factor is the transpose $V^\top$ rather than V.

One application of SVDs is to show the existence of pseudo-inverses:

Definition. A *pseudo-inverse* of an $m \times n$ matrix A is an $n \times m$ matrix A^+ with

$$\boxed{AA^+A = A} \quad \text{and} \quad \boxed{A^+AA^+ = A^+}.$$

Example: If A is invertible, then $A^+ = A^{-1}$ is the pseudo-inverse of A.

Theorem. Every matrix A has a pseudo-inverse, which can be computed as follows. If $\boxed{A = U\Sigma V^\top}$ is a singular value decomposition, and Σ^+ is the matrix formed by transposing Σ and then replacing all of its nonzero entries by their reciprocals. Then $\boxed{A^+ = V\Sigma^+U^\top}$ is a pseudo-inverse for A.

Proof. See lecture notes. Key step is to check that $\Sigma\Sigma^+\Sigma = \Sigma$ and $\Sigma^+\Sigma\Sigma^+ = \Sigma^+$ $\square$

where $\Sigma^+ = \begin{bmatrix} \sigma_1^{-1} & & \\ & \sigma_2^{-1} & \\ & & \ddots \end{bmatrix}$ (same size as A^T)

Example. If

SVD

$$A = U\Sigma V^\top = \overset{U}{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}} \overset{\Sigma}{\begin{bmatrix} \sqrt{360} & 0 & 0 \\ 0 & \sqrt{90} & 0 \end{bmatrix}} \overset{V^T}{\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}}$$

then a pseudo-inverse is provided by

pseudo inverse

$$A^+ = \underset{V}{\begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}} \underset{\Sigma^+}{\begin{bmatrix} 1/\sqrt{360} & 0 \\ 0 & 1/\sqrt{90} \\ 0 & 0 \end{bmatrix}} \underset{U^T}{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}}.$$

One can show that the pseudo-inverse is unique (but we won't prove this).

satisfies $A^+ A A^+ = A^+$ and $A A^+ A = A$