Lecture #24 – MATH 2121

Outline:

1. Review SVD, defs and props
2. What does SVD actually mean

SVD: matrix factorization $A = U\Sigma V^T$
$= USV^T$

1 Last time: singular value decomposition

Let $\boxed{A \text{ be an } m \times n \text{ matrix.}}$

Then $A^\top A$ is a symmetric $n \times n$ matrix, eigenvalues all nonnegative real numbers.

If λ is an eigenvalue of $A^\top A$ and $v \in \mathbb{R}^n$ is a unit vector with $A^\top Av = \lambda v$, then

$$\lambda = \|Av\|^2. \geq 0$$

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 $\boxed{\sigma_i = \sqrt{\lambda_i}}$ for $i = 1, 2, \ldots, n$.

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

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

Def of an SVD

A has m rows (U is $m \times m$)

A has n cols (V is $n \times n$)

Remember that a matrix U is *orthogonal* if U is invertible and $U^{-1} = U^\top$.

Theorem. Suppose $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n$. (r nonzero singular values)

Then $\boxed{\operatorname{rank} A = r}$ and we can write $\boxed{A = U\Sigma V^\top}$ where

- U is some $m \times m$ orthogonal matrix.
- V is some $n \times n$ orthogonal matrix.
- Σ is the $m \times n$ matrix with σ_i in each position (i, i) and zeros elsewhere.

The factorization $A = U\Sigma V^\top$ is called a *singular value decomposition* or *SVD* of A.

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

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

A matrix A may have more than one SVD, but Σ will be the same in all of these.

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

Same size as A,
all unlabeled entries are zero

If A were invertible, then $AA^{-1}A = A$ and $A^{-1}AA^{-1} = A^{-1}$

same size as A^T

A *pseudo-inverse* of an $m \times n$ matrix A is an $n \times m$ matrix A^+ such that

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

If A is a square, invertible matrix, then $A^+ = A^{-1}$ is the pseudo-inverse of A.

Suppose $A = U\Sigma V^\top$ is a singular value decomposition.

Construct $\boxed{\Sigma^+ \text{ by transposing } \Sigma \text{ then replacing } \textbf{nonzero entries by reciprocals}.}$

Then $A^+ = V\Sigma^+U^\top$ is a pseudo-inverse for A.

$x \longmapsto \frac{1}{x}$

Convert an SVD $A = U\Sigma V^T$

to the formula $A^+ = V\Sigma^+ U^T$

where $\Sigma^+ = \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_r} \\ & & & \end{bmatrix}$ same size as A^T all unlabeled entries zero

$\underline{\text{Ex}}$ Suppose $A = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$. What's an SVD for A?

The eigenvalues of $A^T A = A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$ are $\lambda_1 = 4$, $\lambda_2 = 1$

so singular values of A are $\sigma_1 = 2$, $\sigma_2 = 1$ so

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} = \overbrace{\underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}}_{\Sigma}}^{\begin{bmatrix} 0 & 1 \\ -2 & 0 \end{bmatrix}} \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{V^T} = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$$

is an SVD for A.

Permutation matrices are orthogonal, and multiplying columns by ± 1 preserves orthogonality of a matrix

General algorithm to find an SVD:

factoring $\det(A^TA - xI)$

To find a singular value decomposition for an $m \times n$ matrix A:

1. Find the nonnegative eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ of $A^\top A$.

   Find a basis of eigenvectors for $A^\top A$ for each eigenspace. $\rightarrow$ $\mathrm{Nul}(A^TA - \lambda I)$

   Convert each basis to an **orthonormal** basis using the Gram-Schmidt process.

   Combine these to get an $\boxed{\text{orthonormal list of eigenvectors } v_1, v_2, \ldots, v_n.}$

construct V and Σ

2. Let $V = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}$.

   Form Σ as the $m \times n$ matrix with $\boxed{\sigma_i = \sqrt{\lambda_i}}$ in position (i, i) for $i = 1, 2, \ldots, m$.

3. Let $\boxed{u_i = \frac{1}{\sigma_i} A v_i}$ for $i = 1, 2, \ldots, r$ where $r = \operatorname{rank} A$ is maximal with $\sigma_r \neq 0$.

   Find vectors $u_{r+1}, u_{r+2}, \ldots, u_m \in \mathbb{R}^m$ such that $u_1, u_2, \ldots, u_m$ are orthonormal.
   This can be done by finding the pivot columns of $\begin{bmatrix} u_1 & u_2 & \ldots & u_r & e_1 & e_2 & \ldots & e_m \end{bmatrix}$ and then applying the Gram-Schmidt process. Finally let $\boxed{U = \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix}}$.

when $r < m$

step is only complicated when $r = \operatorname{rank}(A) < m$

Finally: $A = U \Sigma V^T$

<u>Ex</u> What's an SVD for $A = \begin{bmatrix} 0&0&1 \\ 1&0&0 \\ 0&1&0 \end{bmatrix}$?

A is a permutation matrix and so $A^{-1} = A^T$. Therefore

$$A^TA = \begin{bmatrix} 1&0&0 \\ 0&1&0 \\ 0&0&1 \end{bmatrix} \text{ has eigvals } \lambda_1 = \lambda_2 = \lambda_3 = 1$$

so A has sing vals $\sigma_1 = \sigma_2 = \sigma_3 = 1$ so an SVD is

$$A = \begin{bmatrix} 0&0&1 \\ 1&0&0 \\ 0&1&0 \end{bmatrix} = \underbrace{\begin{bmatrix} 0&0&1 \\ 1&0&0 \\ 0&1&0 \end{bmatrix}}_{\substack{U \\ \text{orthog}}} \underbrace{\begin{bmatrix} 1&& \\ &1& \\ &&1 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1&& \\ &1& \\ &&1 \end{bmatrix}}_{\substack{V^T \\ \text{orthog}}} = UV^T = U$$

More generally, whenever $A^{-1} = A^T$, one SVD is $A = \overset{U}{A}\,\overset{\Sigma}{I}\,\overset{V^T}{I}$

2 SVDs for symmetric matrices

When we first introduced singular value decompositions we said that they generalized the notion of "orthogonal diagonalization" for symmetric matrices.

How are SVDs are a generalization of the decomposition $\boxed{A = UDU^\top = UDU^{-1} = A^\top}$ that exists for a symmetric matrix?

Suppose $A = A^\top$ is an $n \times n$ symmetric matrix.

We know there are real numbers $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ such that

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

These are the eigenvalues of A. Some of these numbers could be negative.

$DE = \begin{bmatrix} |\lambda_1| & & \\ & |\lambda_2| & \\ & & \ddots \end{bmatrix}$

of $A = A^T$

decreasing abs value

Suppose the eigenvalues are ordered such that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| \geq 0$. Let

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} e_1 & & & \\ & e_2 & & \\ & & \ddots & \\ & & & e_n \end{bmatrix} \text{ where } e_i = \begin{cases} +1 & \text{if } \lambda_i \geq 0 \\ -1 & \text{if } \lambda_i < 0. \end{cases}$$

From previous lectures: there is an orthogonal $n \times n$ matrix U such that $A = UDU^\top$.

Proposition. An SVD for the symmetric matrix $A = A^\top = UDU^\top$ is

$A^2 = UD^2U^{-1} = UD^2U^T$

$$A = U\Sigma V^\top$$

where $\boxed{\Sigma = DE}$ and $\boxed{V = UE.}$ The singular values of A are $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$.

Proof. Need to check that $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_n^2$ are the eigenvalues of $\boxed{A^\top A = A^2.}$ This follows since A^2 is similar to D^2. Checking that $A = U\Sigma V^\top$ is routine. $\square$

Ex If $D = \begin{bmatrix} -5 & & & \\ & 5 & & \\ & & 4 & \\ & & & -1 \end{bmatrix}$ then $E = \begin{bmatrix} -1 & & & \\ & +1 & & \\ & & +1 & \\ & & & -1 \end{bmatrix}$

$\begin{cases} V^TV = E^TU^TUE = E^2 = I \\ U\Sigma V^T = UDEE^TU^T = UDU^T \end{cases}$

3 SVDs for 2×2 matrices

For intuition about what an SVD means, let's consider SVDs for 2×2 matrices.

It is possible describe all 2×2 orthogonal matrices in a simple way:

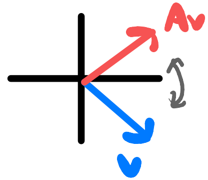
ccw rotation

Proposition. Every 2×2 orthogonal matrix has the form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

reflection

for an angle $0 \le \theta < 2\pi$.



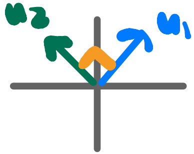
Proof. 1st column must be on the unit circle, so is of the form $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$.

2nd column must be of the unit vectors $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$ or $\begin{bmatrix} \sin\theta \\ -\cos\theta \end{bmatrix}$ orthogonal to 1st.

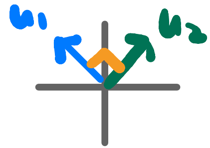
$\square$

$$\begin{bmatrix} x \\ y \end{bmatrix} \perp \begin{bmatrix} -y \\ x \end{bmatrix}$$

$U = [u_1 \; u_2]$ ⟷



or



$\det U = +1$ $\qquad$ $\det U = -1$

Suppose U is a 2×2 orthogonal matrix.

The columns of U are two perpendicular radii of the unit circle.

If the second column is 90 degrees counterclockwise from the first column, then

$$\det U = 1 \qquad \text{and} \qquad U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{for some angle } \theta.$$

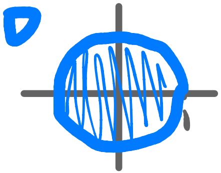
Otherwise, second column must be 90 degrees clockwise from the first column, so

$$\det U = -1 \qquad \text{and} \qquad U = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \quad \text{for some angle } \theta.$$

We can also describe the effect of the mapping $v \mapsto Uv$ for $v \in \mathbb{R}^2$ as follows:

- If $\det U = 1$ then v is rotated counter-clockwise by angle θ. → U is just rotation
- If $\det U = -1$ then v is reflected across $y = x$ then rotated by angle $\theta - \frac{\pi}{2}$.

↳ U combines rotation + reflection



AKA unit circle, unit sphere, unit ball, etc

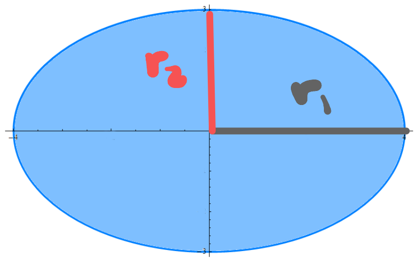
The *unit disc* $\mathcal{D}$ is the set of vectors $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$ with $v_1^2 + v_2^2 \leq 1$. $(v \cdot v \leq 1)$

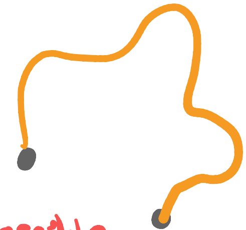
Fix real numbers $r_1, r_2 \geq 0$. Let $\mathcal{E}$ be set of vectors $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$ with

$$\boxed{(v_1/r_1)^2 + (v_2/r_2)^2 \leq 1.}$$

When $r_i = 0$ we consider $(v_i/r_i)^2$ to be zero if $v_i = 0$ and $+\infty$ if $v_i \neq 0$.

We call $\mathcal{E}$ a *(solid)* $\boxed{\textit{standard ellipse}}$.

$\mathcal{E} =$ 



possible to draw $\mathcal{E}$ by hand with string

standard ellipse

unit disc

Proposition. It holds that $\mathcal{E} = \left\{ \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} v : v \in \mathcal{D} \right\}$.

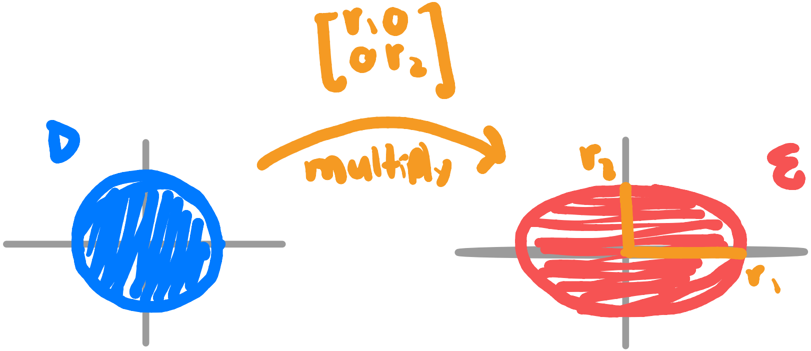
$\mathcal{E} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} \mathcal{D}$

Proof. $\begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} v = \begin{bmatrix} r_1 v_1 \\ r_2 v_2 \end{bmatrix} \in \mathcal{E}$ iff $(r_1 v_1 / r_1)^2 + (r_2 v_2 / r_2)^2 = v_1^2 + v_2^2 \leq 1$. □

The radii of $\mathcal{E}$ are the vectors $\pm \begin{bmatrix} r_1 \\ 0 \end{bmatrix}$ and $\pm \begin{bmatrix} 0 \\ r_2 \end{bmatrix}$.

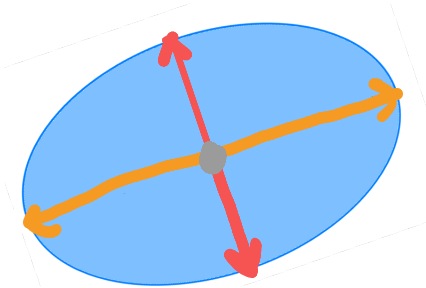
For each radius there is a choice of direction.

But any two orthogonal radii uniquely determine $\mathcal{E}$.



standard ellipse

More generally, we refer to any rotation of the region $\mathcal{E}$ as a *(solid) ellipse*.



The radii of an ellipse formed by rotating $\mathcal{E}$ by some angle θ counterclockwise are

$$\pm \begin{bmatrix} r_1 \cos\theta \\ r_1 \sin\theta \end{bmatrix} \qquad \text{and} \qquad \pm \begin{bmatrix} -r_2 \sin\theta \\ r_2 \cos\theta \end{bmatrix}$$

which are formed by rotating the radii of $\mathcal{E}$ counterclockwise by the same angle.

Any two orthogonal radii once again completely determine the ellipse.

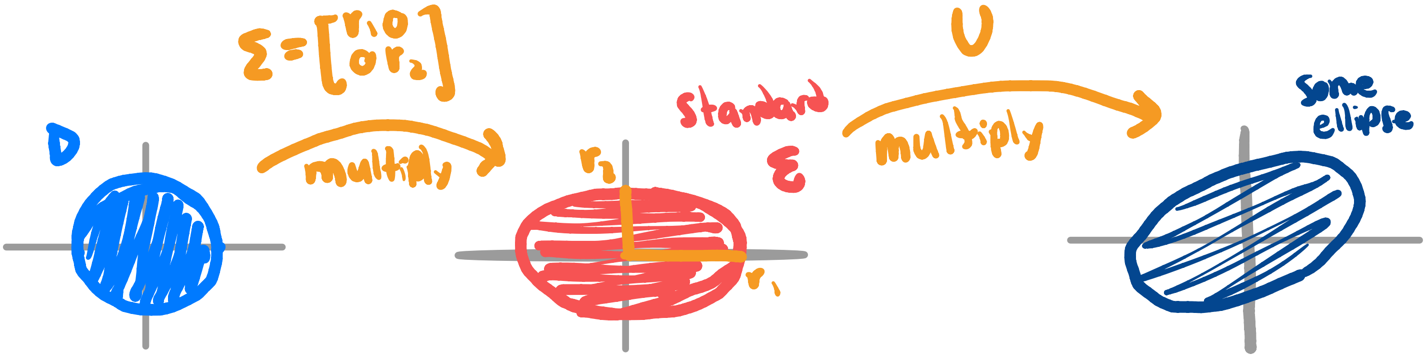
already know that ΣD is a standard ellipse since U is a combin. of rotation & reflection, result follow

Proposition. Suppose U is some orthogonal 2×2 matrix and $\Sigma = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}$. Then the set of vectors

$$\left\{ U\Sigma v \in \mathbb{R}^2 : v \in \mathcal{D} \right\} = U\Sigma D$$

is an ellipse with radii of lengths r_1 and r_2, and every such ellipse arises in this way.

Proof. Reflecting a standard ellipse across the line $y = x$ gives another standard ellipse. The result follows since $\{\Sigma v : v \in \mathcal{D}\}$ is a standard ellipse and U is a rotation matrix times a permutation matrix. $\square$



Cor. Taking $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ in previous prop, shows that $V^T D = D$ when V is orthogonal 2×2 as then V^T is also orthogonal

AD

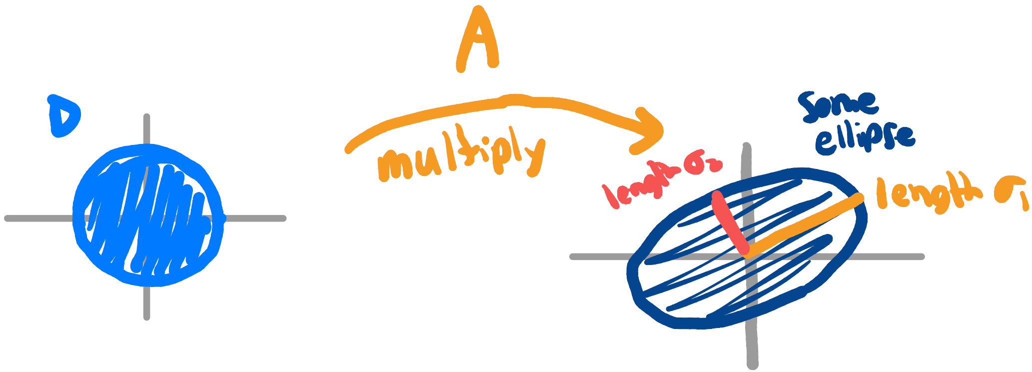
Proposition. Let A be a 2×2 matrix. Then the region $\{Av : v \in \mathcal{D}\}$ is an ellipse. The lengths of the radii of this ellipse are the singular values of A.

Proof. Let $\boxed{A = U\Sigma V^\top}$ be a singular value decomposition.

Then $\{V^\top v : v \in \mathcal{D}\} = \mathcal{D}$ as orthogonal matrices preserve lengths.

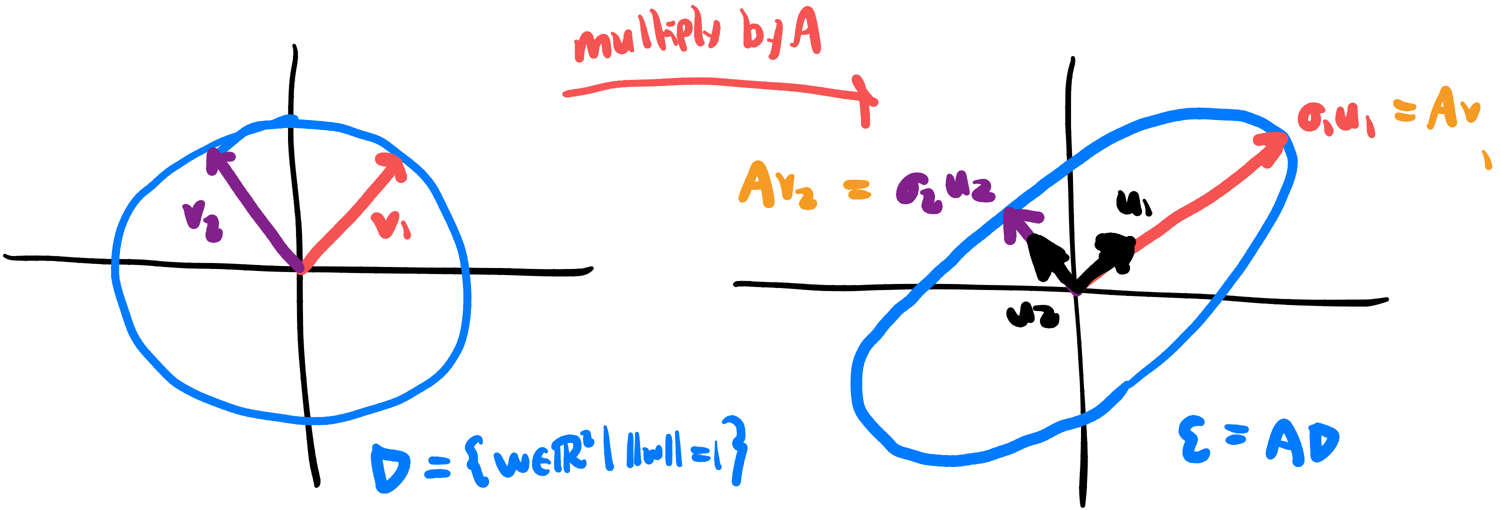
Thus $\{Av : v \in \mathcal{D}\} = \{U\Sigma v : v \in \mathcal{D}\}$ so result follows by previous proposition. $\square$

$AD = U\Sigma V^T D = U \Sigma D$ = ellipse w/ radii σ_1 and σ_2



Suppose an SVD for A is $A = \underbrace{\begin{bmatrix} u_1 & u_2 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} v_1 & v_2 \end{bmatrix}^T}_{V^T = V^{-1}}$

Then we have this picture:



Note: $Av_1 = U\Sigma V^{-1} v_1 = U\Sigma \begin{bmatrix} 1 \\ 0 \end{bmatrix} = U \begin{bmatrix} \sigma_1 \\ 0 \end{bmatrix} = \sigma_1 u_1$ likewise $Av_2 = \sigma_2 u_2$

Let's now try to say what a SVD $A = U\Sigma V^\top$ means physically for a 2×2 matrix.

Suppose the ellipse $\boxed{\mathcal{E} = \{Av : v \in \mathcal{D}\}}$ has radii of lengths $\sigma_1 \geq \sigma_2 \geq 0$.

As noted in the proposition, we then have $\boxed{\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}}$.

rescaling these give columns of U

The columns of $V = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ are two orthogonal radii of the unit disc $\mathcal{D}$.

These vectors have the property that Av_i is a radius of $\mathcal{E}$ with length r_i.

This holds as $Av_1 = U\Sigma V^\top v_1 = U\Sigma V^{-1} v_1 = U\Sigma e_1 = U \begin{bmatrix} \sigma_1 \\ 0 \end{bmatrix}$ and $Av_2 = U \begin{bmatrix} 0 \\ \sigma_2 \end{bmatrix}$.

The matrix U is always an orthogonal matrix whose inverse transforms the ellipse $\mathcal{E}$ back to a standard ellipse (whose radii belong to the x- and y-axes).

If $\det A$ and $\det V$ have the same sign then U is a rotation matrix.

Otherwise U is a rotation matrix with its columns interchanged.

A 2×2 matrix A parametrizes a linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ by telling us the images of the standard basis elements $e_1, e_2 \in \mathbb{R}^2$ (images are the columns of A).

The SVD of A parametrizes a linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ in a different way.

It tells us which orthogonal radii of the unit disc **(the columns of V)** are mapped to which orthogonal radii of the image ellipse (**the columns of** U give the directions of these radii and **the entries of** Σ give their lengths)

We can extend this interpretation of the SVD to higher dimensions, after setting

$$\mathcal{D}^n = \{v \in \mathbb{R}^n : v \bullet v \leq 1\}$$

defining an m-dimensional ellipse to be a set of the form $\{U\Sigma v : v \in \mathcal{D}^n\}$ where U is an orthogonal $m \times m$ matrix and Σ is an $m \times n$ matrix with nonzero entries only on the main diagonal.

If A is $m \times n$, then the first $r = \operatorname{rank} A$ columns of V in an SVD $A = U\Sigma V^\top$ are still orthogonal vectors of the unit disc that are transformed to orthogonal radii of some m-dimensional ellipse (in which $m - r$ radii have length zero), while the last $n - r$ columns are an orthogonal basis for $\operatorname{Nul} A$.