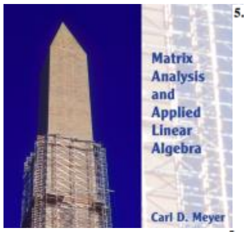
Lecture 13: Rotations and scaling (class)

Admin:



Concepts	Techniques	Decompositions
Vector space Basis	Gaussian elimination	LU decomp.
Inner products/Norm/Orthogonality		
Linear tranformations		
Projections Rank	Gram-Schmidt	QR decomp.

Next

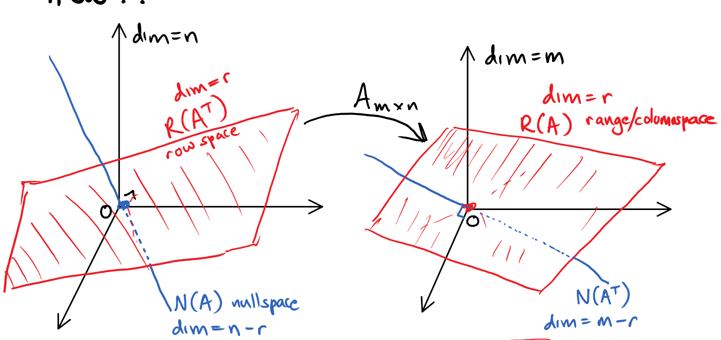
Singular values → Singular-value decomposition
Eigenvalues/vectors → Spectral decomposition

This week: SINGULAR VALUE DECOMPOSITION

Theoretical motivation:

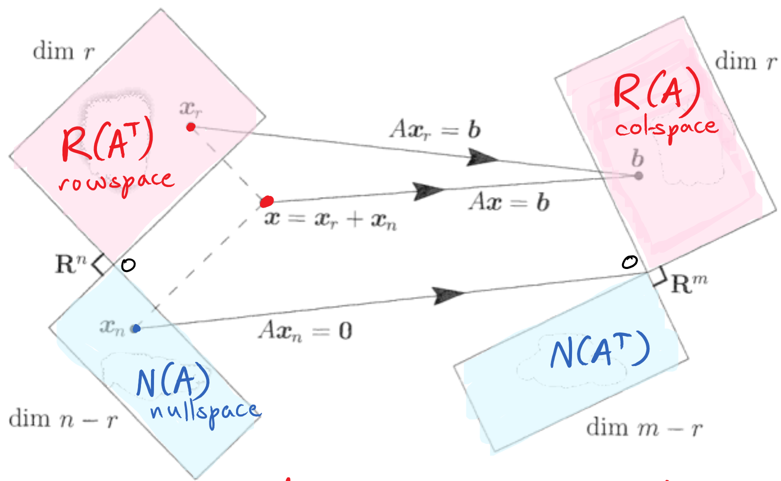
Any linear transformation A maps points in the rowspace $R(A^T)$ to distinct points in the columnspace $R(A)$. [Rank-Nullity thm.]

How??



$$\boxed{\dim R(A^T) = \dim R(A)}$$

$$\boxed{\begin{array}{l}\dim N(A) + \dim R(A^T) \\ = \text{total dimension } n\end{array}} \qquad \boxed{\begin{array}{l}\dim R(A) + \dim N(A^T) \\ = \text{total dimension } m\end{array}}$$



$N(A) = R(A^T)^{\perp}$ $\qquad$ $N(A^T) = R(A)^{\perp}$

Practical motivation: Many applications, including

* Solving linear equations $Ax = b$ → What is the sensitivity, e.g., to numerical errors?
  - → Find the shortest solution
  - → When there is no solution, find x to minimize $\|Ax - b\|$ → Least-squares regression analysis
* Rank minimization → Principal Component Analysis (PCA)
  - Data mining, clustering, recommendation systems, ...

Singular-Value Decomposition (SVD)

Informally: Any linear transformation can be split into:
- a rotation, followed by
- scaling vectors in or out

Before stating the theorem formally, we'll consider these pieces.

Isometries

Definition: An isometry is a linear transformation that preserves length. (iso = same, metric = length/distance)

(That is, $\|A\vec{x}\| = \|\vec{x}\|$ for all $\vec{x}$.)

Examples:

- Identity matrix I
- Rotations, e.g., $\begin{pmatrix} \cos\theta & 0 & +\sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$
- Reflections, e.g., $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ $\qquad$ $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

- Reflections, e.g., $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ $\quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

- Products of ~~rotations and reflections~~ isometries

$\forall x, \|AB\vec{x}\| = \|B\vec{x}\| = \|\vec{x}\|$

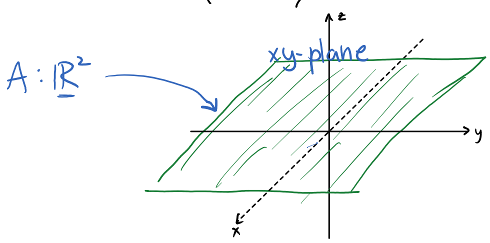
$\mathbb{R}^2 \not\subseteq \mathbb{R}^3$

$\begin{pmatrix} x \\ y \end{pmatrix} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix}$

- Isometric "embeddings", e.g.

$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ puts $\mathbb{R}^2$ into $\mathbb{R}^3$ as the xy-plane

$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \\ 0 & 0 \end{pmatrix}$ ✓



$\begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ also maps $\mathbb{R}^2$ to the xy-plane of $\mathbb{R}^3$ — but does not preserve lengths

Ex: Give an isometric embedding of $\mathbb{R}^2$ into

$$V = N\big((1 \ \ 1 \ \ 1)\big) \subseteq \mathbb{R}^3$$

Answer: $= \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \,\middle|\, x+y+z=0 \right\}$

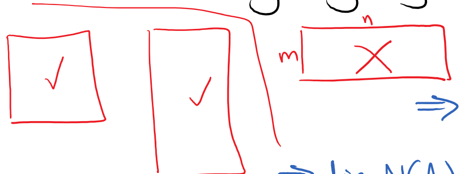
$= \operatorname{Span}\left\{ \underbrace{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}}_{\vec{v}_1}, \underbrace{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}}_{\vec{v}_2} \right\}$

$A = \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}$ This is NOT an isometry

$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$

$A = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{6} \\ 0 & -2/\sqrt{6} \end{pmatrix}$ ✓

- Not an isometry: anything that reduces the dimension



with $m < n$

$\Rightarrow \operatorname{rank}(A) = \#$ lin. indep. rows $\leq m$

$\Rightarrow \dim N(A) = n - \operatorname{rank}(A) > 0$

$\Rightarrow$ lengths of nonzero vectors in $N(A)$ are sent to 0 — not preserved.

Claim: Preserves length $\Rightarrow$ preserves angles.

Proof: Recall the angle θ between real vectors $\vec{x}$ and $\vec{y}$ satisfies $\quad \cos\theta = \dfrac{x \cdot y}{\|x\| \cdot \|y\|}$.

$\Rightarrow$ We have to show that dot products are preserved.

Assume A is an isometry. Goal: $\forall x, y \ \ (Ax)\cdot(Ay) = x \cdot y$ ✓

Trick: Look at $x+y$

$$\|x+y\|^2 = (x+y)\cdot(x+y) = \|x\|^2 + \|y\|^2 + \underbrace{x\cdot y + y \cdot x}_{2\cdot \operatorname{Re}(x\cdot y)}$$

$\|$

$$\|A(\dots)\|^2 = \big(A(x+y)\big)\cdot\big(A(x+y)\big) = \|Ax\|^2 + \|Ay\|^2 + Ax\cdot Ay + Ay \cdot Ax$$

$\|$

$$\|A(x+y)\|^2 = (A(x+y))\cdot(A(x+y)) = \|Ax\|^2 + \|Ay\|^2 + \underbrace{Ax\cdot Ay + Ay\cdot Ax}_{2\mathrm{Re}((Ax)\cdot(Ay))}$$

$$= (Ax+Ay)\cdot(Ax+Ay)$$

$$\Rightarrow \mathrm{Re}(x\cdot y) = \mathrm{Re}(Ax\cdot Ay)$$

$$i(\cancel{a}+bi) - i(\cancel{a}-bi) = -2b$$

$$\|\vec{x}+i\vec{y}\|^2 = \|x\|^2 + \|y\|^2 + \underbrace{i\, x\cdot y - i\, y\cdot x}_{-2\mathrm{Im}(x\cdot y)}$$

$\|$

$$\|A(x+iy)\|^2 = \|Ax\|^2 + \|Ay\|^2 - 2\,\mathrm{Im}(Ax\cdot Ay) \Rightarrow \mathrm{Im}(x\cdot y) = \mathrm{Im}(Ax\cdot Ay)$$

$$\Rightarrow Ax\cdot Ay = x\cdot y \checkmark$$

How to tell if a matrix is an isometry?

$$A = \begin{pmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{pmatrix} = \sum_{j=1}^{n} \vec{v}_j \vec{e}_j^T \qquad A^TA = \sum_{j,k} (v_j e_j^T)^T (v_k e_k^T) = \sum_{j,k} e_j e_k^T \cdot (\vec{v}_j\cdot\vec{v}_k)$$

Claim: A is an isometry $\Leftrightarrow$ The columns are orthonormal $\Updownarrow$ $A^TA = I$

Proof: $\Leftarrow$: Assume $v_j\cdot v_k = \delta_{jk} = \begin{cases} 1 & \text{if } j=k \\ 0 & \text{if } j\neq k \end{cases}$

Consider any $\vec{x} = \sum_j x_j \vec{e}_j$, $\|x\|^2 = \sum_j x_j^2$

$$\|A\vec{x}\|^2 = \|\sum_j x_j \vec{v}_j\|^2 = \sum_j x_j^2 = \|x\|^2 \Rightarrow A \text{ is isometry} \checkmark$$

$\Rightarrow$: Assume A isometry. Goal: Show columns orthonormal.

Why? $\|e_j\| = 1 = \|Ae_j\| = \|v_j\|$

$j\neq k$: $e_j\cdot e_k = 0 = (Ae_j)\cdot(Ae_k) = v_j\cdot v_k$ $\checkmark$

In matrix notation:

$$\underbrace{\begin{pmatrix} \text{—} & v_1^t & \text{—} \\ & \vdots & \\ \text{—} & v_n^t & \text{—} \end{pmatrix}}_{A^t} \underbrace{\begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}}_{A} = \begin{pmatrix} v_1\cdot v_1 & v_1\cdot v_2 & v_1\cdot v_3 & \cdots \\ v_2\cdot v_1 & v_2\cdot v_2 & v_2\cdot v_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & \\ 0 & 1 & 0 & \cdots \\ \vdots & & \ddots & \end{pmatrix} = I$$

Thus an isometry takes one orthonormal set of vectors (the standard basis) into another orthonormal set (the columns).

Exercise: Prove the converse implication:
If the columns of A are orthonormal, then A is an isometry.

$$\boxed{\begin{array}{ccc} A \text{ is an isometry} & \Longleftrightarrow & A \text{ preserves lengths and angles} \\ \Updownarrow & & \Updownarrow \\ A\text{'s columns are orthonormal} & \Longleftrightarrow & A^tA = I \end{array}}$$

Examples: $\begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$

Exercise: Give an isometry from $\mathbb{R}$ to the line $L = \{(x,y,z) \mid x=y=z\} \subset \mathbb{R}^3$.

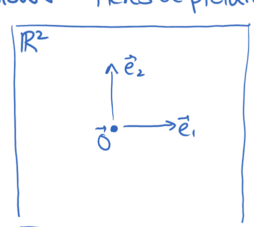
Answer: The line L consists of all multiples of the unit vector $\frac{1}{\sqrt{3}}(1,1,1)$. Therefore, the matrices

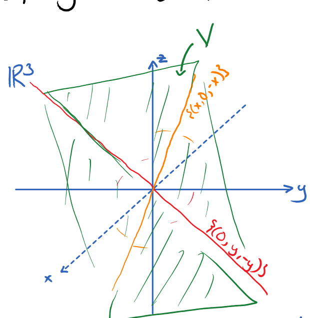
$$\frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \frac{-1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

are both isometries from $\mathbb{R}$ to L.
(And these are the only such isometries.) ✓

Exercise: Give an isometry from $\mathbb{R}^2$ to the plane $V = \{(x,y,z) \mid x+y+z=0\} \subset \mathbb{R}^3$.

Answer: Here's a picture:





To map the plane $\mathbb{R}^2$ isometrically into the plane V, we just need to map $\vec{e}_1$ and $\vec{e}_2$ into two perpendicular unit vectors in V. The isometry will take

$\vec{e}_1 \longmapsto$ first unit vector in $V \equiv \vec{u}$

$\vec{e}_2 \longmapsto$ 2^{nd} unit vector in $V. \equiv \vec{v}$

How to find $\vec{u}$ and $\vec{v}$?

- $\vec{u}$ can be an arbitrary unit vector
  e.g., start with
  $$(1,-1,0) \in V,$$
  and normalize:
  $$\vec{u} = \frac{1}{\sqrt{2}}(1,-1,0).$$
- $\vec{v} = (v_1, v_2, v_3)$ has to lie in V and be perpendicular to $\vec{u}$:
  $$v_1 + v_2 + v_3 = 0 \qquad (\vec{v} \in V)$$
  $$v_1 - v_2 = 0 \qquad (\vec{v} \cdot \vec{u} = 0)$$
  $$\Rightarrow \vec{v} = (1,1,-2)/\underbrace{\sqrt{6}}_{\text{normalization}} \quad \text{works}$$

What is the matrix for our isometry?

$\vec{e}_1 \mapsto \vec{u}$, $\vec{e}_2 \mapsto \vec{v}$

$$A = \begin{pmatrix} | & | \\ \vec{u} & \vec{v} \\ | & | \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{6} \\ 0 & -2/\sqrt{6} \end{pmatrix} \quad ✓$$

(Of course, this answer is not unique. We can also rotate or reflect the plane.)

Short answer:

An orthonormal basis for V is
$$\tfrac{1}{\sqrt{2}}(1,-1,0),\ \tfrac{1}{\sqrt{6}}(1,1,-2).$$
Therefore,
$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{2}{\sqrt{6}} \end{pmatrix} \text{ isometrically maps } \mathbb{R}^2 \text{ onto } V.$$

Of course, this answer is not unique. We could have rotated things around — and used any orthonormal basis for V. ✓

Orthogonal and Unitary Matrices

Definition: An "orthogonal" matrix is a square matrix isometry (ie., $n \times n$).

change of basis matrices that preserve lengths

Recall: The columns of an isometry are orthonormal,
$$A^T A = I.$$

Proposition: The rows of an orthogonal matrix are also orthonormal,
$$A A^T = I.$$

Corollary:

Orthogonal matrix
$$A^T = A^{-1}$$

(10)

rows are not orthogonal for isometric embeddings like $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$

Proof: Let $A = \begin{pmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{pmatrix}$

A maps $\vec{e}_i$ to $\vec{v}_i$

Equivalently, in more compact notation,

$$A = \sum_{j=1}^{n} \vec{v}_j \vec{e}_j^T \qquad \begin{pmatrix} \vec{v}_j \end{pmatrix} \left(- \vec{e}_j^T - \right)$$

Why? Check it:

$$\begin{aligned} A\vec{e}_i &= \Big(\sum_{j=1}^{n} \vec{v}_j \vec{e}_j^T\Big)\vec{e}_i \\ &= \sum_j \vec{v}_j (\vec{e}_j^T \vec{e}_i) \\ &= \sum_j \vec{v}_j \underbrace{(\vec{e}_j \cdot \vec{e}_i)}_{= \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}} \\ &= \vec{v}_j \quad \checkmark \end{aligned}$$

$$\begin{aligned} \Rightarrow A^T A &= \Big(\sum_j v_j e_j^T\Big)^T \Big(\sum_k v_k e_k^T\Big) \\ &= \sum_{j,k} e_j \underbrace{v_j^T v_k}_{= v_j \cdot v_k = \begin{cases} 1 & \text{if } j=k \\ 0 & \text{if } j \neq k \end{cases}} e_k^T \\ &= \sum_j e_j e_j^T \end{aligned}$$

Note: $e_1 e_1^T = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix} (1\ 0\ 0 \cdots) = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \ddots \\ & \vdots & & \end{pmatrix}$

$e_2 e_2^T = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} (0\ 1\ 0 \cdots 0) = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & \ddots \\ & \vdots & & \end{pmatrix}$

$\vdots$

$\Rightarrow e_1 e_1^T + e_2 e_2^T + \cdots + e_n e_n^T = I$ the identity!

Next let's compute AA^T:

$$\begin{aligned} AA^T &= \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \begin{pmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_n^T & - \end{pmatrix} \\ &= \Big(\sum_j v_j e_j^T\Big)\Big(\sum_k v_k e_k^T\Big)^T \\ &= \sum_{j,k} v_j (e_j \cdot e_k) v_k^T \\ &= \sum_{j=1}^{n} v_j v_j^T \end{aligned}$$

Claim: This is the identity again.

Why?

Call it M.

For any $i = 1, 2, \ldots, n$,

$$M\vec{v}_i = \sum_j v_j v_j^T v_i = \vec{v}_i \quad \checkmark$$

so the vectors $\vec{v}_1, \ldots, \vec{v}_n$ are all left alone.

Any other vector can be expanded out in terms of them, like

$$\vec{u} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_n \vec{v}_n$$

$$\Rightarrow M\vec{u} = \alpha_1 M\vec{v}_1 + \cdots + \alpha_n M\vec{v}_n = \vec{u} \checkmark \qquad A^T = A^{-1} \ \boxed{\checkmark}$$

Definition: An $n \times n$ complex isometry is called "unitary".

Orthogonal matrix $A^T = A^{-1}$

Unitary matrix $A^\dagger = A^{-1}$

More examples:

- Permutation matrices, e.g.,
$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} : \quad \begin{array}{l} e_1 \mapsto e_2 \\ e_2 \mapsto e_3 \\ e_3 \mapsto e_4 \\ e_4 \mapsto e_1 \end{array}$$

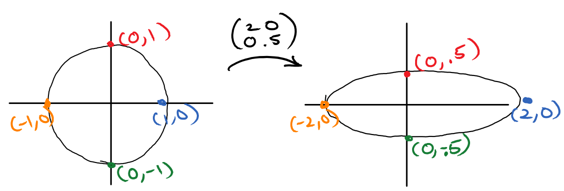
- Rotations, e.g.
$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$ rotates the plane $\mathbb{R}^2$ counterclockwise by angle θ
$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$ rotates $\mathbb{R}^3$ by θ about the z-axis

- $\frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1+i & -1+i \end{pmatrix}$ ✓

SCALING I.

$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ scales every vector up by 2

$\begin{pmatrix} 2 & 0 \\ 0 & .5 \end{pmatrix}$ scales by different amounts



Need not be axis-aligned...

Exercise: Give a 2×2 matrix that maps
$$(1,-1) \longmapsto (2,-2) \ ✓$$
$$\perp \quad (1,1) \longmapsto (\tfrac{1}{2}, \tfrac{1}{2})$$

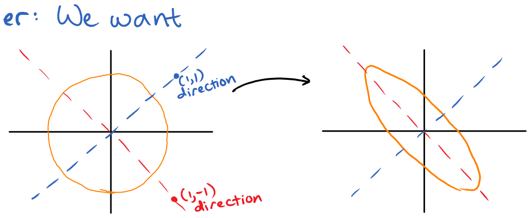
Better answer:

$\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

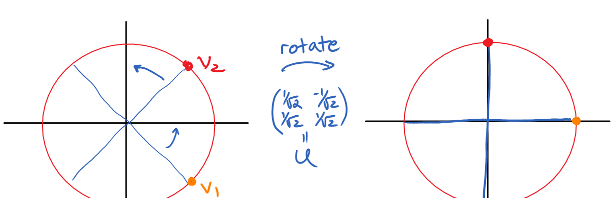
$\vec{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$$A = \boxed{\tfrac{1}{2}\vec{v}_1\vec{v}_1^T + 2\vec{v}_2\vec{v}_2^T} = \tfrac{1}{2}\cdot\tfrac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + 2\cdot\tfrac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \boxed{\begin{pmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{pmatrix}}$$

Answer: We want



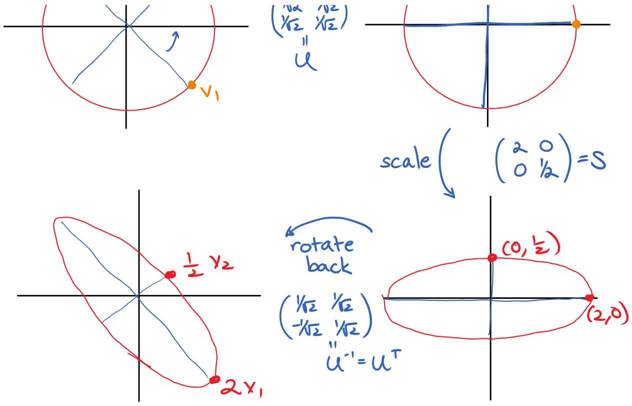
This is the same as above, but rotated by $\pi/4$.



More generally, let $v_1, \dots, v_n$ be an orthonormal basis

let $\sigma_1, \dots, \sigma_n \in \mathbb{R}$

let $A = \sum_i \sigma_i \vec{v}_i \vec{v}_i^T$ ✓ $\boxed{A} = \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{pmatrix}$



$\Rightarrow$ $\boxed{U^T S U}$ works

Alternative answer:

let $\sigma_1, \dots, \sigma_n \in \mathbb{R}$

let $A = \sum_j \sigma_j \vec{v}_j \vec{v}_j^T$ ✓ $\quad [A]_{v_1 \dots v_n} = \begin{pmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \ddots \\ & & & \sigma_n \end{pmatrix}$

Then

$$A\vec{v}_i = \sum_j \sigma_j v_j \underbrace{(v_j^T v_i)}_{\delta_{ij}} = \sigma_i \vec{v}_i$$

SCALING II: MATRIX NORM

Definition: The spectral norm of a linear transformation A is given by

$$\|A\| = \max_{\vec{x} \neq 0} \frac{\|A\vec{x}\|}{\|\vec{x}\|} = \max_{\vec{x},\, \|x\|=1} \|Ax\|$$

$$\frac{\|Ax\|}{\|x\|} = \frac{\|A \cdot c\vec{x}\|}{\|c\vec{x}\|} \quad \text{for any } c \neq 0$$

(It measures the maximum stretch of the matrix.)
(In finite dimensions, the max exists, is $< \infty$.)

Note: Often denoted $\|A\|_2$, for l_2/Euclidean norm.

Properties I:

- For any vector $\vec{x}$ (of appropriate dimension), $\|A\vec{x}\| \leq \|A\| \cdot \|\vec{x}\|$ Why? $\quad \frac{\|Ax\|}{\|x\|} \leq \max_{z \neq 0} \frac{\|Az\|}{\|z\|} = \|A\|$ ✓
- For any real/complex number α, $\|\alpha A\| = |\alpha| \cdot \|A\|$.
- Triangle inequality: $\|A+B\| \leq \|A\| + \|B\|$.

Proof:
$$\begin{aligned} \|A+B\| &= \max_{x : \|x\|=1} \|Ax + Bx\| \\ &\leq \max_{\|x\|=1} (\|Ax\| + \|Bx\|) \quad (\Delta \text{ ineq. for vectors}) \\ &\leq \left(\max_{\|x\|=1} \|Ax\|\right) + \left(\max_{\|y\|=1} \|By\|\right) \\ &= \|A\| + \|B\|. \quad \square \end{aligned}$$

- $\|AB\| \leq \|A\| \cdot \|B\|$

Proof: $\|A(B\vec{x})\| \leq \|A\| \cdot \|Bx\| \leq \|A\| \cdot \|B\| \cdot \|x\|$ ✓

Examples:

- $\|I\| = 1$
  $\|\text{any isometry}\| = 1$
  $\|\text{any projection}\| = 1$, unless the projection is 0
- What is $\left\| \underbrace{\begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix}}_{A} \right\|$? $= 1 + \varepsilon$?

$A = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix} = I + \varepsilon \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

If A and C are isometries
$\|ABC\| \leq \underbrace{\|A\|}_{1} \cdot \|B\| \cdot \underbrace{\|C\|}_{1} = \|B\|$
but $\|ABC\| < \|B\|$ is possible!

If C is unitary/orthogonal, then $\|ABC\| = \|B\|$.

Exercise: Check this.

Exercise: Check this.

$$\|A\| \le \|I\| + |\varepsilon| \cdot \left\|\begin{pmatrix}0&1\\1&0\end{pmatrix}\right\| = 1 + |\varepsilon|$$

$$\|A\| \geq \frac{\|A\vec{x}\|}{\|x\|} \text{ for any } \vec{x} \neq \vec{0}, \quad \|A\| \geq \|A\vec{e}_1\| = \left\|\begin{pmatrix}1\\ \varepsilon\end{pmatrix}\right\| = \sqrt{1+\varepsilon^2} \approx 1 + \tfrac{1}{2}\varepsilon^2 + O(\varepsilon^4)$$

Exact calculation:

$$\|A\|^2 = \max_{\vec{x},\|x\|=1} \|Ax\|^2 = \max_{\theta\in[0,2\pi)} \left\| A\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}\right\|^2$$

$$(c+\varepsilon s)^2 + (\varepsilon c + s)^2 = \begin{pmatrix}c^2+\varepsilon^2 s^2\\ +2\varepsilon cs\end{pmatrix} + \begin{pmatrix}\varepsilon^2 c^2 + s^2\\ +2\varepsilon cs\end{pmatrix}$$

$$= 1 + \varepsilon^2 + 4\varepsilon \cos\theta \cdot \sin\theta$$

$$\theta = \frac{\pi}{4}$$

$$= 1 + \varepsilon^2 + 4|\varepsilon| \cdot \frac{1}{2} = (1+|\varepsilon|)^2$$

$$\Rightarrow \|A\| = 1 + |\varepsilon| \qquad \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$$

Observe: $\left\|A\begin{pmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix}\right\| = 1+\varepsilon > \left\|A\begin{pmatrix}1\\0\end{pmatrix}\right\| = \left\|A\begin{pmatrix}0\\1\end{pmatrix}\right\| = \sqrt{1+\varepsilon^2}$ ✓

Moral: Spreading out is good!

Problem: What are the operator norms of

ⓐ $\begin{pmatrix}1 & 0\\ 0 & 1/100\end{pmatrix}$ ⓑ $\begin{pmatrix}1 & 1/100\\ 0 & 0\end{pmatrix}$?

$$\left\|\begin{pmatrix}a_1 & & & 0\\ & a_2 & & \\ & & a_3 & \\ 0 & & & a_4\end{pmatrix}\right\| = \max_j |a_j| \checkmark$$

Proof: $\|A\| \geq \|A\vec{e}_j\| = |a_j|$

$$\|A\|^2 = \max_{x,\|x\|=1} \|A\vec{x}\|^2 = \max \sum_j a_j^2 x_j^2 \quad \text{s.t. } \sum_j x_j^2 = 1$$

let $p_j = x_j^2$

Observe:

- In (a), you don't want to spread out, since there is no interaction between the two blocks of the matrix.
  - In general,

$$\left\|\begin{pmatrix}A & 0\\ 0 & B\end{pmatrix}\right\| = \max\{\|A\|, \|B\|\}$$

$$\left\|\begin{pmatrix}A & 0 & 0\\ 0 & B & 0\\ 0 & 0 & C\end{pmatrix}\right\| = \max\{\|A\|, \|B\|, \|C\|\}, \text{ etc.}$$

e.g. $\left\|\begin{pmatrix}00 & \varepsilon A\\ 00 & \varepsilon\,1\\ B\,0 & 00\\ 0\,1 & 00\end{pmatrix}\right\| = \max\{\|A\|, \|B\|\}$

$$\left\|\begin{pmatrix}A & \varepsilon C\\ 0 & B\end{pmatrix}\right\| > \max\{\|A\|, \|B\|\}$$

- In (b), even though $\|Ae_2\| = \varepsilon \ll \|Ae_1\| = 1$, you still want to spread between the two columns to maximize the norm.

- Also, in general,
  spectral matrix norm of a $1\times n$ matrix
  $=$ Euclidean norm of the row vector
  (to maximize $|\vec{v}\cdot\vec{x}|$, let $\vec{x} = \vec{v}/\|\vec{v}\|$.)

  & spectral norm of an $n\times 1$ matrix
  $=$ Euclidean norm of the column vector
  (just set $\vec{x} = (1)$.)

Example: What is the spectral norm of the $m\times n$ all-ones matrix?

$$\begin{array}{cc} m & \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \\ & n \end{array}$$

Answer:

① Experiment numerically:

```
octave:1> m = 10;
octave:2> n = 15;
octave:3> A = ones(m,n)
A =

   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1

octave:4> norm(A)
ans =  12.247
octave:5> norm(A)^2
ans =  150.00
```

$\Rightarrow$ maybe $\|A\| = \sqrt{m\cdot n}$?

Mathematica code:

```
In[25]:= Table[
   Norm[ConstantArray[1, {m, n}]]^2,
   {m, 1, 5}, {n, 1, 5}
  ] // MatrixForm
```

t[25]//MatrixForm=

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 6 & 8 & 10 \\ 3 & 6 & 9 & 12 & 15 \\ 4 & 8 & 12 & 16 & 20 \\ 5 & 10 & 15 & 20 & 25 \end{pmatrix}$$

② Guess the best input:

Since the columns are all the same, it makes sense to spread out across them all, and equally:

Let $\vec{x} = \frac{1}{\sqrt{n}}(1, 1, 1, \ldots, 1) \in \mathbb{R}^n$

$\Rightarrow \|\vec{x}\| = 1.$

$$A\vec{x} = \frac{1}{\sqrt{n}}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$= \frac{1}{\sqrt{n}}(n, n, n, \ldots, n) \in \mathbb{R}^m$$

$\Rightarrow \|A\vec{x}\|^2 = m\cdot n$ ✓

$\Rightarrow \|A\| \geq \sqrt{m\cdot n}$

③ Prove that $\|A\| = \sqrt{m\cdot n}$:

One approach is to argue by symmetry that the above $\vec{x}$ is optimal.

Alternatively, notice that $\operatorname{rank}(A) = 1$.

Since all columns are the same,

$\operatorname{rank}(A) = \dim R(A) = \#$ linearly independent columns $= 1$.

A factors as

$$A = \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{u \in \mathbb{R}^m} \underbrace{(1\ 1\ 1\ \cdots\ 1)}_{v^T \in \mathbb{R}^n} = \vec{u}\,\vec{v}^T$$

$\Rightarrow A\vec{x} = (\vec{v} \cdot \vec{x})\,\vec{u}$

$\|Ax\| = |v \cdot x| \cdot \|u\|$,

which reaches its maximum, $\|u\| \cdot \|v\|$,

for $x = \frac{v}{\|v\|}$.

$\Rightarrow \|A\| = \|u\| \cdot \|v\| = \sqrt{m} \cdot \sqrt{n}$. ✓

Observe: Any rank-one matrix A can be factored as

$$A = \vec{u}\,\vec{v}^T$$

for some vectors $\vec{u}$ and $\vec{v}$. Hence $\|A\| = \|u\| \cdot \|v\|$.

Spectral norm

Properties II.

- $\|A\| \geq 0$, and $\|A\| = 0 \Leftrightarrow A = 0$
- $\|A\vec{x}\| \leq \|A\| \cdot \|\vec{x}\|$
  matrix norm   vector/matrix norm
- $\|\alpha A\| = |\alpha| \cdot \|A\|$ for $\alpha \in \mathbb{C}$
- $\|AB\| \leq \|A\| \cdot \|B\|$

  (the amount you can stretch an input by applying AB is at most the stretch from applying B times the stretch from applying A.)
- If U and V are unitary, $\|U\| = \|V\| = 1$ and $\|UAV\| = \|A\|$

  (because unitaries don't change lengths).
- $\left\| \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \right\| = \max\{\|A\|, \|B\|\}$

  e.g., if A is a diagonal matrix, $\|A\| = \max_i |a_{ii}|$.
- If $\operatorname{rank}(A) = 1$, with $A = \vec{u}\,\vec{v}^T$, $\|A\| = \|u\| \cdot \|v\|$.

(extra material)

A fast (crude) estimate for the spectral norm:

Claim: For any $m \times n$ matrix $A = (a_{ij})$,

$$\max_{i,j} |a_{ij}| \leq \|A\| \leq \sqrt{\sum_{i,j} |a_{ij}|^2} \leq \sqrt{m \cdot n} \cdot \max_{i,j} |a_{ij}|.$$

Observe: For the all-ones matrix, the upper bound ($\sqrt{mn}$) is tight, though the lower bound (1) is terrible.

Proof: Start by showing the lower bound, $\|A\| \geq \max_{i,j} |a_{ij}|$.

Let i^*, j^* be such that $|a_{i^*j^*}| = \max_{i,j} |a_{ij}|$.

Let $\vec{x} = \vec{e}_{j^*}$. Then $\|x\|=1$, so

$$
\begin{aligned}
\|A\| &\geq \|Ax\| \qquad \text{column } j^*,\ Ae_{j^*} \\
&= \|(a_{1j^*}, a_{2j^*}, \dots, a_{mj^*})\| \\
&= \sqrt{\sum_i |a_{ij^*}|^2} \\
&\geq \max_i |a_{ij^*}| \\
&= |a_{i^*j^*}| \quad \checkmark
\end{aligned}
$$

Next, let us show the upper bound, $\|A\| \leq \sqrt{m \cdot n} \cdot \max_{i,j} |a_{ij}|$.

$$\|A\|^2 = \max_{x:\|x\|=1} \|Ax\|^2$$

Write $A = \begin{pmatrix} \text{——} r_1 \text{——} \\ \text{——} r_2 \text{——} \\ \vdots \\ \text{——} r_m \text{——} \end{pmatrix}$, so $Ax = \begin{pmatrix} r_1 \cdot x \\ r_2 \cdot x \\ \vdots \\ r_m \cdot x \end{pmatrix}$

$$= \max_{x:\|x\|=1} \sum_{i=1}^{m} |r_i \cdot x|^2$$

Now use Cauchy-Schwarz: $|r_i \cdot x| \leq \|r_i\| \cdot \|x\| = \|r_i\|$

$$
\begin{aligned}
&\leq \sum_{i=1}^{m} \|r_i\|^2 \\
&= \sum_i \sum_j |a_{ij}|^2 \quad \checkmark
\end{aligned}
$$

$\square$

When is a perturbed matrix invertible?

Lemma: If $\|A\| < 1$, then $(I+A)^{-1}$ exists.

Proof:

$$
\begin{aligned}
I+A \text{ is not invertible} &\iff N(I+A) \neq \{0\} \\
&\iff (I+A)x = 0 \text{ for some } x \neq 0 \\
&\implies Ax = -x \\
&\implies \|A\| \geq 1, \text{ a contradiction.} \quad \square
\end{aligned}
$$

Lemma: Let A be an invertible matrix.

If $\|B\| < 1/\|A^{-1}\|$, then $A+B$ is invertible.

Proof: $A+B = A(I + A^{-1}B)$. Now apply the previous lemma, with $\|A^{-1}B\| \leq \|A^{-1}\| \cdot \|B\|$ $\square$

Example: $A = \begin{pmatrix} 1 & 0 \\ 0 & 1/4 \end{pmatrix}$, $A^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$

so if $\|B\| < \frac{1}{4}$, $A+B$ is invertible.

Lemma: If $\|A\| < 1$, then

$$(I+A)^{-1} = I + A + A^2 + A^3 + A^4 + \cdots$$

Proof: Exercise.

Other Matrix Norms (extra material)

Just as we have defined multiple vector norms, like

$\|v\| = \sqrt{\sum_i |v_i|^2}$ Euclidean

$\|v\|_1 = \sum_i |v_i|$ ℓ_1 norm

$\|v\|_p$ $\left(\sum_i |v_i|^p\right)^{1/p}$ ℓ_p norm,

we can define many different matrix norms.

Example:

- $\|A\|_p = \max_{x:\|x\|_p=1} \|Ax\|_p$

see, e.g., arXiv: 1205.4484

Example:

- $\|A\|_p = \max_{x : \|x\|_p = 1} \|Ax\|_p$
- $\|A\|_{p \to q} = \max_{x : \|x\|_p = 1} \|Ax\|_q$ see, eg., arXiv: 1205.4484

Exercise: What is the matrix ℓ_1 norm, $\|A\|_1$, for
$$A = \begin{pmatrix} 5 & 9 \\ -6 & 1 \end{pmatrix} ?$$
What is it in general?

Answer:
$$\|A\|_1 = \max_{x : \|x\|_1 = 1} \|Ax\|_1$$
To evaluate this, there are two steps:

① First, we need to find an upper bound, $\|A\|_1 \leq K$.
② Second, we need to show that this bound is achieved, ie., find x with $\|x\|_1 = 1$ so $\|Ax\|_1 = K$.

① $$\|A\|_1 = \max_{\substack{x_1, x_2 \\ |x_1| + |x_2| = 1}} \left(|5x_1 + 9x_2| + |-6x_1 + 1 x_2| \right)$$
$$\leq \max_{\substack{x_1, x_2 \\ |x_1| + |x_2| = 1}} (5+6)|x_1| + (9+1)|x_2|$$
$$\leq \max \{ 5+6, 9+1 \}$$
$$= 11$$
$\Rightarrow \|A\|_1 \leq 11$

② The bound is achieved, $\|Ax\|_1 = 11$, for $x = (1,0)$.
$\Rightarrow \|A\|_1 = 11$.

In general, $\|A\|_1 = \max_{\text{columns } j} \sum_i |a_{ij}|$
the maximum ℓ_1 norm of a column. ✓

General properties of matrix norms:

All the above norms satisfy:

- $\|A\| \geq 0$, and $\|A\| = 0 \iff A = 0$
- $\|\alpha A\| = |\alpha| \cdot \|A\|$ for all scalars α
- triangle inequality:
  $\|A + B\| \leq \|A\| + \|B\|$ for same-size matrices
- sub-multiplicativity:
  $\|AB\| \leq \|A\| \cdot \|B\|$ whenever AB is defined

Exercise: Is $f(A) = \max_{i,j} |a_{ij}|$ a matrix norm?
That is, does it satisfy the above properties?

Answer: It does satisfy the first three properties.
But sub-multiplicativity is harder
$$A = \begin{pmatrix} 1 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
$$\Rightarrow AB = (2)$$
$$f(AB) = 2 \qquad f(A) = f(B) = 1. ✓$$
So NO, f is not sub-multiplicative.

Example: Frobenius norm
$\|A\| = \sqrt{\sum |a_{ij}|^2}$ ← easy to compute!

Example: Frobenius norm

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$$ ← easy to compute!

Exercise: This does satisfy sub-multiplicativity and the other properties.

Observe: $\|A\|_F^2 = \text{Trace}\,(A^\dagger A)$ (sum of diagonal elements)

$$= \sum_j (A^\dagger A)_{j,j}$$
$$= \sum_{i,j} (A^\dagger)_{ji} (A)_{ij}$$
$$= \sum_{i,j} \underbrace{a_{ij}^* a_{ij}}_{|a_{ij}|^2} \checkmark$$

Fact: The trace is cyclic:

$$\boxed{Tr(AB) = Tr(BA)}$$

Proof: $Tr(AB) = \sum_i (AB)_{i,i}$
$$= \sum_{i,j} a_{ij} b_{ji}$$
$$= \sum_{j,i} b_{ji} a_{ij} = Tr(BA) \;\square$$

Corollary: The Frobenius norm is basis-independent, i.e.,

$$\|A\|_F = \|UAU^\dagger\|_F$$

for any unitary/orthogonal matrix U.

(The Frobenius norm is the same in all orthonormal bases.)

Proof: Since U is unitary, $U^\dagger = U^{-1}$.

$\|UAU^\dagger\|_F = Tr\left((UAU^\dagger)^\dagger (UAU^\dagger)\right)$

$= Tr\left(UA^\dagger U^\dagger UAU^\dagger\right)$ since $(AB)^\dagger = B^\dagger A^\dagger$

$= Tr\left(A^\dagger U^\dagger UAU^\dagger U\right)$ cyclic trace

$= Tr(A^\dagger A)$ since $U^\dagger U = I$ ✓ $\square$

The spectral norm is also basis-independent, $\|A\|_2 = \|UAU^\dagger\|_2$ for any unitary U, since unitaries don't change lengths.

Relationships between matrix norms:

Any two matrix norms are the same up to dimension-dependent factors.

Example: For $A \in \mathbb{R}^{m\times n}$,

$$\|A\|_2 \le \|A\|_F \le \min\{m,n\} \cdot \|A\|_2$$

Proof:

We have already shown $\|A\|_2 \le \|A\|_F$.

$\|A\|_F \le \min\{m,n\} \cdot \|A\|_2$:

We can't prove this yet! Fortunately, it is less important.

Later, it will follow since

$\|A\|_F^2 = Tr(A^\dagger A) =$ sum of eigenvalues of $A^\dagger A$

$\|A\|_2^2 =$ largest eigenvalue of $A^\dagger A$ ...

See http://en.wikipedia.org/wiki/Matrix_norm#Equivalence_of_norms for more.