Lecture 15: Rotations and scaling

Thursday, October 15, 2015 9:30 AM

Admin: Homework 5 out tonight.

Reading: Meyer 5.6 Isometries

5.2 Matrix norms

5.12 Singular-value decomposition

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

Next

Singular values → Singular-value decomposition

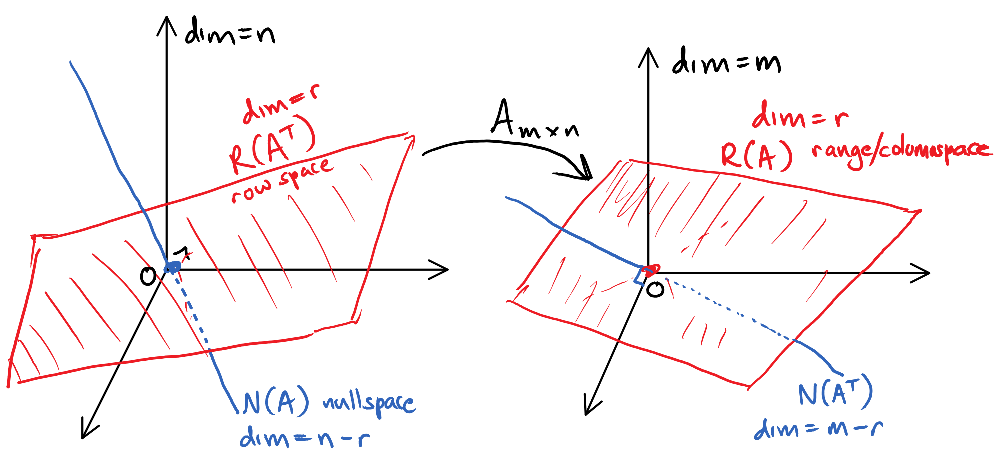
Eigenvalues/vectors Spectral decomposition

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.]

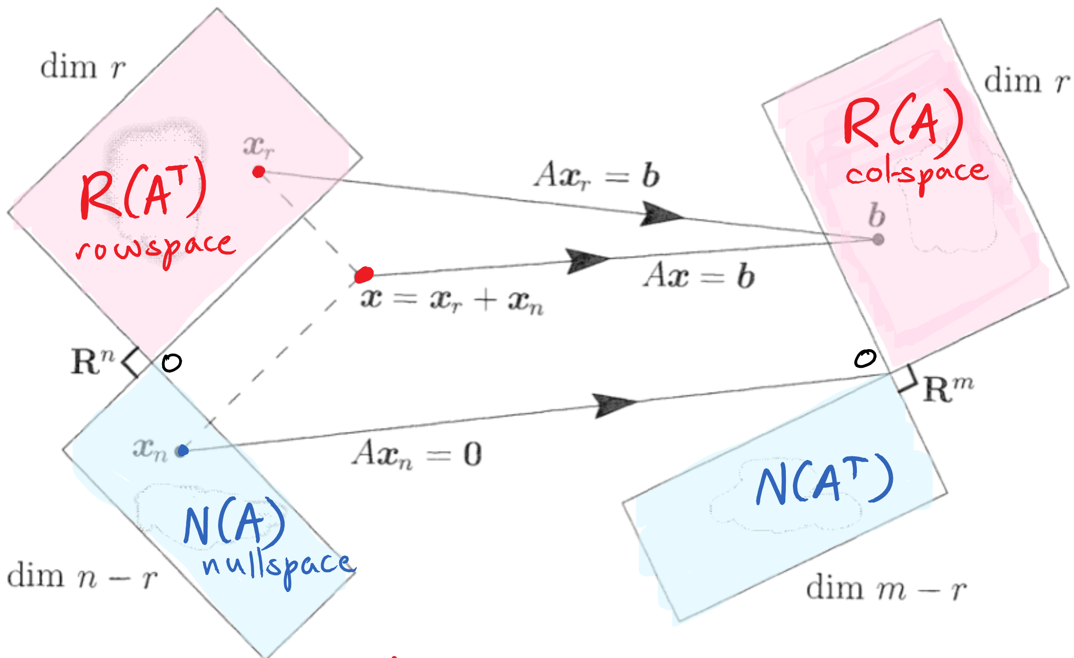
How??



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

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

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



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

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.

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

(iso = same
metric = length/distance)

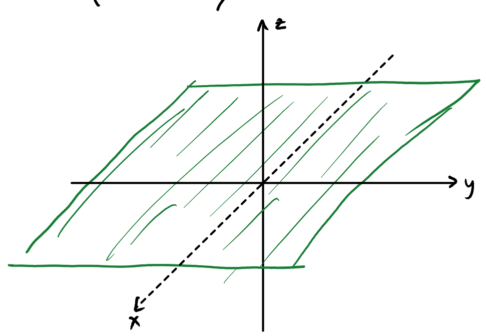
Examples:

- 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., $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$
- Products of rotations and reflections

$$\|ABx\| = \|Bx\| = \|x\|$$

- Isometric "embeddings", e.g.

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



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

- Not an isometry: anything that reduces the dimension

$A = {}_m\left(\quad \overset{n}{} \quad \right)$ with $m < n$

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

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

$\Rightarrow$ 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

$$\cos\theta = \frac{x \cdot y}{\|x\| \cdot \|y\|}.$$

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

Trick: Add the vectors, and use the cross-terms:

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

for reals, $(Ax)\cdot(Ay) = (Ay)\cdot(Ax)$

$$= \|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2x\cdot y$$

$$\Rightarrow (Ax)\cdot(Ay) = x \cdot y \quad \checkmark$$

$\square$

How to tell if a matrix is an isometry?

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} = \sum_{i=1}^{n} v_i e_i^T$$

The columns of an isometry are **orthonormal**

① $A\vec{e}_i = \vec{v}_i \quad \Rightarrow \|v_i\| = 1$ all columns must have length one

② $A\left(\frac{1}{\sqrt{2}}\vec{e}_1 + \frac{1}{\sqrt{2}}\vec{e}_2\right) = \frac{1}{\sqrt{2}}(\vec{v}_1 + \vec{v}_2)$

$$\|v_1 + v_2\|^2 = \|e_1 + e_2\|^2 = 2$$

$$\| \quad (v_1+v_2)\cdot(v_1+v_2)$$
$$= \|v_1\|^2 + \|v_2\|^2 + (v_1 \cdot v_2 + v_2 \cdot v_1)$$
$$\| \quad 2\,\text{Re}(v_1 \cdot v_2)$$

since $v_2 \cdot v_1 =$ complex conj. of $v_1 \cdot v_2$

$$\overset{\shortparallel}{2\,\mathrm{Re}(v_1 \cdot v_2)}$$ since $v_2 \cdot v_1 =$ complex conj. of $v_1 \cdot v_2$

$= \mathrm{Re}(v_1 \cdot v_2) - \mathrm{Im}(v_1 \cdot v_2)$

$\Rightarrow \mathrm{Re}(v_1 \cdot v_2) = 0$

Considering $A(e_1 - e_2)$ gives $\mathrm{Im}(v_1 \cdot v_2) = 0$

$\Rightarrow v_1 \cdot v_2 = 0$ different columns must be perpendicular.

In matrix notation:

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

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

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.

A is an isometry $\Longleftrightarrow$ A preserves lengths and angles

$\Updownarrow$ $\qquad\qquad\qquad\qquad$ $\Updownarrow$

A's columns are orthonormal $\Longleftrightarrow$ $A^\dagger A = I$

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

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

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

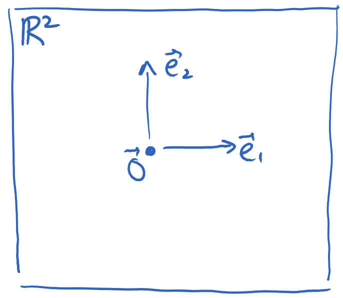
matrices

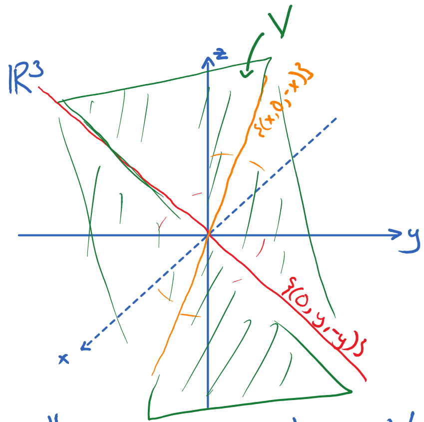
$$\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix} \quad \text{and} \quad -\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}$$

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 $\vec{u}$:
  $$v_1 + v_2 + v_3 = 0 \qquad (\vec{v} \in V)$$
  $$v_1 - v_2 \quad = 0 \qquad (\vec{v} \cdot \vec{u} = 0)$$
  $$\Rightarrow \vec{v} = (1,1,-2)/\sqrt{6} \quad \text{works}$$
  $\uparrow$ normalization

What is the matrix for our isometry?

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

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

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

Short answer:

An orthonormal basis for V is

$\frac{1}{\sqrt{2}}(1,-1,0)$, $\frac{1}{\sqrt{6}}(1,1,-2)$.

Therefore,

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{2}{\sqrt{6}} \end{pmatrix}$$ isometrically maps $\mathbb{R}^2$ 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., $n \times n$).

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:

$$\boxed{\text{Orthogonal matrix} \quad A^T = A^{-1}}$$

[illegible] like $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \cdots \end{pmatrix}$

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,

$$A = \sum_{j=1}^{n} \vec{v}_j \vec{e}_j^{\,T} \qquad \begin{pmatrix} \vec{v}_j \end{pmatrix} \begin{pmatrix} \text{---} \ \vec{e}_j^{\,T} \ \text{---} \end{pmatrix}$$

Why? Check it:

$$A\vec{e}_i = \left(\sum_{j=1}^{n} \vec{v}_j \vec{e}_j^{\,T}\right)\vec{e}_i$$

$$= \sum_j \vec{v}_j (\vec{e}_j^{\,T} \vec{e}_i)$$

$$= \sum_j \vec{v}_j \underbrace{(\vec{e}_j \cdot \vec{e}_i)}_{= \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}}$$

$$= \vec{v}_j \quad \checkmark$$

$$\Rightarrow A^T A = \left(\sum_j v_j e_j^T\right)^T \left(\sum_k v_k e_k^T\right)$$

$$= \sum_{j,k} e_j \underbrace{v_j^T v_k}_{= v_j \cdot v_k = \begin{cases} 1 & \text{if } j=k \\ 0 & \text{if } j \neq k \end{cases}} e_k^T$$

$$= \sum_j e_j e_j^T$$

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

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

$$\vdots$$

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

Next let's compute AA^T:

$$AA^T = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \begin{pmatrix} \text{---} & v_1^T & \text{---} \\ & \vdots & \\ \text{---} & v_n^T & \text{---} \end{pmatrix}$$

$$= \left(\sum_j v_j e_j^T\right)\left(\sum_k v_k e_k^T\right)^T$$

$$= \sum_{j,k} v_j (e_j \cdot e_k) v_k^T$$

$$= \sum_{j=1}^{n} v_j v_j^T$$

Claim: This is the identity again.

Why?

Call it M.

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

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

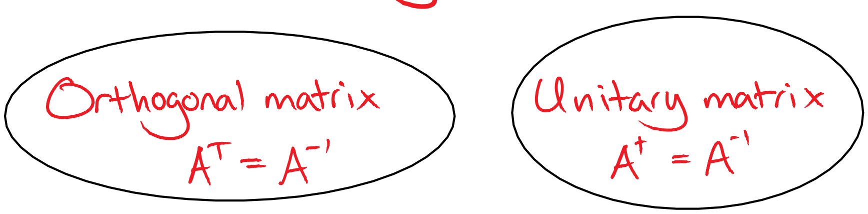
so the vectors $\vec{v}_1, \ldots, \vec{v}_n$ are all left alone.

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

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

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

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



More examples:

- Permutation matrices, e.g.,

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} : \quad \begin{matrix} e_1 \mapsto e_2 \\ e_2 \mapsto e_3 \\ e_3 \mapsto e_4 \\ e_4 \mapsto e_1 \end{matrix}$$

- 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

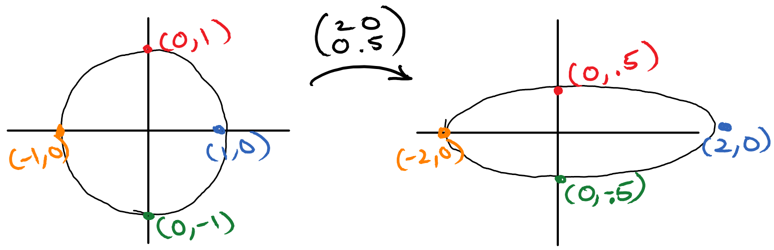
$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$ rotates $\mathbb{R}^3$ by θ about the z-axis

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

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



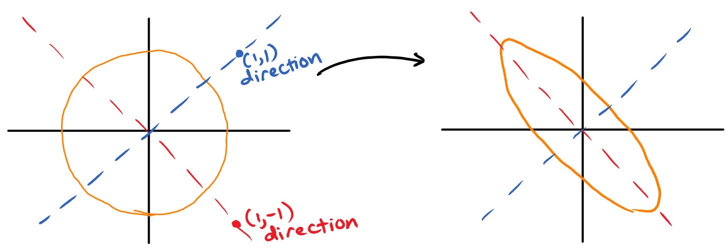
Need not be axis-aligned...

Exercise: Give a 2×2 matrix that maps

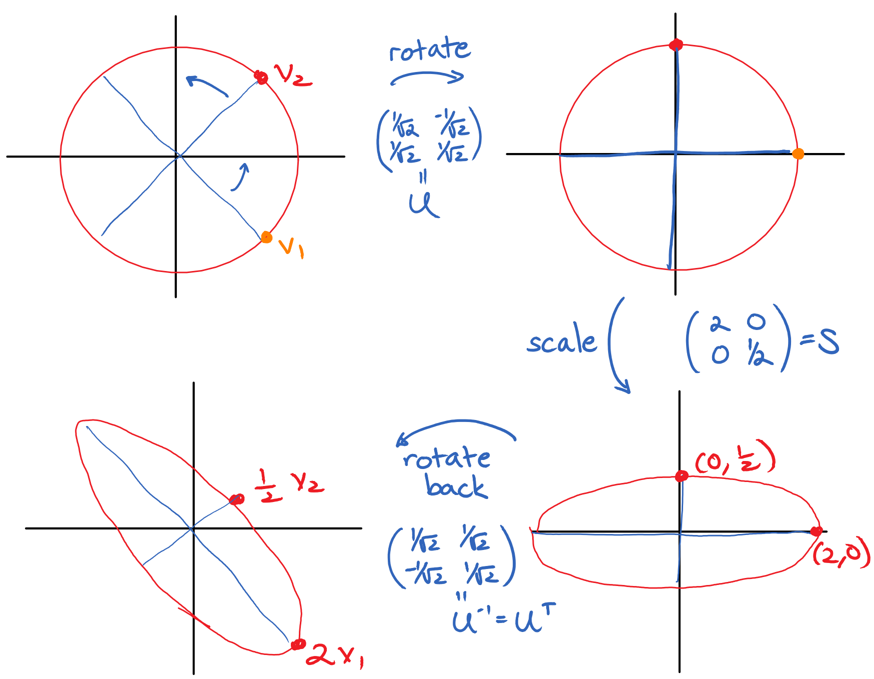
$$(1,-1) \longmapsto (2,-2)$$

$$(1,1) \longmapsto \left(\tfrac{1}{2},\tfrac{1}{2}\right)$$

Answer: We want



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



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

Alternative answer:

Note $\{ \underbrace{\tfrac{1}{\sqrt{2}}(1,-1)}_{= v_1}, \underbrace{\tfrac{1}{\sqrt{2}}(1,1)}_{= v_2} \}$ is an orthonormal basis.

We want $Av_1 = 2v_1$, $Av_2 = \frac{1}{2}v_2$.

Set

$$
\begin{aligned}
A &= 2v_1v_1^T + \tfrac{1}{2}v_2v_2^T \\
&= 2\cdot\tfrac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \tfrac{1}{2}\cdot\tfrac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{pmatrix}.
\end{aligned}
$$

$A(1,-1) = (2,-2)$ ✓
$A(1,1) = (\frac{1}{2}, \frac{1}{2})$

Scaling II: Matrix Norm

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

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

(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),
  $$\|A\vec{x}\| \leq \|A\| \cdot \|\vec{x}\|$$
- For any real/complex number α,
  $$\|\alpha A\| = |\alpha| \cdot \|A\|.$$
- Triangle inequality:
  $$\|A+B\| \leq \|A\| + \|B\|.$$

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

$$\leq \left(\max_{\|x\|=1} \|Ax\|\right) + \left(\max_{\|y\|=1} \|By\|\right)$$
$$= \|A\| + \|B\|. \qquad \square$$

Examples:

- $\|I\| = 1$

  $\|\text{any isometry}\| = 1$

  $\|\text{any projection}\| = 1$, unless the projection is 0

- What is $\left\| \begin{pmatrix} 1 & 1/100 \\ 1/100 & 1 \end{pmatrix} \right\|$?

Answer: $\quad A$ "

① Lower bound

$$\|A\vec{e}_1\| = \left\| \begin{pmatrix} 1 \\ 1/100 \end{pmatrix} \right\| = \sqrt{1 + \frac{1}{100^2}} = \|A\vec{e}_2\|$$
$$\Rightarrow \|A\| \geq \sqrt{1 + \frac{1}{100^2}}$$

② Upper bound $\quad A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{100}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

$$\Rightarrow \|A\| \leq \left\| \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\| + \frac{1}{100}\left\| \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\|$$
$$= 1 + \frac{1}{100}$$

③ Exact $\quad$ We want to find $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, with $\|\vec{x}\| = 1$, to maximize

$$\|Ax\|^2 = \left\| \begin{pmatrix} x_1 + \overset{\varepsilon}{\frac{1}{100}} x_2 \\ \frac{1}{100} x_1 + x_2 \end{pmatrix} \right\|^2$$
$$= (x_1 + \varepsilon x_2)^2 + (\varepsilon x_1 + x_2)^2$$
$$= (1+\varepsilon^2)(x_1^2 + x_2^2) + 4\varepsilon x_1 x_2$$

$\|x\|^2 = x_1^2 + x_2^2 = 1 \quad \Rightarrow x_2 = \pm\sqrt{1 - x_1^2}$

(We may assume that $x_1 \geq 0$ and $x_2 \geq 0$.
Why? If $x_1 < 0$, multiply $\vec{x}$ by -1, leaving $\|x\|$ and $\|Ax\|$ unchanged.
If $x_1 > 0$ and $x_2 < 0$, then we can increase $(x_1 + \varepsilon x_2)^2$ by switching the sign of x_2!)

$$\Rightarrow \|Ax\|^2 = 1 + \varepsilon^2 + 4\varepsilon x_1 \sqrt{1 - x_1^2}$$
$$\Rightarrow x_1 = \frac{1}{\sqrt{2}} \quad \text{(by calculus)}$$
$$\Rightarrow \|A\| = \sqrt{1 + \varepsilon^2 + 2\varepsilon}$$
$$= 1 + \varepsilon \quad \checkmark$$

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

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

Answer:

ⓐ $\left\| \begin{pmatrix} 1 & 0 \\ 0 & \underbrace{1/100}_{\varepsilon} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\|^2 = x_1^2 + \varepsilon^2 x_2^2$

Under the constraint $x_1^2 + x_2^2 = 1$, this is largest for $|x_1| = 1$.

$\Rightarrow \left\| \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix} \right\| = 1$ (if $|\varepsilon| < 1$).

ⓑ $\left\| \begin{pmatrix} 1 & \varepsilon \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\| = |(1, \varepsilon) \cdot (x_1, x_2)|$

To maximize the dot product between $(1, \varepsilon)$ and $\vec{x}$, subject to $\|\vec{x}\| = 1$, we should choose $\vec{x}$ parallel to $(1, \varepsilon)$, i.e., $\vec{x} = \frac{1}{\sqrt{1+\varepsilon^2}}(1, \varepsilon)$.

$\Rightarrow \left\| \begin{pmatrix} 1 & \varepsilon \\ 0 & 0 \end{pmatrix} \right\| = \frac{1+\varepsilon^2}{\sqrt{1+\varepsilon^2}} = \|(1, \varepsilon)\|$

Observe:

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

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

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

- 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

$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$ the $m \times n$ all-ones matrix?

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
```

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

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

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

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

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

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

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

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 $\text{rank}(A) = 1$.

Since all columns are the same,

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

A factors as

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

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

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

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

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

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

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 $\|A\| = 0 \Leftrightarrow A = 0$
- $\|A\vec{x}\| \leq \underbrace{\|A\|}_{\text{matrix norm}} \cdot \underbrace{\|\vec{x}\|}_{\text{vector/matrix norm}}$
- $\|\alpha A\| = |\alpha| \cdot \|A\|$ 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 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 $A = \vec{u}\vec{v}^T$, $\|A\| = \|u\| \cdot \|v\|$.

A fast (crude) estimate for the spectral norm:

Claim: For any $m \times n$ matrix $A = (a_{ij})$,
$$\max_{i,j} |a_{ij}| \le \|A\| \le \sqrt{\sum_{i,j} |a_{ij}|^2}$$
$$\le \sqrt{m \cdot n} \cdot \max_{i,j} |a_{ij}|.$$

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, $\|A\| \ge \max_{i,j} |a_{ij}|$.

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

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

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

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

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

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

m 12

$$\ldots \begin{pmatrix} \vdots \\ r_m \cdot x \end{pmatrix}$$

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

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

$$\le \sum_{i=1}^{m} \|r_i\|^2$$

$$= \sum_i \sum_j |a_{ij}|^2 \quad \checkmark \qquad \square$$

When is a pertubed matrix invertible?

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

Proof:

$I + A$ is not invertible $\iff N(I+A) \ne \{0\}$

$\iff (I+A)x = 0$ for some $x \ne 0$

$\Rightarrow \quad Ax = -x$

$\Rightarrow \quad \|A\| \geq 1$, a contradiction. $\square$

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, $\square$
with $\|A^{-1}B\| \le \|A^{-1}\| \cdot \|B\|$

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

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:

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

$$A = \begin{pmatrix} 5 & 9 \\ -6 & 1 \end{pmatrix} \text{ ?}$$

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

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

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:
  $\|A + B\| \le \|A\| + \|B\|$ for same-size matrices

- triangle inequality:
  $\|A+B\| \leq \|A\| + \|B\|$ for same-size matrices
- sub-multiplicativity:
  $\|AB\| \leq \|A\| \cdot \|B\|$ whenever AB is defined

Exercise: Is $f(A) = \max\limits_{i,j} |a_{ij}|$ a matrix norm?
That is, does it satisfy the above properties?
Answer: It does satisfy the first three properties.
But sub-multiplicativity is harder

$$A = \begin{pmatrix} 1 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\Rightarrow AB = (2)$$

$$f(AB) = 2 \qquad f(A) = f(B) = 1. \checkmark$$

So NO, f is not sub-multiplicative.

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.,

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

for any unitary/orthogonal matrix U.

, $\|A\|_F = \|\mathcal{U} A \mathcal{U}^\dagger\|_F$

for any unitary/orthogonal matrix $\mathcal{U}$.

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

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

$$\begin{aligned}
\|\mathcal{U} A \mathcal{U}^\dagger\|_F &= \mathrm{Tr}\left((\mathcal{U} A \mathcal{U}^\dagger)^\dagger (\mathcal{U} A \mathcal{U}^\dagger)\right) \\
&= \mathrm{Tr}\left(\mathcal{U} A^\dagger \mathcal{U}^\dagger \mathcal{U} A \mathcal{U}^\dagger\right) \quad \text{since } (AB)^\dagger = B^\dagger A^\dagger \\
&= \mathrm{Tr}\left(A^\dagger \mathcal{U}^\dagger \mathcal{U} A \mathcal{U}^\dagger \mathcal{U}\right) \quad \text{cyclic trace} \\
&= \mathrm{Tr}(A^\dagger A) \quad \text{since } \mathcal{U}^\dagger \mathcal{U} = I \quad \checkmark \square
\end{aligned}$$

The spectral norm is also basis-independent, $\|A\|_2 = \|\mathcal{U} A \mathcal{U}^\dagger\|_2$ for any unitary $\mathcal{U}$, 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

$$\|A\|_F^2 = \mathrm{Tr}(A^\dagger A) = \text{sum of eigenvalues of } A^\dagger A$$
$$\|A\|_2^2 = \text{largest eigenvalue of } A^\dagger A \ \ldots$$

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

Key Lemma:

If $\|A\vec{x}\| = \|A\| \cdot \|\vec{x}\|$,

i.e., $\vec{x}$ is stretched maximally by A, achieving the norm,

then

$$A^{\dagger}A\vec{x} = \|A\|^2\vec{x}$$

Corollary: $\|A^{\dagger}\| = \|A\|$.

Proof: For $y = Ax$, $A^{\dagger}y = \|A\|^2 x = \|A\| \cdot \|y\|$

$$\Rightarrow \|A^{\dagger}\| \geq \|A\|.$$

The same inequality with $A \leftrightarrow A^{\dagger}$ switched gives $\|A\| \geq \|A^{\dagger}\|$. $\square$

Corollary: $\|A^{\dagger}A\| = \|A\|^2 = \|AA^{\dagger}\|$.

Proof: We showed before that $\|AB\| \leq \|A\| \cdot \|B\|$. Hence,

$$\|A^{\dagger}A\| \leq \|A^{\dagger}\| \cdot \|A\|$$
$$= \|A\|^2,$$

and by the lemma this norm is achieved by the same $\vec{x}$ that achieves $\|Ax\| = \|A\| \cdot \|x\|$. ✓ $\square$

Proof of the lemma: Scale so $\|x\| = 1$.

Choose a basis for the domain (columns) of A so that the first basis element is x.

Choose a basis for the codomain (rows) so the first basis element is $y = \frac{Ax}{\|Ax\|}$.

A's representation is these bases is

$$\begin{pmatrix} \|A\| & \\ 0 & \\ 0 & \text{stuff} \\ \vdots & \\ 0 & \end{pmatrix}$$

because $Ax = \|A\| y$. Then $A^{\dagger}y$ is the first row of A

$$\Rightarrow A^{\dagger}y = \|A\| x + b$$

where $b \perp x$ (b is everything else in the first row)

We claim that $b = 0$! (So $A^{\dagger}x = \|A\| x$, our goal.)

Because if $b \neq 0$, then spreading out between x and b would increase the norm, a contradiction of $\|Ax\| = \|A\|$.

$$\left\| A \frac{\|A\|x + b}{\sqrt{\|A\|^2 + \|b\|^2}} \right\| \geq \left| y \cdot \left(A \frac{\|A\|x + b}{\sqrt{\|A\|^2 + \|b\|^2}} \right) \right|$$

since $\|u\| \geq |v \cdot u|$ for all vectors v with $\|v\| = 1$

$$= \frac{\|Ax + b\|^2}{\sqrt{\|A\|^2 + \|b\|^2}}$$

$$\sqrt{\|A\|^2+\|b\|^2}$$

$$= \sqrt{\|A\|^2 + \|b\|^2}$$

If $b \neq 0$, then we've found a unit vector that A stretches more than it stretches x, a contradiction. $\square$

Corollary: $\|A^T\| = \|A\|$, even for complex matrices, since $\|A^\dagger\| = \|A^T\|$ — the difference between $A^\dagger$ and A^T is just complex conjugation, which doesn't change any lengths.

Corollary: $\|\underbrace{A \cdot A^\dagger \cdot A \cdot A^\dagger \cdot \ldots \cdot A \cdot A^\dagger}_{m \text{ times}}\| = \|A\|^{2m}$

Corollary: If A is real and symmetric ($A = A^T$) or complex and Hermitian ($A = A^\dagger$), then $\|A^m\| = \|A\|^m$.

Example: For $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, $\|A\| = \sqrt{2}$

but $A^n = A$

since $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$

$\Rightarrow \|A^n\| = \sqrt{2} \ll \sqrt{2}^n$ ✓

Singular-Value Decomposition (SVD)

Informally: Any linear transformation can be split into:

- a rotation, followed by
- scaling vectors in or out

FORMALLY...

Example: recursiono?

look for $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$

eigenvectors? how?

Two ingredients:

Rotations : linear transformations that preserve length

- define isometries
- prove that columns must be orthonormal
- define unitary, orthogonal matrices

Stretches

- examples

SVD examples

- by hand, and in Matlab/Mathematica

Matrix norm

definition

basic properties

examples and intuition (you want to spread out)

Properties: $\|A\| \geq 0$, $\|A\| = 0 \Leftrightarrow A = 0$

$\|\alpha A\| = |\alpha| \cdot \|A\|$

$\|A+B\| \leq \|A\| + \|B\|$

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

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

$\|A\| = \max\limits_{x:\|x\|=1} \|Ax\|$

max exists by compactness

$\|A\| = \sqrt{\max\{\lambda \mid A^{+}A - \lambda I \text{ is singular}\}}$

~~Key Lemma~~ Corollary:

also $\|A\| = \|A^{+}\|$, $\|AA^{+}\| = \|A\|^2$, $\left\|\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right\| = \max\{\|A\|, \|B\|\}$

$\|U^{+}AV\| = \|A\|$ when $UU^{+} = I$, $V^{+}V = I$

for vectors $\|v\|_2 = \|v\|$ as a matrix

eg., for diagonal matrices $\|A\| = \max |a_{ii}|$

Matlab & Mathematica commands

Other matrix norms

Example: $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$

Application of matrix norms: Inverting a matrix that is close to non-singular.

If A^{-1} exists and $\|A^{-1}E\| = \rho < 1$, then $(A+E)^{-1}$ exists, and

$$\|(A+E)^{-1} - A^{-1}\| \leq \frac{\|E\| \cdot \|A^{-1}\|^2}{1-\rho}$$

unit analysis in physics...

similarly, if A is not symmetric & you get an expression like A^2, you probably made a mistake (should be AA^T or A^TA)

Singular-value decomposition

Theorem: (Singular Value Decomposition)
Any linear transformation (over $\mathbb{R}$ or over $\mathbb{C}$) can be expressed as a rotation followed by a scaling (or the opposite, too).

rederive known results using SVD
e.g., $\|A\| = \|A^T\|$
rank-nullity theorem

§5.12 of the text

SVD can be derived
- based on spectral decomposition for A^*A, or
- directly, URV factorization p. 411, using norms, or that is, pull off one dimension at a time, being $x, \|x\|=1$ maximizing $\|Ax\|$ and $y = \frac{Ax}{\|A\|}$
- ~~directly~~ 2×2 matrix at a time
  that is starting with a matrix

where all the entries are potentially nonzero, first rotate the 1st two coords to zero out the two blue entries, then the gold, then green, ...
Actually, this won't work — fixing the gold entries will make the blue entries nonzero again

A^*A is diagonal $\Rightarrow$ use this to introduce eigenvalues?

Pseudoinverse and its least-square properties
Geometric interpretation of SVD
Condition number (sensitivity of solutions of linear equations)
Example 5.12.1, p. 414
Distance to lower-rank matrices

Principal component analysis

2×2 matrices

$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ eigenvalue 1 (multiplicity 2)

eigenvector (1, 0)

singular values $\frac{1}{2}(\sqrt{5} \pm 1)$

matrix norm $\frac{1}{2}(\sqrt{5}+1) \approx 1.62$

= largest singular value

let x maximize $\|Ax\|$ s.t. $\|x\| = 1$

$y = \frac{Ax}{\|Ax\|}$ note $\|Ax\| \neq 0$ unless $A = 0$

$$U = \begin{pmatrix} x & \cdots \end{pmatrix} \qquad V = \begin{pmatrix} y & \cdots \end{pmatrix}$$

Consider $V^+ A U$

$$e_1^+ V^+ A U e_1 = y^+ A x = \frac{\|Ax\|^2}{\|Ax\|} = \|Ax\| = \|A\|$$

$$\begin{aligned} V^+ A U e_1 &= V^+ A x \\ &= \|A\| \cdot V^+ y \\ &= \|A\| \cdot \begin{pmatrix} - \; y^+ \; - \\ \vdots \end{pmatrix} \begin{pmatrix} y \end{pmatrix} \\ &= \|A\| \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} e_1^+ V^+ A U &= y^+ A U \\ &= (\|Ax\|, y^+ A U e_2, \ldots, y^+ A U e_n) \end{aligned}$$

Claim: $y^+ A = \|A\| x^+$

i.e., $A^+ y = \|A\| x$.

Proof: Write $A^+ y = \|A\| x + \delta e$ where $e \perp x$, $\|e\| = 1$

$$\left\| A \frac{\|A\| x + \delta e}{\sqrt{\|A\|^2 + \delta^2}} \right\|^2$$

$$= \frac{1}{\quad} \cdot \left(\|A\|^4 + \delta^2 \|Ae\|^2 + 2\|A\| \ldots \right)$$

$\sqrt{\|A\|^2+\delta^2}$

$$= \frac{1}{(\|A\|^2+\delta^2)} \cdot \left(\|A\|^4 + \delta^2\|Ae\|^2 + 2\|A\|\delta\, e^T A^T A x\right)$$

$$\geq \left| y \cdot A \frac{\|A\|x + \delta e}{\sqrt{\|A\|^2+\delta^2}} \right|^2 = \frac{\|\|A\|x+\delta e\|^4}{\|A\|^2+\delta^2} = \frac{(\|A\|^2+\delta^2)^2}{\|A\|^2+\delta^2} = \|A\|^2+\delta^2$$

$\|(100 \quad 1)\| = \sqrt{100^2+1^2} \Rightarrow$ want to spread out

Natural problems:

1. Solve $Ax=b$ for x
2. Assuming $Ax=b$ has a solution, finding the shortest solution $\arg\min_{x:\, Ax=b} \|x\|$.
3. If $Ax=b$ has no solution, find an x that minimizes $\|Ax-b\|$ (find the shortest such x)
4. Find the sparsest x such that $Ax=b$!!!

We'll solve the 1st 3 problems today.

(We already have an idea for 1, $A\backslash b$:), but still need to cover preconditioners.)