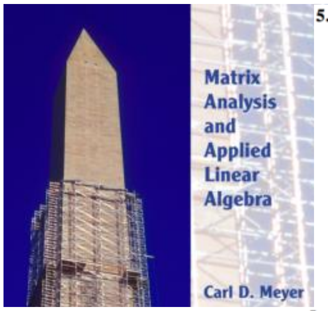
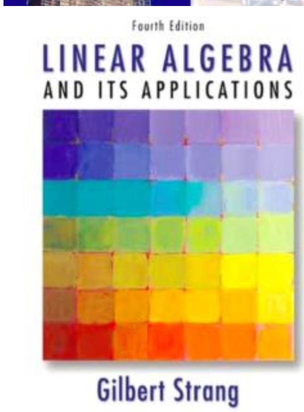
Lecture 13: Rotations and scaling

Admin:





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

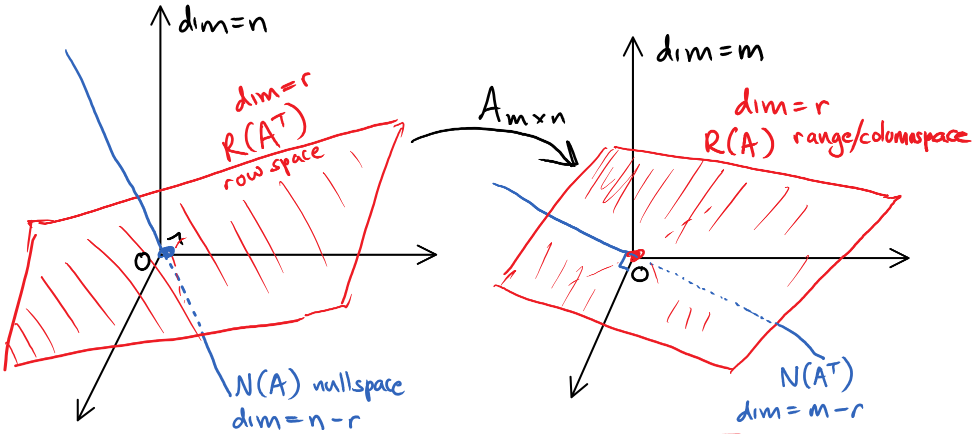
This week: SINGULAR VALUE DECOMPOSITION

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.]

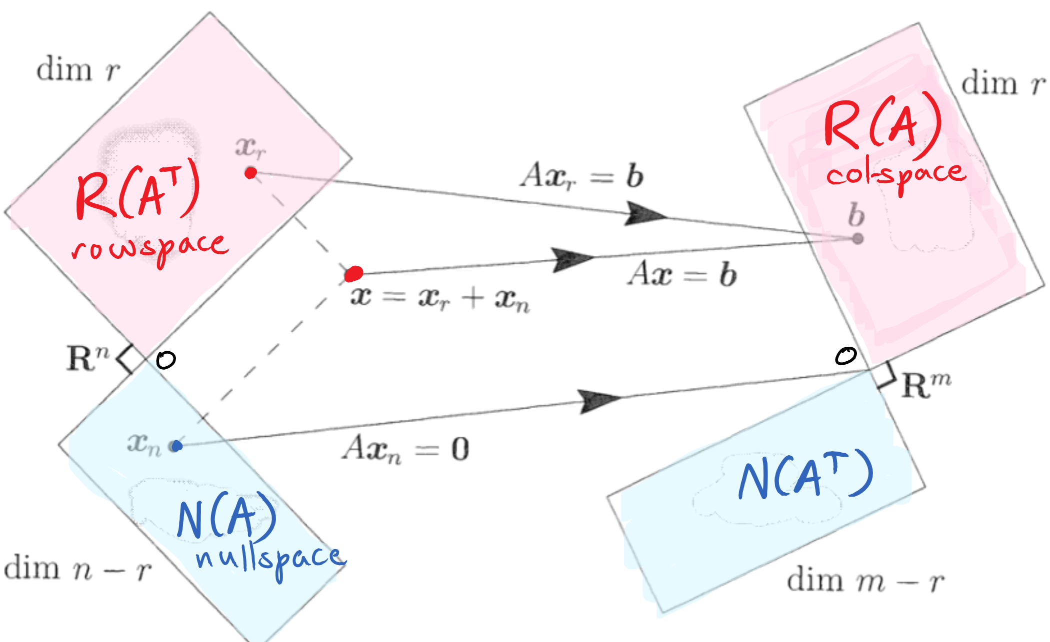
How??

How??



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

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



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

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 $\|Ax-b\|$

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:

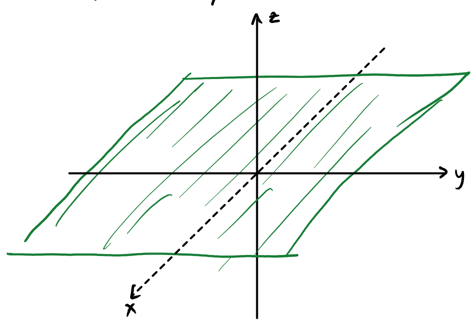
- Identity matrix I
- Rotations, e.g., $\begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$
- Reflections, e.g., $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$
- Products of rotations and reflections

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

- 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

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

$$A = {}_m\Big(\overset{n}{\qquad\qquad} \Big) \quad \text{with } m < n$$

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

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

$\Rightarrow$ 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

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

$\Rightarrow$ 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, $(Ax)\cdot(Ay) = (Ay)\cdot(Ax)$

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

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

$\square$

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

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

isometry are orthonormal

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

② $A(\vec{e}_1 + \vec{e}_2) = \vec{v}_1 + \vec{v}_2$

$$\|\vec{e}_1 + \vec{e}_2\|^2 = 2$$

$$= \|v_1 + v_2\|^2 = (v_1+v_2)\cdot(v_1+v_2)$$

$$= \|v_1\|^2 + \|v_2\|^2 + \underbrace{(v_1 \cdot v_2 + v_2 \cdot v_1)}_{2\,\mathrm{Re}(v_1 \cdot v_2)}$$

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 $A(e_1 + ie_2)$ 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

A is an isometry $\Longleftrightarrow$ A's columns are orthonormal

A's columns are orthonormal $\Longleftrightarrow$ A preserves lengths and angles

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

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

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 $L = \{(x,y,z) \mid x = y = z\} \subset \mathbb{R}^3$.

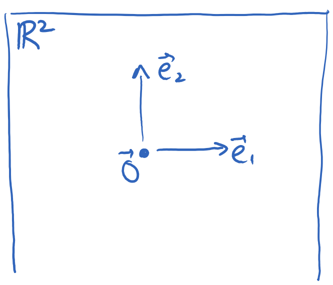
Answer: The line L consists of all multiples of the unit vector $\frac{1}{\sqrt{3}}(1,1,1)$. Therefore, the matrices

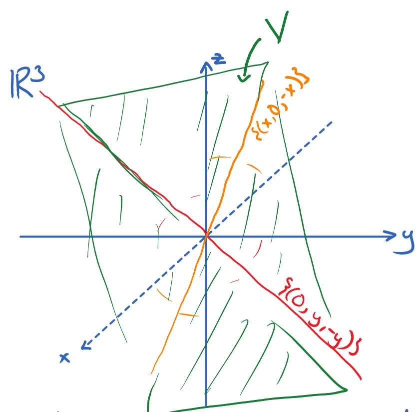
$$\frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \frac{-1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

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

$$\vec{e}_1 \longmapsto \text{first unit vector in } V \equiv \vec{u}$$
$$\vec{e}_2 \longmapsto 2^{nd} \text{ unit vector in } V. \equiv \vec{v}$$

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

- $\vec{u}$ can be an arbitrary unit vector
  eg., 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 $\vec{u}$:

$$\begin{aligned} v_1 + v_2 + v_3 &= 0 \qquad (\vec{v} \in V) \\ v_1 - v_2 &= 0 \qquad (\vec{v} \cdot \vec{u} = 0) \end{aligned}$$

$$\Rightarrow \vec{v} = \underbrace{(1,1,-2)/\sqrt{6}}_{\text{normalization}} \quad \text{works}$$

What is the matrix for our isometry?

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

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

(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,

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{2}{\sqrt{6}} \end{pmatrix} \text{ isometrically maps } \mathbb{R}^2 \text{ onto } V.$$

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., $n \times n$).

Recall: The columns of an isometry are orthonormal,

$$A^T A = I.$$

Proposition: The rows of an orthogonal matrix are also orthonormal,

$$AA^T = I.$$

Corollary:

$$\boxed{\begin{array}{c}\text{Orthogonal matrix}\\ A^T = A^{-1}\end{array}}$$

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:

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

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

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

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

$\vdots$

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

Next let's compute AA^T:

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

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

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

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

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

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

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

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

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

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

More examples:

- Permutation matrices, e.g.,

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

- Rotations, eg.

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$ rotates the plane $\mathbb{R}^2$ counterclockwise by angle θ

$$\begin{pmatrix} \sin\theta & \cos\theta \end{pmatrix}$$ counterclockwise by angle θ

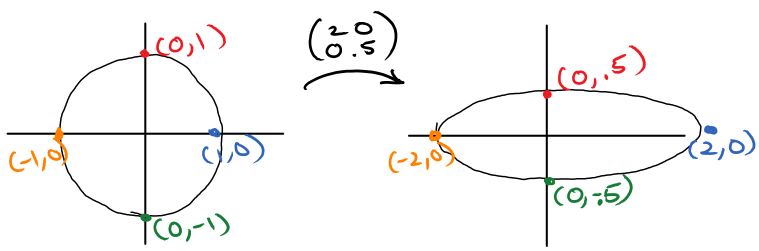
$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$ rotates $\mathbb{R}^3$ by θ about the z-axis

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

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

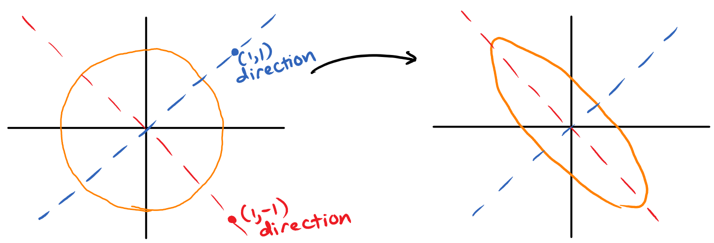


Need not be axis-aligned...

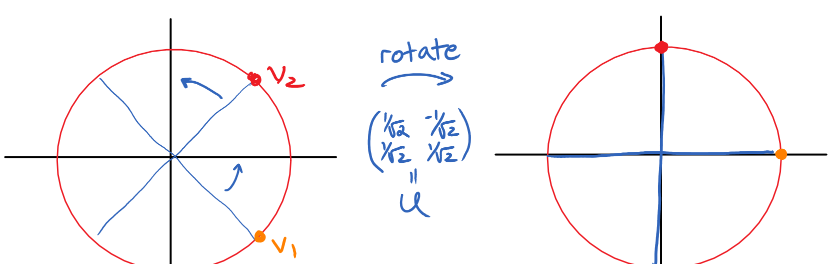
Exercise: Give a 2×2 matrix that maps

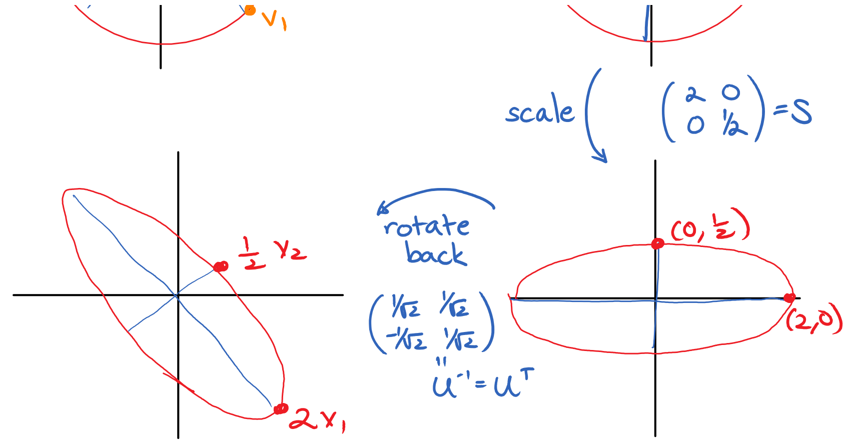
$$(1,-1) \longmapsto (2,-2)$$
$$(1,1) \longmapsto \left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$

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 $A v_1 = 2 v_1$, $A v_2 = \frac{1}{2} v_2$.

Set

$$A = 2 v_1 v_1^T + \frac{1}{2} v_2 v_2^T$$

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

$$= \begin{pmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{pmatrix}.$$

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

Scaling II: Matrix Norm

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

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

(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:
$$\|A+B\| \le \|A\| + \|B\|.$$

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

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: A ″

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

② Upper bound
$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{100}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$\Rightarrow \|A\| \le \left\|\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\| + \frac{1}{100}\left\|\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right\| = 1 + \frac{1}{100}$$

③ Exact We want to find $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, with $\|\vec{x}\| = 1$, to maximize
$$\begin{aligned} \|Ax\|^2 &= \left\| \begin{pmatrix} x_1 + \overset{=\varepsilon}{\frac{1}{100}} x_2 \\ \frac{1}{100} x_1 + x_2 \end{pmatrix} \right\|^2 \\ &= (x_1 + \varepsilon x_2)^2 + (\varepsilon x_1 + x_2)^2 \\ &= (1+\varepsilon^2)(x_1^2 + x_2^2) + 4\varepsilon x_1 x_2 \end{aligned}$$
$$\|x\|^2 = x_1^2 + x_2^2 = 1 \Rightarrow x_2 = \pm\sqrt{1 - x_1^2}$$

(We may assume that $x_1 \ge 0$ and $x_2 \ge 0$.
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!)

$$\Rightarrow \|Ax\|^2 = 1+\varepsilon^2 + 4\varepsilon x_1\sqrt{1-x_1^2}$$

$$\Rightarrow x_1 = \tfrac{1}{\sqrt{2}} \quad \text{(by calculus)}$$

$$\Rightarrow \|A\| = \sqrt{1+\varepsilon^2+2\varepsilon} = 1+\varepsilon \quad \checkmark$$

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:

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

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

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

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

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

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

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

- 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
```

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

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

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

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

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

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

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 rank$(A) = 1$.

Since all columns are the same,

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

A factors as

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

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

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

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

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

(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 $A = \vec{u}\vec{v}^T$, $\|A\| = \|u\| \cdot \|v\| = \sqrt{\sum_i u_i^2 \sum_j v_j^2}$
$$= \sqrt{\sum_{i,j} A_{ij}^2}$$

(extra material)

A fast (crude) estimate for the spectral norm:

Claim: For any $m \times n$ matrix $A = (a_{ij})$,
$$\max_{i,j} |a_{ij}| \le \|A\| \le \sqrt{\sum_{i,j} |a_{ij}|^2} \qquad \text{tight for rank-1}$$
$$\le \sqrt{m \cdot n} \cdot \max_{i,j} |a_{ij}|.$$

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

Consider any fixed row i and column j. Since $\|\vec{e}_j\| = 1$,
$$\|A\|^2 \ge \|A\vec{e}_j\|^2$$
$$= \|(a_{1j}, a_{2j}, \ldots, a_{mj})\|^2$$
$$= \sum_k |a_{kj}|^2 \ge |a_{ij}|^2$$
$$\ge |a_{ij}|^2$$

Since the inequality holds for all i,j, $\|A\| \ge \max_{i,j} |a_{ij}|$. ✓

Next, let us show the upper bound.

Next, let us show the upper bound.

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

Write $A = \begin{pmatrix} \text{———} \\ \text{———} \\ \vdots \\ \text{———} \end{pmatrix}$, 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: $|r_i \cdot x| \le \|r_i\| \cdot \|x\| = \|r_i\|$

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

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

$\square$

Example: Estimate the norm of

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 3 \\ & 5 & 0 & 0 & 6 & \\ & & 9 & 12 & & \\ 0 & & 11 & 12 & & 0 \\ & 7 & & & 8 & \\ 2 & & 0 & & & 4 \end{pmatrix}$$

One approach...

Observe that A splits into 3 blocks:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 3 \\ & 5 & 0 & 0 & 6 & \\ & & 9 & 12 & & \\ 0 & & 11 & 12 & & 0 \\ & 7 & & & 8 & \\ 2 & & 0 & & & 4 \end{pmatrix} \Rightarrow \|A\| = \max\left\{ \left\|\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}\right\|, \left\|\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}\right\|, \left\|\begin{pmatrix} 9 & 12 \\ 11 & 12 \end{pmatrix}\right\| \right\}$$

$$= \underbrace{\left\|\begin{pmatrix} 9 & 12 \\ 11 & 12 \end{pmatrix}\right\|}_{A'}$$

How to estimate $\|A'\|$?

We can try the Δ inequality, $\|B+C\| \le \|B\| + \|C\|$...

$$\|A'\| \le \left\|\begin{pmatrix} 9 & 0 \\ 11 & 0 \end{pmatrix}\right\| + \left\|\begin{pmatrix} 0 & 12 \\ 0 & 12 \end{pmatrix}\right\| = \sqrt{9^2 + 11^2} + 12\sqrt{2} \approx 31.18$$

$$\|A'\| \le \left\|\begin{pmatrix} 9 & 0 \\ 0 & 12 \end{pmatrix}\right\| + \left\|\begin{pmatrix} 0 & 12 \\ 11 & 0 \end{pmatrix}\right\| = 12 + 12 = 24$$

$$\|A'\| \le \left\|\begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}\right\| + \left\|\underbrace{\begin{pmatrix} 9 & 12 \\ 9 & 12 \end{pmatrix}}_{\text{rank-1}}\right\| = 2 + \sqrt{2}\sqrt{9^2 + 12^2} \approx 23.21$$

The Frobenius norm bound is better:

$$\|A'\| \le \|A'\|_F = \sqrt{9^2 + 11^2 + 2 \cdot 12^2} \approx 22.14$$

Software to compute $\|A\|$:

Mathematica/Wolfram Alpha

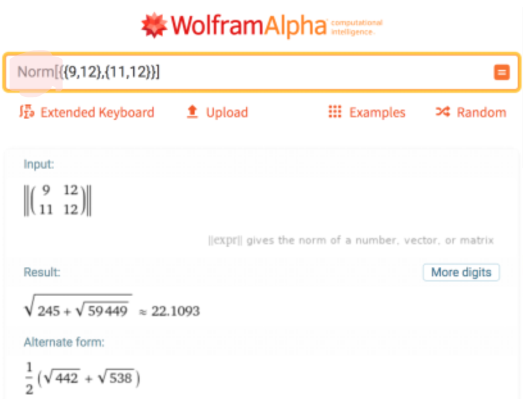
WolframAlpha computational intelligence.

Norm[{{9,12},{11,12}}]

Matlab

```
>> norm([9 12; 11 12])

ans =
```




```
>> norm([9 12; 11 12])

ans =

   22.1093
```

Python

```python
import numpy as np

A = [[9,12],[11,12]]
np.linalg.norm(A, 2), np.linalg.norm(A), np.linalg.norm(A, 'fro')

(22.109311525557523, 22.135943621178654, 22.135943621178654)
```

Note that, unlike Matlab, the default norm is the Frobenius norm.

When is a pertubed matrix invertible?

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

Proof:

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

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

$\implies Ax = -x$

$\implies \|A\| \geq 1$, 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 $\|A^{-1}B\| \leq \|A^{-1}\| \cdot \|B\|$

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 (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

$$\|v\| = \sqrt{\sum_i |v_i|^2} \quad \text{Euclidean}$$
$$\|v\|_1 = \sum_i |v_i| \quad \ell_1 \text{ norm}$$
$$\|v\|_p \quad \left(\sum_i |v_i|^p\right)^{1/p} \quad \ell_p \text{ norm},$$

we can define many different matrix norms.

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

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 $\|A\| = 0 \Leftrightarrow A = 0$

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

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

$$\Rightarrow AB = (2)$$

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

So NO, f is not sub-multiplicative.

Example: Frobenius norm

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

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

Corollary: The Frobenius norm is basis-independent,

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, $U^\dagger = U^{-1}$.

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

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.

Examples: Compute the norms of

$$A = \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 0 & 1 & 0 \\ 2 & 0 & 1 \\ 4 & 0 & 2 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 10i \\ i & 10 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 & 10i & 0 \\ i & 0 & 10 & 0 \\ 0 & 4 & 0 & 5 \end{pmatrix}$$

These all have tricks, but if you get stuck you can always use calculus (or Mathematica or Wolfram Alpha).

Brute-force examples:

You can always use calculus, but it gets ugly fast.

```
A = RandomInteger[10, {2, 2}];
A // MatrixForm
```

MatrixForm=

$$\begin{pmatrix} 7 & 4 \\ 5 & 5 \end{pmatrix}$$

```
Norm[A]
% // N
```

$$\sqrt{\frac{5}{2}\left(23+\sqrt{493}\right)}$$

10.6306

```
v = {1, y};
Av = A.v
Simplify[D[Av.Av/v.v, y]]
Solve[% == 0, y]
Av.Av/v.v /. %[[2]] // Simplify
```

$\{7+4y,\ 5+5y\}$

$$-\frac{2\left(-53+33y+53y^2\right)}{\left(1+y^2\right)^2}$$

minimum maximum

$$\left\{\left\{y\to\frac{1}{106}\left(-33-5\sqrt{493}\right)\right\},\ \left\{y\to\frac{1}{106}\left(-33+5\sqrt{493}\right)\right\}\right\}$$

$$\frac{5}{2}\left(23+\sqrt{493}\right)$$

```
A = RandomInteger[10, {3, 3}];
A // MatrixForm
```

MatrixForm=

$$\begin{pmatrix} 2 & 4 & 2 \\ 3 & 1 & 8 \\ 5 & 9 & 8 \end{pmatrix}$$

```
Norm[A]
% // N[#, 32] &
```

$\sqrt{243.\ldots}$ Root[-400 + 6172 #1 - 268 #1^2 + #1^3 &, 3]

15.574393384086796359704386279789

```
v = {1, y, z};
Av = A.v
Simplify[D[Av.Av/v.v, y]]
ysub = Solve[% == 0, y][[2]]
```

$\{2+4y+2z,\ 3+y+8z,\ 5+9y+8z\}$

$$-\frac{4\left(4y^2\left(7+11z\right)+y\left(-30+68z+17z^2\right)-4\left(7+11z+7z^2+11z^3\right)\right)}{\left(1+y^2+z^2\right)^2}$$

$$\left\{y\to\frac{30-68z-17z^2+\sqrt{4036+5776z+14484z^2+12168z^3+8033z^4}}{8\left(7+11z\right)}\right\}$$

```
Simplify[D[Av.Av/v.v /. ysub, z]]
Solve[% == 0, z] // FullSimplify
(Av.Av/v.v /. ysub) /. %[[1]] // FullSimplify
√% // N[#, 32] &
```

$$-\Big(\big(128\left(7+11z\right)\sqrt{4036+5776z+14484z^2+12168z^3+8033z^4}\,\big(-15780z^2+22769z^3-8\left(4286+71\sqrt{4036+5776z+14484z^2+12168z^3+8033z^4}\right)+3z\left(1014+85\sqrt{4036+5776z+14484z^2+12168z^3+8033z^4}\right)\big)\big)\Big/\Big(4036+12168z^3+8033z^4+30\sqrt{4036+5776z+14484z^2+12168z^3+8033z^4}+z\left(5776-68\sqrt{4036+5776z+14484z^2+12168z^3+8033z^4}\right)-17z^2\left(-852+\sqrt{4036+5776z+14484z^2+12168z^3+8033z^4}\right)\Big)^2\Big)$$

$$\left\{\left\{z\to 1.79\ldots\right\},\ \left\{z\to-\frac{7}{11}\right\}\right\}$$

$243.\ldots$

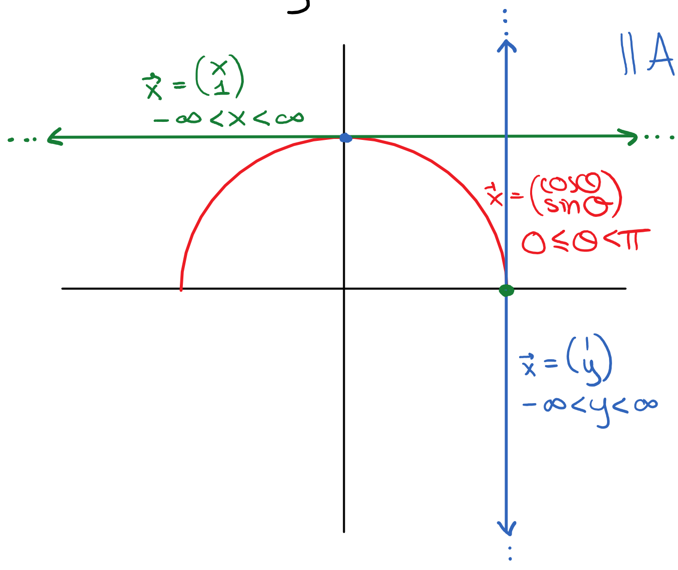
15.574393 Root[-400 + 6172 #1 - 268 #1^2 + #1^3 &, 3]

Different parameterizations

Using $\|A\| = \max\limits_{\vec{x}:\|x\|=1} \|A\vec{x}\|$, we need to optimize over length-one vectors — except $\|A(-\vec{x})\| = \|A\vec{x}\|$.

$\Rightarrow$ for an $m \times 2$ matrix, $\|A\| = \max_{\theta \in [0,\pi)} \left\| A \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \right\|$.

Using $\|A\| = \max_{\vec{x} \neq \vec{0}} \frac{\|A\vec{x}\|}{\|\vec{x}\|}$, we can use any parameterization that covers every direction (± 1).



$$\|A\| = \max \left\{ \max_{y} \frac{\left\| A \begin{pmatrix} 1 \\ y \end{pmatrix} \right\|}{\left\| \begin{pmatrix} 1 \\ y \end{pmatrix} \right\|}, \left\| A \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\| \right\}$$

covers the straight-up direction ("$y = +\infty$")