Lecture 14: Rotations and scaling

Admin: Homework 6 out tonight.
Reading: Meyer 5.6 Isometries
5.2 Matrix norms
5.12 Singular-value decomposition

Concepts
Vector space
Basis
Inner products/Norm/Orthogonality
Linear transformations
Projections
Rank

Techniques
Gaussian elimination
LU decomp.
Gram-Schmidt
QR decomp.

Decompositions

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 row space $R(A^T)$ to distinct points in the column space $R(A)$. [Rank-Nullity Thm.]

How??

\[
\dim R(A^T) = \dim R(A)
\]
\[ \dim R(A^T) = \dim R(A) \]

\[ \dim N(A) + \dim R(A^T) = \text{total dimension } n \]

\[ \dim R(A) + \dim N(A^T) = \text{total dimension } m \]

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, well consider these pieces.

**ISOMETRIES**

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

(That is, \( \|Ax\| = \|x\| \) for all \( x \).)

**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., \( \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \)
- **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
\[ A = m(\ldots) \quad \text{with} \quad m \leq n \]
\[ \Rightarrow \text{rank}(A) = \text{\# lin. indep. rows} \leq m \]
\[ \Rightarrow \dim N(A) = n - \text{rank}(A) > 0 \]
\[ \Rightarrow \text{lengths of nonzero vectors in } N(A) \text{ are sent to } 0 \text{ --- not preserved.} \]

**Claim:** Preserves length \(\Rightarrow\) preserves angles.

**Proof:** Recall the angle \(\Theta\) between real vectors \(\vec{x}\) and \(\vec{y}\) satisfies

\[ \cos(\Theta) = \frac{\vec{x} \cdot \vec{y}}{\|x\| \cdot \|y\|} \]

\[ \Rightarrow \text{We have to show that dot products are preserved.} \]

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

\[ \|A(x+y)\|^2 = (A(x+y)) \cdot (A(x+y)) \]
\[ = (Ax) \cdot (Ax) + (Ay) \cdot (Ay) \]
\[ + (Ax) \cdot (Ay) + (Ay) \cdot (Ax) \]
\[ = \|Ax\|^2 + \|Ay\|^2 + 2 \langle Ax, Ay \rangle \]
\[ = \|x+y\|^2 - \|x\|^2 - \|y\|^2 + 2x \cdot y \]
\[ \Rightarrow \langle Ax, Ay \rangle = x \cdot y \]

**How to tell if a matrix is an isometry?**

\[ A = \left( \begin{array}{c| c| c| c| c} \mid \mid \mid & \mid \mid \mid & \mid \mid \mid & \mid \mid \mid & \mid \mid \mid \\
 v_1 & v_2 & \ldots & v_n \end{array} \right) = \sum_{i=1}^{n} v_i e_i^T \]

1. \(Ae_i = \vec{v}_i \quad \Rightarrow \|v_i\| = 1\) all columns must have length one
2. \(A \left( \frac{1}{\sqrt{2}} e_1 + \frac{1}{\sqrt{2}} e_2 \right) = \frac{1}{\sqrt{2}} (\vec{v}_1 + \vec{v}_2)\)
\[ \|v_1 + v_2\|^2 = \|e_1 + e_2\|^2 = 2 \]
\[ = \|v_1\|^2 + \|v_2\|^2 + (v_1 \cdot v_2) + (v_2 \cdot v_1) \]
\[ = \|v_1\|^2 + \|v_2\|^2 + 2 \text{Re}(v_1 \cdot v_2) \]

since \(v\cdot v = \text{complex conj. of } v \cdot v\)
\[ 2 \Re(v_1 \cdot v_2) \quad \text{since } v_2^* v_1 = \text{complex conj. of } v_1 \cdot v_2 = \Re(v_1 \cdot v_2) - \Im(v_1 \cdot v_2) \]

\[ \Rightarrow \Re(v_1 \cdot v_2) = 0 \]

Considering \( A(e_1, e_2) \) gives \( \Im(v_1 \cdot v_2) = 0 \)

\[ \Rightarrow v_1 \cdot v_2 = 0 \quad \text{different columns must be perpendicular.} \]

In matrix notation:

\[
\begin{pmatrix}
\vdots \\
v_1^* \\
\vdots \\
v_n^*
\end{pmatrix}
\begin{pmatrix}
v_1 \\
\vdots \\
v_n
\end{pmatrix}
= 
\begin{pmatrix}
v_1 \cdot v_1 & v_1 \cdot v_2 & v_1 \cdot v_3 & \cdots \\
v_2 \cdot v_1 & v_2 \cdot v_2 & v_2 \cdot v_3 & \cdots \\
v_3 \cdot v_1 & v_3 \cdot v_2 & v_3 \cdot v_3 & \cdots \\
& & & \ddots
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 & \cdots \\
0 & 1 & 0 & \cdots \\
0 & 0 & 1 & \cdots \\
& & & \ddots
\end{pmatrix}
= I
\]

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.

---

\[ A \text{ is an isometry } \iff A \text{ preserves lengths and angles} \]

\[ \iff A \text{'s columns are orthonormal} \iff A^T A = I \]

**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 \( \mathbb{R}^2 \) to the plane \( V = \{ (x, y, z) | x + y + z = 0 \} \subset \mathbb{R}^3 \).

**Answer:**

An orthonormal basis for \( V \) is

\[
\frac{1}{\sqrt{2}} (1, -1, 0), \quad \frac{1}{\sqrt{2}} (1, 1, -2).
\]

Therefore.
\[
\begin{pmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\
\end{pmatrix}
\]

isometrically maps \( \mathbb{R}^4 \) 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**

**Special cases:** square matrix isometries

- An \( n \times n \) real isometry is called "orthogonal":
  \[ A^T A = I \]

- An \( n \times n \) complex isometry is called "unitary":
  \[ A^* A = I \]

\[
A = \sum_{i=1}^{n} v_i e_i^* \\
\Rightarrow A^* A = \left( \sum_{i=1}^{n} v_i e_i^* \right)^* \left( \sum_{j=1}^{n} v_j e_j^* \right) \\
= \sum_{i \neq j}^n e_i (v_i^* v_j^*) e_j^* + \sum_{i=1}^{n} e_i (v_i^* v_i) e_i^* \\
= \sum_{i \neq j}^n e_i e_i^* + \sum_{i=1}^{n} e_i e_i^* = I_n \\
\]

\[
A A^* = \left( \sum_{i=1}^{n} v_i e_i^* \right) \left( \sum_{j=1}^{n} v_j^* e_j^* \right)^* \\
= \sum_{i \neq j} e_i (v_i e_j^*) v_j^* + \sum_{i=1}^{n} e_i (v_i e_i^*) v_i^* \\
= \sum_{i \neq j} e_i e_i^* + \sum_{i=1}^{n} e_i e_i^* = I_n \\
\]

\[
= \text{Projection onto } R(A) \\
= I_n \\
\]

\Rightarrow The rows of an orthogonal/unitary matrix are orthonormal, too.

**Orthogonal matrix**

\[ A^T = A^{-1} \]

**Unitary matrix**

\[ A^* = A^{-1} \]

rows are not orthogonal for isometric embeddings like

\[
\begin{pmatrix}
0 & 0 & 1 \\
0 & 0 & 0 \\
\end{pmatrix}
\]

More examples:
More examples:

- Permutation matrices, e.g.,

\[
\begin{pmatrix}
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{pmatrix}
\quad \vdots 
\begin{pmatrix}
e_1 & -e_2 \\
e_2 & e_3 \\
e_3 & -e_4 \\
e_4 & e_1
\end{pmatrix}
\]

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

**Scaling I.**

\[
\begin{pmatrix} 2 & 0 \\
0 & 2 \end{pmatrix}
\]
scales every vector up by 2.

\[
\begin{pmatrix} 2 & 0 \\
0 & 0.5 \end{pmatrix}
\]
scales by different amounts.

Need not be axis-aligned...

**Exercise:** Give a 2x2 matrix that maps

\[
\begin{align*}
(1,-1) & \mapsto (2,-2) \\
(1,1) & \mapsto (\frac{1}{2}, \frac{3}{2})
\end{align*}
\]

**Answer:** Note that

\[\begin{pmatrix}
\frac{1}{\sqrt{2}} (1,-1) \\
\frac{1}{\sqrt{2}} (1,1)
\end{pmatrix}\]

is an orthonormal basis.

We want \(A v_1 = 2 v_1\), \(A v_2 = \frac{1}{2} v_2\).

Set

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

= \[
2 \cdot \frac{1}{2}(1,-1) + \frac{1}{2} \cdot \frac{1}{2}(1,1)
\]

= \[
\begin{pmatrix}
5/4 & -3/4 \\
-3/4 & 5/4
\end{pmatrix}
\]

\(A(1,-1) = (2,-2)\)

\(A(1,1) = (\frac{1}{2}, \frac{3}{2})\)

Alternative answer:

The change-of-basis matrix

\[
\begin{pmatrix}
2 & 0 \\
0 & \frac{1}{2}
\end{pmatrix}
\]
$w = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

maps the standard basis to the above basis, and

$w^{-1} = w^T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

goes back.

Hence

$A = w \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} w^{-1}$ works

\[ \text{map } v_1 \rightarrow (1,0) \\
   v_2 \rightarrow (0,1) \]

\[ \text{scale } (1,0) \rightarrow (2,0) \\
   (0,1) \rightarrow (0,\frac{1}{2}) \]

\[ \text{map } (1,0) \rightarrow v_1 \\
   (0,1) \rightarrow v_2 \]

\[ \text{rotate and scale} \\
   \text{rotate back} \]

\[ \frac{1}{2} v_2 \\
   2v_1 \]

---

**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 $\leq \infty$.)

**Note:** Often denoted $\|A\|_2$, for $L_2$/Euclidean norm.

**Properties I:**

- For any vector $x$ (of appropriate dimension),

  $$\|Ax\| \leq \|A\| \cdot \|x\|$$
• For any real/complex number $\alpha$,
  \[ \|\alpha A\| = |\alpha| \|A\|. \]

• **Triangle inequality:**
  \[ \|A + B\| \leq \|A\| + \|B\|. \]

  **Proof:**
  \[
  \|A + B\| = \max_{\|x\| = 1} \|A x + B x\|
  \leq \max_{\|x\| = 1} (\|A x\| + \|B x\|) \quad (\Delta \text{ ineq. for vectors})
  \leq (\max_{\|x\| = 1} \|A x\|) + (\max_{\|y\| = 1} \|B y\|)
  = \|A\| + \|B\|.
  \]

**Examples:**

• \(\|I\| = 1\)
  \(\|\text{any isometry}\| = 1\)
  \(\|\text{any projection}\| = 1\), unless the projection is 0

• What is \(\|(\begin{smallmatrix} 1 & 0 \\ 100 & 1 \end{smallmatrix})\|\)?

  **Answer:**

  1. \(\|A \vec{e}_1\| = \|(\begin{smallmatrix} 1 & 0 \\ 100 & 1 \end{smallmatrix}) \vec{e}_1\| = \sqrt{1 + \frac{1}{100}} = \|A \vec{e}_2\|
      \Rightarrow \|A\| \geq \sqrt{1 + \frac{1}{100}} \)

  2. \(A = (\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}) + \frac{1}{100}(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix})
      \Rightarrow \|A\| \leq \|(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix})\| + \frac{1}{100} \|(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix})\|
      = 1 + \frac{1}{100} \)

  3. We want to find \(\vec{x} = (\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix})\), with \(\|\vec{x}\| = 1\), to maximize
     \[ \|A \vec{x}\|^2 = \| (\begin{smallmatrix} x_1 + \frac{x_2}{100} \\ 100 x_1 + x_2 \end{smallmatrix}) \|^2 \]
     \[ = (x_1 + \varepsilon x_2)^2 + (x_1 + x_2)^2 \]
     \[ = (1 + \varepsilon^2)(x_1^2 + x_2^2) + 4 \varepsilon x_1 x_2 \]
     \[ \|\vec{x}\|^2 = x_1^2 + x_2^2 = 1 \Rightarrow x_2 = \pm \sqrt{1 - x_1^2} \]

     **Why?** If \(x_1 < 0\), multiply \(\vec{x}\) by -1, leaving \(\|\vec{x}\|\) and \(\|A \vec{x}\|\) unchanged.
     If \(x_1 > 0\) and \(x_2 < 0\), then we can increase \((x_1 + \varepsilon x_2)^2\) by switching the sign of \(x_2\).
\[ \|A\|_2 = 1 + \varepsilon^2 + 4\varepsilon x_1 \sqrt{1 - x_1^2} \]

\[ x_1 = \frac{1}{\sqrt{2}} \quad \text{(by calculus)} \]

\[ \|A\| = \sqrt{1 + \varepsilon^2 + 2\varepsilon} = 1 + \varepsilon \sqrt{1 + \varepsilon^2} \]

**Observe:** \( \|A(\frac{1\varepsilon}{\sqrt{2}})\| = 1 + \varepsilon > \|A(\varepsilon)\| = \|A(1)\| = \sqrt{1 + \varepsilon^2} \)

**Moral:** Spreading out is good!

**Problem:** What are the operator norms of

\[ a) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad b) \begin{pmatrix} 1 \\ \varepsilon \end{pmatrix} \]

**Answer:**

\[ a) \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\| = x_1^2 + \varepsilon x_2 \]

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

\[ \Rightarrow \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\| = 1 \quad \text{(if } \varepsilon < 1 \text{).} \]

\[ b) \left\| \begin{pmatrix} 1 \\ \varepsilon \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\| = |(1, \varepsilon) \cdot (x_1, x_2)| \]

To maximize the dot product between \((1, \varepsilon)\) and \(x\), subject to \( \|x\| = 1 \), we should choose \(x\) parallel to \((1, \varepsilon)\), i.e.,

\[ x = \frac{1}{\sqrt{1 + \varepsilon^2}} (1, \varepsilon) \]

\[ \Rightarrow \left\| \begin{pmatrix} 1 \\ \varepsilon \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\| = \frac{1 + \varepsilon}{\sqrt{1 + \varepsilon^2}} = \left\| (1, \varepsilon) \right\| \]

**Observe:**

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

  - In general,
    \[ \left\| \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \right\| = \max \{ \|A\|, \|B\| \} \]
    \[ \left\| \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix} \right\| = \max \{ \|A\|, \|B\|, \|C\| \} \]
    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
    \[ = \text{Euclidean norm of the row vector} \]
(to maximize $|v \cdot x|$, let $x = \frac{v}{\|v\|}$.)

& spectral norm of an $n \times 1$ matrix

$= \text{Euclidean norm of the column vector}$

(just set $x = (1)$)

**Example:** What is the spectral norm of

$$
\mathbf{m} \begin{pmatrix}
1 & 1 & 1 & \ldots & 1 \\
1 & 1 & 1 & \ldots & 1 \\
1 & 1 & 1 & \ldots & 1 \\
1 & 1 & 1 & \ldots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & 1 & \ldots & 1 \\
\end{pmatrix}
$$

the $m \times n$ all-ones matrix?

**Answer:**

1. **Experiment numerically:**

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

\[
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{pmatrix}
\]

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

\[= \text{maybe } \|A\|^2 = \sqrt{m \cdot n} \text{?} \]

2. **Mathematica code:**

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

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

3. **Guess the best input:**

Since the $1 \ldots m$ values are indeed

```
\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 \\
2 & 4 & 6 & 8 & 10 \\
3 & 6 & 9 & 12 & 15 \\
4 & 8 & 12 & 16 & 20 \\
5 & 10 & 15 & 20 & 25 \\
\end{pmatrix}
\]
```
© Guess the best input:
Since the columns are all the same, it makes sense to spread out across them all, and equally:
\[ x = \left( \frac{1}{\sqrt{n}} (1, 1, 1, \ldots, 1) \right) \in \mathbb{R}^n \]
\[ \Rightarrow \| x \| = 1. \]
\[ A x = \frac{1}{\sqrt{n}} \left( \begin{array}{c}
1 \\
1 \\
\vdots \\
1 \\
1
\end{array} \right) \left( \begin{array}{c}
1 \\
1 \\
\vdots \\
1 \\
1
\end{array} \right)
\]
\[ = \frac{1}{\sqrt{n}} (n, n, n, \ldots, n) \in \mathbb{R}^m \]
\[ \Rightarrow \| A x \| = m \cdot n \quad \checkmark \]
\[ \Rightarrow \| A \| \geq \sqrt{m \cdot n} \quad \checkmark \]

© Prove that \( \| A \| < \sqrt{mn} \):
One approach is to argue by symmetry that the above \( x \) is optimal.

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

Since all columns are the same,
\[ \text{rank}(A) = \dim \text{Range}(A) = \# \text{linearly independent columns} = 1. \]

A factors as
\[ A = \left( \begin{array}{c}
1 \\
1 \\
\vdots \\
1 \\
1
\end{array} \right) \left( \begin{array}{c}
1 \\
1 \\
\vdots \\
1 \\
1
\end{array} \right) = v^T \tilde{a} \]
\[ \Rightarrow A x = (v^T x) \tilde{a} \]
\[ \| A x \| = |v^T| \cdot \| x \|, \]
which reaches its maximum, \( \| x \| \cdot \| v \| \),
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
\[ A = \tilde{a} v^T \]
for some vectors \( \tilde{a} \) and \( v \). Hence \( \| A \| = \| u \| \cdot \| v \| \).

**Spectral norm**
**Properties II.**
- \( \| A \| \geq 0 \), and \( \| A \| = 0 \iff A = 0 \)
- \( \| A x \| \leq \| A \| \cdot \| x \| \)
• \( \| A \| = 0 \) and \( \| A \| = 0 \iff A = 0 \)

• \( \| A \mathbf{x} \| \leq \| A \| \cdot \| \mathbf{x} \| \)

\[ \text{matrix norm} \quad \text{vector/matrix norm} \]

• \( \| \lambda A \| = |\lambda| \cdot \| A \| \) for \( \lambda \in \mathbb{C} \)

• \( \| AB \| \leq \| A \| \cdot \| B \| \)

(For 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 
  \[ \| U A V \| = \| A \| \]

(because unitaries don’t change lengths).

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

  e.g. if \( A \) is a diagonal matrix,
  \[ \| A \| = \max_i |a_{ii}|. \]

• If \( \text{rank}(A) = 1 \), with \( A = \mathbf{a} \mathbf{v}^\top \), \( \| A \| = \| \mathbf{u} \| \cdot \| \mathbf{v} \| \).

**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 \( \mathbf{e}_i^*, \mathbf{e}_j^* \) be such that \( |a_{ij}^*| = \max_{i,j} |a_{ij}| \).

Let \( \mathbf{x} = \mathbf{e}_{i,j}^*. \) Then \( \| \mathbf{x} \| = 1 \), so

\[ \| A \| \geq \| A \mathbf{x} \| \]

\[ = \left\| \begin{pmatrix} a_{1j}^* \\ a_{2j}^* \\ \vdots \\ a_{mj}^* \end{pmatrix} \right\| \]

\[ = \sqrt{\sum_i |a_{ij}^*|^2} \]

\[ \geq \max_{i,j} |a_{ij}^*| \]
\[
\max_i |a_{ij}| \leq \sqrt{m} \cdot \max_j |a_{ij}|
\]

Next, let us show the upper bound, \( \|A\| \leq \sqrt{m} \cdot \max_i |a_{ij}| \).

\[
\|A\| = \max_{x: \|x\| = 1} \|Ax\|
\]

Write \( A = \left( \begin{array}{c} r_1 \\ \vdots \\ r_m \end{array} \right) \), so \( Ax = \left( \begin{array}{c} r_1 \cdot x \\ \vdots \\ r_m \cdot x \end{array} \right) \).

\[
\|Ax\| = \max_{\|x\| = 1} \sum_{i=1}^m |r_i \cdot x|^2
\]

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

\[
\sum_{i=1}^m \|r_i\|^2 \leq \sum_{i=1}^m \sum_j |a_{ij}|^2
\]

\[\square\]

**When is a perturbed matrix invertible?**

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

**Proof:**

1. \( I + A \) is not invertible \( \iff \) \( N(I + A) \neq \{0\} \)
2. \( \iff \) \( (I + A)x = 0 \) for some \( x \neq 0 \)
3. \( \iff \) \( Ax = -x \)
4. \( \iff \) \( \|A\| \geq 1 \), a contradiction. \( \square \)

**Lemma:** Let \( A \) be an invertible matrix.

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

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

**Example:** \( A = \left( \begin{array}{cc} 1 & 0 \\ 0 & 1/4 \end{array} \right) \), \( A' = \left( \begin{array}{cc} 1 & 0 \\ 0 & 4 \end{array} \right) \)

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

**Lemma:** If \( \|A\| < 1 \), then

\( (I + A)^{-1} = I - A + A^2 + A^3 + A^4 + \ldots \).

**Proof:** Exercise.

**Other Matrix Norms**

Just as we have defined multiple vector norms, like
**Matrix Norms**

Just as we have defined multiple vector norms, like

\[
\|v\| = \left\{ \sum_i |v_i|^2 \right\}^{1/2} \quad \text{Euclidean}
\]

\[
\|v\|_1 = \sum_i |v_i| \quad \text{ } l_1 \text{ norm}
\]

\[
\|v\|_r = \left( \sum_i |v_i|^r \right)^{1/r} \quad \text{ } l_r \text{ norm}
\]

we can define many different matrix norms.

**Example:**

\[
\|A\|_r = \max_{\|x\|_r = 1} \|Ax\|_r
\]

\[
\|A\|_q = \max_{\|x\|_q = 1} \|Ax\|_q
\]

**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:**

\[
\|A\|_1 = \max_{\|x\|_1 = 1} \|Ax\|_1
\]

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, i.e., find \( x \) with \( \|x\|_1 = 1 \) so \( \|Ax\|_1 = K \).

\[\begin{align*}
\|A\|_1 &= \max_{x_1, x_2} \left( |5x_1 + 9x_2| + |-6x_1 + 1x_2| \right) \\
&= \max_{\|x\|_1 = 1} \left( |5 + 9| + |9 + 1| \right) \\
&= \max \{ 11, 11 \} \\
&= 11
\end{align*}\]

\[\Rightarrow \|A\|_1 \leq 11\]

\[\Rightarrow \|A\|_1 = 11\]

In general, \( \|A\|_1 = \max_{\text{columns } j} \sum_i |a_{i,j}| \)

the maximum \( l_1 \) norm of a column.
General properties of matrix norms:

All the above norms satisfy:

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

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

Answer: It does satisfy the first three properties. But sub-multiplicativity is harder

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

$\Rightarrow AB = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$

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

So NO, $f$ is not sub-multiplicative.

Example: Frobenius norm

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

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

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

$$= \sum (A^\dagger A)_{i,i}$$

$$= \sum (A^\dagger)_{i,j} (A)_{j,i}$$

$$= \sum_i a_{i,i} a_{i,i}$$

$$= \sum_i |a_{i,i}|^2 \quad \checkmark$$

Fact: The trace is cyclic:

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

Proof: $\text{Tr}(AB) = \sum (AB)_{i,i}$
Proof: \( \text{Tr}(AB) = \sum (AB)_{ij} \)  
= \( \sum_{i,j} a_{ij} b_{ij} \)  
= \( \sum_{i} b_{ij} a_{ij} = \text{Tr}(BA) \) \( \Box \)

**Corollary:** The Frobenius norm is basis-independent, i.e.,  
\( \| A \|_F = \| UAU^* \|_F \)  
for any unitary/orthogonal matrix \( U \).  
(The Frobenius norm is the same in all orthonormal bases.)

Proof: Since \( U \) is unitary, \( U^* = U^{-1} \).  
\( \| UAU^* \|_F = \text{Tr} ((UAU^*)^*(UAU^*)) \)  
= \( \text{Tr} (U^*A^*U^*U^*AU^*) \) since \( (AB)^* = B^*A^* \)  
= \( \text{Tr} (A^*U^*A^*U^*U) \) cyclic trace  
= \( \text{Tr} (A^*A) \) since \( U^*U = I \) \( \checkmark \)

The spectral norm is also basis-independent, \( \| A \|_2 = \| UAU^* \|_2 \) 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 \( A \in \mathbb{R}^{m \times n} \),  
\( \| A \|_2 \leq \| A \|_F \leq \min \{m,n\} \cdot \| A \|_2 \)

Proof:  
We have already shown \( \| A \|_2 \leq \| A \|_F \).  
\( \| A \|_F \leq \min \{m,n\} \cdot \| A \|_2 \):  
We can’t prove this yet! Fortunately, it is less important.

Later, it will follow since  
\( \| A \|_F^2 = \text{Tr}(A^*A) = \text{sum of eigenvalues of } A^*A \)  
\( \| A \|_2^2 = \text{largest eigenvalue of } A^*A \) ...

Key Lemma:

If \( \|A\|_2 = \|Ax\|_2 \)

ie, \( x \) is stretched maximally by \( A \), achieving the norm,

then

\[
A^\top A x = \|A\|_2^2 x
\]

Corollary: \( \|A^\top\| = \|A\|_2 \).

Proof: For \( y = Ax \), \( A^\top y = \|A\|_2^2 x = \|A\|_2 \|y\|_2 \)

\[\Rightarrow \|A^\top\| \geq \|A\|_2.\]

The same inequality with \( A \leftrightarrow A^\top \) switched gives \( \|A\|_2 \geq \|A^\top\| \). \( \square \)

Corollary: \( \|A^\top A\| = \|A\|_2^2 = \|AA^\top\| \).

Proof: We showed before that \( \|AB\| \leq \|A\|_1 \cdot \|B\|_1 \). Hence,

\[
\|A^\top A\| = \|A\|_1 \cdot \|A\|_1
\]

\[= \|A\|_2^2,
\]

and by the lemma this norm is achieved by the same \( x \)

that achieves \( \|Ax\| = \|A\|_2 \cdot \|x\| \). \( \checkmark \)

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 in these bases is

\[
\begin{pmatrix}
\|A\|_2 \\
0 \\
\vdots
\end{pmatrix}
\]

because \( Ax = \|A\|_2 y \). Then \( A^\top y \) is the first row of \( A \)

\[\Rightarrow A^\top y = \|A\|_2 x + b
\]

where \( b \perp x \), (\( b \) is everything else in the first row)
We claim that $b = 0$! (So $A^* x = \|A^* x\|$, our goal.)

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

$$\begin{align*}
\left\| A \frac{\|A\| x + b}{\|A\|^2 + \|b\|^2} \right\| & \geq \frac{y \cdot (A \frac{\|A\| x + b}{\|A\|^2 + \|b\|^2})}{\sqrt{\|A\|^2 + \|b\|^2}} \\
& = \frac{\|A x + b\|^2}{\sqrt{\|A\|^2 + \|b\|^2}} \\
& = \sqrt{\|A\|^2 + \|b\|^2}
\end{align*}$$

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

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

**Corollary:** $\|A^T\| = \|A\|$, even for complex matrices,

since $\|A^T\| = \|A^\dagger\|$ — the difference between $A^*$ and $A^T$
is just complex conjugation, which doesn’t change any lengths.

**Corollary:** $\|A \cdot A^* \cdot A \cdot A^* \cdot \ldots \cdot A \cdot A^*\| = \|A\|^{2m}$
m times

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

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

but $A^n = A$
since $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

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

**Singular-Value Decomposition (SVD)**

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

$\begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}$