THE RESILIENCE AND ROBUSTNESS (AND HOPEFULLY RESPONSIBILITY) OF LINEAR ALGEBRA
LECTURE 3

Emily J. King

Mathematics Department and Data Science Research Institute
Colorado State University

IK Interdisciplinary College
March 2024

Course Outline

- Lecture 1
  - Signals and vector spaces
  - Linear combinations
- Lecture 2
  - Inner product
  - Inner product as similarity
  - Moving average
  - Convolution and cross correlation
  - **2*D*** convolution and cross correlation
- **Lecture 3**
  - **Cosine similarity**
  - **Projection**
  - **Change of basis**
  - **Graph theory**
- Lecture 4
  - Eigenvalues and eigenvectors
  - Principal component analysis
  - Responsibility vignette

Lecture Outline

Cosine similarity

Projection

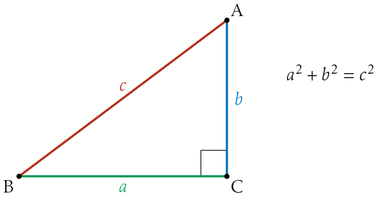
Change of basis

Graph theory

“PYTHAGOREAN” THEOREM

In a right triangle with hypotenuse length c and other two sides of length a and b,

$$a^2 + b^2 = c^2.$$



Pythagoras? This concept was independently discovered by many civilizations with written records up to a thousand years before Pythagoras. We should be more RESPONSIBLE about our naming conventions.

Inspired by this, we define the (Euclidean) norm of vectors in $\mathbb{R}^d$ as

$$\left\| \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} \right\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_d^2}$$

Getting mathy

Let V be a vector space.
A norm is a function $\|\cdot\|: V \to \mathbb{R}$ such that:

- $\|\vec{x}\| \geq 0$ for all $\vec{x} \in V$;
- $\|\vec{x}\| = 0$ if and only if $\vec{x} = \vec{0}$;
- $\|a\vec{x}\| = |a| \, \|\vec{x}\|$ for all $a \in \mathbb{R}$ and $\vec{x} \in V$; and
- $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ for all $\vec{x}, \vec{y} \in V$ (triangle inequality).

ℓ^p Norms

We have the following collection of norms/"norms" on $\mathbb{R}^d$:

- $p = \infty$: $\|\vec{x}\|_\infty = \max_{1 \le j \le d} |x_j| = \lim_{p \to \infty} \|\vec{x}\|_p$
  (ℓ^∞ norm, sup norm, Chebyshev distance)

ℓ^p NORMS

We have the following collection of norms/"norms" on $\mathbb{R}^d$:

- $p = \infty$: $\|\vec{x}\|_\infty = \max_{1 \le j \le d} |x_j| = \lim_{p \to \infty} \|\vec{x}\|_p$
  (ℓ^∞ norm, sup norm, Chebyshev distance)

- $1 \le p < \infty$: $\|\vec{x}\|_p = \left(\sum_{j=1}^d |x_j|^p\right)^{1/p}$
  (ℓ^p norms, $p = 2$ Euclidean, $p = 1$ Manhattan/taxicab metric)

ℓ^p Norms

We have the following collection of norms/"norms" on $\mathbb{R}^d$:

- $p = \infty$: $\|\vec{x}\|_\infty = \max_{1 \le j \le d} |x_j| = \lim_{p \to \infty} \|\vec{x}\|_p$
  (ℓ^∞ norm, sup norm, Chebyshev distance)
- $1 \le p < \infty$: $\|\vec{x}\|_p = \left(\sum_{j=1}^d |x_j|^p\right)^{1/p}$
  (ℓ^p norms, $p = 2$ Euclidean, $p = 1$ Manhattan/taxicab metric)
- $0 < p < 1$: $\|\vec{x}\|_p = \left(\sum_{j=1}^d |x_j|^p\right)^{1/p}$
  (ℓ^p quasinorm, fails triangle inequality)
- $p = 0$: $\|\vec{x}\|_0 = \#\{x_j \neq 0\} = \lim_{p \to 0} \|\vec{x}\|_p^p$
  ("ℓ^0", neither norm nor quasinorm)

"In many cases the mathematical simplicity of ℓ^2 is a misleading fact that diverts engineers from making better choices." – Michael Elad *Sparse and Redundant Representations*

EXAMPLES

$$\left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|_1 = |1| + |1| = 2$$

$$\left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|_2 = \sqrt{|1|^2 + |1|^2} = \sqrt{2} \approx 1.414$$

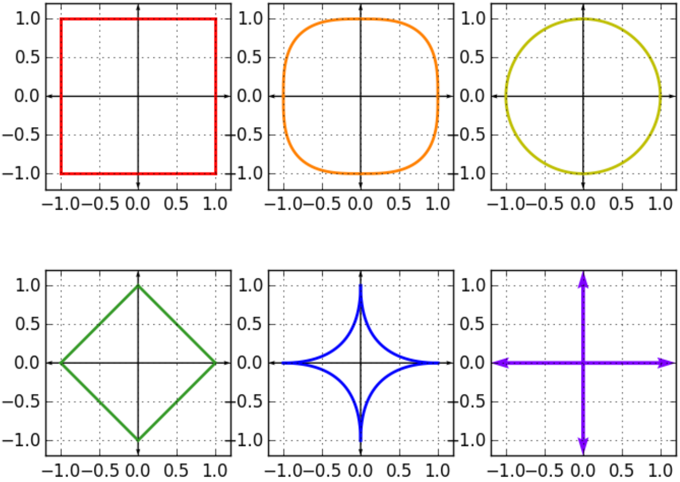
$$\left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|_\infty = \max\{|1|, |1|\} = 1$$

$$\left\| \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix} \right\|_1 = \left|-\sqrt{2}\right| + |0| = \sqrt{2} \approx 1.414$$

$$\left\| \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix} \right\|_2 = \sqrt{\left|-\sqrt{2}\right|^2 + |0|^2} = \sqrt{2} \approx 1.414$$

$$\left\| \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix} \right\|_\infty = \max\{\left|-\sqrt{2}\right|, |0|\} = \sqrt{2} \approx 1.414$$

ℓ^p Spheres



Various ℓ^p unit spheres.

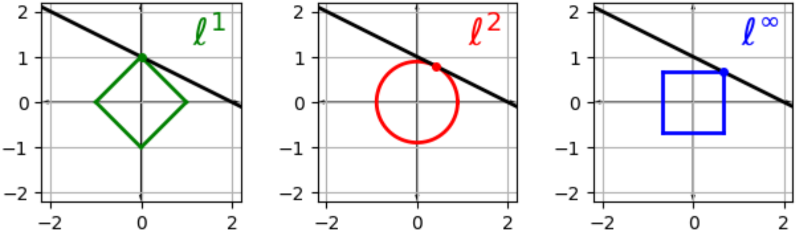
"ℓ^1 is the new ℓ^2" – Stan Osher

ℓ^p distance of $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ is the ℓ^p norm of their difference.

ℓ^p distance of $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ is the ℓ^p norm of their difference.

$$\begin{aligned}
d_1\left(\begin{pmatrix} 0.5 \\ 0.1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix}\right) &= \left\| \begin{pmatrix} 0.5 \\ 0.1 \end{pmatrix} - \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\|_1 \\
&= \left\| \begin{pmatrix} 0.5 - (-2) \\ 0.1 - 1 \end{pmatrix} \right\|_1 \\
&= |2.5| + |-0.9| = 3.4
\end{aligned}$$

$\mathbf{min}_x \| \boldsymbol{x} \|_p$ S.T. $\boldsymbol{Ax} = \boldsymbol{b}$



Plots of solutions to $\mathbf{min} \| x \|_p$ such that $\boldsymbol{Ax} = \boldsymbol{y}$ for (left-to-right) $\boldsymbol{p} = \mathbf{1}, \mathbf{2}, \boldsymbol{\infty}$.

$\min_x \| x \|_p$ s.t. $Ax = b$ but unconstrained

$$\min_{x \in \mathbb{R}^n} \| b - Ax \|_2^2 + \lambda \| Tx \|_p^p,$$

for some $\lambda > 0$ and matrix T.

$\min_x \| x \|_p$ S.T. $Ax = b$ BUT UNCONSTRAINED

$$\min_{x \in \mathbb{R}^n} \| b - Ax \|_2^2 + \lambda \| Tx \|_p^p ,$$

for some $\lambda > 0$ and matrix T.

$p = 2 \;\; \Rightarrow$ Tikhonov regularization

$\mathbf{min}_x \| x \|_p$ S.T. $Ax = b$ BUT UNCONSTRAINED

$$\min_{x \in \mathbb{R}^n} \| b - Ax \|_2^2 + \lambda \| Tx \|_p^p,$$

for some $\lambda > 0$ and matrix T.

$p = 2 \;\; \Rightarrow$ Tikhonov regularization

$p = 2, T = I \;\; \Rightarrow$ L^2-regularization

$\mathbf{min}_x \| x \|_p$ S.T. $Ax = b$ BUT UNCONSTRAINED

$$\min_{x \in \mathbb{R}^n} \| b - Ax \|_2^2 + \lambda \| Tx \|_p^p,$$

for some $\lambda > 0$ and matrix T.

$p = 2 \;\Rightarrow\;$ Tikhonov regularization

$p = 2, T = I \;\Rightarrow\;$ L^2-regularization

$p = 1, T = I \;\Rightarrow\;$ LASSO $\approx$ basis pursuit (denoising)

$\min_x \| x \|_p$ S.T. $Ax = b$ BUT UNCONSTRAINED

$$\min_{x \in \mathbb{R}^n} \| b - Ax \|_2^2 + \lambda \| Tx \|_p^p,$$

for some $\lambda > 0$ and matrix T.

$p = 2 \quad \Rightarrow \quad$ Tikhonov regularization

$p = 2, T = I \quad \Rightarrow \quad$ L^2-regularization

$p = 1, T = I \quad \Rightarrow \quad$ LASSO $\approx$ basis pursuit (denoising)

$p = \infty$ comes up in computer-aided manufacturing (CAM) optimization problems

Fact: For any $a > 0$,
$a\vec{x}$ is the unique vector with norm $a \left\| \vec{x} \right\|$ pointing in the exact same direction as $\vec{x}$.

Fact: For any $a > 0$,
$a\vec{x}$ is the unique vector with norm $a \|\vec{x}\|$ pointing in the exact same direction as $\vec{x}$.

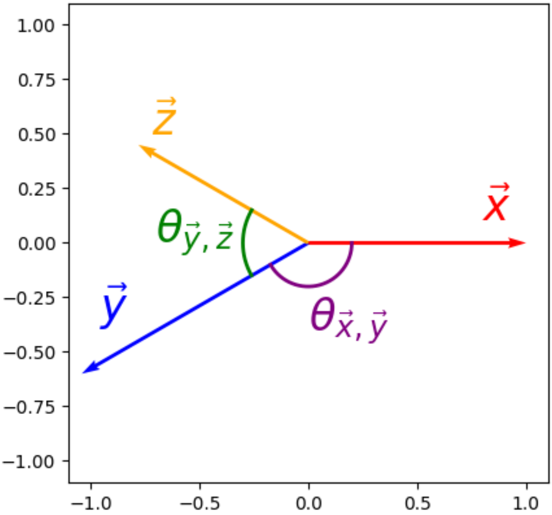
For $\vec{x} \neq \vec{0}$, the normalization of $\vec{x}$ is

$$\frac{\vec{x}}{\|\vec{x}\|}.$$

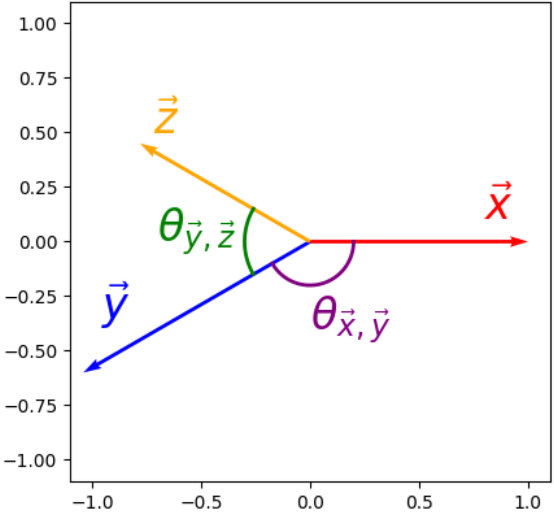
This is the unique vector with norm 1 that points in the exact same direction as $\vec{x}$.

We've seen how a standard inner product gives some scaled measure of similarity between vectors.

Now we consider how to compare data in a scale-free sense.



Three vectors in $\mathbb{R}^2$ and included angles between two pairs of the three.



Let $\vec{x}, \vec{y} \in \mathbb{R}^d \setminus \{\vec{0}\}$, and let $\theta_{\vec{x},\vec{y}}$ be the included angle between them. Then

$$\langle \vec{x}, \vec{y} \rangle = \| \vec{x} \| \, \| \vec{y} \| \cos(\theta_{\vec{x},\vec{y}}).$$

So, the inner product of two vectors tells you about their

norms (“lengths”) and the angle between them.

So, the inner product of two vectors tells you about their
norms (“lengths”) and the angle between them.

This operation that we’ve previously only defined
algebraically is chock full of geometric information!!

Using this, we define the cosine similarity of $\vec{x}$ and $\vec{y}$ as

$$\left\langle \frac{\vec{x}}{\|\vec{x}\|}, \frac{\vec{y}}{\|\vec{y}\|} \right\rangle = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \, \|\vec{y}\|} = \cos(\theta_{\vec{x},\vec{y}}).$$

If $\vec{x}$ or $\vec{y}$ is the zero vector, then the cosine similarity is defined to be equal to 0.

Cosine similarity is always between -1 and 1.

Inner product vs. cosine similarity

(Geogebra worksheet)

`https://www.geogebra.org/m/hfsc8dwg`

Application of cosine similarity

	Thumbprint cookies	Espresso shortbread	Soft pull-apart rolls	Roasted cauliflower
Jam	158	0	0	0
Butter	113	227	85.5	0
White sugar	200	0	0	0
Brown sugar	0	100	0	0
Egg	50	0	150	0
Vanilla extract	13	4.3	0	0
Finely ground, uncooked oatmeal	100	0	0	0
All-purpose flour	156	281	300	0
Baking soda	2.4	0	0	0
Salt	1.5	1.5	25.6	0.7
Powdered instant espresso	0	2	0	0
100% hydration sourdough starter	0	0	375	0
Milk	0	0	360	0
Whole wheat flour	0	0	400	0
Bread flour	0	0	400	0
Honey	0	0	170	0
Vegetable oil	0	0	80	27
Cauliflower	0	0	0	588
Ground cayenne	0	0	0	0.9

Application of cosine similarity

	Thumbprint cookies	Espresso shortbread	Soft pull-apart rolls	Roasted cauliflower
Jam	158	0	0	0
Butter	113	227	85.5	0
White sugar	200	0	0	0
Brown sugar	0	100	0	0
Egg	50	0	150	0
Vanilla extract	13	4.3	0	0
Finely ground, uncooked oatmeal	100	0	0	0
All-purpose flour	156	281	300	0
Baking soda	2.4	0	0	0
Salt	1.5	1.5	25.6	0.7
Powdered instant espresso	0	2	0	0
100% hydration sourdough starter	0	0	375	0
Milk	0	0	360	0
Whole wheat flour	0	0	400	0
Bread flour	0	0	400	0
Honey	0	0	170	0
Vegetable oil	0	0	80	27
Cauliflower	0	0	0	588
Ground cayenne	0	0	0	0.9

Cosine similarity of thumbprint cookie vector with …

… the espresso shortbread vector $\approx \mathbf{0.55}$

… the soft pull-apart rolls vectors $\approx \mathbf{0.22}$

… the roasted cauliflower $\approx \mathbf{5.2 \times 10^{-6}}$

The cosine similarity of the thumbprint cookie vector with the espresso shortbread vector is $\approx \mathbf{0.55}$

with the soft pull-apart rolls vectors is $\approx \mathbf{0.22}$,

and with the roasted cauliflower is $\approx \mathbf{5.2 \times 10^{-6}}$

Doc 1	The first half of the Wikipedia article on Dianne O'Leary, an important researcher in computational linear algebra
Doc 2	The full Wikipedia article on Dianne O'Leary
Doc 3	The full Wikipedia article on Margaret Wright, an important researcher in computational linear algebra
Doc 4	The full Wikipedia article on Flower Travellin' Band, a Japanese progressive and psychedelic rock band from the 1970's
Doc 5	The full Wikipedia article on Satori, the Flower Travellin' Band's debut album of original material
Doc 6	Half of the Wikipedia article on Sonic Hedgehog protein, a signaling molecule that plays an important role in many cell processes starting in fetal development

We convert the articles to vectors by first throwing out the stop words (like "the" and "is") and making vectors with the counts of each word in each document.

There are **1469** total words (or character sequences parsed like words by Matlab) amongst the six documents, so each document is assigned a vector in $\mathbb{R}^{1469}$, where an entry is **0** if the corresponding word does not appear in the given article.

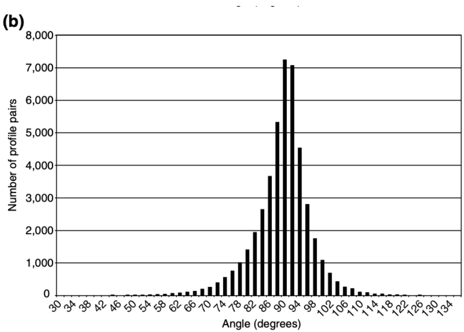
We replace the non-zero entries with the term frequency-inverse document frequency (tf-idf)[1], which takes into account how common a word is within a document versus across the compared documents.

(Go to Matlab)

[1] We won't define tf-idf in this lecture.

Matrix of cosine similarities

$$
\begin{array}{cc}
 & \begin{array}{cccccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \end{array} \\
\begin{array}{c} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \\ \mathbf{5} \\ \mathbf{6} \end{array} &
\begin{bmatrix}
\mathbf{1.0000} & \mathbf{0.6223} & \mathbf{0.1259} & \mathbf{0.0059} & \mathbf{0.0024} & \mathbf{0.0023} \\
\mathbf{0.6223} & \mathbf{1.0000} & \mathbf{0.1849} & \mathbf{0.0046} & \mathbf{0.0046} & \mathbf{0.0035} \\
\mathbf{0.1259} & \mathbf{0.1849} & \mathbf{1.0000} & \mathbf{0.0297} & \mathbf{0.0104} & \mathbf{0.0093} \\
\mathbf{0.0059} & \mathbf{0.0046} & \mathbf{0.0297} & \mathbf{1.0000} & \mathbf{0.3489} & \mathbf{0.0053} \\
\mathbf{0.0024} & \mathbf{0.0046} & \mathbf{0.0104} & \mathbf{0.3489} & \mathbf{1.0000} & \mathbf{0.0045} \\
\mathbf{0.0023} & \mathbf{0.0035} & \mathbf{0.0093} & \mathbf{0.0053} & \mathbf{0.0045} & \mathbf{1.0000}
\end{bmatrix}
\end{array}
$$



A histogram of the angles between the vectors in **44850** pairs formed from a set of **300** expression vectors taken from the Rosetta Inpharmatics website. Source: Kuruvilla, Park, Schreiber 2004

Cosine similarity of two different gene vectors give a sense of how similar the corresponding genes are expressed across all of the experiments. The distribution of values of cosine similarities of pairs of experiment vectors is also informative.

Warning about cosine similarity

Some authors define cosine similarity between $\vec{x}$ and $\vec{y}$ to be

$$\frac{\left\langle \vec{x} - \mu_{\vec{x}}, \vec{y} - \mu_{\vec{y}} \right\rangle}{\| \vec{x} - \mu_{\vec{x}} \| \, \| \vec{y} - \mu_{\vec{y}} \|},$$

where $\mu_{\vec{x}}$ and $\mu_{\vec{y}}$ are the respective average values of the components of $\vec{x}$ and $\vec{y}$,

WARNING ABOUT COSINE SIMILARITY

Some authors define cosine similarity between $\vec{x}$ and $\vec{y}$ to be

$$\frac{\langle \vec{x} - \mu_{\vec{x}}, \vec{y} - \mu_{\vec{y}} \rangle}{\| \vec{x} - \mu_{\vec{x}} \| \, \| \vec{y} - \mu_{\vec{y}} \|},$$

where $\mu_{\vec{x}}$ and $\mu_{\vec{y}}$ are the respective average values of the components of $\vec{x}$ and $\vec{y}$,

but this already has a name:
correlation (coefficient)!

WARNING ABOUT COSINE SIMILARITY

Some authors define cosine similarity between $\vec{x}$ and $\vec{y}$ to be

$$\frac{\langle \vec{x} - \mu_{\vec{x}}, \vec{y} - \mu_{\vec{y}} \rangle}{\| \vec{x} - \mu_{\vec{x}} \| \, \| \vec{y} - \mu_{\vec{y}} \|},$$

where $\mu_{\vec{x}}$ and $\mu_{\vec{y}}$ are the respective average values of the components of $\vec{x}$ and $\vec{y}$,

but this already has a name:
correlation (coefficient)[2]!

[2] Pearson was a piece of shit who tried to use mathematics and statistics to give a scientific veneer to eugenics. He also wasn't even the first person to write about the quantity. Quoting Pearson: "My view – and I think it may be called the scientific view of a nation is that of an organized whole, kept up to a high pitch of internal efficiency by insuring that its numbers are substantially recruited from the better stocks, and kept up to a high pitch of external efficiency by contest, chiefly by way of war with inferior races."

Lecture Outline

Cosine similarity

Projection

Change of basis

Graph theory

We've seen how the

inner product $\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \rangle$

and cosine similarity $\left\langle \frac{\vec{\boldsymbol{x}}}{\|\vec{\boldsymbol{x}}\|}, \frac{\vec{\boldsymbol{y}}}{\|\vec{\boldsymbol{y}}\|} \right\rangle$

yield symmetric measures of similarity of vectors $\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}$,

with the former being scaled and the latter unscaled.

We've seen how the

inner product $\langle \vec{x}, \vec{y} \rangle$

and cosine similarity $\left\langle \frac{\vec{x}}{\|\vec{x}\|}, \frac{\vec{y}}{\|\vec{y}\|} \right\rangle$

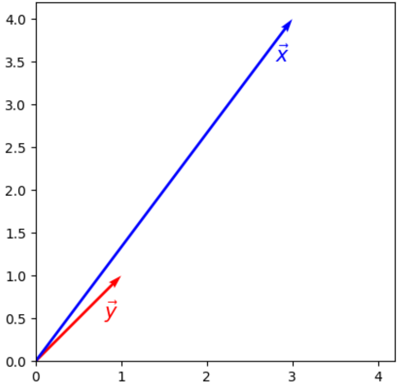
yield symmetric measures of similarity of vectors $\vec{x}, \vec{y}$,

with the former being scaled and the latter unscaled.

We finish our journey with the inner product with an asymmetric comparison.

Scalar Projection

"how much" of one vector's structure is present in the other



$$\left\langle \vec{x}, \frac{\vec{y}}{\| \vec{y} \|} \right\rangle = \left\langle \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \right\rangle$$

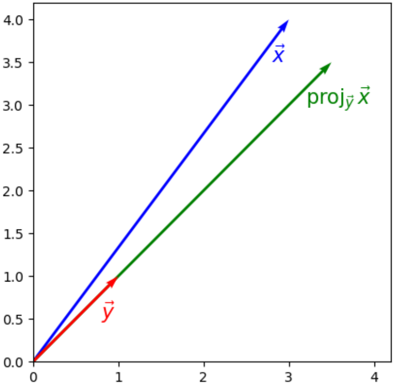
$$= 3/\sqrt{2} + 4/\sqrt{2}$$

$$\approx 4.9497$$

Note: $\| \vec{x} \| = 5$.

Vector Projection / Orthogonal Projection onto Line

"what part" of one vector is has the same structure as the other



$$
\begin{aligned}
\mathbf{proj}_{\vec{y}}\,\vec{x} &= \left\langle \vec{x}, \frac{\vec{y}}{\|\vec{y}\|} \right\rangle \frac{\vec{y}}{\|\vec{y}\|} \\
&= \left\langle \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \right\rangle \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \\
&= 7/\sqrt{2} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \\
&= \begin{pmatrix} 3.5 \\ 3.5 \end{pmatrix}
\end{aligned}
$$

A LITTLE JANET JACKSON

(Go to YouTube)

`https://www.youtube.com/watch?v=OAwaNWGLM0c`

Lecture Outline

Cosine similarity

Projection

Change of basis

Graph theory

Orthonormal Bases

$\{\vec{u}_i\}_{i=1}^{d}$ is an orthonormal basis (onb) for $\mathbb{R}^d$ if

$$\langle \vec{u}_i, \vec{u}_j \rangle = \begin{cases} 1 & ; \quad i = j \\ 0 & ; \quad i \neq j \end{cases}$$

Orthonormal Bases

$\{\vec{u}_i\}_{i=1}^d$ is an orthonormal basis (onb) for $\mathbb{R}^d$ if

$$\langle \vec{u}_i, \vec{u}_j \rangle = \begin{cases} 1 & ; \quad i = j \\ 0 & ; \quad i \neq j \end{cases}$$

If $\{\vec{u}_i\}_{i=1}^d$ is an onb, then for any $\vec{x} \in \mathbb{R}^d$,

$$\vec{x} = \sum_{i=1}^{d} \langle \vec{x}, \vec{u}_i \rangle \, \vec{u}_i \tag{1}$$

Orthonormal Bases

$\{\vec{u}_i\}_{i=1}^d$ is an orthonormal basis (onb) for $\mathbb{R}^d$ if

$$\langle \vec{u}_i, \vec{u}_j \rangle = \begin{cases} 1 & ; \quad i = j \\ 0 & ; \quad i \neq j \end{cases}$$

If $\{\vec{u}_i\}_{i=1}^d$ is an onb, then for any $\vec{x} \in \mathbb{R}^d$,

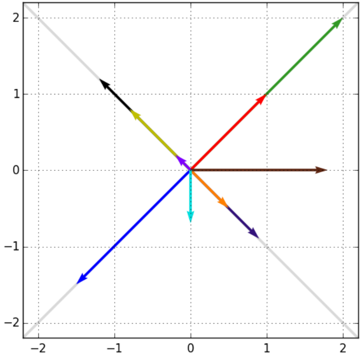
$$\vec{x} = \sum_{i=1}^{d} \langle \vec{x}, \vec{u}_i \rangle \, \vec{u}_i \tag{1}$$

We can view (1) as a two-stage process

$$\vec{x} \mapsto \{\langle \vec{x}, \vec{u}_i \rangle\}_i^d$$

$$\{\alpha_i\}_i^d \mapsto \sum_{i=1}^{d} \alpha_i \vec{u}_i.$$

Example: Lossless Compression



$$\vec{e}_1 = \begin{pmatrix}1\\0\end{pmatrix}, \vec{e}_2 = \begin{pmatrix}0\\1\end{pmatrix}, \vec{u}_1 = \begin{pmatrix}\sqrt{2}/2\\\sqrt{2}/2\end{pmatrix}, \vec{u}_2 = \begin{pmatrix}-\sqrt{2}/2\\\sqrt{2}/2\end{pmatrix}$$

$\langle \vec{x}, \vec{e}_1 \rangle$	$\langle \vec{x}, \vec{e}_2 \rangle$	$\langle \vec{x}, \vec{u}_1 \rangle$	$\langle \vec{x}, \vec{u}_2 \rangle$
·	·	·	0
·	·	·	0
·	·	·	0
·	·	0	·
·	·	0	·
·	·	0	·
·	·	0	·
·	·	0	·
0	·	·	·
·	0	·	·

$\cdot \neq 0$

More common = smaller storage ≈ Huffmann coding, ZIP, PNG, etc.

Applied Harmonic Analysis

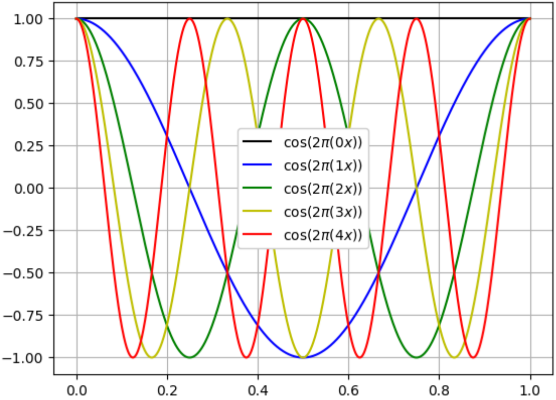
$$\underset{\substack{\text{e.g.: image,}\\ \text{function, signal}}}{\overset{\text{object}}{\boxed{\vec{x}}}} \xrightarrow[\text{analysis}]{\Phi\vec{x}=\{\langle\vec{x},\vec{u}_i\rangle\}_i} \overset{\substack{\text{transformed}\\ \text{object}}}{\boxed{\Phi\vec{x}}} \xrightarrow[\text{processing}]{\substack{\textbf{store, transmit,}\\ \textbf{modify, etc.}}} \boxed{\widetilde{\Phi\vec{x}}} \xrightarrow[\text{synthesis}]{\sum_i(\widetilde{\Phi\vec{x}})_i\vec{u}_i} \boxed{\tilde{\vec{x}}}$$

or

$$\underset{\substack{\text{e.g.: image,}\\ \text{function, signal}}}{\overset{\text{object}}{\boxed{\vec{x}}}} \xrightarrow[\text{analysis}]{\Phi\vec{x}=\{\langle\vec{x},\vec{u}_i\rangle\}_i} \overset{\substack{\text{transformed}\\ \text{object}}}{\boxed{\Phi\vec{x}}} \xrightarrow[\text{(non-linear map)}]{\text{determine characteristics}} \boxed{T\Phi\vec{x}}$$

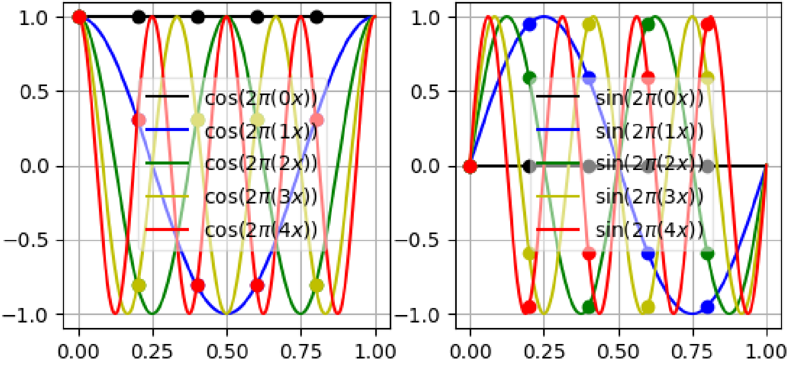
Goal: Find the best building blocks $\{\vec{u}_i\}$ for the task at hand.

Cosine Functions of Varying Frequencies



Graphs of the functions $\cos(2\pi m x)$ for $m \in \{0, 1, 2, 3, 4\}$

Vectors



Graphs of the functions $\cos(2\pi mx)$ (left) and $\sin(2\pi mx)$ (right) for $m \in \{0, 1, 2, 3, 4\}$, sampled at the points $\{0/5, 1/5, 2/5, 3/5, 4/5\}$.

What Is a Fourier Transform?

- The basic idea is that you decompose a function/signal/image/data into frequency components.

[3] For certain infinite dimensional domains, the corresponding Fourier transform only looks like an inner product.

What Is a Fourier Transform?

- The basic idea is that you decompose a function/signal/image/data into frequency components.

- General Fourier transform: $f : G \to \mathbb{C}$

$$\hat{f}(\gamma) = \left\langle f, e^{2\pi i \gamma \cdot} \right\rangle, \quad \gamma \in \widehat{G},^{3}$$

where $e^{2\pi i x y} = \cos(2\pi x y) + i \sin(2\pi x y), \quad x, y \in \mathbb{R}$.

[3] For certain infinite dimensional domains, the corresponding Fourier transform only looks like an inner product.

What Is a Fourier Transform?

- The basic idea is that you decompose a function/signal/image/data into frequency components.

- General Fourier transform: $f : G \to \mathbb{C}$

$$\hat{f}(\gamma) = \left\langle f, e^{2\pi i \gamma \cdot} \right\rangle, \quad \gamma \in \widehat{G},^{3}$$

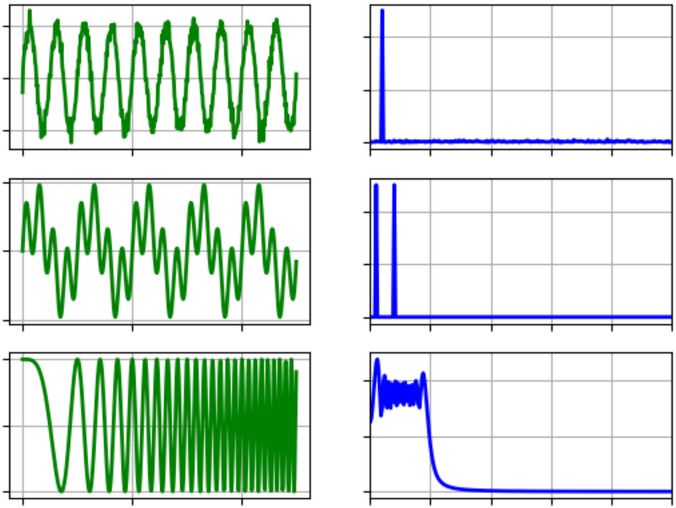
where $e^{2\pi i x y} = \cos(2\pi x y) + i \sin(2\pi x y), \quad x, y \in \mathbb{R}$.

- Essentially,

$$\left| \left\langle f, e^{2\pi i \gamma \cdot} \right\rangle \right|$$

is "how much" frequency γ is in signal f.

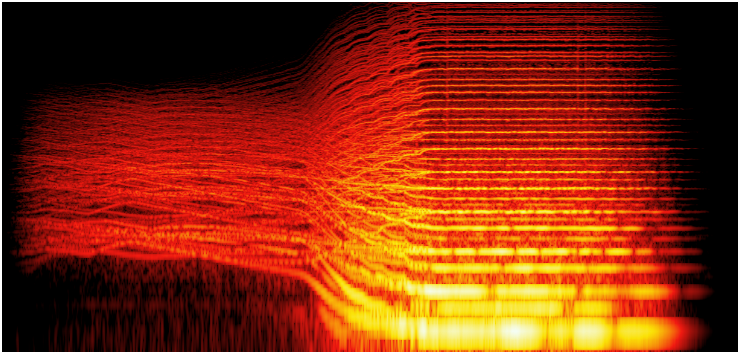
[3] For certain infinite dimensional domains, the corresponding Fourier transform only looks like an inner product.



Left-to-right: The function (centered at the non-zero values) and the absolute value of the Fourier transform (with only non-negative frequencies plotted).

Hey, now we can compute those linear combinations from Lecture 1!

THX Deep Note



THX Deep Note Spectrogram. Image source:
`https://beautifulspectrograms.tumblr.com/`

Rows = convolutions of signal with modulated window
Columns = Fourier transform of windowed signal

EXAMPLE: LOSSY COMPRESSION

JPEG uses DCT-II (basically Fourier transform)

Playing with music

(Go to Matlab)

The Matlab script performed an orthogonal projection onto the subspace spanned by the high frequency vectors.

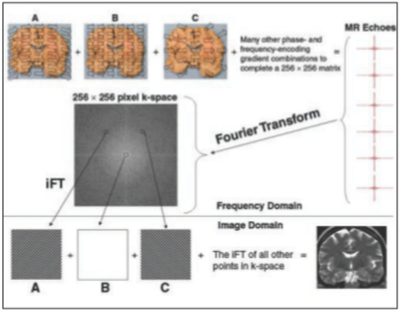
This is a generalization of $\mathbf{proj}_{\vec{y}}\,\vec{x}$, which projected $\vec{x}$ onto the subspace spanned by $\vec{y}$.

Example: magnetic resonance imaging (MRI)

MRI measurements are of spatial frequencies in k-space, a Fourier domain.

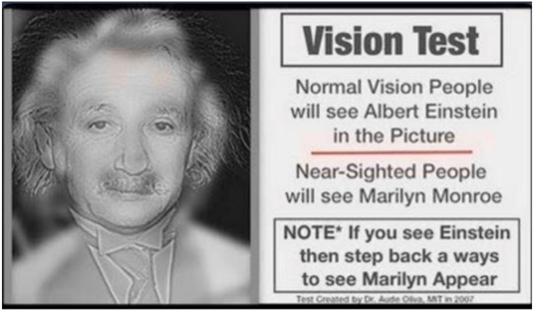
An inverse Fourier transform must be applied to obtain the final image.

The idea here: Due to the physics of the measurement device, we must transform the data to make it usable.



Source: Gallagher et al. 2008

DATA FUSION



Fusion of the high frequency Fourier coefficients of an image of Albert Einstein and the low frequency Fourier coefficients of an image of Marilyn Monroe.
Download Source: `http://media.carbonated.tv/102215_story__ph1.png`, Original Source: Dr. Aude Oliva MIT

Lecture Outline

Cosine similarity

Projection

Change of basis

Graph theory

How do we multiply matrices?

$$\begin{pmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{4} & \mathbf{5} & \mathbf{6} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} \end{pmatrix}$$

Six Degrees of Kevin Bacon

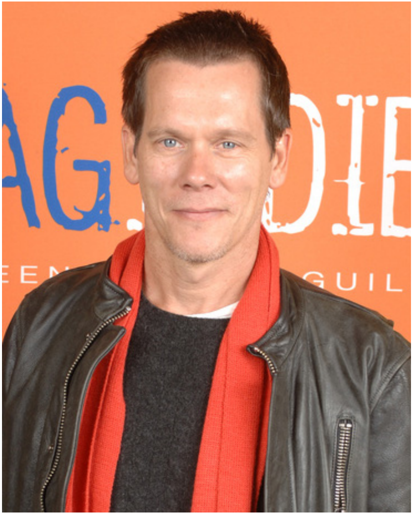


Image source:
https://upload.wikimedia.org/wikipedia/commons/d/d2/Kevin_Bacon.jpg

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.

Acting chain generated on The Oracle of Bacon.

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.
- Name another actor that acted in the same movie as the first.

Acting chain generated on The Oracle of Bacon.

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.
- Name another actor that acted in the same movie as the first.
- Name another actor that acted in the same movie as the second.

Acting chain generated on The Oracle of Bacon.

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.
- Name another actor that acted in the same movie as the first.
- Name another actor that acted in the same movie as the second.
- Continue on until you have reached Kevin Bacon as quickly as possible.

Acting chain generated on The Oracle of Bacon.

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.
- Name another actor that acted in the same movie as the first.
- Name another actor that acted in the same movie as the second.
- Continue on until you have reached Kevin Bacon as quickly as possible.
- This should not take more than 6 steps.

Acting chain generated on The Oracle of Bacon.

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.
- Name another actor that acted in the same movie as the first.
- Name another actor that acted in the same movie as the second.
- Continue on until you have reached Kevin Bacon as quickly as possible.
- This should not take more than 6 steps.

E.g.:
Millie Bobby Brown

Acting chain generated on The Oracle of Bacon.

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.
- Name another actor that acted in the same movie as the first.
- Name another actor that acted in the same movie as the second.
- Continue on until you have reached Kevin Bacon as quickly as possible.
- This should not take more than 6 steps.

E.g.:
Millie Bobby Brown was in *Godzilla: King of the Monsters* with

Acting chain generated on The Oracle of Bacon.

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.
- Name another actor that acted in the same movie as the first.
- Name another actor that acted in the same movie as the second.
- Continue on until you have reached Kevin Bacon as quickly as possible.
- This should not take more than 6 steps.

E.g.:
Millie Bobby Brown was in *Godzilla: King of the Monsters* with Charles Dance, who was in *White Mischief* with

Acting chain generated on The Oracle of Bacon.

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.
- Name another actor that acted in the same movie as the first.
- Name another actor that acted in the same movie as the second.
- Continue on until you have reached Kevin Bacon as quickly as possible.
- This should not take more than 6 steps.

E.g.:
Millie Bobby Brown was in *Godzilla: King of the Monsters* with
Charles Dance, who was in *White Mischief* with
John Hurt, who was in *Jayne Mansfield's Car* with

Acting chain generated on The Oracle of Bacon.

Six Degrees of Kevin Bacon

Game rules:

- Start with an actor.
- Name another actor that acted in the same movie as the first.
- Name another actor that acted in the same movie as the second.
- Continue on until you have reached Kevin Bacon as quickly as possible.
- This should not take more than 6 steps.

E.g.:
Millie Bobby Brown was in *Godzilla: King of the Monsters* with
Charles Dance, who was in *White Mischief* with
John Hurt, who was in *Jayne Mansfield's Car* with
Kevin Bacon.

Acting chain generated on The Oracle of Bacon.

~~Six~~ Less Than Four Degrees of Kevin Bacon

Surprising fact:

Although the number of actors listed on IMDB is very high relative to the number of actors any particular actor has acted with, the average distance between actors is 3.65!

E.g., the distance from Millie Bobby Brown to Kevin Bacon is 3.

Such a structure is called a small-world graph or small-world network.

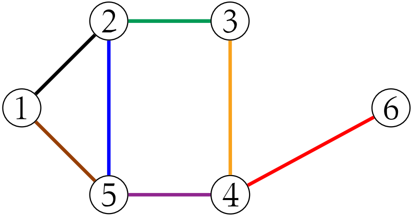
Watts & Strogatz 1998

Graphs / Networks

A graph (also called network) is set of vertices (also called nodes or points) and a set of edges (also called arcs, links, or lines) that connect two vertices.

The degree of a vertex is the number of edges that touch it.

Example 2:

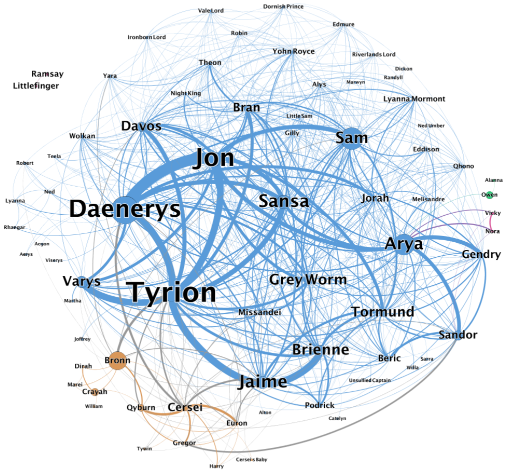


Example 1:

Vertices = actors on IMDB

Edges = when two actors have acted together

Network of Thrones

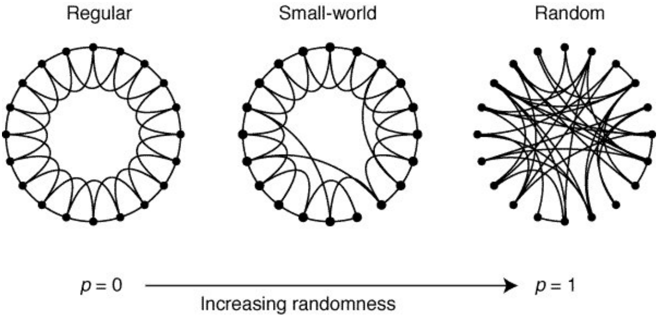


Vertices = GoT characters. Edges = character interactions.

Image source: `networkofthrones.wordpress.com`

Small-World Graphs

A small-world graph has “small“ average degree, “short” average distance between vertices, and “large” cliquishness.



Fixed average degree. Moving left-to-right from large average distance/large cliquishness to small average distance/large cliquishness to small average distance/small cliquishness

Image source: Watts & Strogatz 1998

SMALL-WORLD GRAPHS

A small-world graph has “small“ average degree, “short” average distance between vertices, and “large” cliquishness.

Watts & Strogatz 1998; Backstrom, Boldi, Rosa, Ugander, Vigna 2011

Small-World Graphs

A small-world graph has “small“ average degree, “short” average distance between vertices, and “large” cliquishness.

Examples include:

IMDB: vertices = actors, edges if acted together
(ave. dist. = 3.65)

C. elegans: vertices = neurons in roundworm, edges = synapses
(ave. dist. = 2.65)

Facebook: vertices = users, edges if FB friends
(ave. dist. = 4.74)

US power grid: vertices = generators, transformers and substations, edges if high-voltage transmission lines between them
(ave. dist. = 18.7)

and many more …

Watts & Strogatz 1998; Backstrom, Boldi, Rosa, Ugander, Vigna 2011

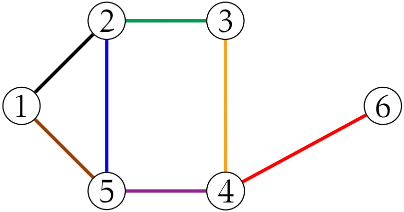
Why do we care?

(Mis-)information, disease, etc.

spread more quickly in small-world graphs.

One way to study graphs and their properties is coding them as a matrix and using linear algebra tools.

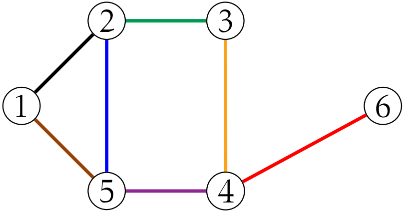
One such matrix is the adjacency matrix.



$$
\begin{array}{c|c}
 & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\
\begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} &
\begin{pmatrix}
0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\end{array}
$$

One way to study graphs and their properties is coding them as a matrix and using linear algebra tools.

One such matrix is the adjacency matrix.

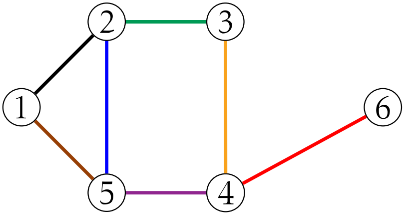


$$
\begin{array}{c c}
 & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} &
\begin{pmatrix}
0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\end{array}
$$

The adjacency matrix $\mathbf{A}$ of an (undirected, unweighted, simple) graph has

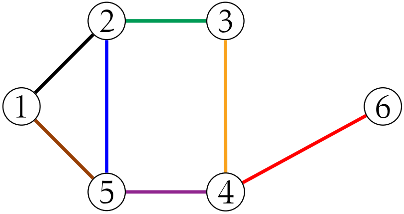
$A_{i,j} = 1$ when there is an edge between vertex i and vertex j

$A_{i,j} = 0$ otherwise.



$$
\begin{array}{c|cccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
1 & 0 & 1 & 0 & 0 & 1 & 0 \\
2 & 1 & 0 & 1 & 0 & 1 & 0 \\
3 & 0 & 1 & 0 & 1 & 0 & 0 \\
4 & 0 & 0 & 1 & 0 & 1 & 1 \\
5 & 1 & 1 & 0 & 1 & 0 & 0 \\
6 & 0 & 0 & 0 & 1 & 0 & 0
\end{array}
$$

$$\mathbf{A}\begin{pmatrix}\mathbf{1}\\\mathbf{1}\\\vdots\\\mathbf{1}\end{pmatrix}=\begin{pmatrix}\mathbf{2}\\\mathbf{3}\\\mathbf{2}\\\mathbf{3}\\\mathbf{3}\\\mathbf{1}\end{pmatrix}.$$



$$
\begin{array}{c c}
 & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} &
\begin{pmatrix}
0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\end{array}
$$

$$
\begin{pmatrix}
0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
=
\begin{pmatrix}
2 & 1 & 1 & 1 & 1 & 0 \\
1 & 3 & 0 & 2 & 1 & 0 \\
1 & 0 & 2 & 0 & 2 & 1 \\
1 & 2 & 0 & 3 & 0 & 0 \\
1 & 1 & 2 & 0 & 3 & 1 \\
0 & 0 & 1 & 0 & 1 & 1
\end{pmatrix}.
$$

If $\mathbf{A}$ is an adjacency matrix of a graph, the $\boldsymbol{i}$th entry of

$$\mathbf{A}\begin{pmatrix}\mathbf{1}\\ \mathbf{1}\\ \vdots\\ \mathbf{1}\end{pmatrix}$$

is the number of neighbors of the $\boldsymbol{i}$th vertex.

If $\mathbf{A}$ is an adjacency matrix of a graph, the ith entry of

$$\mathbf{A}\begin{pmatrix}\mathbf{1}\\ \mathbf{1}\\ \vdots\\ \mathbf{1}\end{pmatrix}$$

is the number of neighbors of the ith vertex.

Also, the (i, j) entry of $\mathbf{A}^k$ counts the number of walks of k steps from vertex i to vertex j.

(Go to Matlab)

More cool uses of linear algebra on graphs requires eigenvalues, coming in the next lecture.

DIRECTED GRAPHS

Graphs are called directed if edges have a start and end and undirected otherwise.

In networks like Instagram and TikTok, individual users are vertices.

If user `@mathdeptatcsu` decides to follow user `@csucamtheram`,

then there is a directed edge from `@mathdeptatcsu` to `@csucamtheram`.

However, `@csucamtheram` is not required to follow `@mathdeptatcsu`;

so, that edge does not go both ways.

Directed graphs

Graphs are called directed if edges have a start and end and undirected otherwise.

In networks like Instagram and TikTok, individual users are vertices.

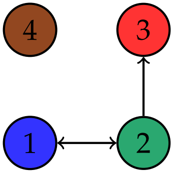
If user `@mathdeptatcsu` decides to follow user `@csucamtheram`,

then there is a directed edge from `@mathdeptatcsu` to `@csucamtheram`.

However, `@csucamtheram` is not required to follow `@mathdeptatcsu`;

so, that edge does not go both ways.

Example of directed graph:



And its adjacency matrix:

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

Weighted graphs

If edges have weights that describe the strength of the relationship between two vertices, it is called weighted.

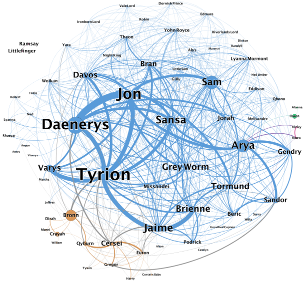


Image source: `networkofthrones.wordpress.com`

The adjacency matrix **A** of a (undirected) weighted graph has

$A_{i,j} = w_{i,j}$ when there is an edge between vertex i and vertex j with weight $w_{i,j}$

$A_{i,j} = 0$ otherwise.