# Linear Algebra and Graphs <br> IGERT Data and Network Science Bootcamp 

Victor Amelkin 〈victor@cs.ucsb.edu〉

UC Santa Barbara

September 11, 2015

## Table of Contents

Linear Algebra: Review of Fundamentals
Matrix Arithmetic
Inversion and Linear Systems Vector Spaces
Geometry
Eigenproblem

Linear Algebra and Graphs
Graphs: Definitions, Properties, Representation Spectral Graph Theory

## Materials

Copy of These Slides
http://cs.ucsb.edu/~victor/pub/ucsb/igert-2015/slides.pdf

MATLAB Source Code for Examples
http://cs.ucsb.edu/~victor/pub/ucsb/igert-2015/examples.tar.gz

## Vectors and Matrices

Notation

- $x \in \mathbb{C}$ - scalar (often, $\mathbb{R}$ is good enough)
- $\boldsymbol{v} \in \mathbb{C}^{n}$ - $n$-dimensional (column-)vector
- $\boldsymbol{A} \in \mathbb{C}^{n \times m}$ - matrix with $n$ rows and $m$ columns
- $\boldsymbol{A}_{i j}=\boldsymbol{A}_{i, j}$ - element of $\boldsymbol{A}$ in $i$ 'th row and $j$ 'th column
- $\boldsymbol{A}^{\top}$ - transpose of $\boldsymbol{A}\left(\boldsymbol{A}_{i j}^{\top}=\boldsymbol{A}_{j i}\right)$
- $\boldsymbol{A}^{\mathrm{H}}$ - Hermitian transpose of $\boldsymbol{A}\left(\boldsymbol{A}_{i j}^{\mathrm{H}}=\overline{\boldsymbol{A}}_{j i}\right)^{1}$

Examples

$$
\boldsymbol{A}=\left[\begin{array}{lll}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}\right] \in \mathbb{R}^{2 \times 3} \quad \boldsymbol{A}^{\top}=\left[\begin{array}{ll}
1 & 4 \\
2 & 5 \\
3 & 6
\end{array}\right] \in \mathbb{R}^{3 \times 2}
$$

[^0]
## Block Matrices

$$
A=\begin{gathered}
\\
n_{1} \\
\vdots \\
n_{q}
\end{gathered}\left[\begin{array}{ccc}
m_{1} & \ldots & m_{p} \\
A_{11} & \ldots & A_{1 p} \\
\vdots & \ddots & \vdots \\
A_{q 1} & \ldots & A_{q p}
\end{array}\right]
$$

Definition
Block matrix - a "matrix of matrices". $\boldsymbol{A}_{i j}$ - block at $i$ 'th row and $j$ 'th column of partitioned matrix $\boldsymbol{A}$. Blocks of the same row (column) have the same number of rows (columns).

## Matrix Arithmetic (1)

Multiplication by scalar
Performed elementwise: $\alpha \cdot \boldsymbol{A}_{n \times m}=\left\{\alpha \cdot A_{i j}\right\}_{n \times m}$
Addition
Performed elementwise: $\boldsymbol{A}_{n \times m}+\boldsymbol{B}_{n \times m}=\left\{A_{i j}+B_{i j}\right\}_{n \times m}$
Examples

$$
2 \cdot\left[\begin{array}{lll}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}\right]+\left[\begin{array}{ccc}
-2 & 0 & -1 \\
2 & 5 & 8
\end{array}\right]=\left[\begin{array}{ccc}
0 & 4 & 5 \\
10 & 15 & 20
\end{array}\right]
$$

## Matrix Arithmetic (2)

Multiplication

$$
\begin{aligned}
\boldsymbol{A}_{n \times m} \boldsymbol{B}_{m \times k} & =\boldsymbol{C}_{n \times k} \\
{[]_{1 \times 0}[]_{0 \times 1} } & =[0]_{1 \times 1} \\
{[a]_{1 \times 1}[b]_{1 \times 1} } & =[a \cdot b]_{1 \times 1}(a, b \in \mathbb{C}) \\
{\left[\begin{array}{ll}
\boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\
\boldsymbol{A}_{21} & \boldsymbol{A}_{22}
\end{array}\right]\left[\begin{array}{ll}
\boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\
\boldsymbol{B}_{21} & \boldsymbol{B}_{22}
\end{array}\right] } & =\left[\begin{array}{ll}
\boldsymbol{A}_{11} \boldsymbol{B}_{11}+\boldsymbol{A}_{12} \boldsymbol{B}_{21} & \boldsymbol{A}_{11} \boldsymbol{B}_{12}+\boldsymbol{A}_{12} \boldsymbol{B}_{22} \\
\boldsymbol{A}_{21} \boldsymbol{B}_{11}+\boldsymbol{A}_{22} \boldsymbol{B}_{21} & \boldsymbol{A}_{21} \boldsymbol{B}_{12}+\boldsymbol{A}_{22} \boldsymbol{B}_{22}
\end{array}\right]
\end{aligned}
$$

Corollary

$$
\boldsymbol{A}_{n \times m} \boldsymbol{B}_{m \times k}=\left\{\sum_{\ell=1}^{m} A_{i \ell} B_{\ell j}\right\}_{n \times k}
$$

## Matrix Arithmetic (2) - Examples

- Example 1 (matrix-matrix (MM) multiplication):

$$
\left[\begin{array}{lll}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}\right]\left[\begin{array}{cc}
0 & 1 \\
-1 & 3 \\
5 & 2
\end{array}\right]=\left[\begin{array}{ll}
13 & 13 \\
25 & 31
\end{array}\right]
$$

- Example 2 (matrix-vector (MV) multiplication):

$$
\left[\begin{array}{lll}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}\right]\left[\begin{array}{l}
1 \\
3 \\
2
\end{array}\right]=\left[\begin{array}{l}
13 \\
31
\end{array}\right]
$$

- Example 3 (MV block multiplication):

$$
\left[\begin{array}{l|l|l}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}\right]_{1 \times 3}\left[\frac{1}{\frac{1}{2}}\right]_{3 \times 1}=\left[\left[\begin{array}{l}
1 \\
4
\end{array}\right] \cdot[1]+\left[\begin{array}{l}
2 \\
5
\end{array}\right] \cdot[3]+\left[\begin{array}{l}
3 \\
6
\end{array}\right] \cdot[2]\right]_{1 \times 1}
$$

## Matrix Arithmetic (2) - More Examples

- Example 4 (MV block multiplication):

$$
\begin{aligned}
& {\left[\begin{array}{lll}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}\right]_{1 \times 1}\left[\begin{array}{c|c}
0 & 1 \\
-1 & 3 \\
5 & 2
\end{array}\right]_{1 \times 2}=} \\
& \left.\left[\begin{array}{lll}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}\right]\left[\begin{array}{c}
0 \\
-1 \\
5
\end{array}\right]\left[\begin{array}{lll}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}\right]\left[\begin{array}{l}
1 \\
3 \\
2
\end{array}\right]\right] \left._{1 \times 2}=\left[\begin{array}{l}
13 \\
25
\end{array}\right] \right\rvert\,\left[\begin{array}{l}
13 \\
31
\end{array}\right]
\end{aligned}
$$

- Example 5 (row scaling ${ }^{2}$ ):

$$
\left[\begin{array}{ccc}
d_{1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & d_{n}
\end{array}\right]_{n \times n}\left[\begin{array}{c}
-\boldsymbol{A}_{1}- \\
\vdots \\
-\boldsymbol{A}_{n}-
\end{array}\right]_{n \times n}=\left[\begin{array}{c}
-d_{1} \boldsymbol{A}_{1}- \\
\vdots \\
-d_{n} \boldsymbol{A}_{n}-
\end{array}\right]_{n \times n}
$$

- Example 6 (permutation of rows ${ }^{3}$ ):

$$
\begin{aligned}
& 3 \\
& 1 \\
& 2
\end{aligned}\left[\begin{array}{lll}
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{array}\right]\left[\begin{array}{c}
-\boldsymbol{A}_{1}- \\
-\boldsymbol{A}_{2}- \\
-\boldsymbol{A}_{3}-
\end{array}\right]_{3 \times m}=\left[\begin{array}{c}
-\boldsymbol{A}_{3}- \\
-\boldsymbol{A}_{1}- \\
-\boldsymbol{A}_{2}-
\end{array}\right]_{3 \times m}
$$

[^1]Left Inverse
$\boldsymbol{A}^{-L}$ is a left inverse of $\boldsymbol{A}$ if $\boldsymbol{A}^{-L} \boldsymbol{A}=\boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix ( $\boldsymbol{I}_{i i}=1, \boldsymbol{I}_{i j}=0$ if $i \neq j$ ).

Right Inverse
$\boldsymbol{A}^{-R}$ is a right inverse of $\boldsymbol{A}$ if $\boldsymbol{A} \boldsymbol{A}^{-R}=I$.
Inverse
$\boldsymbol{A}^{-1}$ is the inverse of $\boldsymbol{A}$ if $\boldsymbol{A}^{-1}=\boldsymbol{A}^{-L}=\boldsymbol{A}^{-R}$. If inverse exists, it is unique; if it does not, then the (Moore-Penrose) pseudoinverse is the closest substitute.

## Inversion - Examples (matrix_inversion.m)

- Left Inverse (no right inverse for skinny matrices):

$$
\left[\begin{array}{ll}
1 & 4 \\
2 & 5 \\
3 & 6
\end{array}\right]^{-L} \approx\left[\begin{array}{ccc}
-0.94 & -0.11 & 0.72 \\
0.44 & 0.11 & -0.22
\end{array}\right]
$$

- Right Inverse (no left inverse for fat matrices):

$$
\left[\begin{array}{lll}
1 & 2 & 3 \\
4 & 5 & 6
\end{array}\right]^{-R} \approx\left[\begin{array}{cc}
-0.94 & 0.44 \\
-0.11 & 0.11 \\
0.72 & -0.22
\end{array}\right]
$$

- Inverse (may exist only for square matrices):

$$
\left[\begin{array}{lll}
1 & 2 & 3 \\
2 & 3 & 1 \\
3 & 1 & 2
\end{array}\right]^{-1} \approx\left[\begin{array}{ccc}
-0.28 & 0.06 & 0.39 \\
0.06 & 0.39 & -0.28 \\
0.39 & -0.28 & 0.06
\end{array}\right]
$$

## Linear Systems

- Here, "a linear system" = "a system of linear algebraic equations".
- Solving $\boldsymbol{A} \boldsymbol{x}=\boldsymbol{b}$ w.r.t. $\boldsymbol{x} \in \mathbb{R}^{n}$ is one of two fundamental problems of linear algebra (the other one is eigenproblem).
- Unique solution exists iff $A$ is non-singular $(\operatorname{det}(A) \neq 0)$.
- Problem is related to matrix inversion (i.e., $x=\boldsymbol{A}^{-1} \boldsymbol{b}$ ).
- A system with singular $\boldsymbol{A}$ either has no or infinitely many solutions.


## LU Factorization

- The method to directly solve linear systems - (a kind of) Gaussian elimination.
- Bad ideas: Cramer's rule; inversion followed by multiplication.
- One kind of Gaussian elimination - LU factorization / decomposition (a.k.a. Gaussian elimination with partial pivoting).


## Theorem (LU factorization)

For any $n$-by-m matrix $\boldsymbol{A}$, there exist a permutation matrix $\boldsymbol{P}$ such that $\boldsymbol{P} \boldsymbol{A}=\boldsymbol{L} \boldsymbol{U}$, where $\boldsymbol{L}$ is lower-triangular with units on the main diagonal and $\boldsymbol{U}$ is a block-matrix of the form

$$
U=\underset{(n-r)}{r}\left[\begin{array}{cc}
r & (m-r) \\
\boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\
\mathbf{0} & \mathbf{0}
\end{array}\right]
$$

where $U_{11}$ is upper-triangular with non-zero diagonal entries. Integer $r$ is the rank of $\boldsymbol{A}$.

## Solving Linear Systems with LU (linear_systems.m)

- Problem: solve $\boldsymbol{A} \boldsymbol{x}=\boldsymbol{b}$ w.r.t. $\boldsymbol{x} \in \mathbb{R}^{n \times 1}$.
- Step 1: decompose $\boldsymbol{A}=\boldsymbol{P}^{-1} \boldsymbol{L} \boldsymbol{U}\left(O\left(n^{3}\right)\right)$.
- Step 2: $\left(\boldsymbol{P}^{-1} \boldsymbol{L} \boldsymbol{U}\right) \boldsymbol{x}=\boldsymbol{b} \Longleftrightarrow(\boldsymbol{L} \boldsymbol{U}) \boldsymbol{x}=\boldsymbol{b}^{\prime}$, where $\boldsymbol{b}^{\prime}=\boldsymbol{P b}(\mathcal{b}(n))$.
- Step 3: solve $\boldsymbol{L} \boldsymbol{y}=\boldsymbol{b}^{\prime}$ w.r.t. $\boldsymbol{y}$ using forward substitution $\left(\mathcal{O}\left(n^{2}\right)\right)$.
- Step 4: solve $\boldsymbol{U} \boldsymbol{x}=\boldsymbol{y}$ w.r.t. $\boldsymbol{x}$ using back substitution $\left(\mathcal{O}\left(n^{2}\right)\right)$.


## Vector Space

A vector space consists of a set (a field) of scalars $\mathbb{F}$ (often, $\mathbb{C}$ or $\mathbb{R}$ ), a set of vectors $\mathcal{V}$ (sequences, matrices, functions, ...) and a pair of operations, vector addition $+: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$ and scalar multiplication $\times: \mathbb{F} \times \mathcal{V} \rightarrow \mathcal{V}$, such that $\forall \alpha, \beta \in \mathbb{F} \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{V}:$

- $\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$ (commutativity of addition),
- $\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})=(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$ (associativity of addition),
- $\exists \mathbf{0} \in \mathcal{V}: \boldsymbol{x}+\mathbf{0}=\boldsymbol{x}$ (existence of additive identity),
- $\exists-\boldsymbol{x} \in \mathcal{V}: \boldsymbol{x}+(-\boldsymbol{x})=\mathbf{0}$ (existence of additive inverse),
- $\alpha(\beta \boldsymbol{x})=(\alpha \beta) \boldsymbol{x}$ (multiplicative associativity),
- $1 x=x$ (unit scaling),
- $\alpha(\boldsymbol{x}+\boldsymbol{y})=\alpha \boldsymbol{x}+\alpha \boldsymbol{y}$ (distributivity),
- $(\alpha+\beta) \boldsymbol{x}=\alpha \boldsymbol{x}+\beta \boldsymbol{x}$ (distributivity).


## Vector Space

A vector space consists of a set (a field) of scalars $\mathbb{F}$ (often, $\mathbb{C}$ or $\mathbb{R}$ ), a set of vectors $\mathcal{V}$ (sequences, matrices, functions, ...) and a pair of operations, vector addition $+: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$ and scalar multiplication $\times: \mathbb{F} \times \mathcal{V} \rightarrow \mathcal{V}$, such that $\forall \alpha, \beta \in \mathbb{F} \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{V}:$

- $\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$ (commutativity of addition),
- $\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})=(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$ (associativity of addition),
- $\exists \mathbf{0} \in \mathcal{V}: \boldsymbol{x}+\mathbf{0}=\boldsymbol{x}$ (existence of additive identity),
- $\exists-\boldsymbol{x} \in \mathcal{V}: \boldsymbol{x}+(-\boldsymbol{x})=\mathbf{0}$ (existence of additive inverse),
- $\alpha(\beta \boldsymbol{x})=(\alpha \beta) \boldsymbol{x}$ (multiplicative associativity),
- $1 x=x$ (unit scaling),
- $\alpha(\boldsymbol{x}+\boldsymbol{y})=\alpha \boldsymbol{x}+\alpha \boldsymbol{y}$ (distributivity),
- $(\alpha+\beta) \boldsymbol{x}=\alpha \boldsymbol{x}+\beta \boldsymbol{x}$ (distributivity).
(Notation abuse: instead of $\langle\mathcal{V}, \mathbb{F},+, \times\rangle$, we usually refer to a vector space simply as $\mathcal{V}$ or $\mathcal{V}$ over $\langle\bullet, \bullet\rangle$ when we want to emphasize how the operations + and $\times$ are defined.)


## Vector Space

A vector space consists of a set (a field) of scalars $\mathbb{F}$ (often, $\mathbb{C}$ or $\mathbb{R}$ ), a set of vectors $\mathcal{V}$ (sequences, matrices, functions, ...) and a pair of operations, vector addition $+: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$ and scalar multiplication $\times: \mathbb{F} \times \mathcal{V} \rightarrow \mathcal{V}$, such that $\forall \alpha, \beta \in \mathbb{F} \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathcal{V}:$

- $\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$ (commutativity of addition),
- $\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})=(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}$ (associativity of addition),
- $\exists \mathbf{0} \in \mathcal{V}: \boldsymbol{x}+\mathbf{0}=\boldsymbol{x}$ (existence of additive identity),
- $\exists-\boldsymbol{x} \in \mathcal{V}: \boldsymbol{x}+(-\boldsymbol{x})=\mathbf{0}$ (existence of additive inverse),
- $\alpha(\beta \boldsymbol{x})=(\alpha \beta) \boldsymbol{x}$ (multiplicative associativity),
- $1 x=x$ (unit scaling),
- $\alpha(\boldsymbol{x}+\boldsymbol{y})=\alpha \boldsymbol{x}+\alpha \boldsymbol{y}$ (distributivity),
- $(\alpha+\beta) \boldsymbol{x}=\alpha \boldsymbol{x}+\beta \boldsymbol{x}$ (distributivity).
(Notation abuse: instead of $\langle\mathcal{V}, \mathbb{F},+, \times\rangle$, we usually refer to a vector space simply as $\mathcal{V}$ or $\mathcal{V}$ over $\langle\bullet, \bullet\rangle$ when we want to emphasize how the operations + and $\times$ are defined.)

A subset $\mathcal{W}$ of $\mathcal{V}$ is a subspace of $\mathcal{V}$ if $\mathcal{W}$ is a vector space on its own. Alternatively, $\mathcal{W}$ is a subspace iff it is closed under $\langle+, \times\rangle$.

## Linear Independence

- Linear combination: $a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}$ (where $a_{i} \in \mathbb{F}, \boldsymbol{x}_{j} \in \mathcal{V}$ ) - linear combination of vectors $\left\{\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right\}$ with coefficients $\left\{a_{1}, \ldots, a_{n}\right\}$.


## Linear Independence

- Linear combination: $a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}$ (where $a_{i} \in \mathbb{F}, \boldsymbol{x}_{j} \in \mathcal{V}$ ) - linear combination of vectors $\left\{\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right\}$ with coefficients $\left\{a_{1}, \ldots, a_{n}\right\}$.
- Important observation:

$$
a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}=\left[\boldsymbol{x}_{1}|\ldots| \boldsymbol{x}_{n}\right]\left[\begin{array}{c}
a_{1} \\
\vdots \\
a_{n}
\end{array}\right]
$$

If we think about $\boldsymbol{x}_{i}$ not as abstract elements of $\mathcal{V}$, but as (column-)vectors, it becomes clear that the result of a matrix-vector multiplication is a linear combination of the matrix' columns.

## Linear Independence

- Linear combination: $a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}$ (where $a_{i} \in \mathbb{F}, \boldsymbol{x}_{j} \in \mathcal{V}$ ) - linear combination of vectors $\left\{\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right\}$ with coefficients $\left\{a_{1}, \ldots, a_{n}\right\}$.
- Important observation:

$$
a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}=\left[\boldsymbol{x}_{1}|\ldots| \boldsymbol{x}_{n}\right]\left[\begin{array}{c}
a_{1} \\
\vdots \\
a_{n}
\end{array}\right]
$$

If we think about $\boldsymbol{x}_{i}$ not as abstract elements of $\mathcal{V}$, but as (column-)vectors, it becomes clear that the result of a matrix-vector multiplication is a linear combination of the matrix' columns.
$-\operatorname{span}\left(\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right)=\left\{a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n} \mid a_{i} \in \mathbb{F}\right\}-$ a span of a set of vectors is the set of all their possible linear combinations.

## Linear Independence

- Linear combination: $a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}$ (where $a_{i} \in \mathbb{F}, \boldsymbol{x}_{j} \in \mathcal{V}$ ) - linear combination of vectors $\left\{\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right\}$ with coefficients $\left\{a_{1}, \ldots, a_{n}\right\}$.
- Important observation:

$$
a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}=\left[\boldsymbol{x}_{1}|\ldots| \boldsymbol{x}_{n}\right]\left[\begin{array}{c}
a_{1} \\
\vdots \\
a_{n}
\end{array}\right]
$$

If we think about $\boldsymbol{x}_{i}$ not as abstract elements of $\mathcal{V}$, but as (column-)vectors, it becomes clear that the result of a matrix-vector multiplication is a linear combination of the matrix' columns.
$-\operatorname{span}\left(\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right)=\left\{a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n} \mid a_{i} \in \mathbb{F}\right\}-$ a span of a set of vectors is the set of all their possible linear combinations.

- $\left\{\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right\}$ are linearly independent if $a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}=0$ iff $\left[\begin{array}{lll}a_{1} & \ldots & a_{n}\end{array}\right]=\mathbf{0}$.


## Linear Independence

- Linear combination: $a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}$ (where $a_{i} \in \mathbb{F}, \boldsymbol{x}_{j} \in \mathcal{V}$ ) - linear combination of vectors $\left\{\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right\}$ with coefficients $\left\{a_{1}, \ldots, a_{n}\right\}$.
- Important observation:

$$
a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}=\left[\boldsymbol{x}_{1}|\ldots| \boldsymbol{x}_{n}\right]\left[\begin{array}{c}
a_{1} \\
\vdots \\
a_{n}
\end{array}\right]
$$

If we think about $\boldsymbol{x}_{i}$ not as abstract elements of $\mathcal{V}$, but as (column-)vectors, it becomes clear that the result of a matrix-vector multiplication is a linear combination of the matrix' columns.
$-\operatorname{span}\left(\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right)=\left\{a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n} \mid a_{i} \in \mathbb{F}\right\}-$ a span of a set of vectors is the set of all their possible linear combinations.

- $\left\{\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right\}$ are linearly independent if $a_{1} \boldsymbol{x}_{1}+\cdots+a_{n} \boldsymbol{x}_{n}=0$ iff $\left[\begin{array}{lll}a_{1} & \ldots & a_{n}\end{array}\right]=\mathbf{0}$.
- A set $\left\{\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right\}$ of linearly independent vectors is a basis of subspace $\mathcal{W}$ if $\mathcal{W}=\operatorname{span}\left(\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{n}\right)$. $\mathcal{W}$ 's dimension $\operatorname{dim} \mathcal{W}$ is $n$.


## Fundamental Subspaces of a Matrix

- The nullspace (kernel) of $\boldsymbol{A}_{n \times m}$ is

$$
\mathcal{N}(\boldsymbol{A})=\{\boldsymbol{x} \mid \boldsymbol{A x}=\mathbf{0}\} .
$$

- The range (column space, image) of $\boldsymbol{A}_{n \times m}$ is

$$
\mathcal{R}(\boldsymbol{A})=\operatorname{colspan}(\boldsymbol{A})=\{\boldsymbol{y} \mid \boldsymbol{y}=\boldsymbol{A} \boldsymbol{x}\}
$$

- $\mathcal{N}\left(\boldsymbol{A}^{\mathrm{H}}\right)$ - left nullspace of $\boldsymbol{A}$.
- $\mathcal{R}\left(\boldsymbol{A}^{\mathrm{H}}\right)$ - row space of $\boldsymbol{A}$.


## Fundamental Subspaces of a Matrix - "The Big Picture"



## Norm

- Lengths in a vector space are measured using a norm $\|\cdot\|$. A vector space augmented with a norm is a normed (vector) space.
- A norm defined on a vector space $\mathcal{V}$ over field $\mathbb{F}$ is a mapping $\|\cdot\|: \mathcal{V} \rightarrow \mathbb{R}$, such that $\forall \alpha \in \mathbb{F} \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$ the following norm axioms hold
- $\boldsymbol{x} \in \mathcal{V}:\|x\| \geq 0$ (non-negativity ${ }^{4}$ ),
- $\|\boldsymbol{x}\|=0 \rightarrow \boldsymbol{x}=\mathbf{0}$ (positive definiteness),
- $\|\alpha \boldsymbol{x}\|=|\alpha|\|\boldsymbol{x}\|$ (homogeneity),
- $\|\boldsymbol{x}+\boldsymbol{y}\| \leq\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$ (subadditivity / triangle inequality).
- Examples:
- $\ell_{p}$-norm: $\|\boldsymbol{x}\|_{p}=\left(\sum_{i}\left|x_{i}\right|^{p}\right)^{1 / p}$,
- $L_{p}$-norm: $\|\boldsymbol{f}\|_{p}=\left(\int_{D}|\boldsymbol{f}|^{p} \mathrm{~d} \mu\right)^{1 / p}<\infty$.

[^2]
## Convexity and Norm

- A unit sphere is defined as $\{\boldsymbol{x} \mid\|\boldsymbol{x}\|=1\}$.
- A set $S \subseteq \mathcal{V}$ is convex if $\forall \boldsymbol{x}, \boldsymbol{y} \in S \forall 0 \leq \lambda \leq 1: \lambda \boldsymbol{x}+(1-\lambda) \boldsymbol{y} \in S$. In other words, for any two points of $S$, all the points on the line between $\boldsymbol{x}$ and $\boldsymbol{y}$ are also in $S$.

- A function $f: \mathcal{V} \rightarrow \mathbb{R}$ is convex if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V} \forall 0 \leq \lambda \leq 1$ : $f(\lambda \boldsymbol{x}+(1-\lambda) \boldsymbol{y}) \leq \lambda f(\boldsymbol{x})+(1-\lambda) f(\boldsymbol{y})$
- Norms are convex.


## Convexity and Norm - $\ell_{p}$ (norms_and_convexity.m)

- Function $f_{p}(\boldsymbol{x})=\left(\sum_{i}\left|x_{i}\right|^{p}\right)^{1 / p}$ mapping $\mathbb{R}^{n}$ to $\mathbb{R}$ is a norm iff $1 \leq p \leq \infty$.
- Alternatively, $f_{p}$ is a norm iff the unit sphere $\left\{\boldsymbol{x} \mid f_{p}(\boldsymbol{x})=1\right\}$ induced by $f_{p}$ is convex.



## Common Vector Norms

## Euclidean norm

$$
\|\boldsymbol{x}\|_{2}=\sqrt{x_{1}^{2}+\cdots+x_{n}^{2}}
$$

Taxicab / Manhattan norm

$$
\|\boldsymbol{x}\|_{1}=\left|x_{1}\right|+\cdots+\left|x_{n}\right|
$$

Chebyshev norm

$$
\|\boldsymbol{x}\|_{\infty}=\max \left\{\left|x_{1}\right|, \ldots,\left|x_{n}\right|\right\}
$$

## Norm Equivalence

## Norm Equivalence

Two norms $\|\cdot\|_{\alpha}$ and $\|\cdot\|_{\beta}$ are equivalent if there exist two positive constants $c_{1}, c_{2}<\infty$ such that $\forall \boldsymbol{x} \in \mathcal{V}$ :

$$
c_{1}\|\boldsymbol{x}\|_{\alpha} \leq\|\boldsymbol{x}\|_{\beta} \leq c_{2}\|\boldsymbol{x}\|_{\alpha} .
$$

Theorem
In finite-dimensional vector spaces, all norms are equivalent.
Examples for $\boldsymbol{x} \in \mathbb{C}^{n}$

$$
\begin{aligned}
& \|\boldsymbol{x}\|_{1} \leq \sqrt{n}\|\boldsymbol{x}\|_{2} \\
& \|\boldsymbol{x}\|_{1} \leq n\|\boldsymbol{x}\|_{\infty} \\
& \|\boldsymbol{x}\|_{2} \leq \sqrt{n}\|\boldsymbol{x}\|_{\infty} \\
& \quad \text { and, more generally, for } 0<p<q \\
& \|\boldsymbol{x}\|_{q} \leq\|\boldsymbol{x}\|_{p} \leq n^{(1 / p-1 / q)}\|\boldsymbol{x}\|_{q} .
\end{aligned}
$$

## Matrix Norms (operator_norm.m)

- Frobenius Norm: $\|\boldsymbol{A}\|_{F}=\sqrt{\sum_{i, j=1}^{n} A_{i j}^{2}}=\operatorname{trace}\left(\boldsymbol{A}^{\mathrm{H}} \boldsymbol{A}\right)$, where trace () of a matrix is the sum of the elements on its main diagonal.
- Operator Norm (Induced p-norm): $\|\boldsymbol{A}\|=\sup _{\boldsymbol{x} \neq 0} \frac{\|\boldsymbol{A} \boldsymbol{x}\|_{p}}{\|\boldsymbol{x}\|_{p}}=\sup _{\|\boldsymbol{x}\|_{p}=1}\|\boldsymbol{A} \boldsymbol{x}\|_{p}$ An operator norm measures the maximum degree of distortion / amount of stretch of a unit sphere under transformation by $\boldsymbol{A}$.

- For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{n \times n},\|\boldsymbol{A} \boldsymbol{B}\| \leq\|\boldsymbol{A}\|\|\boldsymbol{B}\|$ (submultiplicativity).
- Inner product (scalar product) of $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}:\langle\boldsymbol{x}, \boldsymbol{y}\rangle=\boldsymbol{x}^{\mathrm{H}} \boldsymbol{y}=\sum_{i} \bar{x}_{i} y_{i}$.
- Inner product (scalar product) of $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}:\langle\boldsymbol{x}, \boldsymbol{y}\rangle=\boldsymbol{x}^{\mathrm{H}} \boldsymbol{y}=\sum_{i} \bar{x}_{i} y_{i}$.
- Hölder's inequality: $\langle\boldsymbol{x}, \boldsymbol{y}\rangle \leq\|\boldsymbol{x}\|_{p}\|\boldsymbol{y}\|_{q}$ for $\frac{1}{p}+\frac{1}{q}=1$.
- Cauchy-Bunyakovsky-Schwarz (CBS) inequality: $\langle\boldsymbol{x}, \boldsymbol{y}\rangle \leq\|\boldsymbol{x}\|_{2}\|\boldsymbol{y}\|_{2}$
- Inner product (scalar product) of $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}:\langle\boldsymbol{x}, \boldsymbol{y}\rangle=\boldsymbol{x}^{\mathrm{H}} \boldsymbol{y}=\sum_{i} \bar{x}_{i} y_{i}$.
- Hölder's inequality: $\langle\boldsymbol{x}, \boldsymbol{y}\rangle \leq\|\boldsymbol{x}\|_{p}\|\boldsymbol{y}\|_{q}$ for $\frac{1}{p}+\frac{1}{q}=1$.
- Cauchy-Bunyakovsky-Schwarz (CBS) inequality: $\langle\boldsymbol{x}, \boldsymbol{y}\rangle \leq\|\boldsymbol{x}\|_{2}\|\boldsymbol{y}\|_{2}$
- CBS inequality inspires the following definition of an angle $\theta$ between vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n}$ :

$$
\cos \theta=\frac{\langle\boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}
$$

- Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal $(\boldsymbol{x} \perp \boldsymbol{y})$ if the cosine of the angle between them is 0 .
- Inner product (scalar product) of $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}:\langle\boldsymbol{x}, \boldsymbol{y}\rangle=\boldsymbol{x}^{\mathrm{H}} \boldsymbol{y}=\sum_{i} \bar{x}_{i} y_{i}$.
- Hölder's inequality: $\langle\boldsymbol{x}, \boldsymbol{y}\rangle \leq\|\boldsymbol{x}\|_{p}\|\boldsymbol{y}\|_{q}$ for $\frac{1}{p}+\frac{1}{q}=1$.
- Cauchy-Bunyakovsky-Schwarz (CBS) inequality: $\langle\boldsymbol{x}, \boldsymbol{y}\rangle \leq\|\boldsymbol{x}\|_{2}\|\boldsymbol{y}\|_{2}$
- CBS inequality inspires the following definition of an angle $\theta$ between vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n}$ :

$$
\cos \theta=\frac{\langle\boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}
$$

- Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal $(\boldsymbol{x} \perp \boldsymbol{y})$ if the cosine of the angle between them is 0 .
- The length of the orthogonal projection of $\boldsymbol{x}$ upon $\boldsymbol{y}$ is $\left\langle\boldsymbol{x}, \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}\right\rangle$


## Inner Product, Angle, Projection - Example

- Problem: Given a matrix $A \in \mathbb{R}^{n \times n}$, find the amount of stretch caused by $\boldsymbol{A}$ to vectors along a direction defined by a vector $\boldsymbol{x}$ in $\ell_{2}$.



## Inner Product, Angle, Projection - Example

- Problem: Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, find the amount of stretch caused by $\boldsymbol{A}$ to vectors along a direction defined by a vector $\boldsymbol{x}$ in $\ell_{2}$.

- Given an arbitrary $\boldsymbol{x}$, normalize it, so that its length is 1 :

$$
\overline{\boldsymbol{x}}=\boldsymbol{x} /\|\boldsymbol{x}\| \quad\left(\|\overline{\boldsymbol{x}}\|=\left\|\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right\|=\frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|}=1\right)
$$

## Inner Product, Angle, Projection - Example

- Problem: Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, find the amount of stretch caused by $\boldsymbol{A}$ to vectors along a direction defined by a vector $\boldsymbol{x}$ in $\ell_{2}$.

- Given an arbitrary $\boldsymbol{x}$, normalize it, so that its length is 1 :

$$
\overline{\boldsymbol{x}}=\boldsymbol{x} /\|\boldsymbol{x}\| \quad\left(\|\overline{\boldsymbol{x}}\|=\left\|\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right\|=\frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|}=1\right)
$$

- The amount of stretch caused by $\boldsymbol{A}$ along $\boldsymbol{x}$ :

$$
\frac{\text { new size }}{\text { original size }}=\frac{\left|\operatorname{proj}_{\overline{\boldsymbol{x}}} \boldsymbol{A} \overline{\boldsymbol{x}}\right|}{\|\overline{\boldsymbol{x}}\|}=\left|\operatorname{proj}_{\overline{\boldsymbol{x}}} \boldsymbol{A} \overline{\boldsymbol{x}}\right|=\langle\overline{\boldsymbol{x}}, \boldsymbol{A} \overline{\boldsymbol{x}}\rangle=\frac{\langle\boldsymbol{x}, \boldsymbol{A} \boldsymbol{x}\rangle}{\|\boldsymbol{x}\|^{2}}=\frac{\boldsymbol{x}^{\top} \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^{\top} \boldsymbol{x}}
$$

## Inner Product, Angle, Projection - Example

- Problem: Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, find the amount of stretch caused by $\boldsymbol{A}$ to vectors along a direction defined by a vector $\boldsymbol{x}$ in $\ell_{2}$.

- Given an arbitrary $\boldsymbol{x}$, normalize it, so that its length is 1 :

$$
\overline{\boldsymbol{x}}=\boldsymbol{x} /\|\boldsymbol{x}\| \quad\left(\|\overline{\boldsymbol{x}}\|=\left\|\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right\|=\frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|}=1\right)
$$

- The amount of stretch caused by $\boldsymbol{A}$ along $\boldsymbol{x}$ :

$$
\frac{\text { new size }}{\text { original size }}=\frac{\left|\operatorname{proj}_{\overline{\boldsymbol{x}}} \boldsymbol{A} \overline{\boldsymbol{x}}\right|}{\|\overline{\boldsymbol{x}}\|}=\left|\operatorname{proj}_{\overline{\boldsymbol{x}}} \boldsymbol{A} \overline{\boldsymbol{x}}\right|=\langle\overline{\boldsymbol{x}}, \boldsymbol{A} \overline{\boldsymbol{x}}\rangle=\frac{\langle\boldsymbol{x}, \boldsymbol{A} \boldsymbol{x}\rangle}{\|\boldsymbol{x}\|^{2}}=\frac{\boldsymbol{x}^{\top} \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^{\top} \boldsymbol{x}}
$$

- Derived Rayleigh quotient (generally, $x^{H}$ is used instead of $x^{\top}$ ).


## Linear Systems (Revisited)

- If the columns of matrix $\boldsymbol{A}$ of a linear system $\boldsymbol{A} \boldsymbol{x}=\boldsymbol{b}$ span the entire space, then $\boldsymbol{b}$ can be uniquely "explained" in terms of these columns ( $\boldsymbol{b}$ has a unique representation in this basis).
- If the columns of $\boldsymbol{A}$ span a subspace, then either $\boldsymbol{b}$ has infinitely many representations (if it belongs to the column (sub)space) or it has no precise representation in terms of $\boldsymbol{A}$ 's columns.
- Even if $\boldsymbol{b}$ is out of $\boldsymbol{A}$ 's range, we can replace $\boldsymbol{b}$ by the next best thing - its projection upon the column (sub)space.
- $\boldsymbol{U} \boldsymbol{U}^{\dagger}$ is a projector upon subspace spanned by $\boldsymbol{U}$ 's columns, where $\boldsymbol{U}^{\dagger}=\left(\boldsymbol{U}^{\mathrm{H}} \boldsymbol{U}\right)^{-1} \boldsymbol{U}^{\mathrm{H}}$ is $\boldsymbol{U}^{\prime}$ 's pseudo-inverse.



## Determinants - Definition

- Permutation $\left\langle p_{1}, \ldots, p_{n}\right\rangle$ of numbers $\langle 1, \ldots, n\rangle$ is their rearrangement.
- Sign $\sigma(p)$ of permutation $p$ is 1 if $p$ has an even number of element interchanges; otherwise, it is -1 (e.g., $\sigma(\langle 1,3,2\rangle=-1), \sigma(\langle 3,1,2\rangle=1)$ ).
- Determinant $\operatorname{det}\left(\boldsymbol{A}_{n \times n}\right)=|A|=\sum_{p} \sigma(p) A_{1, p_{1}} \ldots A_{n, p_{n}} \in \mathbb{C}$ (Leibniz).
- Simple determinants:

$$
\begin{aligned}
\operatorname{det}\left(\boldsymbol{A}_{1 \times 1}\right) & =A_{11} \\
\operatorname{det}\left(\boldsymbol{A}_{2 \times 2}\right) & =A_{11} A_{22}-A_{12} A_{21}
\end{aligned}
$$

- Expansion along a row (similarly, along a column):

$$
\begin{aligned}
\left|\begin{array}{lll}
1 & 2 & 0 \\
4 & 3 & 5 \\
1 & 1 & 2
\end{array}\right| & =(+1) \cdot 1 \cdot\left|\begin{array}{cc}
3 & 5 \\
1 & 2
\end{array}\right|+(-1) \cdot 2 \cdot\left|\begin{array}{cc}
4 & 5 \\
1 & 2
\end{array}\right|+(+1) \cdot 0 \cdot\left|\begin{array}{cc}
4 & 3 \\
1 & 1
\end{array}\right|= \\
& =1 \cdot(3 \cdot 2-1 \cdot 5)-2 \cdot(4 \cdot 2-1 \cdot 5)+0 \cdot(4 \cdot 1-1 \cdot 3)=-5
\end{aligned}
$$

## Determinants - Properties, Computation

## Properties

- $\operatorname{det}\left(\boldsymbol{A}^{\mathrm{T}}\right)=\operatorname{det}(\boldsymbol{A})$,
- Adding rows to each other does not change $\operatorname{det}(\boldsymbol{A})$.
- Multiplying any row by $\alpha \neq 0$ scales $\operatorname{det}(\boldsymbol{A})$ by $\alpha$.
- Even \# of row swaps does not change $\operatorname{det}()$; odd number - changes sign.
- $\operatorname{det}(\boldsymbol{A B})=\operatorname{det}(\boldsymbol{A}) \operatorname{det}(\boldsymbol{B})$
- For (block-)triangular matrices

$$
\left|\begin{array}{cccc}
\boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \ldots & \boldsymbol{A}_{1 n} \\
0 & \boldsymbol{A}_{22} & \ldots & \boldsymbol{A}_{2 n} \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \boldsymbol{A}_{n n}
\end{array}\right|=\prod_{i=1}^{n} \operatorname{det}\left(\boldsymbol{A}_{i i}\right)
$$

Computing $\operatorname{det}(\boldsymbol{A})$ for large $\boldsymbol{A}$

$$
\operatorname{det}(\boldsymbol{A})=\operatorname{det}(\boldsymbol{P} \boldsymbol{L} \boldsymbol{U})=\underbrace{\operatorname{det}(\boldsymbol{P})}_{\sigma(P)} \times \underbrace{\operatorname{det}(\boldsymbol{L})}_{1} \times \underbrace{\operatorname{det}(\boldsymbol{U})}_{\prod_{i=1}^{n} U_{i i}}
$$

## Determinants - Important Facts

Matrix Singularity

- An invertible $\boldsymbol{A}_{n \times n}$ is called non-singular. Otherwise, it is singular.
- $\boldsymbol{A}$ is singular iff $\operatorname{det}(\boldsymbol{A})=0$. $(\operatorname{det}(\boldsymbol{A}) \approx 0$ does not mean "almost singular".)


## Determinants - Important Facts

## Matrix Singularity

- An invertible $\boldsymbol{A}_{n \times n}$ is called non-singular. Otherwise, it is singular.
- $\boldsymbol{A}$ is singular iff $\operatorname{det}(\boldsymbol{A})=0 .(\operatorname{det}(A) \approx 0$ does not mean "almost singular".)

Thus,

- All the columns (rows) of $\boldsymbol{A}$ are linearly independent iff $\operatorname{det}(\boldsymbol{A}) \neq 0$. (In this case, we say that matrix $\boldsymbol{A}$ is full-rank, i.e., $\operatorname{rank}\left(\boldsymbol{A}_{n \times n}\right)=n$ ).
- Linear system $\boldsymbol{A x}=\boldsymbol{b}$ has a unique solution iff $\operatorname{det}(\boldsymbol{A}) \neq 0$.
- Homogeneous linear system $\boldsymbol{A x}=\mathbf{0}$ has non-trivial solutions iff $\operatorname{det}(\boldsymbol{A})=0$.


## Eigenproblem

## Definition

For a square matrix $\boldsymbol{A}_{n \times n}$, we are interested in those non-trivial vectors $\boldsymbol{x} \neq \mathbf{0}$ that do not change their direction under transformation by $\boldsymbol{A}$ :

$$
\boldsymbol{A x}=\lambda \boldsymbol{x}, \lambda \in \mathbb{C} .
$$

These $\boldsymbol{x}$ are eigenvectors ${ }^{5}$ of $\boldsymbol{A}$, and the corresponding scaling factors $\lambda$ are eigenvalues of $\boldsymbol{A}$. Pairs $\langle\lambda, \boldsymbol{x}\rangle$ of corresponding eigenvalues and eigenvectors are eigenpairs. Distinct eigenvalues of matrix $\boldsymbol{A}$ comprise its spectrum $\sigma(\boldsymbol{A})$. Spectral radius $\rho(\boldsymbol{A})=\max \left\{\left|\sigma_{i}(\boldsymbol{A})\right|\right\}$


[^3]
## Characteristic Polynomial, Its Roots and Coefficients

## Definition

- Goal: find non-trivial solutions of $\boldsymbol{A}_{n \times n} \boldsymbol{x}=\lambda \boldsymbol{x} \Longleftrightarrow(\boldsymbol{A}-\lambda \boldsymbol{I}) \boldsymbol{x}=\mathbf{0}$, where $\boldsymbol{I}$ is the $n$-by- $n$ identity matrix $\left(I_{i i}=1, I_{i j}=0(i \neq j)\right.$.


## Characteristic Polynomial, Its Roots and Coefficients

## Definition

- Goal: find non-trivial solutions of $\boldsymbol{A}_{n \times n} \boldsymbol{x}=\lambda \boldsymbol{x} \Longleftrightarrow(\boldsymbol{A}-\lambda \boldsymbol{I}) \boldsymbol{x}=\mathbf{0}$, where $\boldsymbol{I}$ is the $n$-by- $n$ identity matrix ( $I_{i i}=1, I_{i j}=0(i \neq j)$.
- Homogeneous system $(\boldsymbol{A}-\lambda \boldsymbol{I}) \boldsymbol{x}=0$ has non-trivial solutions iff its matrix is singular, that is, if $\operatorname{det}(\boldsymbol{A}-\lambda \boldsymbol{I})=0$.


## Characteristic Polynomial, Its Roots and Coefficients

## Definition

- Goal: find non-trivial solutions of $\boldsymbol{A}_{n \times n} \boldsymbol{x}=\lambda \boldsymbol{x} \Longleftrightarrow(\boldsymbol{A}-\lambda \boldsymbol{I}) \boldsymbol{x}=\mathbf{0}$, where $\boldsymbol{I}$ is the $n$-by- $n$ identity matrix ( $I_{i i}=1, I_{i j}=0(i \neq j)$.
- Homogeneous system $(\boldsymbol{A}-\lambda \boldsymbol{I}) \boldsymbol{x}=0$ has non-trivial solutions iff its matrix is singular, that is, if $\operatorname{det}(\boldsymbol{A}-\lambda \boldsymbol{I})=0$.
- $p(\lambda)=\operatorname{det}(\boldsymbol{A}-\lambda \boldsymbol{I})$ is the characteristic polynomial (in $\lambda$, of degree $n$ ) of matrix $\boldsymbol{A}$; its roots are $\boldsymbol{A}$ 's eigenvalues, and multiplicity of each root is the algebraic multiplicity of the corresponding eigenvalue. $p(\lambda)=0$ is the characteristic equation for $\boldsymbol{A}$.


## Characteristic Polynomial, Its Roots and Coefficients

## Definition

- Goal: find non-trivial solutions of $\boldsymbol{A}_{n \times n} \boldsymbol{x}=\lambda \boldsymbol{x} \Longleftrightarrow(\boldsymbol{A}-\lambda \boldsymbol{I}) \boldsymbol{x}=\mathbf{0}$, where $\boldsymbol{I}$ is the $n$-by- $n$ identity matrix ( $I_{i i}=1, I_{i j}=0(i \neq j)$.
- Homogeneous system $(\boldsymbol{A}-\lambda \boldsymbol{I}) \boldsymbol{x}=0$ has non-trivial solutions iff its matrix is singular, that is, if $\operatorname{det}(\boldsymbol{A}-\lambda \boldsymbol{I})=0$.
- $p(\lambda)=\operatorname{det}(\boldsymbol{A}-\lambda \boldsymbol{I})$ is the characteristic polynomial (in $\lambda$, of degree $n$ ) of matrix $\boldsymbol{A}$; its roots are $\boldsymbol{A}$ 's eigenvalues, and multiplicity of each root is the algebraic multiplicity of the corresponding eigenvalue. $p(\lambda)=0$ is the characteristic equation for $\boldsymbol{A}$.


## Useful Facts

- From the fundamental theorem of algebra, any $n$-by- $n$ square matrix always has $n$ (not necessarily distinct) complex eigenvalues.
- From the complex conjugate root theorem, if $a+i \cdot b \in \mathbb{C}$ is an eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, then $a-i \cdot b$ is also its eigenvalue.
- From Vieta's theorem applied to $p(\lambda)=\lambda^{n}+c_{1} \lambda^{n-1}+\cdots+c_{n-1} \lambda+c_{n}$,
- $\lambda_{1}+\lambda_{2}+\cdots+\lambda_{n}=-c_{1}=\operatorname{trace}(\boldsymbol{A})$
- $\lambda_{1} \lambda_{2} \ldots \lambda_{n}=(-1)^{n} c_{n}=\operatorname{det}(\boldsymbol{A})$


## Resolving Eigenproblem Directly

## Algorithm

- Step 0: estimate where eigenvalues are located.
- Step 1: solve the characteristic equation $\operatorname{det}(\boldsymbol{A}-\lambda \boldsymbol{I})=0$ using the estimates from Step 0, and find eigenvalues.
- Step 2: for each eigenvalue $\lambda_{i}$, find the corresponding eigenspace by solving homogeneous system $\left(\boldsymbol{A}-\lambda_{i} \boldsymbol{I}\right) \boldsymbol{x}=0$. This eigenspace is comprised of non-trivial members of $\mathcal{N}\left(\boldsymbol{A}-\lambda_{i} \boldsymbol{I}\right)$.


## Resolving Eigenproblem Directly - Example (eigen_direct.m)

$$
\begin{gathered}
\boldsymbol{A}=\left[\begin{array}{rrr}
2 \frac{2}{3} & -\frac{2}{3} & -1 \\
-\frac{1}{3} & 2 \frac{2}{3} & -1 \\
\frac{1}{3} & -\frac{1}{3} & 1
\end{array}\right], \\
p(\lambda)=\operatorname{det}(\boldsymbol{A}-\lambda \boldsymbol{I})=-\lambda^{3}+6 \lambda^{2}-11 \lambda+6, \\
\lambda_{1,2,3}=1,2,3, \\
\lambda_{1}=1: \mathcal{N}\left(\boldsymbol{A}-\lambda_{1} \boldsymbol{I}\right)=\operatorname{span}(\underbrace{\left[\begin{array}{lll}
1 & 1 & 1
\end{array}\right]^{\top}}_{\boldsymbol{e}_{1}}), \\
\lambda_{2}=2: \mathcal{N}\left(\boldsymbol{A}-\lambda_{2} \boldsymbol{I}\right)=\operatorname{span}(\underbrace{\left[\begin{array}{lll}
1 & 1 & 0
\end{array}\right]^{\top}}_{\boldsymbol{e}_{2}}), \\
\lambda_{3}=3: \mathcal{N}\left(\boldsymbol{A}-\lambda_{3} \boldsymbol{I}\right)=\operatorname{span}(\underbrace{\left[\begin{array}{lll}
1 & -2 & 1
\end{array}\right]^{\top}}_{\boldsymbol{e}_{3}}) .
\end{gathered}
$$

## Resolving Eigenproblem - What Actually Works (eigen_arnoldi.m)

- Bad news: solving an equation $p(\lambda)=0$ for high-degree $p$ is very hard.


## Resolving Eigenproblem - What Actually Works (eigen_arnoldi.m)

- Bad news: solving an equation $p(\lambda)=0$ for high-degree $p$ is very hard.
- Most real-world eigensolvers use the idea of Krylov sequences $\left\{\boldsymbol{x}, \boldsymbol{A} \boldsymbol{x}, \boldsymbol{A}^{2} \boldsymbol{x}, \ldots\right\}$ and subspaces spanned by them.


## Resolving Eigenproblem - What Actually Works (eigen_arnoldi.m)

- Bad news: solving an equation $p(\lambda)=0$ for high-degree $p$ is very hard.
- Most real-world eigensolvers use the idea of Krylov sequences $\left\{\boldsymbol{x}, \boldsymbol{A} \boldsymbol{x}, \boldsymbol{A}^{2} \boldsymbol{x}, \ldots\right\}$ and subspaces spanned by them.
- A popular eigensolver for sparse matrices - Arnoldi/Lancsoz iteration (an advanced version of the power method). It allows to quickly compute several (largest, smallest, closest to a given value) eigenvalues and the corresponding eigenvectors of a sparse matrix, mostly, using matrix-vector multiplication. This method is used by MATLAB's eigs and by Python's scipy.sparse.linalg.eigs.
- For dense matrices, eigensolvers based on Schur or Cholesky decomposition may be used.


## Eigenvalue Localization

Sometimes, it may be enough to have a good estimate of where eigenvalues are, without actually computing them. That estimation is referred to as eigenvalue localization.

## Tools

- Crude bound: $|\lambda(A)|<\|A\|$.
- Cauchy's Interlacing Theorem: for real symmetric $n$-by- $n$ matrix

$$
\begin{aligned}
& \boldsymbol{A}=\left[\begin{array}{ll}
\boldsymbol{B} & \boldsymbol{c} \\
\boldsymbol{c}^{\top} & \delta
\end{array}\right], \text { where } \delta \in \mathbb{R}, \\
& \\
& \lambda_{n}(\boldsymbol{A}) \leq \lambda_{n-1}(\boldsymbol{B}) \leq \ldots \lambda_{k}(\boldsymbol{B}) \leq \lambda_{k}(\boldsymbol{A}) \leq \lambda_{k-1}(\boldsymbol{B}) \leq \cdots \leq \lambda_{1}(\boldsymbol{B}) \leq \lambda_{1}(\boldsymbol{A}) .
\end{aligned}
$$

- Gerschgorin Circles The eigenvalues of $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ are trapped inside the union of Gerschgorin circles $\left|z-A_{i i}\right|<r_{i}$, where $r_{i}=\min \left\{\sum_{\substack{j=1 \\ j \neq i}}^{n}\left|A_{i j}\right|, \sum_{\substack{j=1 \\ j \neq i}}^{n}\left|A_{j i}\right|,\right\}, i=1, \ldots, n$. A $k$ Gerschgorin circles disjoint from others contain exactly $k$ eigenvalues.


## Gerschgorin Circles - Example



Eigendecomposition (eigendecomp.m)

- $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is diagonalizable if there is an invertible $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1} \boldsymbol{A} \boldsymbol{P}$ is diagonal.
- If a real-valued matrix is symmetric, then it is diagonalizable. (Though, invertible matrices do not have to be symmetric in general.)

Eigendecomposition (eigendecomp.m)

- $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is diagonalizable if there is an invertible $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1} \boldsymbol{A} \boldsymbol{P}$ is diagonal.
- If a real-valued matrix is symmetric, then it is diagonalizable. (Though, invertible matrices do not have to be symmetric in general.)
- Each diagonalizable $\boldsymbol{A}$ permits (eigen)decomposition:

$$
\boldsymbol{A}=\underbrace{\left[\begin{array}{ccc}
\mid & \cdots & \mid \\
q_{1} & \cdots & q_{n} \\
\mid & \cdots & \mid
\end{array}\right]}_{\boldsymbol{Q}} \underbrace{\left[\begin{array}{cccc}
\lambda_{1} & 0 & \cdots & 0 \\
0 & \lambda_{2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_{n}
\end{array}\right]}_{\boldsymbol{\Lambda}}\left[\begin{array}{ccc}
\mid & \ldots & \mid \\
q_{1} & \cdots & q_{n} \\
\mid & \cdots & \mid
\end{array}\right]^{-1}=\boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{-1} .
$$

- Analog for non-diagonalizable matrices - Jordan normal form.

Eigendecomposition - MV Multiplication for Real Symmetric Matrices

- Assume $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ - symmetric.
- $\boldsymbol{A}$ is diagonalizable, and its eigenvectors are orthogonal.
- For orthogonal matrix $A, A^{-1}=A^{\top}$.

$$
\boldsymbol{A} \boldsymbol{x}=\left(\boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{-1}\right) \boldsymbol{x}=\left(\sum_{i=1}^{n} \lambda_{i} \boldsymbol{q}_{i} \boldsymbol{q}_{i}^{\top}\right) \boldsymbol{x}=\sum_{i=1}^{n} \lambda_{i} \boldsymbol{q}_{i}\left(\boldsymbol{q}_{i}^{\top} \boldsymbol{x}\right)=\sum_{i=1}^{n} \lambda_{i} \underbrace{\left\langle\boldsymbol{x}, \boldsymbol{q}_{i}\right\rangle}_{\mid \text {proj }_{\boldsymbol{q}_{i}} \boldsymbol{x} \mid} \boldsymbol{q}_{i}
$$



## Singular Value Decomposition (SVD)

Theorem
For every (rectangular) matrix $\boldsymbol{A} \in \mathbb{C}^{n \times m}$, there are two unitary ${ }^{6}$ matrices $\boldsymbol{U} \in \mathbb{C}^{m \times m}$ and $\boldsymbol{V} \in \mathbb{C}^{n \times n}$, as well as a matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ of the form

$$
\boldsymbol{\Sigma}=\left[\begin{array}{ccc}
\sigma_{1} & 0 & \cdots \\
0 & \sigma_{2} & \ddots \\
\vdots & \ddots & \ddots
\end{array}\right]
$$

with $\sigma_{1} \geq \sigma_{2} \geq \cdots \geq \sigma_{\min (n, m)} \geq 0$, such that $\boldsymbol{A}=\boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^{\mathrm{H}}$. Diagonal values of $\Sigma$ - singular values of $\boldsymbol{A}$, columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ - left and right singular vectors of $A$, respectively.

(Notice redundant columns in $U T$ and rows (or columns) in $\Sigma$.)
${ }^{6}$ A matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is unitary if $\boldsymbol{A}^{\mathrm{H}} \boldsymbol{A}=\boldsymbol{A} \boldsymbol{A}^{\mathrm{H}}=\boldsymbol{I}$. Unitary matrices play a role similar to the role a scalar 1 plays ("a size-preserving transform".)

## Table of Contents

Linear Algebra: Review of Fundamentals
Matrix Arithmetic
Inversion and Linear Systems Vector Spaces
Geometry
Eigenproblem

Linear Algebra and Graphs
Graphs: Definitions, Properties, Representation Spectral Graph Theory

- An (edge-weighted) graph is a tuple $G=\langle V, E, w\rangle$, where $V$ is a set of nodes, $E \subseteq V \times V$ is a set of edges between the nodes, and $w: E \rightarrow \mathbb{R}$ defined edge weights. If $w$ is not specified, then edge weights can assumed to be equal 1 .
- If $E$ is a symmetric relation (and $w$, if specified, is a symmetric function), then $G$ is said to be undirected; otherwise, it is directed.
- A graph may have weights on its nodes rather than the edges (or on both). A node-weighted graph can be transformed into an edge-weighted graph, or vice versa.
- A graph $G=\langle V, E, w\rangle$ is (weakly) connected if for any $v_{1}, v_{2} \in V$, $v_{1} \neq v_{2}$, we can reach $v_{2}$ from $v_{1}$ by walking along the adjacent edges $E$ (ignoring their direction). A (weakly) disconnected graph $G$ consists of connected components (CC), which are maximal (weakly) connected subgraphs.
- When we take into account direction of edges, the notion of connectedness extends to the notion of strong connectedness (strongly connected components (SCC) are, then, defined similarly to weakly connected components).


## Graphs - Examples

Undirected, (weakly) connected


Undirected, (weakly) disconnected


Directed, (weakly) connected, strongly disconnected


Directed, strongly connected


## Graphs - Special Graphs



Star


Complete Graph (a.k.a. Clique)


Bipartite Graph $\langle L \cup R, E\rangle$


## Representation of Graphs - Adjacency Matrix

- All definitions are given for a graph $G=\langle V, E, w\rangle$ having $|V|=n$ nodes and $|E|=m$ edges.
- The most popular representation of a graph is its adjacency matrix $\boldsymbol{A}$ :

$$
A_{i j}= \begin{cases}w(\langle i, j\rangle)=w_{i j} & \text { if }\langle i, j\rangle \in E \\ 0 & \text { otherwise }\end{cases}
$$

If weights $w$ are not specified, then $\boldsymbol{A}$ is a binary matrix.

## Representation of Graphs - Adjacency Matrices of Special Graphs

(Undirected) Chain


Star


Complete Graph (a.k.a. Clique)


Bipartite Graph $\langle L \cup R, E\rangle$


$$
\begin{gathered}
\quad\left[\begin{array}{lll:l}
0 & 0 & 0 & L \rightarrow R \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{array}\right]
\end{gathered}
$$

## Representation of Graphs - Degree Matrix

- The in-degree $d_{i}^{\text {in }}$ of node $v_{i}$ is the sum of the weights on all of its incoming edges. The out-degree $d_{i}^{\text {out }}$ of node $v_{i}$ is similarly defined via $v_{i}$ 's outgoing edges. For undirected graphs, both in-degree and out-degree are equal $d_{i}$ referred to as simply the degree of node $v_{i}$. For unweighted graphs, the degree measures the number of a node's (in-, our-, or all) neighbors.
- The degree matrix $\boldsymbol{D}=\operatorname{diag}\left(\left\{d_{i}\right\}_{1 \ldots n}\right)$ of a graph $G$ is a diagonal matrix with node degrees on its main diagonal. For directed graphs, in-degree and out-degree matrices can be similarly defined using the appropriate degree definitions.


## Laplacian of Undirected Graphs

(Combinatorial) Laplacian

$$
L=D-A
$$

- Weighted graph:

$$
L_{i j}= \begin{cases}D_{i i}=d_{i}=\sum_{\langle i, \ell\rangle \in E} w_{i, \ell} & \text { if } i=j, \\ -w_{i j} & \text { if }\langle i, j\rangle \in E \\ 0 & \text { otherwise }\end{cases}
$$

- Unweighted graph:

$$
L_{i j}= \begin{cases}d_{i}=\sum_{\langle i, \ell\rangle \in E} 1 & \text { if } i=j, \\ -1 & \text { if }\langle i, j\rangle \in E, \\ 0 & \text { otherwise }\end{cases}
$$

Other Laplacians

$$
\begin{aligned}
\boldsymbol{L}^{s y m} & =\boldsymbol{D}^{-1 / 2} \boldsymbol{L} \boldsymbol{D}^{1 / 2}=\boldsymbol{I}-\boldsymbol{D}^{-1 / 2} \boldsymbol{A} \boldsymbol{D}^{1 / 2} \\
\boldsymbol{L}^{r w} & =\boldsymbol{D}^{-1} \boldsymbol{A}
\end{aligned}
$$

## Spectral Graph Theory

## Spectral Graph Theory

Graphs are usually represented with matrices ${ }^{7}$. Spectral graph theory attempts to connect spectral properties of these matrices with the corresponding graphs' structural properties.

## Limitations

Most results of spectral graph theory are obtained for undirected and unweighted graphs, i.e., graphs having binary symmetric adjacency matrices. If a result applies to weighted graphs, it will be explicitly stated.

[^4]
## Spectrum of Adjacency Matrix - Walks in Graphs

- $\boldsymbol{A} \in\{0,1\}^{n \times n}$ - adjacency matrix of an undirected unweighted graph $G$.
- $A_{i j}$ - number of walks of length 1 in $G$ between nodes $v_{i}$ and $v_{j}$.
- $\left(\boldsymbol{A}^{k}\right)_{i j}$ - number of walks of length $k$ in $G$ between nodes $v_{i}$ and $v_{j}$.
- $\left(\boldsymbol{A}^{k} \mathbb{1}\right)_{i}$ - number of walks of length $k$ ending at $v_{i}$.

- $\mathbb{1}^{\top} \boldsymbol{A}^{k} \mathbb{1}$ - number of walks of length $k$ in $G$.


## Largest ${ }^{8}$ Eigenvalue of Adjacency Matrix $\mu_{1}=\mu_{\max }$

- Connection to $\mu_{\max }$ (undirected, unweighted, connected $G$ ):

$$
\mathbb{1}^{\top} \boldsymbol{A}^{k} \mathbb{1}=(\text { since } \boldsymbol{A} \text { is real and symmetric })=\mathbb{1}^{\top}\left(\boldsymbol{Q} \operatorname{diag}\left(\mu_{i}\right) \boldsymbol{Q}^{-1}\right)^{k} \mathbb{1}=
$$

$=($ since $\boldsymbol{Q}$ is orthogonal $)=\mathbb{1}^{\top} \boldsymbol{Q} \operatorname{diag}\left(\mu_{i}^{k}\right) \boldsymbol{Q}^{-1} \mathbb{1}=$

$$
=\mathbb{1}^{\top}\left(\sum_{i=1}^{n} \mu_{i}^{k} \boldsymbol{q}_{i} \boldsymbol{q}_{i}^{\top}\right) \mathbb{1}=\sum_{i=1}^{n} \mu_{i}^{k}\left(\mathbb{1}^{\top} \boldsymbol{q}_{i}\right)\left(\mathbb{1}^{\top} \boldsymbol{q}_{i}\right)^{\top}=\sum_{i=1}^{n} \mu_{i}^{k}\left\langle\boldsymbol{q}_{i}, \mathbb{1}\right\rangle^{2}=
$$

$$
=\left(\mathbb{1}=\alpha_{1} \boldsymbol{q}_{1}+\cdots+\alpha_{n} \boldsymbol{q}_{n}\right)=\sum_{i=1}^{n} \mu_{i}^{k}\left(\sum_{j=1}^{n} \alpha_{j}\left\langle\boldsymbol{q}_{i}, \boldsymbol{q}_{j}\right\rangle\right)^{2}=\sum_{i=1}^{n} \mu_{i}^{k}\left(\alpha_{i}\right)^{2} ;
$$

$$
\lim _{k \rightarrow \infty}\left(\mathbb{1}^{\top} \boldsymbol{A}^{k_{1}} \mathbb{1}\right)^{1 / k}=\lim _{k \rightarrow \infty}\left(\sum_{i=1}^{n} \mu_{i}^{k} \alpha_{i}^{2}\right)^{1 / k}=
$$

$$
=\lim _{k \rightarrow \infty} \mu_{\max }\left(\alpha_{\max }^{2}+\sum_{i \neq \max ^{\prime \prime}}\left(\frac{\mu_{i}}{\mu_{\max }}\right)^{k} \alpha_{i}^{2}\right)^{1 / k}=\mu_{\max }(=\|A\|) .
$$

- Thus, $\mu_{\max }^{k}=\|\boldsymbol{A}(G)\|^{k}$ is $\approx$ the number of walks of length $k$ in $G$.

[^5]
## Largest Eigenvalue of Adjacency Matrix $\mu_{1}=\mu_{\max }$ - Summary

## Derived

- $\mu_{\max }^{k}=\|\boldsymbol{A}(G)\|^{k}$ is $\approx$ the number of walks of length $k$ in $G$.
- For directed $\boldsymbol{A}$, the meaning of $\lim _{k \rightarrow \infty} \boldsymbol{A}^{k} \mathbb{1}=\boldsymbol{q}_{\text {max }}$ is close to the one of PageRank.


## Largest Eigenvalue of Adjacency Matrix $\mu_{1}=\mu_{\max }$ - Summary

## Derived

- $\mu_{\text {max }}^{k}=\|\boldsymbol{A}(G)\|^{k}$ is $\approx$ the number of walks of length $k$ in $G$.
- For directed $\boldsymbol{A}$, the meaning of $\lim _{k \rightarrow \infty} \boldsymbol{A}^{k} \mathbb{1}=\boldsymbol{q}_{\max }$ is close to the one of PageRank.

Beyond Walks ( $(\dagger)$ - applies to weighted graphs)

- ( $\dagger$ ) If graph $G$ is connected, $\mu_{\max }$ has multiplicity 1 , and its eigenvector is positive (all its entries are strictly positive).
- $(\dagger) d_{\text {avg }} \leq \mu_{1} \leq d_{\max }\left(d_{\text {avg }}, d_{\max }-\right.$ mean and maximum node degrees).
- $\max \left\{d_{\text {avg }}, \sqrt{d_{\text {max }}}\right\} \leq \mu_{\text {max }} \leq d_{\text {max }}$.
- ( $\dagger$ ) If $G$ is connected, and $\mu_{\text {max }}=d_{\text {max }}$, then $\forall i: d_{i}=d_{\text {max }}$.
- ( $\dagger$ ) A connected graph is bipartite iff $\mu_{\text {min }}=-\mu_{\text {max }}$.
- A graph is bipartite iff its spectrum is symmetric about 0 .
- $\chi(G) \geq 1+\mu_{\min } / \mu_{\max }$.


## Smallest Eigenvalues of Combinatorial Laplacian

- $L=D-A$.
- Eigenvalues of $\boldsymbol{L}$ are non-negative: $0 \leq \lambda_{1} \leq \lambda_{2} \leq \cdots \leq \lambda_{n}$.


## Smallest Eigenvalues of Combinatorial Laplacian

- $L=D-A$.
- Eigenvalues of $\boldsymbol{L}$ are non-negative: $0 \leq \lambda_{1} \leq \lambda_{2} \leq \cdots \leq \lambda_{n}$.
- $L$ is singular $\Rightarrow 0$ is its eigenvalue corresponding to eigenvector $\mathbb{1}$.
- If $G$ is connected, $\lambda_{1}=0$ has multiplicity 1 .


## Smallest Eigenvalues of Combinatorial Laplacian

- $L=D-A$.
- Eigenvalues of $\boldsymbol{L}$ are non-negative: $0 \leq \lambda_{1} \leq \lambda_{2} \leq \cdots \leq \lambda_{n}$.
- $L$ is singular $\Rightarrow 0$ is its eigenvalue corresponding to eigenvector $\mathbb{1}$.
- If $G$ is connected, $\lambda_{1}=0$ has multiplicity 1 .
- If $G$ has $k$ connected components, $\lambda_{1}=0$ 's multiplicity is $k$.


## Smallest Eigenvalues of Combinatorial Laplacian

- $L=D-\boldsymbol{A}$.
- Eigenvalues of $\boldsymbol{L}$ are non-negative: $0 \leq \lambda_{1} \leq \lambda_{2} \leq \cdots \leq \lambda_{n}$.
- $L$ is singular $\Rightarrow 0$ is its eigenvalue corresponding to eigenvector $\mathbb{1}$.
- If $G$ is connected, $\lambda_{1}=0$ has multiplicity 1 .
- If $G$ has $k$ connected components, $\lambda_{1}=0$ 's multiplicity is $k$.
- The harder it is to disconnect $G$ by removing its edges, the larger the gap between $\lambda_{1}=0$ and $\lambda_{2}>0$ is.


## Smallest Eigenvalues of Combinatorial Laplacian

- $L=D-A$.
- Eigenvalues of $\boldsymbol{L}$ are non-negative: $0 \leq \lambda_{1} \leq \lambda_{2} \leq \cdots \leq \lambda_{n}$.
- $L$ is singular $\Rightarrow 0$ is its eigenvalue corresponding to eigenvector $\mathbb{1}$.
- If $G$ is connected, $\lambda_{1}=0$ has multiplicity 1 .
- If $G$ has $k$ connected components, $\lambda_{1}=0$ 's multiplicity is $k$.
- The harder it is to disconnect $G$ by removing its edges, the larger the gap between $\lambda_{1}=0$ and $\lambda_{2}>0$ is.
- $\lambda_{2}$ - algebraic connectivity (a.k.a. Fiedler value, spectral gap) - a measure of graph connectedness.
- $\boldsymbol{q}_{2}$ - Fiedler vector - eigenvector associated with $\lambda_{2}$ - solution to a relaxed min-cut (sparsest cut) in $G$. The same eigenvector of a normalized Laplacian $L^{\text {sym }}$ - solution to a relaxed normalized min-cut ("edge-balanced sparsest cut") in $G$.


## Spectral Bisection (spectral_bisection.m)



18 cut edges
Figure : Example of spectral bisection with Fiedler vector.
(The "Tapir" graph as well as the plotting functions come from meshpart
toolbox by John R. Gilbert and Shang-Hua Teng.)

## Spectral Clustering (spectral_clustering.m)



Figure : Example of spectral clustering using normalized Laplacian and k-means.
(The "Tapir" graph as well as the plotting functions come from meshpart toolbox by John R. Gilbert and Shang-Hua Teng.)

## What is next

## Relevant Courses at UCSB

- ECE/CS211A Matrix Analysis - a decent overview of most fundamentals of linear algebra, from the definition of block-matrix arithmetic to spectral theory.
- CS290H Graph Laplacians and Spectra - this course is focused on the study of spectra of graph Laplacians as well as on the accompanying computational problems (extracting eigenpairs of Laplacians, solving Laplacian linear systems).


## Reading - Linear Algebra

- "Core Matrix Analysis" by Shiv Chandrasekaran - a textbook for ECE/CS211A. Provides an overview of most necessary fundamentals.
- "Introduction to Linear Algebra" (any edition) by Gilbert Strang - an entry-level book about fundamentals of linear algebra; great exposition.
- "Matrix Analysis and Applied Linear Algebra" by Carl Meyer - an advanced linear algebra textbook; pick this one if Strang's textbook feels too easy to read.


## What is next

## Reading - "Linear Algebra of Graphs"

- "Spectral Graph Theory" by Fan Chung (1997).
- "Complex Graphs and Networks" by Fan Chung (2006).
- Dan Spielman's course on spectral graph theory.
- "Eigenvalues of Graphs" by László Lovász (2007).
- Luca Trevisan's course on spectral graph theory.
- "Algebraic connectivity of graphs" by Miroslav Fiedler (1973).
- "A tutorial on spectral clustering" by Ulrike von Luxburg (2007).


## $\sim$ Thanks $\sim$

Copyright (c) Victor Amelkin, 2015


[^0]:    ${ }^{1} \overline{a+i \cdot b}=a-i \cdot b \in \mathbb{C}$

[^1]:    ${ }^{2}$ For column scaling, apply diagonal matrix from the right.
    ${ }^{3}$ For permutation of columns, apply $P^{\top}$ from the right.

[^2]:    ${ }^{4}$ Non-negativity axiom is redundant, as it can be derived from other axioms of a norm.

[^3]:    ${ }^{5}$ If $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$, then $\alpha \cdot \boldsymbol{x}$ is also an eigenvector. Thus, an eigenvector defines an entire "direction" or, more generally, a subspace, referred to as eigenspace, whose elements do not change direction when transformed by $\boldsymbol{A}$. Thus, an eigenspace is an invariant subspace of its matrix.

[^4]:    ${ }^{7}$ Some may even go as far as to claim that graphs and matrices are the same thing.

[^5]:    ${ }^{8}$ The largest eigenvalue is such w.r.t. its absolute value.

