Linear Algebra and Graphs 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

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
- $A \in \mathbb{C}^{n imes m}$ matrix with n rows and m columns
- $A_{ij} = A_{i,j}$ element of A in i'th row and j'th column
- $\boldsymbol{A}^{\mathsf{T}}$ transpose of \boldsymbol{A} $(\boldsymbol{A}_{ij}^{\mathsf{T}} = \boldsymbol{A}_{ji})$
- ▶ $oldsymbol{A}^{\mathsf{H}}$ Hermitian transpose of $oldsymbol{A}$ $(oldsymbol{A}_{ij}^{\mathsf{H}} = ar{oldsymbol{A}}_{ji})^1$

Examples

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2 \times 3} \qquad \boldsymbol{A}^{\mathsf{T}} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \in \mathbb{R}^{3 \times 2}$$

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

Block Matrices

$$A = \begin{array}{ccc} & & & & & & & & & & \\ n_1 & & & & & & & \\ A_{11} & & & & & & & \\ \vdots & \ddots & & \vdots & & & \\ A_{q1} & & & & & & & A_{qp} \end{array} \right]$$

Definition

Block matrix – a "matrix of matrices". A_{ij} – block at *i*'th row and *j*'th column of partitioned matrix A. Blocks of the same row (column) have the same number of rows (columns).

Multiplication by scalar

Performed elementwise: $\alpha \cdot A_{n \times m} = \{\alpha \cdot A_{ij}\}_{n \times m}$

Addition

Performed elementwise: $A_{n \times m} + B_{n \times m} = \{A_{ij} + B_{ij}\}_{n \times m}$

Examples

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

Matrix Arithmetic (2)

Multiplication

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

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):

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 3 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} 13 & 13 \\ 25 & 31 \end{bmatrix}$$

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

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

Example 3 (MV block multiplication):

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}_{1\times 3} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{2} \end{bmatrix}_{3\times 1} = \begin{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} 2 \end{bmatrix}_{1\times 1}$$

Matrix Arithmetic (2) - More Examples

Example 4 (MV block multiplication):

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}_{1 \times 1} \begin{bmatrix} 0 & | & 1 \\ -1 & | & 3 \\ 2 & | & 2 \end{bmatrix}_{1 \times 2} = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 0 & | & 1 \\ -1 \\ 5 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}_{1 \times 2} = \begin{bmatrix} 13 & 2 & 2 \\ 25 & 25 \end{bmatrix} | \begin{bmatrix} 13 & 2 & 2 \\ 31 & 2 \end{bmatrix}$$

Example 5 (row scaling²):

$$\begin{bmatrix} d_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d_n \end{bmatrix}_{n \times n} \begin{bmatrix} -\mathbf{A}_{1-} \\ \vdots \\ -\mathbf{A}_{n-} \end{bmatrix}_{n \times n} = \begin{bmatrix} -d_1\mathbf{A}_{1-} \\ \vdots \\ -d_n\mathbf{A}_{n-} \end{bmatrix}_{n \times n}$$

Example 6 (permutation of rows³):

$$\begin{array}{c} {}_{3} \\ {}_{1} \\ {}_{2} \end{array} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array} \end{bmatrix} \begin{bmatrix} -\boldsymbol{A}_{1} - \\ -\boldsymbol{A}_{2} - \\ -\boldsymbol{A}_{3} - \end{array} \end{bmatrix}_{3 \times m} = \begin{bmatrix} -\boldsymbol{A}_{3} - \\ -\boldsymbol{A}_{1} - \\ -\boldsymbol{A}_{2} - \\ 3 \times m \end{bmatrix}_{3 \times m}$$

²For column scaling, apply diagonal matrix from the right.

³For permutation of columns, apply P^{T} from the right.

Inversion

Left Inverse A^{-L} is a left inverse of A if $A^{-L}A = I$, where I is the identity matrix $(I_{ii} = 1, I_{ij} = 0 \text{ if } i \neq j).$

Right Inverse A^{-R} is a right inverse of A if $AA^{-R} = I$.

Inverse

 A^{-1} is the inverse of A if $A^{-1} = A^{-L} = A^{-R}$. 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):

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}^{-L} \approx \begin{bmatrix} -0.94 & -0.11 & 0.72 \\ 0.44 & 0.11 & -0.22 \end{bmatrix}$$

Right Inverse (no left inverse for fat matrices):

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^{-R} \approx \begin{bmatrix} -0.94 & 0.44 \\ -0.11 & 0.11 \\ 0.72 & -0.22 \end{bmatrix}$$

Inverse (may exist only for square matrices):

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}^{-1} \approx \begin{bmatrix} -0.28 & 0.06 & 0.39 \\ 0.06 & 0.39 & -0.28 \\ 0.39 & -0.28 & 0.06 \end{bmatrix}$$

Linear Systems

- ▶ Here, "a linear system" = "a system of linear algebraic equations".
- ▶ Solving Ax = b w.r.t. $x \in \mathbb{R}^n$ is one of two fundamental problems of linear algebra (the other one is eigenproblem).
- Unique solution exists iff A is non-singular $(\det(A) \neq 0)$.
- Problem is related to matrix inversion (i.e., $x = A^{-1}b$).
- ► A system with singular *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 A, there exist a permutation matrix P such that PA = LU, where L is lower-triangular with units on the main diagonal and U is a block-matrix of the form

$$U = \frac{r}{(n-r)} \begin{bmatrix} r & (m-r) \\ \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

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

- Problem: solve Ax = b w.r.t. $x \in \mathbb{R}^{n \times 1}$.
- Step 1: decompose $\boldsymbol{A} = \boldsymbol{P}^{-1} \boldsymbol{L} \boldsymbol{U} (\mathcal{O}(n^3)).$
- ▶ Step 2: $(P^{-1}LU)x = b \iff (LU)x = b'$, where b' = Pb (O(n)).
- ▶ Step 3: solve Ly = b' w.r.t. y using forward substitution $(\mathcal{O}(n^2))$.
- ▶ Step 4: solve Ux = y w.r.t. x using back substitution $(O(n^2))$.

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 $+: \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and *scalar multiplication* $\times: \mathbb{F} \times \mathcal{V} \to \mathcal{V}$, such that $\forall \alpha, \beta \in \mathbb{F} \ \forall x, y, z \in \mathcal{V}$:

- x + (y + z) = (x + y) + z (associativity of addition),
- $\exists \mathbf{0} \in \mathcal{V} : \mathbf{x} + \mathbf{0} = \mathbf{x}$ (existence of additive identity),
- ▶ $\exists -x \in \mathcal{V} : x + (-x) = 0$ (existence of additive inverse),
- $\alpha(\beta x) = (\alpha \beta) x$ (multiplicative associativity),
- ▶ 1x = x (unit scaling),
- $\alpha(x + y) = \alpha x + \alpha y$ (distributivity),
- $(\alpha + \beta)\mathbf{x} = \alpha \mathbf{x} + \beta \mathbf{x}$ (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 $+: \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and *scalar multiplication* $\times: \mathbb{F} \times \mathcal{V} \to \mathcal{V}$, such that $\forall \alpha, \beta \in \mathbb{F} \ \forall x, y, z \in \mathcal{V}$:

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

•
$$\alpha(x + y) = \alpha x + \alpha y$$
 (distributivity),

• $(\alpha + \beta)\mathbf{x} = \alpha \mathbf{x} + \beta \mathbf{x}$ (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 $+: \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and *scalar multiplication* $\times: \mathbb{F} \times \mathcal{V} \to \mathcal{V}$, such that $\forall \alpha, \beta \in \mathbb{F} \ \forall x, y, z \in \mathcal{V}$:

•
$$x + y = y + x$$
 (commutativity of addition),

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

•
$$\alpha(x + y) = \alpha x + \alpha y$$
 (distributivity),

• $(\alpha + \beta)\mathbf{x} = \alpha \mathbf{x} + \beta \mathbf{x}$ (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 combination: $a_1x_1 + \cdots + a_nx_n$ (where $a_i \in \mathbb{F}$, $x_j \in \mathcal{V}$) – linear combination of vectors $\{x_1, \ldots, x_n\}$ with coefficients $\{a_1, \ldots, a_n\}$.

- ▶ Linear combination: $a_1x_1 + \cdots + a_nx_n$ (where $a_i \in \mathbb{F}$, $x_j \in \mathcal{V}$) linear combination of vectors $\{x_1, \ldots, x_n\}$ with coefficients $\{a_1, \ldots, a_n\}$.
- Important observation:

$$a_1 \boldsymbol{x}_1 + \dots + a_n \boldsymbol{x}_n = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

If we think about 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 combination: $a_1x_1 + \cdots + a_nx_n$ (where $a_i \in \mathbb{F}$, $x_j \in \mathcal{V}$) linear combination of vectors $\{x_1, \ldots, x_n\}$ with coefficients $\{a_1, \ldots, a_n\}$.
- Important observation:

$$a_1 \boldsymbol{x}_1 + \dots + a_n \boldsymbol{x}_n = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

If we think about 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.

▶ span $(x_1, ..., x_n) = \{a_1x_1 + \cdots + a_nx_n | a_i \in \mathbb{F}\}$ – a span of a set of vectors is the set of all their possible linear combinations.

- ▶ Linear combination: $a_1x_1 + \cdots + a_nx_n$ (where $a_i \in \mathbb{F}$, $x_j \in \mathcal{V}$) linear combination of vectors $\{x_1, \ldots, x_n\}$ with coefficients $\{a_1, \ldots, a_n\}$.
- Important observation:

$$a_1 \boldsymbol{x}_1 + \dots + a_n \boldsymbol{x}_n = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

If we think about 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.

- ▶ span $(x_1, ..., x_n) = \{a_1x_1 + \cdots + a_nx_n | a_i \in \mathbb{F}\}$ a span of a set of vectors is the set of all their possible linear combinations.
- ► { $x_1, ..., x_n$ } are linearly independent if $a_1x_1 + \cdots + a_nx_n = 0$ iff [$a_1 \dots a_n$] = 0.

- ▶ Linear combination: $a_1x_1 + \cdots + a_nx_n$ (where $a_i \in \mathbb{F}$, $x_j \in \mathcal{V}$) linear combination of vectors $\{x_1, \ldots, x_n\}$ with coefficients $\{a_1, \ldots, a_n\}$.
- Important observation:

$$a_1 \boldsymbol{x}_1 + \dots + a_n \boldsymbol{x}_n = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

If we think about 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.

- ▶ span $(x_1, ..., x_n) = \{a_1x_1 + \cdots + a_nx_n | a_i \in \mathbb{F}\}$ a span of a set of vectors is the set of all their possible linear combinations.
- ▶ { $x_1, ..., x_n$ } are linearly independent if $a_1x_1 + \cdots + a_nx_n = 0$ iff [$a_1 \ldots a_n$] = 0.
- ► A set {x₁,..., x_n} of *linearly independent* vectors is a **basis** of subspace W if W = span(x₁,..., x_n). W's dimension dim W is n.

Fundamental Subspaces of a Matrix

• The nullspace (kernel) of $A_{n \times m}$ is

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

• The range (column space, image) of $A_{n \times m}$ is

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

- $\mathcal{N}(\mathbf{A}^{\mathsf{H}})$ left nullspace of \mathbf{A} .
- $\blacktriangleright \ \mathcal{R}(\mathbf{A}^{\mathsf{H}}) \mathbf{row} \ \mathbf{space} \ \mathbf{of} \ \mathbf{A}.$

Fundamental Subspaces of a Matrix - "The Big Picture"



(from Gilbert Strang's "Introduction to Linear Algebra", 4'th edition, 2009)

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 $\|\cdot\|:\mathcal{V}\to\mathbb{R}$, such that $\forall \alpha\in\mathbb{F}$ $\forall x,y\in\mathcal{V}$ the following *norm axioms* hold

•
$$\boldsymbol{x} \in \mathcal{V} : \|\boldsymbol{x}\| \ge 0$$
 (non-negativity⁴),

•
$$\|\boldsymbol{x}\| = 0 \rightarrow \boldsymbol{x} = \boldsymbol{0}$$
 (positive definiteness),

•
$$\|\alpha x\| = |\alpha| \|x\|$$
 (homogeneity),

•
$$\|x + y\| \le \|x\| + \|y\|$$
 (subadditivity / triangle inequality).

Examples:

•
$$\ell_p$$
-norm: $\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$,

•
$$L_p$$
-norm: $\|\boldsymbol{f}\|_p = \left(\int_D |\boldsymbol{f}|^p \mathrm{d}\mu\right)^{1/p} < \infty.$

⁴Non-negativity axiom is redundant, as it can be derived from other axioms of a norm.

Convexity and Norm

• A unit sphere is defined as $\{x \mid ||x|| = 1\}$.

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



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

Convexity and Norm $-\ell_p$ (norms_and_convexity.m)

Function
$$f_p(\boldsymbol{x}) = \left(\sum_i |x_i|^p\right)^{1/p}$$
 mapping \mathbb{R}^n to \mathbb{R} is a norm iff $1 \le p \le \infty$.

• Alternatively, f_p is a norm iff the unit sphere $\{x \mid f_p(x) = 1\}$ induced by f_p is convex.



Common Vector Norms

Euclidean norm

$$\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + \dots + x_n^2}$$

Taxicab / Manhattan norm

$$\|\boldsymbol{x}\|_1 = |x_1| + \dots + |x_n|$$

Chebyshev norm

$$\|\boldsymbol{x}\|_{\infty} = \max\{|x_1|,\ldots,|x_n|\}$$

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 $oldsymbol{x} \in \mathbb{C}^n$

$$\begin{split} \| \boldsymbol{x} \|_1 &\leq \sqrt{n} \| \boldsymbol{x} \|_2, \\ \| \boldsymbol{x} \|_1 &\leq n \| \boldsymbol{x} \|_{\infty}, \\ \| \boldsymbol{x} \|_2 &\leq \sqrt{n} \| \boldsymbol{x} \|_{\infty}, \\ \text{ and, more generally, for } 0$$

Matrix Norms (operator_norm.m)

Frobenius Norm:
$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j=1}^n A_{ij}^2} = \mathbf{trace}(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}),$$

where trace() of a matrix is the sum of the elements on its main diagonal.

• Operator Norm (Induced p-norm): $\|A\| = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \sup_{\|x\|_p=1} \|Ax\|_p$

An operator norm measures the maximum degree of distortion / amount of stretch of a unit sphere under transformation by A.



▶ For $A, B \in \mathbb{C}^{n \times n}$, $||AB|| \le ||A|| ||B||$ (submultiplicativity).

▶ Inner product (scalar product) of $x, y \in \mathcal{V}$: $\langle x, y \rangle = x^{\mathsf{H}}y = \sum_{i} \bar{x}_{i}y_{i}$.

▶ Inner product (scalar product) of $x, y \in \mathcal{V}$: $\langle x, y \rangle = x^{\mathsf{H}}y = \sum_{i} \bar{x}_{i}y_{i}$.

- ► 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 $x, y \in \mathcal{V}$: $\langle x, y \rangle = x^{\mathsf{H}}y = \sum_{i} \bar{x}_{i}y_{i}$.

- ► 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: $\langle x, y \rangle \leq \|x\|_2 \|y\|_2$
- ► CBS inequality inspires the following definition of an **angle** θ between vectors $x, y \in \mathbb{R}^n$:

$$\cos heta = rac{\langle oldsymbol{x}, oldsymbol{y}
angle}{\|oldsymbol{x}\| \|oldsymbol{y}\|}$$

• Vectors x and y are **orthogonal** $(x \perp y)$ if the cosine of the angle between them is 0.

• Inner product (scalar product) of $x, y \in \mathcal{V}$: $\langle x, y \rangle = x^{\mathsf{H}}y = \sum_{i} \bar{x}_{i}y_{i}$.

- ► 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: $\langle x, y \rangle \leq \|x\|_2 \|y\|_2$
- ► CBS inequality inspires the following definition of an **angle** θ between vectors $x, y \in \mathbb{R}^n$:

$$\cos heta = rac{\langle oldsymbol{x}, oldsymbol{y}
angle}{\|oldsymbol{x}\| \|oldsymbol{y}\|}$$

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

Inner Product, Angle, Projection – Example

▶ Problem: Given a matrix $A \in \mathbb{R}^{n \times n}$, find the amount of stretch caused by A to vectors along a direction defined by a vector x in ℓ_2 .

 $(\|\mathbf{x}\| = 1)$ relative length?

Inner Product, Angle, Projection – Example

▶ Problem: Given a matrix $A \in \mathbb{R}^{n \times n}$, find the amount of stretch caused by A to vectors along a direction defined by a vector x in ℓ_2 .



• Given an arbitrary x, normalize it, so that its length is 1:

$$\bar{\boldsymbol{x}} = \boldsymbol{x} / \|\boldsymbol{x}\| \quad \left(\|\bar{\boldsymbol{x}}\| = \left\|\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right\| = \frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = 1\right)$$
Inner Product, Angle, Projection – Example

▶ Problem: Given a matrix $A \in \mathbb{R}^{n \times n}$, find the amount of stretch caused by A to vectors along a direction defined by a vector x in ℓ_2 .



• Given an arbitrary x, normalize it, so that its length is 1:

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

▶ The amount of stretch caused by *A* along *x*:

$$\frac{\mathsf{new size}}{\mathsf{original size}} = \frac{|\operatorname{proj}_{\bar{\boldsymbol{x}}} \boldsymbol{A} \bar{\boldsymbol{x}}|}{\|\bar{\boldsymbol{x}}\|} = |\operatorname{proj}_{\bar{\boldsymbol{x}}} \boldsymbol{A} \bar{\boldsymbol{x}}| = \langle \bar{\boldsymbol{x}}, \boldsymbol{A} \bar{\boldsymbol{x}} \rangle = \frac{\langle \boldsymbol{x}, \boldsymbol{A} \boldsymbol{x} \rangle}{\|\boldsymbol{x}\|^2} = \frac{\boldsymbol{x}^{\mathsf{T}} \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^{\mathsf{T}} \boldsymbol{x}}$$

Inner Product, Angle, Projection – Example

▶ Problem: Given a matrix $A \in \mathbb{R}^{n \times n}$, find the amount of stretch caused by A to vectors along a direction defined by a vector x in ℓ_2 .



Given an arbitrary x, normalize it, so that its length is 1:

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

▶ The amount of stretch caused by *A* along *x*:

$$\frac{\mathsf{new size}}{\mathsf{original size}} = \frac{|\operatorname{proj}_{\bar{\boldsymbol{x}}} \boldsymbol{A} \bar{\boldsymbol{x}}|}{\|\bar{\boldsymbol{x}}\|} = |\operatorname{proj}_{\bar{\boldsymbol{x}}} \boldsymbol{A} \bar{\boldsymbol{x}}| = \langle \bar{\boldsymbol{x}}, \boldsymbol{A} \bar{\boldsymbol{x}} \rangle = \frac{\langle \boldsymbol{x}, \boldsymbol{A} \boldsymbol{x} \rangle}{\|\boldsymbol{x}\|^2} = \frac{\boldsymbol{x}^{\mathsf{T}} \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^{\mathsf{T}} \boldsymbol{x}}$$

• Derived Rayleigh quotient (generally, x^{H} is used instead of x^{T}).

Linear Systems (Revisited)

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



Determinants – Definition

• **Permutation** $\langle p_1, \ldots, p_n \rangle$ 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** det
$$(\mathbf{A}_{n \times n}) = |A| = \sum_{p} \sigma(p) A_{1,p_1} \dots A_{n,p_n} \in \mathbb{C}$$
 (Leibniz).

Simple determinants:

$$det(\mathbf{A}_{1\times 1}) = A_{11}, det(\mathbf{A}_{2\times 2}) = A_{11}A_{22} - A_{12}A_{21}$$

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

$$\begin{vmatrix} 1 & 2 & 0 \\ 4 & 3 & 5 \\ 1 & 1 & 2 \end{vmatrix} = (+1) \cdot 1 \cdot \begin{vmatrix} 3 & 5 \\ 1 & 2 \end{vmatrix} + (-1) \cdot 2 \cdot \begin{vmatrix} 4 & 5 \\ 1 & 2 \end{vmatrix} + (+1) \cdot 0 \cdot \begin{vmatrix} 4 & 3 \\ 1 & 1 \end{vmatrix} = \\ = 1 \cdot (3 \cdot 2 - 1 \cdot 5) - 2 \cdot (4 \cdot 2 - 1 \cdot 5) + 0 \cdot (4 \cdot 1 - 1 \cdot 3) = -5.$$

Determinants - Properties, Computation

Properties

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

$$\begin{vmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ 0 & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{nn} \end{vmatrix} = \prod_{i=1}^{n} \det(A_{ii}).$$

Computing $\det(A)$ for large A

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

Determinants – Important Facts

Matrix Singularity

- An invertible $A_{n \times n}$ is called **non-singular**. Otherwise, it is **singular**.
- ► A is singular iff det(A) = 0. (det(A) ≈ 0 does not mean "almost singular".)

Determinants – Important Facts

Matrix Singularity

- An invertible $A_{n \times n}$ is called **non-singular**. Otherwise, it is **singular**.
- ► A is singular iff det(A) = 0. (det(A) ≈ 0 does not mean "almost singular".)

Thus,

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

Eigenproblem

Definition

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

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

These x are eigenvectors⁵ of A, and the corresponding scaling factors λ are eigenvalues of A. Pairs $\langle \lambda, x \rangle$ of corresponding eigenvalues and eigenvectors are eigenpairs. Distinct eigenvalues of matrix A comprise its spectrum $\sigma(A)$. Spectral radius $\rho(A) = \max\{|\sigma_i(A)|\}$



⁵If x is an eigenvector of A, then $\alpha \cdot x$ 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 A. Thus, an eigenspace is an *invariant subspace* of its matrix.

Definition

► Goal: find non-trivial solutions of $A_{n \times n} x = \lambda x \iff (A - \lambda I) x = 0$, where I is the *n*-by-*n* identity matrix $(I_{ii} = 1, I_{ij} = 0 (i \neq j))$.

Definition

- ► Goal: find non-trivial solutions of $A_{n \times n} x = \lambda x \iff (A \lambda I) x = 0$, where I is the *n*-by-*n* identity matrix $(I_{ii} = 1, I_{ij} = 0 (i \neq j))$.
- Homogeneous system (A − λI)x = 0 has non-trivial solutions iff its matrix is singular, that is, if det(A − λI) = 0.

Definition

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

Definition

- ► Goal: find non-trivial solutions of $A_{n \times n} x = \lambda x \iff (A \lambda I) x = 0$, where I is the *n*-by-*n* identity matrix $(I_{ii} = 1, I_{ij} = 0 (i \neq j))$.
- ► Homogeneous system (A λI)x = 0 has non-trivial solutions iff its matrix is singular, that is, if det(A λI) = 0.
- ▶ p(λ) = det(A − λI) is the characteristic polynomial (in λ, of degree n) of matrix A; its roots are A's eigenvalues, and multiplicity of each root is the algebraic multiplicity of the corresponding eigenvalue. p(λ) = 0 is the characteristic equation for 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 $A \in \mathbb{R}^{n \times n}$, then $a i \cdot b$ is also its eigenvalue.
- From Vieta's theorem applied to $p(\lambda) = \lambda^n + c_1 \lambda^{n-1} + \dots + c_{n-1} \lambda + c_n$,
 - $\lambda_1 + \lambda_2 + \dots + \lambda_n = -c_1 = \mathbf{trace}(\mathbf{A})$
 - $\lambda_1 \lambda_2 \dots \lambda_n = (-1)^n c_n = \det(\mathbf{A})$

Resolving Eigenproblem Directly

Algorithm

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

Resolving Eigenproblem Directly - Example (eigen_direct.m)

$$\boldsymbol{A} = \begin{bmatrix} 2\frac{2}{3} & -\frac{2}{3} & -1\\ -\frac{1}{3} & 2\frac{2}{3} & -1\\ \frac{1}{3} & -\frac{1}{3} & 1 \end{bmatrix},$$
$$p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = -\lambda^3 + 6\lambda^2 - 11\lambda + 6,$$
$$\lambda_{1,2,3} = 1, 2, 3,$$
$$\lambda_1 = 1: \mathcal{N}(\boldsymbol{A} - \lambda_1 \boldsymbol{I}) = \operatorname{span}(\underbrace{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}}_{\boldsymbol{e}_1}^{\mathsf{T}}),$$
$$\lambda_2 = 2: \mathcal{N}(\boldsymbol{A} - \lambda_2 \boldsymbol{I}) = \operatorname{span}(\underbrace{\begin{bmatrix} 1 & 1 & 0 \end{bmatrix}}_{\boldsymbol{e}_2}^{\mathsf{T}}),$$
$$\lambda_3 = 3: \mathcal{N}(\boldsymbol{A} - \lambda_3 \boldsymbol{I}) = \operatorname{span}(\underbrace{\begin{bmatrix} 1 & -2 & 1 \end{bmatrix}}_{\boldsymbol{e}_3}^{\mathsf{T}}).$$

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 $\{x, Ax, A^2x, \ldots\}$ and subspaces spanned by them.

- ▶ 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 $\{x, Ax, A^2x, \ldots\}$ 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.

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

$$oldsymbol{A} = \begin{bmatrix} oldsymbol{B} & oldsymbol{c} \\ oldsymbol{c}^{\mathsf{T}} & \delta \end{bmatrix}$$
, where $\delta \in \mathbb{R}$,
 $\lambda_n(oldsymbol{A}) \leq \lambda_{n-1}(oldsymbol{B}) \leq \dots \lambda_k(oldsymbol{B}) \leq \lambda_k(oldsymbol{A}) \leq \lambda_{k-1}(oldsymbol{B}) \leq \dots \leq \lambda_1(oldsymbol{B}) \leq \lambda_1(oldsymbol{A}).$

▶ Gerschgorin Circles The eigenvalues of $A \in \mathbb{C}^{n \times n}$ are trapped inside the union of Gerschgorin circles $|z - A_{ii}| < r_i$, where $r_i = \min \{\sum_{\substack{j=1 \ j \neq i}}^n |A_{ij}|, \sum_{\substack{j=1 \ j \neq i}}^n |A_{ji}|, \}, i = 1, ..., n. A k$ Gerschgorin circles

disjoint from others contain exactly k eigenvalues.

Gerschgorin Circles - Example



Eigendecomposition (eigendecomp.m)

- $A \in \mathbb{C}^{n \times n}$ is diagonalizable if there is an invertible P such that $P^{-1}AP$ 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)

- $A \in \mathbb{C}^{n \times n}$ is diagonalizable if there is an invertible P such that $P^{-1}AP$ 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 A permits (eigen)decomposition:

$$\boldsymbol{A} = \underbrace{\left[\begin{array}{cccc} | & \dots & | \\ q_1 & \dots & q_n \\ | & \dots & | \end{array}\right]}_{\boldsymbol{Q}} \underbrace{\left[\begin{array}{cccc} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{array}\right]}_{\boldsymbol{\Lambda}} \left[\begin{array}{cccc} | & \dots & | \\ q_1 & \dots & q_n \\ | & \dots & | \end{array}\right]^{-1} = \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{-1}.$$

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

Eigendecomposition – MV Multiplication for Real Symmetric Matrices

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

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



Singular Value Decomposition (SVD)

Theorem

For every (rectangular) matrix $A \in \mathbb{C}^{n \times m}$, there are two unitary⁶ matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$, as well as a matrix $\Sigma \in \mathbb{R}^{m \times n}$ of the form

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \dots \\ 0 & \sigma_2 & \ddots \\ \vdots & \ddots & \ddots \\ \vdots & \ddots & \ddots \end{bmatrix},$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(n,m)} \geq 0$, such that $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\mathsf{H}}$. Diagonal values of Σ – singular values of \mathbf{A} , columns of \mathbf{U} and \mathbf{V} – left and right singular vectors of \mathbf{A} , respectively.



(Notice redundant columns in U and rows (or columns) in Σ .) ⁶A matrix $A \in \mathbb{C}^{n \times n}$ is unitary if $A^{H}A = AA^{H} = I$. 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

Graphs: definitions, properties

- An (edge-weighted) graph is a tuple G = ⟨V, E, w⟩, where V is a set of nodes, E ⊆ V × V is a set of edges between the nodes, and w : E → 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 = ⟨V, E, w⟩ is (weakly) connected if for any v₁, v₂ ∈ V, v₁ ≠ v₂, we can reach v₂ from v₁ 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



Graphs – Special Graphs



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 A:

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

If weights w are not specified, then A is a binary matrix.

Representation of Graphs - Adjacency Matrices of Special Graphs





 $\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ L & R & R & R & L \end{bmatrix}$

Representation of Graphs - Degree Matrix

- ► The in-degree dⁱⁿ_i of node v_i is the sum of the weights on all of its incoming edges. The **out-degree** d^{out}_i 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 D = diag({d_i}_{1...n}) 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:

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

Unweighted graph:

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

Other Laplacians

$$L^{sym} = D^{-1/2}LD^{1/2} = I - D^{-1/2}AD^{1/2},$$

 $L^{rw} = D^{-1}A.$

Spectral Graph Theory

Graphs are usually represented with matrices⁷. 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.

⁷Some may even go as far as to claim that graphs and matrices are the same thing.

Spectrum of Adjacency Matrix - Walks in Graphs

- ▶ $A \in \{0,1\}^{n \times n}$ adjacency matrix of an undirected unweighted graph G.
- A_{ij} number of walks of length 1 in G between nodes v_i and v_j .
- $(\mathbf{A}^{k})_{ij}$ number of walks of length k in G between nodes v_i and v_j .
- $(\mathbf{A}^k \mathbb{1})_i$ number of walks of length k ending at v_i .



• $\mathbb{1}^T A^k \mathbb{1}$ – number of walks of length k in G.

Largest⁸ Eigenvalue of Adjacency Matrix $\mu_1 = \mu_{max}$

• Connection to μ_{max} (undirected, unweighted, connected G):

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

 $=(\mathsf{since}\; \boldsymbol{Q}\;\mathsf{is\;orthogonal})=\mathbb{1}^{\mathsf{T}}\boldsymbol{Q}\operatorname{diag}\left(\boldsymbol{\mu}_{i}^{k}\right)\boldsymbol{Q}^{-1}\mathbb{1}=$

$$= \mathbb{1}^{\mathsf{T}} \left(\sum_{i=1}^{n} \mu_i^k \boldsymbol{q}_i \boldsymbol{q}_i^{\mathsf{T}} \right) \mathbb{1} = \sum_{i=1}^{n} \mu_i^k (\mathbb{1}^{\mathsf{T}} \boldsymbol{q}_i) (\mathbb{1}^{\mathsf{T}} \boldsymbol{q}_i)^{\mathsf{T}} = \sum_{i=1}^{n} \mu_i^k \langle \boldsymbol{q}_i, \mathbb{1} \rangle^2 =$$
$$= (\mathbb{1} = \alpha_1 \boldsymbol{q}_1 + \dots + \alpha_n \boldsymbol{q}_n) = \sum_{i=1}^{n} \mu_i^k \left(\sum_{j=1}^{n} \alpha_j \langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle \right)^2 = \sum_{i=1}^{n} \mu_i^k (\alpha_i)^2;$$
$$\lim \left(\mathbb{1}^{\mathsf{T}} \boldsymbol{A}^k \mathbb{1} \right)^{1/k} = \lim \left(\sum_{j=1}^{n} \mu_i^k \alpha_i^2 \right)^{1/k} =$$

$$\lim_{k \to \infty} \left(\mathbb{I} \ \mathbf{A} \ \mathbb{I} \right)^{-1} = \lim_{k \to \infty} \left(\sum_{i=1}^{k} \mu_i \alpha_i \right)^{-1} =$$
$$= \lim_{k \to \infty} \mu_{max} \left(\alpha_{max}^2 + \sum_{i \neq max''} \left(\frac{\mu_i}{\mu_{max}} \right)^k \alpha_i^2 \right)^{1/k} = \mu_{max} (= ||A||).$$

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

⁸The *largest* eigenvalue is such w.r.t. its absolute value.

Largest Eigenvalue of Adjacency Matrix $\mu_1 = \mu_{max}$ – Summary

Derived

- $\mu_{max}^k = \|\mathbf{A}(G)\|^k$ is \approx the number of walks of length k in G.
- ► For directed A, the meaning of $\lim_{k\to\infty} A^k \mathbb{1} = q_{max}$ is close to the one of **PageRank**.

Largest Eigenvalue of Adjacency Matrix $\mu_1 = \mu_{max}$ – Summary

Derived

- $\mu_{max}^k = \|\mathbf{A}(G)\|^k$ is \approx the number of walks of length k in G.
- ► For directed A, the meaning of $\lim_{k\to\infty} A^k \mathbb{1} = q_{max}$ is close to the one of PageRank.

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

- ► (†) If graph G is connected, µ_{max} has multiplicity 1, and its eigenvector is positive (all its entries are strictly positive).
- ▶ (†) $d_{avg} \le \mu_1 \le d_{max}$ (d_{avg} , d_{max} mean and maximum node degrees).
- $\max\{d_{avg}, \sqrt{d_{max}}\} \le \mu_{max} \le d_{max}$.
- (†) If G is connected, and $\mu_{max} = d_{max}$, then $\forall i : d_i = d_{max}$.
- (†) A connected graph is bipartite iff $\mu_{min} = -\mu_{max}$.
- A graph is bipartite iff its spectrum is symmetric about 0.
- $\chi(G) \ge 1 + \mu_{min}/\mu_{max}$.
- $\blacktriangleright L = D A.$
- Eigenvalues of L are non-negative: $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$.

- $\blacktriangleright L = D A.$
- Eigenvalues of L are non-negative: $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$.
- L is singular $\Rightarrow 0$ is its eigenvalue corresponding to eigenvector 1.
- If G is connected, $\lambda_1 = 0$ has multiplicity 1.

- $\blacktriangleright L = D A.$
- Eigenvalues of L are non-negative: $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$.
- L is singular $\Rightarrow 0$ is its eigenvalue corresponding to eigenvector 1.
- If G is connected, $\lambda_1 = 0$ has multiplicity 1.
- If G has k connected components, $\lambda_1 = 0$'s multiplicity is k.

- $\blacktriangleright L = D A.$
- Eigenvalues of L are non-negative: $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$.
- L is singular $\Rightarrow 0$ is its eigenvalue corresponding to eigenvector 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.

- $\blacktriangleright L = D A.$
- Eigenvalues of L are non-negative: $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$.
- L is singular $\Rightarrow 0$ is its eigenvalue corresponding to eigenvector 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.
- ▶ λ₂ algebraic connectivity (a.k.a. Fiedler value, spectral gap) a measure of graph connectedness.
- q₂ Fiedler vector eigenvector associated with λ₂ solution to a relaxed min-cut (sparsest cut) in G. The same eigenvector of a normalized Laplacian L^{sym} solution to a relaxed normalized min-cut ("edge-balanced sparsest cut") in G.

Spectral Bisection (spectral_bisection.m)



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

