UVA CS 4774: Machine Learning

Lecture: Algebra and Calculus Review

Dr. Yanjun Qi

University of Virginia   Department of Computer Science

$$\begin{pmatrix} -1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \textbf{?}$$

$$A = \begin{pmatrix} 0 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix} \quad A^T = \textbf{?}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix}$$

C = A – B = ?

C = A + B = ?

$$\left(\begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right)^T = \textbf{?}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

C = A B = ?

C = B A = ?

Minimum minimum requirement test

Notation

- Inputs
  - X, X_j (jth element of vector X) : random variables written in capital letter
  - p #input variables, n #observations
  - $\mathbf{X}$: matrix written in bold capital
  - Vectors are assumed to be column vectors
  - Discrete inputs often described by characteristic vector (dummy variables)
- Outputs
  - quantitative Y
  - qualitative C (for categorical)
- Observed variables written in lower case
  - The i-th observed value of X_j is $x_{j,i}$ (a scalar)

DEFINITIONS - SCALAR

- a scalar is a number
  - (denoted with regular type: 1 or 22)

DEFINITIONS - VECTOR

◆Vector: a single row or column of numbers

- denoted with bold small letters
- row vector

a = $$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

- column vector (default)

b = $$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$$

DEFINITIONS - VECTOR

- Vector in real space R^n is an ordered set of n real numbers.
  - e.g. $v = (1,6,3,4)^T$ is in R^4
  - A column vector: $\begin{pmatrix} 1 \\ 6 \\ 3 \\ 4 \end{pmatrix}$
  - v^T as a row vector: $\begin{pmatrix} 1 & 6 & 3 & 4 \end{pmatrix}$

DEFINITIONS - MATRIX

- m-by-n matrix in $R^{m\times n}$ with m rows and n columns, each entry filled with a (typically) real number:

- e.g. 3*3 matrix

$$\begin{pmatrix} 1 & 2 & 8 \\ 4 & 78 & 6 \\ 9 & 3 & 2 \end{pmatrix}$$

Square matrix

DEFINITIONS - MATRIX

- We normally write the entry of a matrix as

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

- Denoted with a Capital letter
- All matrices have an order (or dimension):
  that is, the number of rows * the number of columns.

So, A is 2 by 3 or (2 * 3).

- A square matrix is a matrix that has the same number of rows and columns (n * n)

Special matrices

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$ diagonal

$$\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$$ upper-triangular

$$\begin{pmatrix} a & b & 0 & 0 \\ c & d & e & 0 \\ 0 & f & g & h \\ 0 & 0 & i & j \end{pmatrix}$$ tri-diagonal

$$\begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix}$$ lower-triangular

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$ I (identity matrix)

Special matrices: Symmetric Matrices

$$A = A^T \ (a_{ij} = a_{ji})$$

e.g.:

$$\begin{bmatrix} 4 & 5 & -3 \\ 5 & 7 & 2 \\ -3 & 2 & 10 \end{bmatrix}$$

Column or Row Views to Denote

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- We denote the jth column of A by a_j or $A_{:,j}$:

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}.$$

- We denote the ith row of A by a_i^T or $A_{i,:}$:

$$A = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix}.$$

- Note that these definitions are ambiguous (for example, the a_1 and a_1^T in the previous two definitions are *not* the same vector). Usually the meaning of the notation should be obvious from its use.

Review of MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

(1) Transpose

Transpose: You can think of it as

- "flipping" the rows and columns

e.g. $\begin{pmatrix} a \\ b \end{pmatrix}^T = \begin{pmatrix} a & b \end{pmatrix}$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A+B)^T = A^T + B^T$

(2) Matrix Addition/Subtraction

- Matrix addition/subtraction
  - Matrices must be of same size.
  - Entry-wise operation across all entries

(2) Matrix Addition/Subtraction An Example

- If we have

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix}$$

then we can calculate **C** = **A** + **B** by

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} 8 & 12 \\ 11 & 15 \\ 14 & 18 \end{bmatrix}$$

(2) Matrix Addition/Subtraction An Example

- Similarly, if we have

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix}$$

then we can calculate **C** = **A** - **B** by

$$\mathbf{C} = \mathbf{A} - \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} - \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} -6 & -8 \\ -5 & -7 \\ -4 & -6 \end{bmatrix}$$

OPERATION on MATRIX

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

(3) Products of Matrices

- We write the multiplication of two matrices **A** and **B** as **AB**

- This is referred to either as
  - pre-multiplying **B** by **A**

    or
  - post-multiplying **A** by **B**

- So for matrix multiplication **AB**, **A** is referred to as the *premultiplier* and **B** is referred to as the *postmultiplier*

Products of Matrices

- If we have $A_{(3x3)}$ and $B_{(3x2)}$ then

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \mathrm{x} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} = \mathbf{C}$$

where

$$\begin{aligned} c_{11} &= a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ c_{12} &= a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ c_{21} &= a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ c_{22} &= a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \\ c_{31} &= a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} \\ c_{32} &= a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} \end{aligned}$$

Matrix Multiplication
An Example

- If we have

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

then

$$\mathbf{AB} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \text{x} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} = \begin{bmatrix} 30 & 66 \\ 36 & 81 \\ 42 & 96 \end{bmatrix}$$

where

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} = 1(1) + 4(2) + 7(3) = 30$$

$$c_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} = 1(4) + 4(5) + 7(6) = 66$$

$$c_{21} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} = 2(1) + 5(2) + 8(3) = 36$$

$$c_{22} = a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} = 2(4) + 5(5) + 8(6) = 81$$

$$c_{31} = a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} = 3(1) + 6(2) + 9(3) = 42$$

$$c_{32} = a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} = 3(4) + 6(5) + 9(6) = 96$$

Products of Matrices

$$\underset{m \times n}{\begin{bmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{22} & . & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & . & a_{mn} \end{bmatrix}} \underset{q \times p}{\begin{bmatrix} b_{11} & b_{12} & . & b_{1p} \\ b_{21} & b_{22} & . & b_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ b_{q1} & b_{q2} & . & b_{qp} \end{bmatrix}} = \underset{m \times p}{\begin{bmatrix} c_{11} & c_{12} & . & c_{1p} \\ c_{21} & c_{22} & . & c_{2p} \\ \cdots & \cdots & c_{ij} & \cdots \\ c_{m1} & c_{m2} & . & c_{mp} \end{bmatrix}}$$

Condition: n = q

$$c_{ij} = \sum_{k=1}^{\mathbf{n}} a_{ik} b_{kj}$$

$$AB \neq BA$$

Products of Matrices: Conformable

- In order to multiply matrices, they must be conformable (the number of columns in the premultiplier must equal the number of rows in postmultiplier)

- Note that
  - an (m x n) x (n x p) = (m x p)
  - an (m x n) x (p x n) = cannot be done
  - a (1 x n) x (n x 1) = a scalar (1 x 1)

Some Properties of Matrix Multiplication

- Note that
  - Even if conformable, **AB** does not necessarily equal **BA** (i.e., matrix multiplication is *not commutative*)
  - Matrix multiplication can be extended beyond two matrices
  - matrix multiplication is *associative*, i.e., **A**(**BC**) = (**AB**)**C**

Some Properties of Matrix Multiplication

- Multiplication and transposition

  $(AB)^T = B^T A^T$

- Multiplication with Identity Matrix

$$AI = IA = A, \text{ where } I = \begin{bmatrix} 1 & 0 & . & 0 \\ 0 & 1 & . & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & . & 1 \end{bmatrix}$$

Special Uses for Matrix Multiplication

- Products of Scalars & Matrices ➔ Example, If we have

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad b = 3.5$$

then we can calculate b**A** by

$$\mathbf{bA} = 3.5 \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 3.5 & 7.0 \\ 10.5 & 14.0 \\ 17.5 & 21.0 \end{bmatrix}$$

◆ Note that b**A** = **A**b if b is a scalar

Special Uses for Matrix Multiplication

- Dot (or Inner) Product of two Vectors
  - Premultiplication of a column vector **a** by conformable row vector **b** yields a single value called the *dot product* or *inner product*
  - \- If

$$\mathbf{b} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix} \qquad \mathbf{a}^{\mathbf{T}} = \begin{bmatrix} 3 & 4 & 6 \end{bmatrix}$$

then **their inner product** gives us

$$\mathbf{a}^{\mathbf{T}}\mathbf{b} = \mathbf{a} \bullet \mathbf{b} = \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix} = 3(5) + 4(2) + 6(8) = 71 = \mathbf{b}^{\mathbf{T}}\mathbf{a}$$

which is the sum of products of elements in similar positions for the two vectors

Special Uses for Matrix Multiplication

- Outer Product of two Vectors
  - Postmultiplication of a column vector **a** by conformable row vector **b** yields a matrix containing the products of each pair of elements from the the two matrices (called the *outer product*) - If

$$\mathbf{a} = \begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix}$$

then $\mathbf{ab}^\mathsf{T}$ gives us

$$\mathbf{ab}^\mathsf{T} = \begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix} \begin{bmatrix} 5 & 2 & 8 \end{bmatrix} = \begin{bmatrix} 15 & 6 & 24 \\ 20 & 8 & 32 \\ 30 & 12 & 48 \end{bmatrix}$$

Special Uses for Matrix Multiplication

- Outer Product of two Vectors, e.g. a special case :

As an example of how the outer product can be useful, let $\mathbf{1} \in \mathbb{R}^n$ denote an n-dimensional vector whose entries are all equal to 1. Furthermore, consider the matrix $A \in \mathbb{R}^{m \times n}$ whose columns are all equal to some vector $x \in \mathbb{R}^m$. Using outer products, we can represent A compactly as,

$$A = \begin{bmatrix} | & | & & | \\ x & x & \cdots & x \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x_1 & x_1 & \cdots & x_1 \\ x_2 & x_2 & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_m & \cdots & x_m \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} = x\mathbf{1}^T.$$

Special Uses for Matrix Multiplication

•Matrix-Vector Products (I)

Given a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $x \in \mathbb{R}^n$, their product is a vector $y = Ax \in \mathbb{R}^m$.

If we write A by rows, then we can express Ax as,

$$y = Ax = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}.$$

Special Uses for Matrix Multiplication

•Matrix-Vector Products (II)

Alternatively, let's write A in column form. In this case we see that,

$$y = Ax = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \\ a_1 \\ \\ \end{bmatrix} x_1 + \begin{bmatrix} \\ a_2 \\ \\ \end{bmatrix} x_2 + \ldots + \begin{bmatrix} \\ a_n \\ \\ \end{bmatrix} x_n \ .$$

In other words, y is a ***linear combination*** of the *columns* of A, where the coefficients of the linear combination are given by the entries of x.

Special Uses for Matrix Multiplication

•Matrix-Vector Products (III)

to multiply on the left by a row vector. This is written, $y^T = x^T A$ for $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^m$, and $y \in \mathbb{R}^n$.

$$y^T = x^T A = x^T \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x^T a_1 & x^T a_2 & \cdots & x^T a_n \end{bmatrix}$$

which demonstrates that the ith entry of y^T is equal to the inner product of x and the ith *column* of A.

Special Uses for Matrix Multiplication

- Matrix-Vector Products (IV)

$$
\begin{aligned}
y^T &= x^T A \\
&= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \\
&= x_1 \begin{bmatrix} — & a_1^T & — \end{bmatrix} + x_2 \begin{bmatrix} — & a_2^T & — \end{bmatrix} + \ldots + x_n \begin{bmatrix} — & a_n^T & — \end{bmatrix}
\end{aligned}
$$

so we see that y^T is a linear combination of the *rows* of A, where the coefficients for the linear combination are given by the entries of x.

MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

(4) Vector norms

A norm of a vector $||x||$ is informally a measure of the "length" of the vector.

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

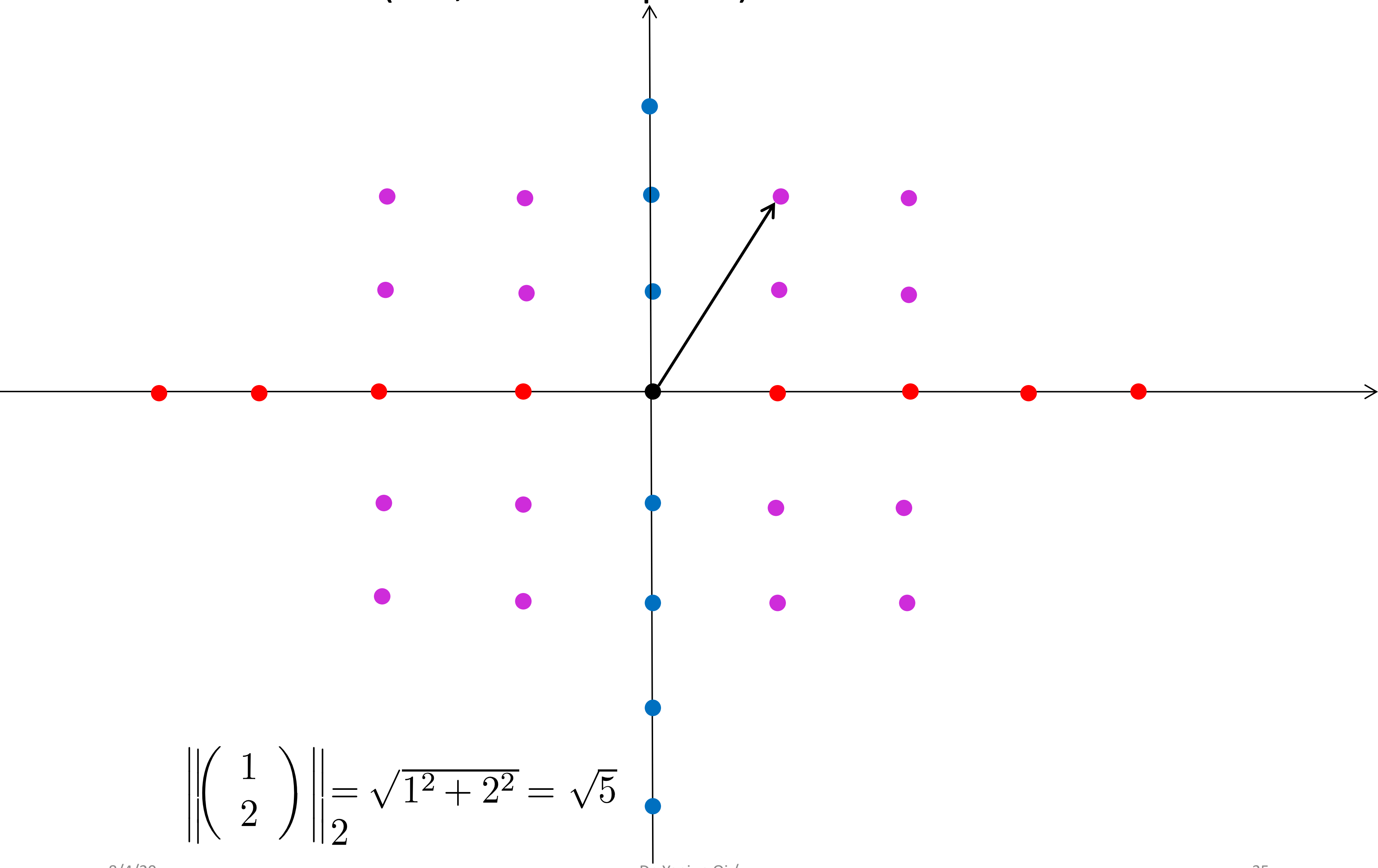
– Common norms: L_1, L_2 (Euclidean)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^T x}$$

– $L_{infinity}$

$$\|x\|_\infty = \max_i |x_i|$$

Vector Norm (L2, when p=2)



$$\left\| \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\|_2 = \sqrt{1^2 + 2^2} = \sqrt{5}$$

Special Uses for Matrix Multiplication

- Sum the Squared Elements of a Vector
  - Premultiply a column vector **a** by its transpose – If

$$\mathbf{a} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix}$$

then premultiplication by a row vector $\mathbf{a}^{T}$

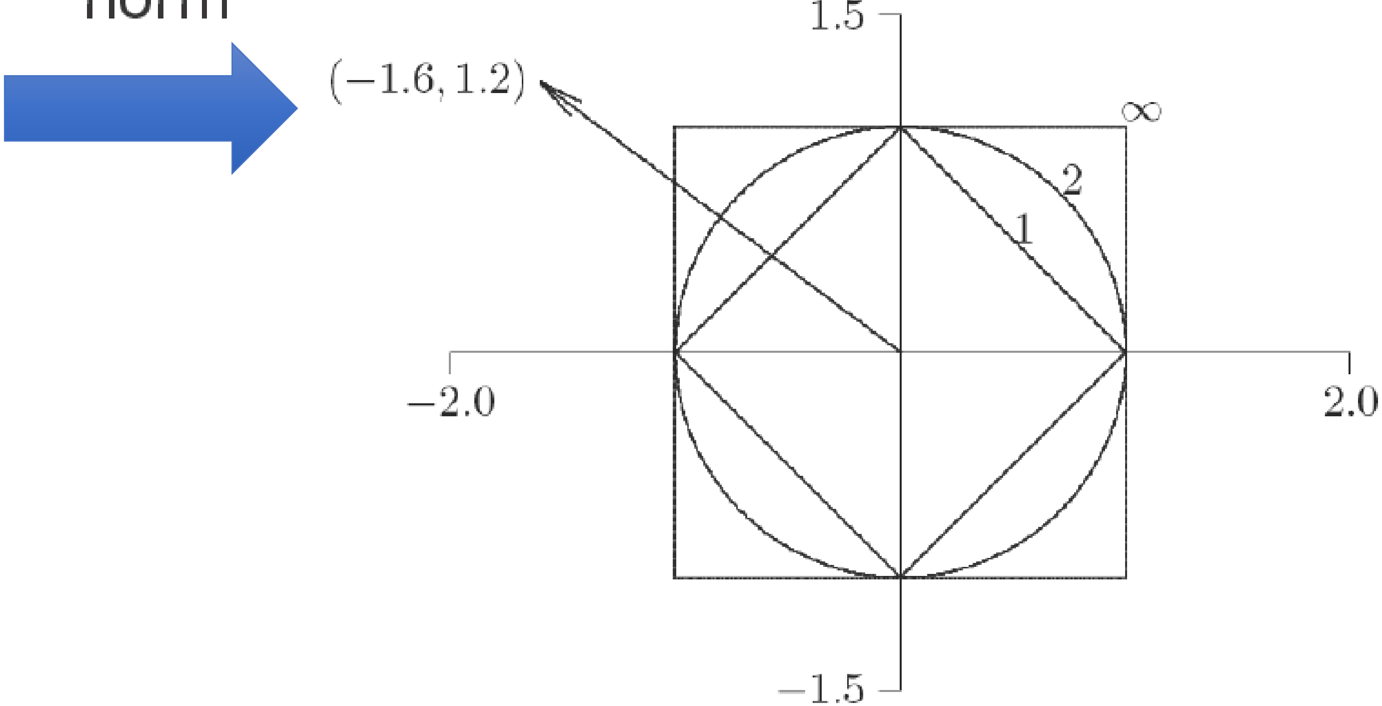
$$a^{T} = \begin{bmatrix} 5 & 2 & 8 \end{bmatrix}$$

will yield the sum of the squared values of elements for **a**, i.e.

$$a^{T}\mathbf{a} = \begin{bmatrix} 5 & 2 & 8 \end{bmatrix}\begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix} = 5^2 + 2^2 + 8^2 = 93$$

Vector Norms (e.g.,)

Drawing shows unit sphere in two dimensions for each norm



Norms have following values for vector shown

$$\|\boldsymbol{x}\|_1 = 2.8 \quad \|\boldsymbol{x}\|_2 = 2.0 \quad \|\boldsymbol{x}\|_\infty = 1.6$$

In general, for any vector $\boldsymbol{x}$ in $\mathbb{R}^n$, $\|\boldsymbol{x}\|_1 \geq \|\boldsymbol{x}\|_2 \geq \|\boldsymbol{x}\|_\infty$

More General : Norm

- A norm is any function g() that maps vectors to real numbers that satisfies the following conditions:

- **Non-negativity**: for all $\boldsymbol{x} \in \mathbb{R}^D$, $g(\boldsymbol{x}) \geq 0$
- **Strictly positive**: for all $\boldsymbol{x}$, $g(\boldsymbol{x}) = 0$ implies that $\boldsymbol{x} = \mathbf{0}$
- **Homogeneity**: for all $\boldsymbol{x}$ and a, $g(a\boldsymbol{x}) = |a|\, g(\boldsymbol{x})$, where $|a|$ is the absolute value.
- **Triangle inequality**: for all $\boldsymbol{x}, \boldsymbol{y}$, $g(\boldsymbol{x} + \boldsymbol{y}) \leq g(\boldsymbol{x}) + g(\boldsymbol{y})$

Orthogonal & Orthonormal

Inner Product defined between column vector **x** and **y** , as

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ x_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i . = x \bullet y$$

If u•v=0, $||u||_2$!= 0, $||v||_2$!= 0

→ *u and v are orthogonal*

If u•v=0, $||u||_2 = 1$, $||v||_2 = 1$

→ *u and v are orthonormal*

Orthogonal matrices

- Notation:

$$A = \begin{bmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{22} & . & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & . & a_{mn} \end{bmatrix}, \Rightarrow \begin{array}{l} u_1^T = [a_{11}\ a_{12}\ \cdots\ a_{1n}] \\ u_2^T = [a_{21}\ a_{22}\ \cdots\ a_{2n}] \\ \quad\cdots \\ u_m^T = [a_{m1}\ a_{m2}\ \cdots\ a_{mn}] \end{array} \Rightarrow A = \begin{bmatrix} u_1^T \\ u_2^T \\ \cdots \\ u_m^T \end{bmatrix}$$

- A is orthogonal if:

$$(1)\ u_k . u_k = 1 \text{ or } ||u_k|| = 1, \text{ for every } k$$

$$(2)\ u_j . u_k = 0, \text{ for every } j \neq k \ (u_j \text{ is perpendicular to } u_k)$$

Example: $\begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix}$

Orthogonal matrices

- If square *A* is orthogonal, it is easy to find its inverse:

$$AA^T = A^T A = I \quad \text{(i.e., } A^{-1} = A^T\text{)}$$

Property: $\|Av\| = \|v\|$ (does not change the magnitude of v)

Matrix Norm

- **Definition**: Given a vector norm $||x||$, the **matrix norm** defined by the vector norm is given by:

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

- What does a matrix norm represent?
- It represents the maximum "stretching" that A does to a vector **x** -> (A**x**).

Matrix 1- Norm

Theorem A: The matrix norm corresponding to 1-norm is maximum absolute column sum:

$$\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|$$

Proof: From previous slide, we can have $\|A\|_1 = \max_{\|x\|=1} \|Ax\|_1$

Also, $$Ax = x_1 A_1 + x_2 A_2 + \cdots + x_n A_n = \sum_{j=1}^{n} x_j A_j$$

where A_j is the j-th column of A.

MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

(5) Inverse of a Matrix

- The inverse of a matrix $\mathbf{A}$ is commonly denoted by $\mathbf{A}^{-1}$ or inv $\mathbf{A}$.
- The inverse of an $n \times n$ matrix $\mathbf{A}$ is the matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$
- The matrix inverse is analogous to a scalar reciprocal
- A matrix which has an inverse is called *nonsingular*

(5) Inverse of a Matrix

- For some *n x n* matrix **A** , an inverse matrix $\mathbf{A}^{-1}$ *may not exist*.
- A matrix which does not have an inverse is singular.
- An inverse of *n x n* matrix **A** exists iff |**A**| not 0

THE DETERMINANT OF A MATRIX

- The determinant of a matrix A is denoted by |A| (or det(A) or det A).
- Determinants exist only for square matrices.

- E.g. If A =

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$|A| = a_{11}a_{22} - a_{12}a_{21}$$

THE DETERMINANT OF A MATRIX

2 x 2

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

3 x 3

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

n x n

$$det(A) = \sum_{j=1}^{m} (-1)^{j+k} a_{jk} det(A_{jk}), \text{ for any } k\text{: } 1 \leq k \leq m$$

THE DETERMINANT OF A MATRIX

$$det(AB) = det(A)det(B)$$

$$det(A + B) \neq det(A) + det(B)$$

diagonal matrix:

$$\text{If } A = \begin{bmatrix} a_{11} & 0 & . & 0 \\ 0 & a_{22} & . & 0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & . & a_{nn} \end{bmatrix}, \text{ then } det(A) = \prod_{i=1}^{n} a_{ii}$$

HOW TO FIND INVERSE MATRIXES? An example,

- If
- $$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$ and |A| not 0

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Matrix Inverse

- The inverse A^{-1} of a matrix A has the property:

$$AA^{-1}=A^{-1}A=I$$

- A^{-1} exists if only if $det(A) \neq 0$

- Terminology
  - Singular matrix: A^{-1} does not exist
  - Ill-conditioned matrix: A is close to being singular

PROPERTIES OF INVERSE MATRICES

- $$\left(\boldsymbol{AB}\right)^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$
- $$\left(\boldsymbol{A}^{\boldsymbol{T}}\right)^{-1} = \left(\boldsymbol{A}^{-1}\right)^{T}$$
- $$\left(\boldsymbol{A}^{-1}\right)^{-1} = \boldsymbol{A}$$

Inverse of special matrix

- For diagonal matrices $\mathbf{D}^{-1} = \mathrm{diag}\{d_1^{-1}, \ldots, d_n^{-1}\}$

- For orthogonal matrices $\mathbf{A}^{-1} = \mathbf{A}^\top$

  - a square matrix with real entries whose columns and rows are orthogonal unit vectors (i.e., orthonormal vectors)

Pseudo-inverse

- The pseudo-inverse A^+ of a matrix A (could be non-square, e.g., m x n) is given by:

$$A^+ = (A^T A)^{-1} A^T$$

- It can be shown that:

$$A^+ A = I \quad \text{(provided that } (A^T A)^{-1} \text{ exists)}$$

MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

(6) Rank: Linear independence

- A set of vectors is linearly independent if none of them can be written as a linear combination of the others.

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

$$x_3 = -2\, x_1 + x_2$$

➔ NOT linearly independent

(6) Rank: Linear independence

- <u>Alternative definition:</u> Vectors $v_1,\ldots,v_k$ are linearly independent if $c_1v_1+\ldots+c_kv_k = 0$ implies $c_1=\ldots=c_k=0$

$$\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

e.g.

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

(u,v)=(0,0), i.e. the columns are linearly independent.

(6) Rank of a Matrix

- rank(A) (the rank of a m-by-n matrix A) is
  = The maximal number of linearly independent columns
  =The maximal number of linearly independent rows

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$$

Rank=? **Rank=?**

- If A is n by m, then
  - rank(A)<= min(m,n)
  - If n=rank(A), then A has full row rank
  - If m=rank(A), then A has full column rank

(6) Rank of a Matrix

- Equal to the dimension of the largest square sub-matrix of A that has a non-zero determinant.

Example:

$$\begin{bmatrix} 4 & 5 & 2 & 14 \\ 3 & 9 & 6 & 21 \\ 8 & 10 & 7 & 28 \\ 1 & 2 & 9 & 5 \end{bmatrix}$$

has rank 3

$$det(A) = 0, \text{ but } det(\begin{bmatrix} 4 & 5 & 2 \\ 3 & 9 & 6 \\ 8 & 10 & 7 \end{bmatrix}) = 63 \neq 0$$

(6) Rank and singular matrices

rank(AB) <= min(rank(A), rank(B))

If A is nxn, $rank(A) = n$ iff A is nonsingular (i.e., invertible).

If A is nxn, $rank(A) = n$ iff $det(A) \neq 0$ (**full rank**).

If A is nxn, $rank(A) < n$ iff A is singular

We can use row reduction to calculating Rank of a matrix

From Wiki

The following complexity figures assume that arithmetic with individual elements has complexity $O(1)$, as is the case with fixed-preci
operations on a finite field.

Operation	Input	Output	Algorithm	Complexity
Matrix multiplication	Two $n \times n$ matrices	One $n \times n$ matrix	Schoolbook matrix multiplication	$O(n^3)$
			Strassen algorithm	$O(n^{2.807})$
			Coppersmith–Winograd algorithm	$O(n^{2.376})$
			Optimized CW-like algorithms[14][15][16]	$O(n^{2.373})$
Matrix multiplication	One $n \times m$ matrix & one $m \times p$ matrix	One $n \times p$ matrix	Schoolbook matrix multiplication	$O(nmp)$
Matrix inversion*	One $n \times n$ matrix	One $n \times n$ matrix	Gauss–Jordan elimination	$O(n^3)$
			Strassen algorithm	$O(n^{2.807})$
			Coppersmith–Winograd algorithm	$O(n^{2.376})$
			Optimized CW-like algorithms	$O(n^{2.373})$
Singular value decomposition	One $m \times n$ matrix	One $m \times m$ matrix, one $m \times n$ matrix, & one $n \times n$ matrix		$O(mn^2)$ $(m \le n)$
		One $m \times r$ matrix, one $r \times r$ matrix, & one $n \times r$ matrix		
Determinant	One $n \times n$ matrix	One number	Laplace expansion	$O(n!)$
			Division-free algorithm[17]	$O(n^4)$
			LU decomposition	$O(n^3)$
			Bareiss algorithm	$O(n^3)$
			Fast matrix multiplication[18]	$O(n^{2.373})$
Back substitution	Triangular matrix	n solutions	Back substitution[19]	$O(n^2)$

MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

Review: Derivative of a Function

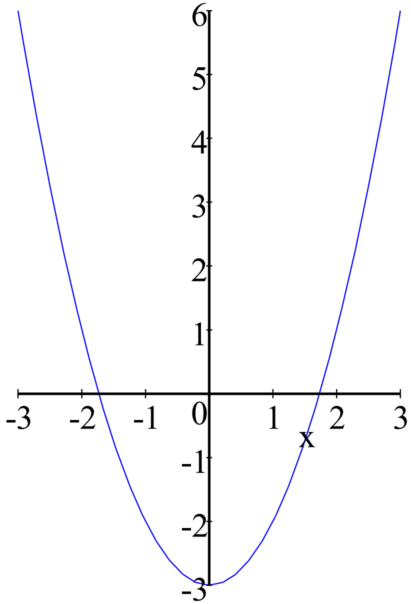
$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$ is called the derivative of f at a.

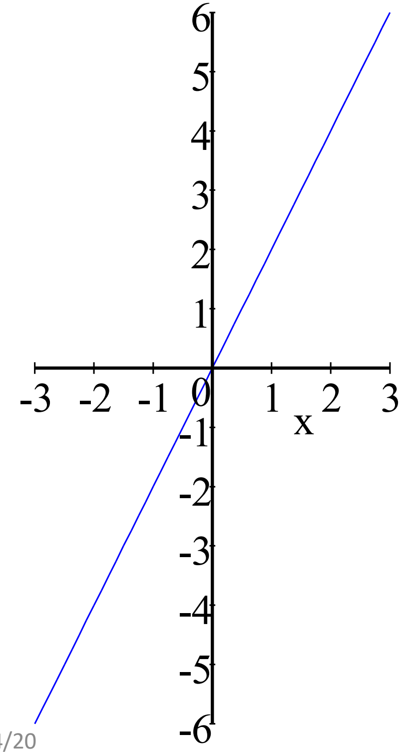
We write: $$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

"The derivative of f with respect to X is …"

There are many ways to write the derivative of $y = f(x)$

➔ e.g. define the slope of the curve y=f(x) at the point x





Review: Derivative of a Quadratic Function

$$y = x^2 - 3$$

$$y' = \lim_{h \to 0} \frac{(x+h)^2 - 3 - (x^2 - 3)}{h}$$

$$y' = \lim_{h \to 0} \frac{\cancel{x^2} + 2x\cancel{h} + h^{\cancel{2}} - \cancel{x^2}}{\cancel{h}}$$

$$y' = \lim_{h \to 0} 2x + \cancelto{0}{h}$$

$$y' = 2x$$

Single Var-Func to Multivariate

Single Var-Function	**Multivariate Calculus**
Derivative Second-order derivative	Partial Derivative Gradient Directional Partial Derivative Vector Field Contour map of a function Surface map of a function Hessian matrix Jacobian matrix (vector in / vector out)

Some important rules for taking (partial) derivatives

- Scalar multiplication: $\partial_x[af(x)] = a[\partial_x f(x)]$
- Polynomials: $\partial_x[x^k] = kx^{k-1}$
- Function addition: $\partial_x[f(x) + g(x)] = [\partial_x f(x)] + [\partial_x g(x)]$
- Function multiplication: $\partial_x[f(x)g(x)] = f(x)[\partial_x g(x)] + [\partial_x f(x)]g(x)$
- Function division: $\partial_x\left[\frac{f(x)}{g(x)}\right] = \frac{[\partial_x f(x)]g(x) - f(x)[\partial_x g(x)]}{[g(x)]^2}$
- Function composition: $\partial_x[f(g(x))] = [\partial_x g(x)][\partial_x f](g(x))$
- Exponentiation: $\partial_x[e^x] = e^x$ and $\partial_x[a^x] = \log(a)e^x$
- Logarithms: $\partial_x[\log x] = \frac{1}{x}$

Review: Definitions of gradient (Matrix_calculus / Scalar-by-matrix)

Suppose that $f : \mathbb{R}^{m\times n} \rightarrow \mathbb{R}$ is a function that takes as input a matrix A of size $m \times n$ and returns a real value. Then the ***gradient*** of f (with respect to $A \in \mathbb{R}^{m\times n}$) is the matrix of

➔ ***Denominator layout***

$$\nabla_A f(A) \in \mathbb{R}^{m\times n} = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \cdots & \frac{\partial f(A)}{\partial A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}} \end{bmatrix}$$

In principle, gradients are a natural extension of partial derivatives to functions of multiple variables.

Review: Definitions of gradient
(Matrix_calculus / Scalar-by-vector)

- Size of gradient is always the same as the size of variable

➔ *Denominator layout*

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n \quad \text{if } x \in \mathbb{R}^n$$

For Examples

$$\boxed{\frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}}$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T$$

$$\frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{x}$$

Exercise: a simple example

$$f(w) = w^T a = [w_1, w_2, w_3]\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = w_1 + 2w_2 + 3w_3$$

→ ***Denominator layout***

$$\frac{\partial f}{\partial w_1} = 1$$

$$\frac{\partial f}{\partial w_2} = 2$$

$$\frac{\partial f}{\partial w_3} = 3$$

$$\frac{\partial f}{\partial w} = \frac{\partial w^T a}{\partial w} = a = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Even more general Matrix Calculus: Types of Matrix Derivatives

	Scalar	Vector	Matrix
Scalar	$\frac{df}{dx}$	$\frac{dF}{dx} = [\frac{\partial F_i}{\partial x}]$	$\frac{dF}{dx} = [\frac{\partial F_{ij}}{\partial x}]$
Vector	$\frac{df}{dX} = [\frac{df}{dX_i}]$	$\frac{dF}{dX} = [\frac{\partial F_i}{\partial X_j}]$	
Matrix	$\frac{df}{dX} = [\frac{df}{dX_{ij}}]$		

Adapted from Thomas Minka. Old and New Matrix Algebra Useful for Statistics

Review: Hessian Matrix / n=2 case

Singlevariate → multivariate

$$f(x,z)$$

- 1st derivative to gradient,

$$g = \nabla f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial z} \end{pmatrix}$$

- 2nd derivative to Hessian

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial x \partial z} & \frac{\partial^2 f}{\partial z^2} \end{pmatrix}$$

Review: Hessian Matrix

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function that takes a vector in $\mathbb{R}^n$ and returns a real number. Then the ***Hessian*** matrix with respect to x, written $\nabla_x^2 f(x)$ or simply as H is the $n \times n$ matrix of partial derivatives,

$$\nabla_x^2 f(x) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

Hessian PD/PSD (Extra)

Let $f : D \to \mathbb{R}$ be a function on non-singular, convex domain $D \subseteq \mathbb{R}^d$ and let us assume the second-order derivatives of f exist. It is well known that f is convex if and only if its Hessian $\nabla^2 f(x)$ is positive semi-definite for all $x \in D$. It is also known that if $\nabla^2 f(x)$ is positive definite for all $x \in D$, we may conclude that f is strictly convex (for a reference, see Boyd and Vandenberghe, 2004).

On the other hand, if f is strictly convex, we still merely know that $\nabla^2 f(x)$ is positive semi-definite for all $x \in D$. That is, there may be $x \in D$ such that $y^T \nabla^2 f(x) y = 0$ for some $y \neq 0$.

As an example, consider $f(x) = x^4$. In this case, f is strictly convex, but $f''(x) = 12x^2$ and, hence, $yf''(x)y = 0$ for $x = 0$ and $yf''(x)y > 0$ for all $x \neq 0$.

http://people.seas.harvard.edu/~yaron/AM221/lecture_notes/AM221_lecture10.pdf

Theorem 2. *Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be twice continuos differentiable on S.*

1. *If $H_f(\mathbf{x})$ is positive semi-definite for any $\mathbf{x} \in S$ then f is convex on S.*
2. *If $H_f(\mathbf{x})$ is positive definite for any $\mathbf{x} \in S$ then f is strongly convex on S.*
3. *If S is open and f is convex, then $H_f(\mathbf{x})$ is positive semi-definite $\forall \mathbf{x} \in S$.*

Today Recap

- **Linear Algebra and Matrix Calculus Review**

0) Basic Calculus
1) Transposition
2) Addition and Subtraction
3) Multiplication

MUST KNOW

---

4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus

Good to KNOW

Extra

- The following topics are covered by handout, but not by this slide (some will be covered …)
  - Trace()
  - Eigenvalue / Eigenvectors
  - Positive definite matrix , Gram matrix
  - Quadratic form
  - Projection (vector on a plane, or on a vector)

Best Place to Review: Khan Academy

‹ MATH

Linear algebra

Vectors and spaces

0 of 45 complete

Let's get our feet wet by thinking in terms of vectors and spaces.

- Vectors
- Linear combinations and spans
- Linear dependence and independence
- Subspaces and the basis for a subspace
- Vector dot and cross products
- Matrices for solving systems by elimina...
- Null space and column space

Matrix transformations

0 of 58 complete

Understanding how we can map one set of vectors to another set. Matrices used to define linear transformations.

- Functions and linear transformations
- Linear transformation examples
- Transformations and matrix multiplicati...
- Inverse functions and transformations
- Finding inverses and determinants
- More determinant depth
- Transpose of a matrix

Alternate coordinate systems (bases)

0 of 39 complete

We explore creating and moving between various coordinate systems.

- Orthogonal complements
- Orthogonal projections
- Change of basis
- Orthonormal bases and the Gram-Sch...
- Eigen-everything

Best Place to Review: Khan Academy

< MATH

Multivariable calculus

Thinking about multivariable functions

2 of 22 complete

The only thing separating multivariable calculus from ordinary calculus is this newfangled word "multivariable". It means we will deal with functions whose inputs or outputs live in two or more dimensions. Here we lay the foundations for thinking about and visualizing multivariable functions.

- Introduction to multivariable calculus
- Visualizing scalar-valued functions
- Visualizing vector-valued functions
- Transformations
- Visualizing multivariable functions (artic...

Derivatives of multivariable functions

6 of 72 complete

What does it mean to take the derivative of a function whose input lives in multiple dimensions? What about when its output is a vector? Here we go over many different ways to extend the idea of a derivative to higher dimensions, including partial derivatives, directional derivatives, the gradient, vector derivatives, divergence, curl, etc.

- Partial derivatives
- Gradient and directional derivatives
- Partial derivative and gradient (articles)
- Differentiating parametric curves
- Multivariable chain rule
- Curvature
- Partial derivatives of vector-valued fun...
- Differentiating vector-valued functions (...
- Divergence
- Curl
- Divergence and curl (articles)
- Laplacian
- Jacobian

Applications of multivariable derivatives

1 of 37 complete

The tools of partial derivatives, the gradient, etc. can be used to optimize and approximate multivariable functions. These are very useful in practice, and to a large extent this is why people study multivariable calculus.

- Tangent planes and local linearization
- Quadratic approximations
- Optimizing multivariable functions
- Optimizing multivariable functions (artic...
- Lagrange multipliers and constrained o...
- Constrained optimization (articles)

From Khan Academy

- Matrix representing linear transformation of the basic space (each column of the matrix is the new basis)
- Matrix determinant (therefore representing the transformed unit square’s area, the bigger, means the bigger transformation)
- Jacobian matrix determinant therefore representing the speed/amount of func change at each point
- Laplacian of a function is the trace of its Hessian
- Harmonic func means a function’s laplacian is 0 in every point ➔ some level of function stability / because curvature or hessian diag means on average how the neighbor points are higher than me or NOT

References

- ❑ http://www.cs.cmu.edu/~zkolter/course/linalg/index.html
- ❑Prof. James J. Cochran's tutorial slides "Matrix Algebra Primer II"
- ❑ http://www.cs.cmu.edu/~aarti/Class/10701/recitation/LinearAlgebra_Matlab_Rev iew.ppt
- ❑ Prof. Alexander Gray's slides
- ❑ Prof. George Bebis' slides
- ❑ Prof. Hal Dauḿe III' notes
- ❑ Khan Academy