Brief overview of linear algebra

Hamed Shahbazi, Fall 2018, CS-534

Notation

- Vector (Column Vector):

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- Vector Transpose (Row vector):

$$x^T = [x_1, x_2, \dots . x_n]$$

- Matrix $A_{m\times n}$:

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

Notation, Transpose

The ***transpose*** of a matrix results from "flipping" the rows and columns. Given a matrix $A \in \mathbb{R}^{m \times n}$, its transpose, written $A^T \in \mathbb{R}^{n \times m}$, is the $n \times m$ matrix whose entries are given by

$$(A^T)_{ij} = A_{ji}.$$

$$\begin{pmatrix} 5 & 4 & 3 \\ 4 & 0 & 4 \\ 7 & 10 & 3 \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} 5 & 4 & 7 \\ 4 & 0 & 10 \\ 3 & 4 & 3 \end{pmatrix}$$

Notation

- 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 $A_{i,:}$:

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

Definition, Identity Matrix

The ***identity matrix***, denoted $I \in \mathbb{R}^{n \times n}$, is a square matrix with ones on the diagonal and zeros everywhere else. That is,

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

It has the property that for all $A \in \mathbb{R}^{m \times n}$,

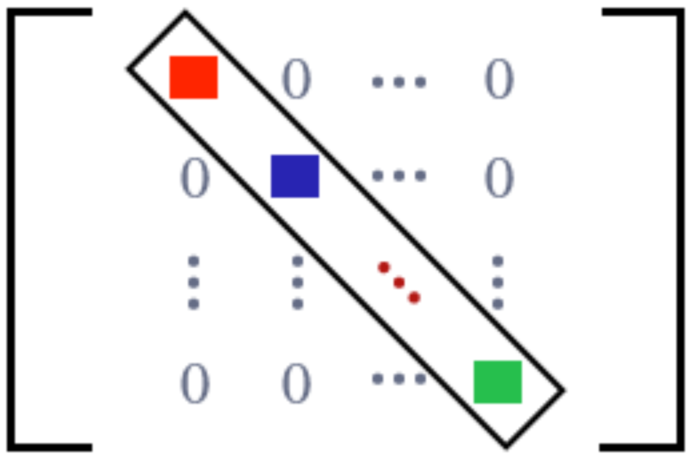
$$AI = A = IA.$$

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

Definition, Diagonal Matrix

A ***diagonal matrix*** is a matrix where all non-diagonal elements are 0. This is typically denoted $D = \text{diag}(d_1, d_2, \ldots, d_n)$, with

$$D_{ij} = \begin{cases} d_i & i = j \\ 0 & i \neq j \end{cases}$$



Clearly, $I = \text{diag}(1, 1, \ldots, 1)$.

Definition, Symmetric Matrix

A square matrix $A \in \mathbb{R}^{n \times n}$ is ***symmetric*** if $A = A^T$. It is ***anti-symmetric*** if $A = -A^T$.

$$\begin{bmatrix} 9 & 13 & 5 & 2 \\ 1 & 11 & 7 & 6 \\ 3 & 7 & 4 & 1 \\ 6 & 0 & 7 & 10 \end{bmatrix}$$

Not Symmetrical

$$\begin{bmatrix} 9 & 13 & 3 & 6 \\ 13 & 11 & 7 & 6 \\ 3 & 7 & 4 & 7 \\ 6 & 6 & 7 & 10 \end{bmatrix}$$

Symmetrical

for any matrix $A \in \mathbb{R}^{n \times n}$, the matrix $A + A^T$ is symmetric

matrix $A - A^T$ is anti-symmetric.

Definition, Trace

A square matrix $A \in \mathbb{R}^{n\times n}$

$$\mathrm{tr}A = \sum_{i=1}^{n} A_{ii}.$$

$$\begin{bmatrix} \boxed{3} & 8 & 5 \\ 6 & \boxed{-2} & 7 \\ 3 & 4 & \boxed{1} \end{bmatrix}$$

trace = 3+(-2)+1 = 2

Definition, Norm

- Eucledian Norm (L_2 norm)

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

- L_1 Norm

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

- Infinity Norm

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

- L_p Norm

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

- Frobenius norm (of a Matrix)

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2} = \sqrt{\mathrm{tr}(A^T A)}.$$

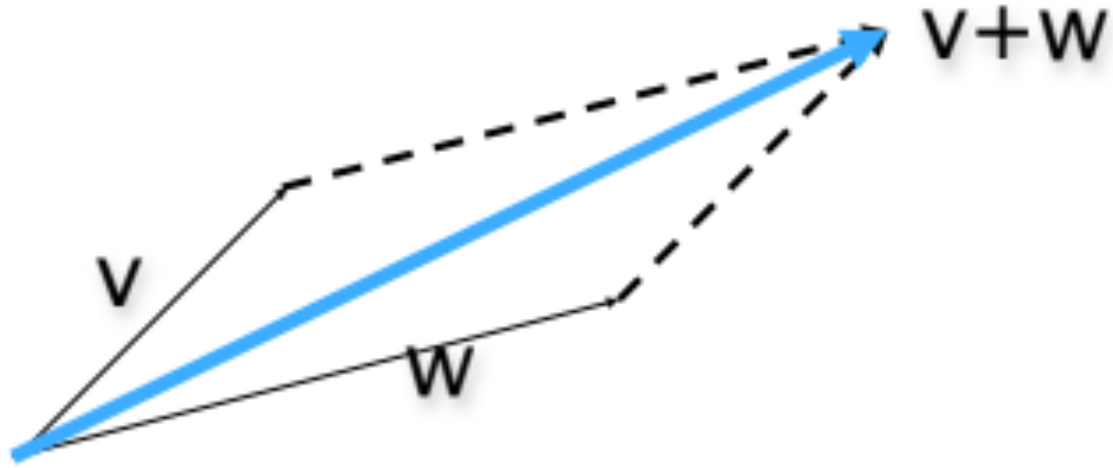
Norm, Continue

More formally, a norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies 4 properties:

1. For all $x \in \mathbb{R}^n$, $f(x) \geq 0$ (non-negativity).
2. $f(x) = 0$ if and only if $x = 0$ (definiteness).
3. For all $x \in \mathbb{R}^n$, $t \in \mathbb{R}$, $f(tx) = |t| f(x)$ (homogeneity).
4. For all $x, y \in \mathbb{R}^n$, $f(x + y) \leq f(x) + f(y)$ (triangle inequality).

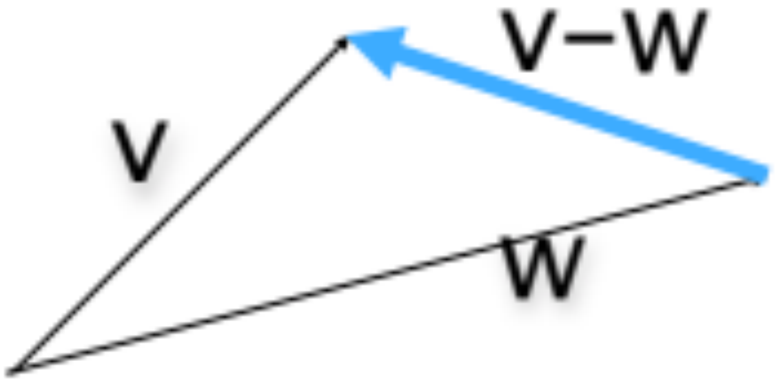
Operation, Vector Summation

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \end{bmatrix}$$



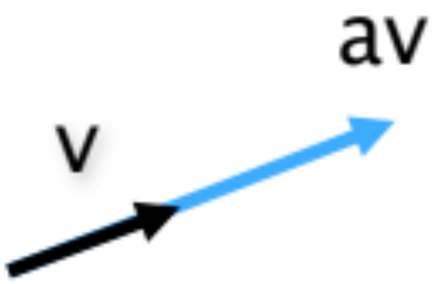
Operation, Vector Subtraction

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} - \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} v_1 - w_1 \\ v_2 - w_2 \end{bmatrix}$$



Operation, Vector Scaler Product

$$a\begin{bmatrix}v_1\\v_2\end{bmatrix}=\begin{bmatrix}av_1\\av_2\end{bmatrix}$$



Operation, Matrix Summation

Sum: $C_{n\times m} = A_{n\times m} + B_{n\times m}$ $\qquad c_{ij} = a_{ij} + b_{ij}$

A and B must have the same dimensions!

Example: $\begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 6 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 8 & 7 \\ 4 & 6 \end{bmatrix}$

Operation, Transpose

The ***transpose*** of a matrix results from "flipping" the rows and columns. Given a matrix $A \in \mathbb{R}^{m \times n}$, its transpose, written $A^T \in \mathbb{R}^{n \times m}$, is the $n \times m$ matrix whose entries are given by

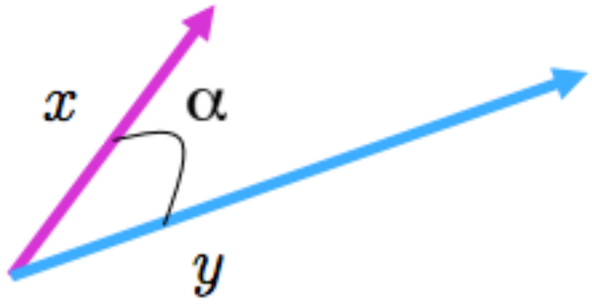
$$(A^T)_{ij} = A_{ji}.$$

$$\begin{pmatrix} 5 & 4 & 3 \\ 4 & 0 & 4 \\ 7 & 10 & 3 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 5 & 4 & 7 \\ 4 & 0 & 10 \\ 3 & 4 & 3 \end{pmatrix}$$

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

Operation, Inner Product (Dot Product)

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i.$$



$$\cos(\alpha) = \frac{x^T y}{\| x \| \| y \|}$$

The inner product is a SCALAR!

Orthogonal vectors x and y $\longrightarrow$ $x^T y = 0$

Operation, Outer Product

$$xy^T \in \mathbb{R}^{m\times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}.$$

Example:

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

Operation, Matrix-Vector Product

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

$$y_i = a_i^T x.$$

Operation, Matrix-Vector Product (linear combination)

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

Operation, Matrix-Vector Product (other ways)

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

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

Operation, Matrix-Matrix Product

The product of two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ is the matrix

$$C = AB \in \mathbb{R}^{m \times p},$$

where

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}.$$

Operation, Matrix-Matrix Product Ways

1)

$$C = AB = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix}$$

2)

$$C = AB = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} - & b_1^T & - \\ - & b_2^T & - \\ & \vdots & \\ - & b_n^T & - \end{bmatrix} = \sum_{i=1}^{n} a_i b_i^T$$

3)

$$C = AB = A \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab_1 & Ab_2 & \cdots & Ab_p \\ | & | & & | \end{bmatrix}$$

4)

$$C = AB = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} B = \begin{bmatrix} - & a_1^T B & - \\ - & a_2^T B & - \\ & \vdots & \\ - & a_m^T B & - \end{bmatrix}$$

Operation, Matrix-Matrix Product, Example

$$A = \begin{bmatrix} 4 & -1 & -2 & 5 \\ 0 & 1 & 2 & 3 \\ -3 & 6 & 7 & 8 \end{bmatrix}, B = \begin{bmatrix} 4 & -1 \\ 0 & 9 \\ -3 & 6 \\ 10 & 1 \end{bmatrix}$$

1)

$$C = AB = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix}$$

$$C = \begin{bmatrix} & & \\ & C_{2,2} & \\ & & \end{bmatrix}$$

$$C_{2,2} = a_2^T b_2 = [0, 1, 2, 3] \begin{bmatrix} -1 \\ 9 \\ 6 \\ 1 \end{bmatrix} = 24$$

$$= \begin{bmatrix} 4\times4 + -1\times0 + -2\times-3 + 5\times10 & 4\times-1 + -1\times9 + -2\times6 + 5\times1 \\ 0\times4 + 1\times0 + 2\times-3 + 3\times10 & 0\times-1 + 1\times9 + 2\times6 + 3\times1 \\ -3\times4 + 6\times0 + 7\times-3 + 8\times10 & -3\times-1 + 6\times9 + 7\times6 + 8\times1 \end{bmatrix}$$

$$C = \begin{bmatrix} 72 & -20 \\ 24 & 24 \\ 47 & 107 \end{bmatrix}$$

Operation, Matrix-Matrix Product, Example

$$A = \begin{bmatrix} 4 & -1 & -2 & 5 \\ 0 & 1 & 2 & 3 \\ -3 & 6 & 7 & 8 \end{bmatrix}, B = \begin{bmatrix} 4 & -1 \\ 0 & 9 \\ -3 & 6 \\ 10 & 1 \end{bmatrix}$$

2)

$$C = AB = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & b_1^T & — \\ — & b_2^T & — \\ & \vdots & \\ — & b_n^T & — \end{bmatrix} = \sum_{i=1}^{n} a_i b_i^T$$

$$C = a_1\, b_1{}^T + a_2\, b_2{}^T + a_3\, b_3{}^T + a_4\, b_4{}^T =$$

$$\begin{bmatrix} 4 \\ 0 \\ -3 \end{bmatrix} \begin{bmatrix} 4 & -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 6 \end{bmatrix} \begin{bmatrix} 0 & 9 \end{bmatrix} + \begin{bmatrix} -2 \\ 2 \\ 7 \end{bmatrix} \begin{bmatrix} -3 & 6 \end{bmatrix} + \begin{bmatrix} 5 \\ 3 \\ 8 \end{bmatrix} \begin{bmatrix} 10 & 1 \end{bmatrix}$$

Matrix-Matrix Product, Example

$$A = \begin{bmatrix} 4 & -1 & -2 & 5 \\ 0 & 1 & 2 & 3 \\ -3 & 6 & 7 & 8 \end{bmatrix}, B = \begin{bmatrix} 4 & -1 \\ 0 & 9 \\ -3 & 6 \\ 10 & 1 \end{bmatrix}$$

3)

$$C = AB = A \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab_1 & Ab_2 & \cdots & Ab_p \\ | & | & & | \end{bmatrix}$$

$$\downarrow$$

$$\left[\begin{bmatrix} 4 & -1 & -2 & 5 \\ 0 & 1 & 2 & 3 \\ -3 & 6 & 7 & 8 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ -3 \\ 10 \end{bmatrix}, \begin{bmatrix} 4 & -1 & -2 & 5 \\ 0 & 1 & 2 & 3 \\ -3 & 6 & 7 & 8 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ -3 \\ 10 \end{bmatrix} \right]$$

Matrix-Matrix Product, Example

$$A = \begin{bmatrix} 4 & -1 & -2 & 5 \\ 0 & 1 & 2 & 3 \\ -3 & 6 & 7 & 8 \end{bmatrix}, B = \begin{bmatrix} 4 & -1 \\ 0 & 9 \\ -3 & 6 \\ 10 & 1 \end{bmatrix}$$

4)

$$C = AB = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_m^T & - \end{bmatrix} B = \begin{bmatrix} - & a_1^T B & - \\ - & a_2^T B & - \\ & \vdots & \\ - & a_m^T B & - \end{bmatrix}$$

$$\begin{bmatrix} \begin{bmatrix} 4 & -1 & -2 & 5 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ 0 & 9 \\ -3 & 6 \\ 10 & 1 \end{bmatrix} \\ \begin{bmatrix} 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ 0 & 9 \\ -3 & 6 \\ 10 & 1 \end{bmatrix} \\ \begin{bmatrix} -3 & 6 & 7 & 8 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ 0 & 9 \\ -3 & 6 \\ 10 & 1 \end{bmatrix} \end{bmatrix}$$

Matrix-Matrix Product, Properties

- Matrix multiplication is associative: $(AB)C = A(BC)$.
- Matrix multiplication is distributive: $A(B + C) = AB + AC$.
- Matrix multiplication is, in general, *not* commutative; that is, it can be the case that $AB \neq BA$. (For example, if $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times q}$, the matrix product BA does not even exist if m and q are not equal!)

$$
\begin{aligned}
((AB)C)_{ij} &= \sum_{k=1}^{p} (AB)_{ik} C_{kj} = \sum_{k=1}^{p} \left(\sum_{l=1}^{n} A_{il} B_{lk} \right) C_{kj} \\
&= \sum_{k=1}^{p} \left(\sum_{l=1}^{n} A_{il} B_{lk} C_{kj} \right) = \sum_{l=1}^{n} \left(\sum_{k=1}^{p} A_{il} B_{lk} C_{kj} \right) \\
&= \sum_{l=1}^{n} A_{il} \left(\sum_{k=1}^{p} B_{lk} C_{kj} \right) = \sum_{l=1}^{n} A_{il} (BC)_{lj} = (A(BC))_{ij}.
\end{aligned}
$$

Matrix-Matrix Product, Application Example

- For $A \in \mathbb{R}^{n\times n}$, $\mathrm{tr}A = \mathrm{tr}A^T$.
- For $A, B \in \mathbb{R}^{n\times n}$, $\mathrm{tr}(A+B) = \mathrm{tr}A + \mathrm{tr}B$.
- For $A \in \mathbb{R}^{n\times n}$, $t \in \mathbb{R}$, $\mathrm{tr}(tA) = t\,\mathrm{tr}A$.
- For A, B such that AB is square, $\mathrm{tr}AB = \mathrm{tr}BA$.
- For A, B, C such that ABC is square, $\mathrm{tr}ABC = \mathrm{tr}BCA = \mathrm{tr}CAB$, and so on for the product of more matrices.

$$\begin{aligned}
\mathrm{tr}AB &= \sum_{i=1}^{m}(AB)_{ii} = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}B_{ji}\right) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} B_{ji}A_{ij} \\
&= \sum_{j=1}^{n}\left(\sum_{i=1}^{m} B_{ji}A_{ij}\right) = \sum_{j=1}^{n}(BA)_{jj} = \mathrm{tr}BA.
\end{aligned}$$

Definition, Linear Independence

A set of vectors $\{x_1, x_2, \ldots x_n\} \subset \mathbb{R}^m$ is said to be ***(linearly) independent*** if no vector can be represented as a linear combination of the remaining vectors. Conversely, if one vector belonging to the set *can* be represented as a linear combination of the remaining vectors, then the vectors are said to be ***(linearly) dependent***. That is, if

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

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

$$x_3 = -2x_1 + x_2.$$

Definition, Matrix Rank

The number of linearly independent columns (Rows) of A

- For $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) \leq \min(m, n)$. If $\text{rank}(A) = \min(m, n)$, then A is said to be ***full rank***.
- For $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = \text{rank}(A^T)$.
- For $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$.
- For $A, B \in \mathbb{R}^{m \times n}$, $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$.

Definition, Orthogonal Matrix

Two vectors $x, y \in \mathbb{R}^n$ are ***orthogonal*** if $x^T y = 0$. A vector $x \in \mathbb{R}^n$ is ***normalized*** if $\|x\|_2 = 1$. A square matrix $U \in \mathbb{R}^{n \times n}$ is ***orthogonal*** (note the different meanings when talking about vectors versus matrices) if all its columns are orthogonal to each other and are normalized (the columns are then referred to as being ***orthonormal***).

It follows immediately from the definition of orthogonality and normality that

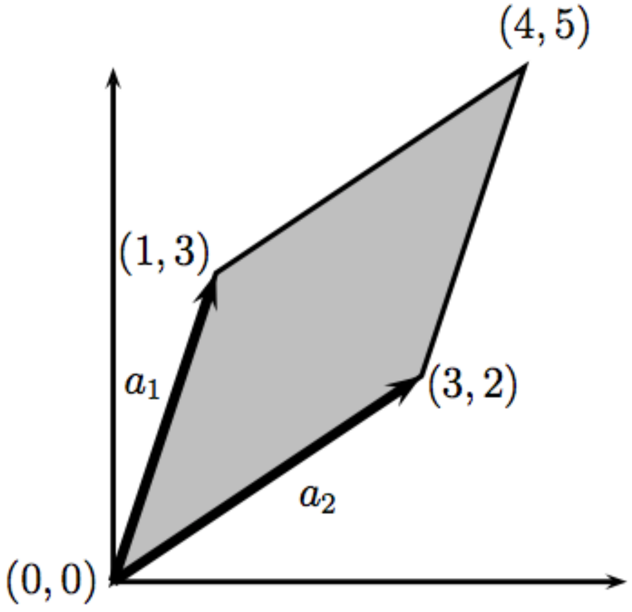
$$U^T U = I = U U^T.$$

The <u>inverse</u> of an orthogonal matrix is its <u>transpose</u>.

The Determinant

The ***determinant*** of a square matrix $A \in \mathbb{R}^{n\times n}$, is a function $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$, and is denoted $|A|$ or $\det A$ (like the trace operator, we usually omit parentheses).

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix} \qquad a_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad a_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$



|A| = Area given by $\alpha_1 a_1 + \alpha_2 a_2 \quad 0 \le \alpha_1 \le 1, \quad 0 \le \alpha_2 \le 1$

The Determinant, continue

1. The determinant of the identity is 1, $|I| = 1$. (Geometrically, the volume of a unit hypercube is 1).

2. Given a matrix $A \in \mathbb{R}^{n \times n}$, if we multiply a single row in A by a scalar $t \in \mathbb{R}$, then the determinant of the new matrix is $t|A|$,

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

   (Geometrically, multiplying one of the sides of the set S by a factor t causes the volume to increase by a factor t.)

3. If we exchange any two rows a_i^T and a_j^T of A, then the determinant of the new matrix is $-|A|$, for example

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

The Determinant, continue

- For $A \in \mathbb{R}^{n \times n}$, $|A| = |A^T|$.
- For $A, B \in \mathbb{R}^{n \times n}$, $|AB| = |A||B|$.
- For $A \in \mathbb{R}^{n \times n}$, $|A| = 0$ if and only if A is singular (i.e., non-invertible). (If A is singular then it does not have full rank, and hence its columns are linearly dependent. In this case, the set S corresponds to a "flat sheet" within the n-dimensional space and hence has zero volume.)
- For $A \in \mathbb{R}^{n \times n}$ and A non-singular, $|A^{-1}| = 1/|A|$.

$$
\begin{aligned}
|A| &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{\setminus i, \setminus j}| \quad \text{(for any } j \in 1, \ldots, n) \\
&= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{\setminus i, \setminus j}| \quad \text{(for any } i \in 1, \ldots, n)
\end{aligned}
$$

The Determinant, Example

$$\begin{aligned}
|[a_{11}]| &= a_{11} \\
\left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| &= a_{11}a_{22} - a_{12}a_{21} \\
\left|\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right| &= \begin{array}{l} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ \quad -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{array}
\end{aligned}$$

Definition, Matrix Inverse

The ***inverse*** of a square matrix $A \in \mathbb{R}^{n\times n}$ is denoted A^{-1}, and is the unique matrix such that

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

- Non-square matrices, do not have inverses by definition.
- For some square matrices A, inverse may not exist.
- A is invertible or non-singular if A^{-1} exists and non-invertible or singular otherwise

The following are properties of the inverse; all assume that $A, B \in \mathbb{R}^{n\times n}$ are non-singular:

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^{-1})^T = (A^T)^{-1}$. For this reason this matrix is often denoted A^{-T}.

Definition, Matrix Inverse

$$A^{-1} = \frac{1}{|A|}\mathrm{adj}(A)$$

$$\mathrm{adj}(A) \in \mathbb{R}^{n\times n}, \quad (\mathrm{adj}(A))_{ij} = (-1)^{i+j}|A_{\setminus j,\setminus i}|$$

classical adjoint

Example: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^{-1} = \frac{1}{-2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix}$

Eigenvectors

Eigenvector and Eigenvalue

A eigenvalue λ and eigenvector $\mathbf{u}$ satisfies

$$\mathbf{Au} = \lambda \mathbf{u}$$

where $\mathbf{A}$ is a square matrix.

- Multiplying $\mathbf{u}$ by $\mathbf{A}$ scales $\mathbf{u}$ by λ

Eigenvector and Eigenvalue, Continue

Rearranging the previous equation gives the system

$$\mathbf{Au} - \lambda\mathbf{u} = (\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = 0$$

which has a solution if and only if $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

- The eigenvalues are the roots of this determinant which is polynomial in λ.
- Substitute the resulting eigenvalues back into $\mathbf{Au} = \lambda\mathbf{u}$ and solve to obtain the corresponding eigenvector.

Eigenvector and Eigenvalue, Example

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$$

$$|\mathbf{A} - \lambda \cdot \mathbf{I}| = \left| \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0$$

$$\left| \begin{bmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{bmatrix} \right| = \lambda^2 + 3\lambda + 2 = 0$$

characteristic polynomial

$\lambda_1=-1$, $\lambda_2=-2$

$$\mathbf{A} \cdot \mathbf{v}_1 = \lambda_1 \cdot \mathbf{v}_1$$

$$(\mathbf{A} - \lambda_1) \cdot \mathbf{v}_1 = 0$$

$$\begin{bmatrix} -\lambda_1 & 1 \\ -2 & -3-\lambda_1 \end{bmatrix} \cdot \mathbf{v}_1 = 0$$

$$\begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \cdot \mathbf{v}_1 = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \cdot \begin{bmatrix} v_{1,1} \\ v_{1,2} \end{bmatrix} = 0$$

$v_{1,1} + v_{1,2} = 0$, so

$v_{1,1} = -v_{1,2}$

$-2 \cdot v_{1,1} + -2 \cdot v_{1,2} = 0$, so again

$v_{1,1} = -v_{1,2}$

$$\mathbf{v}_1 = k_1 \begin{bmatrix} +1 \\ -1 \end{bmatrix}$$

$$\mathbf{A} \cdot \mathbf{v}_2 = \lambda_2 \cdot \mathbf{v}_2$$

$$(\mathbf{A} - \lambda_2) \cdot \mathbf{v}_2 = \begin{bmatrix} -\lambda_2 & 1 \\ -2 & -3-\lambda_2 \end{bmatrix} \cdot \mathbf{v}_2 = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_{2,1} \\ v_{2,2} \end{bmatrix} = 0 \quad \text{so}$$

$2 \cdot v_{2,1} + 1 \cdot v_{2,2} = 0$ (or from bottom line: $-2 \cdot v_{2,1} - 1 \cdot v_{2,2} = 0$)

$2 \cdot v_{2,1} = -v_{2,2}$

$$\mathbf{v}_2 = k_2 \begin{bmatrix} +1 \\ -2 \end{bmatrix}$$

Eigenvector and Eigenvalue

- The trace of a A is equal to the sum of its eigenvalues,

$$\mathrm{tr} A = \sum_{i=1}^{n} \lambda_i.$$

- The determinant of A is equal to the product of its eigenvalues,

$$|A| = \prod_{i=1}^{n} \lambda_i.$$

- The rank of A is equal to the number of non-zero eigenvalues of A.
- If A is non-singular then $1/\lambda_i$ is an eigenvalue of A^{-1} with associated eigenvector x_i, i.e., $A^{-1}x_i = (1/\lambda_i)x_i$. (To prove this, take the eigenvector equation, $Ax_i = \lambda_i x_i$ and left-multiply each side by A^{-1}.)
- The eigenvalues of a diagonal matrix $D = \mathrm{diag}(d_1, \ldots d_n)$ are just the diagonal entries $d_1, \ldots d_n$.

Eigenvector and Eigenvalue, continue

We can write all the eigenvector equations simultaneously as

$$AX = X\Lambda$$

where the columns of $X \in \mathbb{R}^{n\times n}$ are the eigenvectors of A and Λ is a diagonal matrix whose entries are the eigenvalues of A, i.e.,

$$X \in \mathbb{R}^{n\times n} = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix}, \quad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

Matrix Calculus

Derivatives

$$\frac{\partial f}{\partial x}(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

This essentially says that the derivative of f with respect to x, evaluated at a point x_0, is the rate of change of f at x_0. It is fairly common to see $\partial f / \partial x$ denoted by f'.

- **Constant Rule:** $f(x) = c$ then $f'(x) = 0$
- **Constant Multiple Rule:** $g(x) = c \cdot f(x)$ then $g'(x) = c \cdot f'(x)$
- **Power Rule:** $f(x) = x^n$ then $f'(x) = nx^{n-1}$
- **Sum and Difference Rule:** $h(x) = f(x) \pm g(x)$ then $h'(x) = f'(x) \pm g'(x)$
- **Product Rule:** $h(x) = f(x)g(x)$ then $h'(x) = f'(x)g(x) + f(x)g'(x)$
- **Quotient Rule:** $h(x) = \frac{f(x)}{g(x)}$ then $h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$
- **Chain Rule:** $h(x) = f(g(x))$ then $h'(x) = f'(g(x))g'(x)$

Derivatives

- **Exponential Derivatives**
  - $f(x) = a^x$ then $f'(x) = \ln(a)a^x$
  - $f(x) = e^x$ then $f'(x) = e^x$
  - $f(x) = a^{g(x)}$ then $f'(x) = \ln(a)a^{g(x)}g'(x)$
  - $f(x) = e^{g(x)}$ then $f'(x) = e^{g(x)}g'(x)$

- **Logarithm Derivatives**
  - $f(x) = \log_a(x)$ then $f'(x) = \dfrac{1}{\ln(a)x}$
  - $f(x) = \ln(x)$ then $f'(x) = \dfrac{1}{x}$
  - $f(x) = \log_a(g(x))$ then $f'(x) = \dfrac{g'(x)}{\ln(a)g(x)}$
  - $f(x) = \ln(g(x))$ then $f'(x) = \dfrac{g'(x)}{g(x)}$

Derivatives, Example

$$
\begin{aligned}
\partial_m J(m,b) &= \partial_m \left(\sum_{n=1}^{N} \left[(mx_n + b) - y_n \right]^2 \right) \\
&= \sum_{n=1}^{N} \partial_m \left[(mx_n + b) - y_n \right]^2 \\
&= \sum_{n=1}^{N} \left[2 \left[(mx_n + b) - y_n \right] \right] \partial_m \left[(mx_n + b) - y_n \right] \\
&= \sum_{n=1}^{N} \left[2 \left[(mx_n + b) - y_n \right] \right] x_n
\end{aligned}
$$

Matrix Calculus, The Gradient

Suppose that $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ 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

partial derivatives, defined as:

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

i.e., an $m \times n$ matrix with

$$(\nabla_A f(A))_{ij} = \frac{\partial f(A)}{\partial A_{ij}}.$$

a vector $x \in \mathbb{R}^n$

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

Gradient, Example

For $x \in \mathbb{R}^n$, let $f(x) = b^T x$ for some known vector $b \in \mathbb{R}^n$. Then

$$f(x) = \sum_{i=1}^{n} b_i x_i$$

so

$$\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^{n} b_i x_i = b_k.$$

$$\nabla_x b^T x = b.$$

Gradient, Example

$$f(x) = x^T A x \text{ for } A \in \mathbb{S}^n.$$

$$\begin{aligned}
\frac{\partial f(x)}{\partial x_k} &= \frac{\partial}{\partial x_k} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j \\
&= \frac{\partial}{\partial x_k} \left[\sum_{i \neq k} \sum_{j \neq k} A_{ij} x_i x_j + \sum_{i \neq k} A_{ik} x_i x_k + \sum_{j \neq k} A_{kj} x_k x_j + A_{kk} x_k^2 \right] \\
&= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2 A_{kk} x_k \\
&= \sum_{i=1}^{n} A_{ik} x_i + \sum_{j=1}^{n} A_{kj} x_j = 2 \sum_{i=1}^{n} A_{ki} x_i,
\end{aligned}$$

$$\downarrow$$

$$\nabla_x x^T A x = 2Ax.$$

Gradient, Example

$$f(z) = z^T z, \text{ such that } \nabla_z f(z) = 2z.$$

Gradient, Example

$$
\begin{aligned}
\|Ax - b\|_2^2 &= (Ax-b)^T(Ax-b) \\
&= x^TA^TAx - 2b^TAx + b^Tb
\end{aligned}
$$

Taking the gradient with respect to x we have, and using the properties we derived in the previous section

$$
\begin{aligned}
\nabla_x(x^TA^TAx - 2b^TAx + b^Tb) &= \nabla_x x^TA^TAx - \nabla_x 2b^TAx + \nabla_x b^Tb \\
&= 2A^TAx - 2A^Tb
\end{aligned}
$$

Setting this last expression equal to zero and solving for x gives the normal equations

$$x = (A^TA)^{-1}A^Tb$$