# Linear Algebra Review

Sanjay Singh[⋆†]

[⋆]Department of Information and Communication Technology
Manipal Institute of Technology, Manipal University
Karnataka-576104, INDIA
sanjay.singh@manipal.edu

[†]Centre for Artificial and Machine Intelligence (CAMI)
Manipal University, Karnataka-576104, INDIA

January 28, 2017



## Notation: Sets and Graphs

- $\mathbb{A}$: A set
- $\mathbb{R}$: Set of real numbers
- $\{0, 1, 2, \ldots, n\}$: Set of all integers between 0 and $n$
- $[a, b]$: Real intervals including $a$ and $b$
- $(a, b]$: Interval excluding $a$ but including $b$
- $\mathbb{A} \backslash \mathbb{B}$: Set subtraction
- $\mathcal{G}$: A graph
- $Pa_{\mathcal{G}}(x_i)$: Parent of $x_i$ in $G$

## Indexing Notation

- $a_i$: Element $i$ of vector $\boldsymbol{a}$, with indexing starting at 1
- $a_{-i}$: All elements of $\boldsymbol{a}$ except for element $i$
- $\boldsymbol{A}_{i,:}$: Row $i$ of matrix $\boldsymbol{A}$
- $\boldsymbol{A}_{:,j}$: Column $j$ of matrix $\boldsymbol{A}$
- $\boldsymbol{A}_{i,j,k}$: Element $(i, j, k)$ of a 3-D tensor $A$
- $\boldsymbol{A}_{:,:,k}$: 2-D slice of a 3-D tensor



## Linear Algebra Operations

- $I_n$: Identity matrix with $n$ rows and $n$ columns
- $A^T$: Transpose of matrix $A$
- $A^+$: Moore-Pensrose pseudoinverse of $A$
- $A \odot B$: Element-wise (Hadamard) product of $A$ and $B$
- $det(A)$: Determinant of $A$
- $x \in \mathbb{R}^n$: A vector with $n$ entries
- $A \in \mathbb{R}^{m \times n}$: A matrix with $m$ rows and $n$ columns, entries of $A$ are real numbers

# Scalars, Vectors, Matrices and Tensors

Study of linear algebra involves following types of mathematical objects

- Scalars- just a single number
- Vectors-it is an array of numbers
- Matrices-a 2-D array of numbers
- Tensors-are useful for representing higher order relations, for example when we need array with more than two axes



- 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

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

- Given two vectors $x, y \in \mathbb{R}^n$, quantity $x^T y$ called **inner product** or dot product of vectors, is a real number given by

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

- Note that, $x^T y = y^T x$
- Given vectors $x \in \mathbb{R}^m, y \in \mathbb{R}^n$ (not of same size), $xy^T \in \mathbb{R}^{m \times n}$ is called the **outer product**

$$xy^T \in \mathbb{R}^{m \times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \ldots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \ldots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \ldots & x_m y_n \end{bmatrix}$$

Sanjay Singh Linear Algebra Review

# Usage of Outer Product

- Let $\mathbf{1} \in \mathbb{R}^n$, and $A \in \mathbb{R}^{m \times n}$ whose columns are all equal to some vector $x \in \mathbb{R}^m$
- Using outer product, we can represent $A$ compactly as,

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



# Matrix-Vector Products

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

- $y = Ax = \begin{bmatrix} - & a_1^T & - \\ - & a_2^T & - \\ & \vdots & \\ - & a_1^T & - \end{bmatrix} x = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x. \end{bmatrix}$
- $y = Ax = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} =$
  $[a_1]x_1 + [a_2]x_2 + \dots + [a_n]x_n.$
- $y^T = x^T A = x^T \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x^T a_1 & x^T a_2 & \dots & x^T a_n \end{bmatrix}$

- Identity matrix $I \in \mathbb{R}^{n \times n}$, defined as

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

- For all $A \in \mathbb{R}^{m \times n}$, $AI = A = IA$
- Diagonal matrix is a matrix where all non-diagonal elements are 0, and denoted $D = diag(d_1, d_2, \ldots, d_n)$ with

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

- $I = diag(1, 1, \ldots, 1)$
- $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$, and it is **anti-symmetric** if $A = -A^T$
- $A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$
- $A \in \mathbb{S}^n$ means that $A$ is a symmetric $n \times n$ matrix



# Trace

**Trace** of a square matrix $A \in \mathbb{R}^{n \times n}$, denoted by tr(A) (or trA) is the sum of diagonal elements in the matrix

- $trA = \sum_{i=1}^{n} A_{ii}$
- For $A, B, C \in \mathbb{R}^{n \times n}$, it has following properties
  - $trA = trA^T$
  - $tr(A + B) = trA + trB$
  - $\alpha \in \mathbb{R}, tr(\alpha A) = \alpha trA$
  - For $A, B$ such that $AB$ is square $trAB = trBA$
  - For $A, B, C$ such that $ABC$ is square $trABC = trBCA = trCAB$

# Norms

A **norm** of a vector $||x||$ is a measure of the "length" of the vectors

- Commonly used Euclidean or $L_2$ norm is defined as

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

- Note that $||x||_2^2 = x^T x$
- A norm is any function $f : \mathbb{R}^n \to \mathbb{R}$ that satisfies the following properties
  - For all $x \in \mathbb{R}^n, f(x) \geq 0$ (non-negative)
  - $f(x) = 0$ iff $x = 0$ (definiteness)
  - For all $x \in \mathbb{R}^n, \alpha \in \mathbb{R}, f(\alpha x) = |\alpha| f(x)$ (homogeneity)
  - For all $x, y \in \mathbb{R}^n, f(x+y) \leq f(x) + f(y)$ (triangle inequality)



- $L_1$ and $L_\infty$ norm is defined as

$$||x||_1 = \sum_{i=1}^{n} |x_i|, \quad ||x||_\infty = max_i |x_i|$$

- In general, $L_p$ norm is defined as

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

- Norm of a matrix, such as Frobenius norm is defined as

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

## Linear Independence and Rank

- Set of vectors $\{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^m$ is said to be (linearly) independent if no vectors can be presented as a linear combination of remaining vectors
- Conversely, they are called as dependent, i,e $x_n = \sum_{i=1}^{n-1} \alpha_i x_i$, for some $\alpha_i \in \mathbb{R}$
- Column rank of a matrix $A \in \mathbb{R}^{m \times n}$ is the number of independent columns of $A$
- Row rank is the largest number of rows of $A$ that constitute a linearly independent set
- For any matrix $A \in \mathbb{R}^{m \times n}$ the column rank is equal to the row rank, and collectively called as rank of $A$, denoted by rank(A)
- Basic properties of rank:
  - For $A \in \mathbb{R}^{m \times n}$, $rank(A) \leq min(m, n)$. If $rank(A) = min(m, n)$, then $A$ is said to be full rank
  - For $A \in \mathbb{R}^{m \times n}$, $rank(A) = rank(A^T)$
  - For $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, $rank(AB) \leq min(rank(A), rank(B))$
  - For $A, B \in \mathbb{R}^{m \times n}$, $rank(A + B) \leq rank(A) + rank(B)$



## Inverse

- Inverse of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted by $A^{-1}$, $A^{-1}A = I = AA^{-1}$
- $A$ is invertible iff it is a non-singular i.e $det(A) \neq 0$
- $A$ to be invertible, it must be full rank
- Properties of inverse, all assume that $A, B \in \mathbb{R}^{n \times n}$
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^T = (A^T)^{-1}$

# Orthogonal Matrices

- 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 if all its columns are orthogonal to each other and are normalized (the columns are then referred to as being **orthonormal**)
- From orthogonality and normality, $U^T U = I = UU^T$
- If $U$ is not square but its columns are still orthonormal, then $U^T U = I \neq UU^T$
- Operating on a vector with an orthogonal matrix will not change its Euclidean norm i.e., $\|Ux\|_2 = \|x\|_2$, for any $x \in \mathbb{R}^n, U \in \mathbb{R}^{n \times n}$



# Range and Nullspace of a Matrix

- **Span** of a set of vectors $\{x_1, x_2, \ldots, x_n\}$ is the set of all vectors that can be expressed as a linear combination of $\{x_1, x_2, \ldots, x_n\}$ i.e.,

$$span(\{x_1, x_2, \ldots, x_n\}) = \left\{ v : v = \sum_{i=1}^{n} \alpha_i x_i, \quad \alpha_i \in \mathbb{R} \right\}$$

- **Projection** of a vector $y \in \mathbb{R}^m$ onto span of $\{x_1, x_2, \ldots, x_n\}$ is the vector $v \in span(\{x_1, x_2, \ldots, x_n\})$, such that $v$ is as close to $y$ as measured by the Euclidean norm $\|v - y\|_2$
- Projection is denoted as

$$Proj(y; \{x_1, x_2, \ldots, x_n\}) = argmin_{v \in span(\{x_1, x_2, \ldots, x_n\})} \|y - v\|_2$$

- **Range** of a matrix $A \in \mathbb{R}^{m \times n}$, denoted by $\mathcal{R}(A)$ is the span of columns of $A$

$$\mathcal{R}(A) = \{v \in \mathbb{R}^m : v = Ax, x \in \mathbb{R}^n\}$$

- **Nullspace** of a matrix $A \in \mathbb{R}^{m \times n}$, denoted as $\mathcal{N}(A)$ is the set of all vectors that equal 0 when multiplied by $A$ i.e.,

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$$

- Vectors in $\mathcal{R}(A)$ are of size $m$, while vectors in $\mathcal{N}(A)$ are of size $n$
- Vectors in $\mathcal{R}(A^T)$ and $\mathcal{N}(A)$ are both in $\mathbb{R}^n$ i.e

$$\{w : w = u+v, u \in \mathcal{R}(A^T), v \in \mathcal{N}(A)\} = \mathbb{R}^n, \quad \mathcal{R}(A^T) \cap \mathcal{N}(A) = \{0\}$$

- $\mathcal{R}(A^T)$ and $\mathcal{N}(A)$ are disjoint subsets that together span the entire space of $\mathbb{R}^n$
- Sets of this type are called orthogonal complements, and denoted by $\mathcal{R}(A^T) = \mathcal{N}(A)^\perp$



## Determinant

- Determinant of $A \in \mathbb{R}^{n \times n}$ is a function $det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ and is denoted by $|A|$ or det A
- Determinant satisfies the following properties
  - Determinant of the identity is 1, $|I| = 1$
  - Given a matrix $A \in \mathbb{R}^{n \times n}$, if we multiply a single row in $A$ by a scalar $\alpha \in \mathbb{R}$, then $\alpha|A|$
  - If we exchange any two rows of A, then the determinant of the new matrix is $-|A|$
- For $A \in \mathbb{R}^{n \times n}$, $|A| = |A^T|$
- For $A, B \in \mathbb{R}^{n \times n}$, $|AB| = |A||B|$
- For $A \in \mathbb{R}^{n \times n}$, $|A| = 0$ iff A is singular
- For $A \in \mathbb{R}^{n \times n}$ and $A$ non-singular, $|A^{-1}| = 1/|A|$
- $|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{\setminus i, \setminus j}|$, where $A_{\setminus i, \setminus j}$ is the matrix that results from deleting ith row and jth column from $A$

## Quadratic Forms and Positive Semidefinitve Matrices

- Given $A \in \mathbb{R}^{n\times n}$ and a vector $x \in \mathbb{R}^n$, the scalar value $x^T Ax$ is called a **quadratic form**
- $x^T Ax = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \left( \sum_{j=1}^{n} A_{ij} x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$
- Some definitions
  - A symmetric matrix $A \in \mathbb{S}^n$ is **positive definite** (PD) if for all non-zero vectors $x \in \mathbb{R}^n, x^T Ax > 0$
  - A symmetric matrix $A \in \mathbb{S}^n$ is **positive semidefinite** (PSD) if for all vectors $x \in \mathbb{R}^n, x^T Ax \geq 0$
  - A symmetric matrix $A \in \mathbb{S}^n$ is **negative definite** (ND) if for all vectors $x \in \mathbb{R}^n, x^T Ax < 0$
  - A symmetric matrix $A \in \mathbb{S}^n$ is **negative semidefinite** (NSD) if for all vectors $x \in \mathbb{R}^n, x^T Ax \leq 0$
  - Finally, a symmetric matrix $A \in \mathbb{S}^n$ is **indefinite** if it is neither PSD nor NSD.
  - Given $A \in \mathbb{R}^{n\times n}$, the matrix $G = A^T A$ (also known as **Gram matrix**) is always positive semidefinite



## Eigenvalues and Eigenvectors

- Given $A \in \mathbb{R}^{n\times n}$, we say that $\lambda \in \mathbb{C}$ is an eigenvalue of $A$ and $x \in \mathbb{C}^n$ is the corresponding eigenvector if

$$Ax = \lambda x, \quad x \neq 0$$

- To obtain eigenvector corresponding to the eigenvalue $\lambda_i$, we solve linear equation $(\lambda_i I - A)x = 0$
- Properties of eigenvalues and eigenvectors
  - $trA = \sum_{i=1}^{n} \lambda_i$
  - $|A| = \prod_{i=1}^{n} \lambda_i$
  - Rank of $A$ is equal to the number of non-zero eigenvalues of $A$
  - For $|A| \neq 0$, $1/\lambda_i$ is an eigenvalue of $A^{-1}$ with associated eigenvector $x_i$, i.e $A^{-1}x_i = (1/\lambda_i)x_i$
  - Eigenvlaues of a diagonal matrix $D = diag(d_1, \ldots, d_n)$ are the diagonal entries $d_1, \ldots, d_n$

- We can write eigenvector equations simultaneously as

$$AX = X\Lambda$$

, where

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

- If the eigenvectors are linearly independent, then $X$ will be invertible so $A = X\Lambda X^{-1}$
- A matrix that can be written in this form is called **diagonalizable**



## Matrix Calculus

- Matrix calculus???
- It is the extension of calculus to the vector setting
- Actual calculus is relatively trivial, however, notations are bit different
- We'll particularly focus on
  - Gradient
  - Hessian
  - Gradient and Hessian of quadratic and linear functions
  - Least Squares
  - Gradient of determinant

- Suppose $F : \mathbb{R}^{m\times n} \to \mathbb{R}$, then the gradient of $f$ (wrt $A \in \mathbb{R}^{m\times n}$) is a matrix of partial derivatives, defined as

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

- An $m \times n$ matrix with

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

- If $A$ is just 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 of a function is only defined if the function is real-valued, i.e., it returns a scalar value
- We cannot take the gradient of $Ax, A \in \mathbb{R}^{n\times n}$ wrt $x$, as it is vector-valued
- From the properties of partial derivatives
  - $\nabla_x(f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x)$
  - For $\alpha \in \mathbb{R}, \nabla_x(\alpha f(x)) = \alpha \nabla_x f(x)$
- Gradients are natural extension of partial derivative for multi-variate functions
- Working with gradients can be tricky due to notational reasons

**Example**

Let $f : \mathbb{R}^m \mapsto \mathbb{R}$ be the function defined by

$$f(z) = z^T z$$

then $\nabla_z f(z) = 2z$.

## Hessian

- Suppose that $f : \mathbb{R}^n \mapsto \mathbb{R}$
- Hessian matrix wrt $x$, written as $\nabla_x^2 f(x)$ is given by

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

- $\nabla_x^2 f(x) \in \mathbb{R}^{n\times n}$, with

$$(\nabla_x^2 f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

- Hessian is always symmetric, since $\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}$
- Gradient is the analogue of first derivative for function
- Hessian as the analogue of second derivative, but with few caveats



## Gradient and Hessian of Quadratic and Linear Function

- For some $x \in \mathbb{R}^n$, let $f(x) = b^T x$ for some $b \in \mathbb{R}^n$, then

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

so

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

- We can say that $\nabla_x b^T x = b$

- Consider a quadratic function $f(x) = x^T Ax$ for $A \in \mathbb{S}^n$, which can be written as

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

- To take partial derivative, we'll consider term including $x_k$ and $x_k^2$ factors separately

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



- The $k$th entry of $\nabla_x f(x)$ is the inner product of the $k$th row of $A$ and $x$ $a/b$
- $\nabla_x x^T Ax = 2Ax$ (think of $\partial/(\partial x) ax^2 = 2ax$)
- Hessian of the quadratic function $f(x) = x^T Ax$

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{\partial}{\partial x_k} \left[ \frac{\partial f(x)}{\partial x_l} \right] = \frac{\partial}{\partial x_k} \left[ 2 \sum_{i=1}^{n} A_{li} x_i \right] = 2A_{lk} = 2A_{kl}$$

- So, $\nabla_x^2 x^T Ax = 2A$
- Recap
  - $\nabla_x b^T x = b$
  - $\nabla_x x^T Ax = 2Ax$ (if $A$ symmetric)
  - $\nabla_x^2 x^T Ax = 2A$ (if $A$ symmetric)

- Consider $A \in \mathbb{R}^{n\times n}$, we want to find $\nabla_A |A|$, since

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} |A_{\backslash i,\backslash j}|$$

so

$$\frac{\partial}{\partial A_{kl}} |A| = \frac{\partial}{\partial A_{kl}} \sum_{i=1}^{n} (-1)^{i+j} A_{ij} |A_{\backslash i,\backslash j}| = (-1)^{k+l} |A_{\backslash k,\backslash l}| = (adj(A))_{lk}$$

- $\nabla_A |A| = (adj(A))^T = |A| A^{-T}$



- Consider a function $f : \mathbb{S}^n_{++} \mapsto \mathbb{R}$, $f(A) = \log |A|$
- $\dfrac{\partial \log |A|}{\partial A_{ij}} = \dfrac{\partial \log |A|}{\partial |A|} \dfrac{\partial |A|}{\partial A_{ij}} = \dfrac{1}{|A|} \dfrac{\partial |A|}{\partial A_{ij}}$
- Now it is obvious that

$$\nabla_A \log |A| = \frac{1}{|A|} \nabla_A |A| = A^{-1}$$