CSCI 567 Discussion
Linear Algebra

(Slides adapted from Bhavya Vasudeva and Sampad Mohanty's slides for CSCI567 in 2022 Fall)

Mengxiao Zhang

2024.1.12

Outline

- Basic Concepts and Notation
- Matrix Multiplications
- Operations and Properties
- Matrix Calculus

Basic Concepts and Notation

Basic Notation

- By $x \in \mathbb{R}^n$, we denote a vector with n entries.

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

- By $A \in \mathbb{R}^{m \times n}$, we denote a matrix with m rows and n columns.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \text{---} & a_1^T & \text{---} \\ \text{---} & a_2^T & \text{---} \\ & \vdots & \\ \text{---} & a_m^T & \text{---} \end{bmatrix}$$

Special Matrices

Identity matrix

$I_n \in \mathbb{R}^{n \times n}$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

Property: for all $A \in \mathbb{R}^{m \times n}$, $AI_n = A = I_m A$

Diagonal matrix

$D = diag(d_1, \dots, d_n)$

$$\begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & d_n \end{bmatrix}$$

Clearly, $I =$
$diag(1,1, \dots, 1)$

Matrix Multiplication

Vector-Vector Product

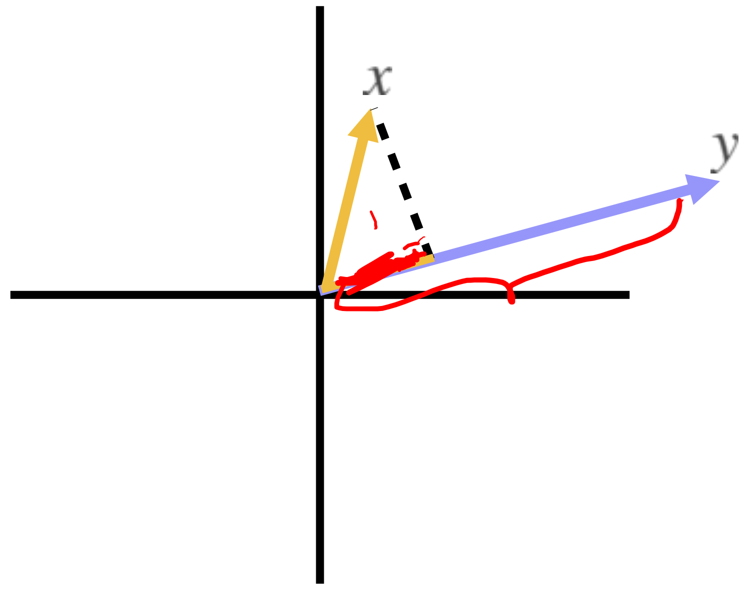
Inner Product / Dot Product

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

Intuitio

n

$x^T y =$ (Length of projected x)·(Length of y)



Vector-Vector Product

Outer Product

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

$$\begin{bmatrix} y_1 \begin{matrix} x_1 \\ \vdots \\ x_m \end{matrix} & y_2 \begin{matrix} x_1 \\ \vdots \\ x_m \end{matrix} & \begin{matrix} \cdots \\ \ddots \\ \cdots \end{matrix} & y_n \begin{matrix} x_1 \\ \vdots \\ x_m \end{matrix} \end{bmatrix} \qquad \begin{bmatrix} x_1 & (\cdots & y^T & \cdots) \\ x_2 & (\cdots & y^T & \cdots) \\ \vdots & \vdots & \vdots & \vdots \\ x_m & (\cdots & y^T & \cdots) \end{bmatrix}$$

Matrix-Vector Product

View 1: Write A by rows

$\mathbb{R}^{m \times n}$ ($\cancel{n \times m}$) $\quad \mathbb{R}^{n \times 1}$

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

$a_i^T \cdot x$

Set of inner products with each row vector

Matrix-Vector Product

View 2: Write A by columns

$$y = Ax = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} | \\ a^1 \\ | \end{bmatrix} x_1 + \begin{bmatrix} | \\ a^2 \\ | \end{bmatrix} x_2 + \ldots + \begin{bmatrix} | \\ a^n \\ | \end{bmatrix} x_n .$$

Linear combination of column vectors

Vector-Matrix Product

View 1: Write A by columns

$x \in \mathbb{R}^{m \times 1}$ $\quad \mathbb{R}^{m \times n}$ $\quad x^T a_i$

$$\underbrace{y^T}_{1 \times n} = \underbrace{x^T A}_{1 \times m \quad m \times n} = x^T \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x^T a^1 & x^T a^2 & \cdots & x^T a^n \end{bmatrix}$$

Set of inner products with each column vector

Vector-Matrix Product

View 2: Write A by rows

$$y^T = x^T A = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} \text{---} & a_1^T & \text{---} \\ \text{---} & a_2^T & \text{---} \\ & \vdots & \\ \text{---} & a_m^T & \text{---} \end{bmatrix}$$

$$= x_1 \begin{bmatrix} \text{---} & a_1^T & \text{---} \end{bmatrix} + x_2 \begin{bmatrix} \text{---} & a_2^T & \text{---} \end{bmatrix} + \ldots + x_m \begin{bmatrix} \text{---} & a_m^T & \text{---} \end{bmatrix}$$

Linear combination of row vectors

Matrix-Matrix Multiplication

View 1: Set of inner products

$$C = AB = \begin{bmatrix} \text{---} & a_1^T & \text{---} \\ \text{---} & a_2^T & \text{---} \\ & \vdots & \\ \text{---} & a_m^T & \text{---} \end{bmatrix} \begin{bmatrix} | & | & & | \\ b^1 & b^2 & \cdots & b^n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^T b^1 & a_1^T b^2 & \cdots & a_1^T b^n \\ a_2^T b^1 & a_2^T b^2 & \cdots & a_2^T b^n \\ \vdots & \vdots & \ddots & \vdots \\ a_m^T b^1 & a_m^T b^2 & \cdots & a_m^T b^n \end{bmatrix}$$

$A \in \mathbb{R}^{m \times d}$ $\quad B \in \mathbb{R}^{d \times n}$

$C \in \mathbb{R}^{m \times n}$ $\quad c_{ij} = a_i^T b_j$

Matrix-Matrix Multiplication

View 2: Sum of outer products

$$C = AB = \begin{bmatrix} | & | & & | \\ a^1 & a^2 & \cdots & a^n \\ | & | & & | \end{bmatrix} \begin{bmatrix} --- & b_1^T & --- \\ --- & b_2^T & --- \\ & \vdots & \\ --- & b_n^T & --- \end{bmatrix} = a^1 b_1^T + a^2 b_2^T + \cdots a^n b_n^T = \sum_{i=1}^{n} a^i b_i^T$$

$a_i \quad b_j^T$

Matrix-Matrix Multiplication

Properties

- Associative: $(AB)C = A(BC)$.
- Distributive: $A(B + C) = AB + AC$
- In general, **not** commutative; it can be the case that $AB \neq BA$.
- Counterexample: $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. $AB = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ but $BA = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$

Exercise

$A v = \lambda v$

- Suppose $x_1, \ldots, x_N$ are all D-dimensional vectors and $X \in \mathbb{R}^{N \times D}$ is a matrix where the n^{th} row is x_n^T. Then which of the following identities are correct?

$X^T \in \mathbb{R}^{D \times N}$

A. $X^\top X = \sum_{n=1}^{N} x_n x_n^\top$

B. $X^\top X = \sum_{n=1}^{N} x_n^\top x_n$

C. $XX^\top = \sum_{n=1}^{N} x_n x_n^\top$

D. $XX^\top = \sum_{n=1}^{N} x_n^\top x_n$

$D \times 1$

$$(X^T X)_{ij} = \sum_{n=1}^{N} (X^T)_{in} (X)_{nj} = \sum_{n=1}^{N} X_{ni} X_{nj}$$

$$x_n = \begin{bmatrix} x_{n1} \\ \vdots \\ x_{nD} \end{bmatrix} \quad [x_{n1} \cdots x_{nD}]$$

$$\left(\sum_{n=1}^{N} x_n x_n^T\right)_{ij} = \sum_{n=1}^{N} (x_n x_n^T)_{ij} = \sum_{n=1}^{N} (x_{ni} x_{nj})$$

$$(XX^T)_{ij} = \sum_{k=1}^{D} X_{ik} \cdot (X^T)_{kj} = \sum_{k=1}^{D} X_{ik} \cdot X_{jk} = x_i^T x_j$$

$$\Rightarrow \begin{bmatrix} x_1^T x_1 & x_1^T x_2 & \cdots & x_1^T x_N \\ \vdots & & & \vdots \\ x_N^T x_1 & & \cdots & x_N^T x_N \end{bmatrix}$$

Operations and Properties

Transpose

The transpose of a matrix results from 'flipping' the rows and columns.

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

- Properties:
  - $(A^\top)^\top = A$
  - $(AB)^\top = B^\top A^\top$
  - $(A+B)^\top = A^\top + B^\top$

$$(AB)^T_{ij} = (AB)_{ji} = \sum_{d=1}^{D} A_{jd} \cdot B_{di}$$

$$(B^T A^T)_{ij} = \sum_{d=1}^{D} (B^T)_{id} \cdot (A^T)_{dj}$$

$$= \sum_{d=1}^{D} B_{di} \cdot A_{jd}$$

- If $A = A^\top$, then A is a symmetric matrix

- If $A = -A^\top$, then A is an anti-symmetric matrix

Trace

The trace of a square matrix is the sum of its diagonal elements

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

Properties ($A, B, C \in \mathbb{R}^{n\times n}$):

- $tr(A) = tr(A^\top)$
- $tr(A+B) = tr(A) + tr(B)$
- $tr(\alpha A) = \alpha \cdot tr(A)$
- $tr(AB) = tr(BA)$
- $tr(ABC) = tr(BCA) = tr(CAB)$

$tr(AB)$

$= \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \cdot B_{ji}$

$= \sum_{j=1}^{n} \sum_{i=1}^{n} B_{ji} A_{ij}$

$= \sum_{j=1}^{n} (BA)_{jj} = tr(BA)$

Inverse of a Square Matrix

- The inverse of a square matrix $A \in \mathbb{R}^n$, denoted A^{-1}, is the unique matrix such that $A^{-1}A = I_n = AA^{-1}$
- A must be full rank for its inverse to exist

- Properties (suppose that A and B are invertible):
  - $(A^{-1})^{-1} = A$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^\top = (A^\top)^{-1}$, denoted by $A^{-\top}$

$B^{-1}\underbrace{A^{-1} \cdot (A}_{I.}B) = I$

$(AB) \cdot B^{-1}A^{-1} = I.$

$B^{-1}A^{-1}AB$

$= B^{-1}(A^{-1}A)B$

$= B^{-1} I B$

$= B^{-1} B = I$

Exercise

- Which identities are NOT correct for real-valued matrices A, B, and C? Assume that inverse exists and multiplications are legal.

A. $(AB)^{-1} = B^{-1}A^{-1}$ ✓

B. $(I + A)^{-1} = I - A$ ✗ $\quad (I-A)(I+A) = I - \cancel{AI} + \cancel{IA} - A^2 = I - A^2 \quad \times$

C. $tr(AB) = tr(BA)$ ✓

D. $\underbrace{(AB)^{\top} = A^{\top}B^{\top}}_{= B^{\top}A^{\top}}$ ✗

Matrix Calculus

Gradient

- Suppose $f: \mathbb{R}^{m\times n} \mapsto \mathbb{R}$ is a scalar function that takes as input a matrix $A \in \mathbb{R}^{m\times n}$
- The gradient of f with respect to A is the matrix of partial derivatives in $\mathbb{R}^{m\times n}$.

$\frac{\partial f}{\partial A_{ij}}$

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

Gradient

- If the input is just a vector $x \in \mathbb{R}^n$,

$$\frac{\partial (f+g)}{\partial x_1} = \frac{\partial f}{\partial x_1} + \frac{\partial g}{\partial x_1}$$

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

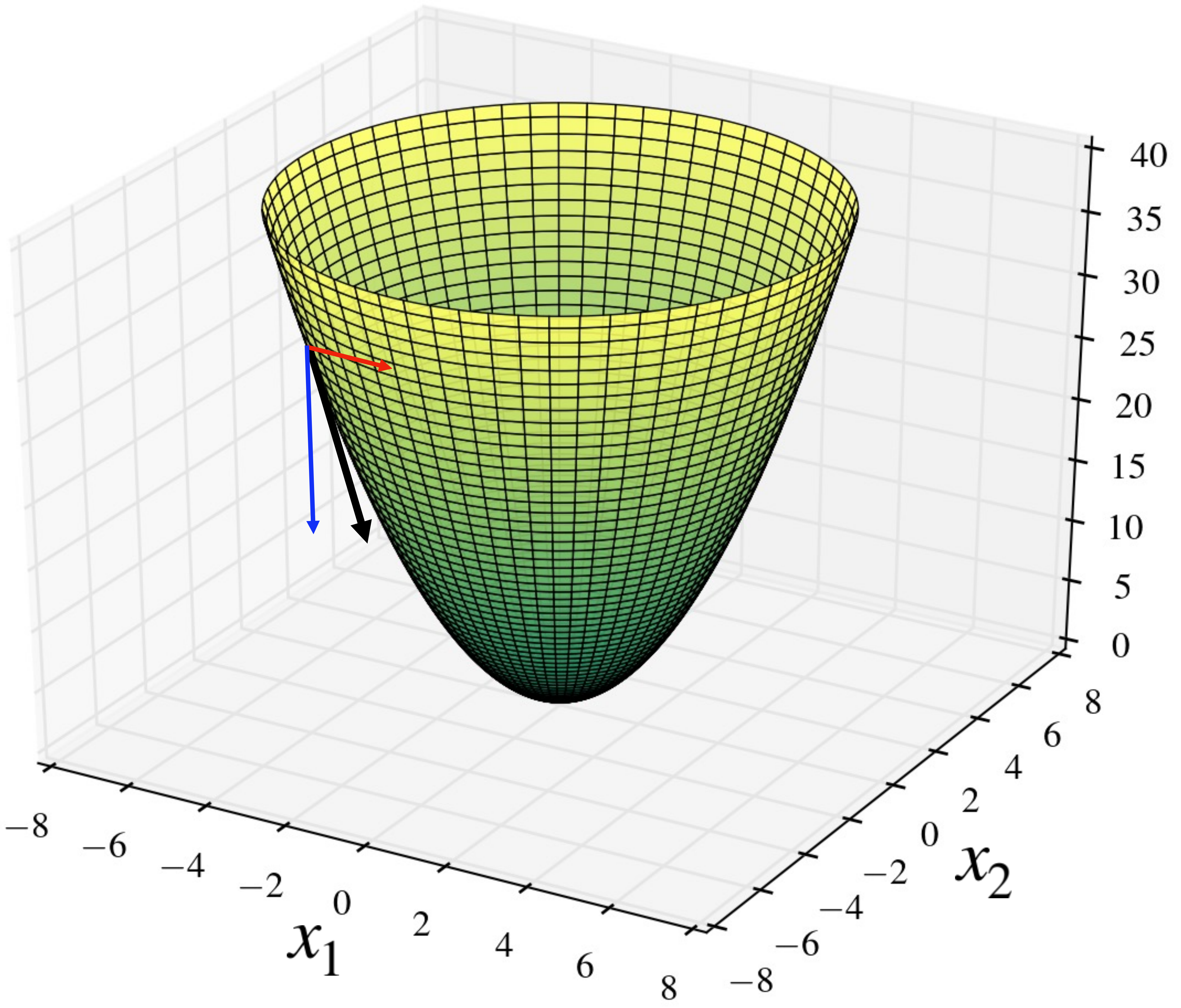
$$\nabla g = \begin{bmatrix} \frac{\partial g(x)}{\partial x_1} \\ \\ \\ \end{bmatrix}$$

- Properties of partial derivatives extend here (can be derived via scalar derivatives):
  - $\nabla_x\big(f(x) + g(x)\big) = \nabla_x f(x) + \nabla_x g(x)$
  - For $t \in \mathbb{R}$, $\nabla_x\big(tf(x)\big) = t\nabla_x f(x)$

Gradient

Visual Example

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



Hessian

- Suppose $f: \mathbb{R}^n \mapsto \mathbb{R}$ is a scalar function that takes as input a vector $x \in \mathbb{R}^n$
- The Hessian of f with respect to x is the second-order partial derivative matrix in $\mathbb{R}^{n\times n}$:

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

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j x_i}$$

- It is symmetric given the continuity of second-order partial derivative.

Examples: Gradient of a Linear Function

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

$f(x_1 \ldots x_n)$

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

$(\nabla f(x))_i = \frac{\partial f}{\partial x_i} = b_i$

$= b_i x_i + \cdots + b_n x_n$

- This gives:

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

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

- Analogous to single variable calculus, where $\frac{\partial (ax)}{\partial x} = a$

Exercise

- Which of the following are correct chain rules: ($g, g_1, \ldots, g_d$ are functions from $\mathbb{R}$ to $\mathbb{R}$)?

$g(w) = w^2$ $\quad f(x) = x^2$

A. ✓ For a composite function $f(g(w))$, $\frac{\partial f}{\partial w} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial w}$

$f(g(w)) = w^4 \Rightarrow 4 \cdot w^3$

$\frac{\partial f}{\partial g} = 2g(w) = 2 \cdot w^2$

B. ✗ For a composite function $f(g(w))$, $\frac{\partial f}{\partial w} = \frac{\partial f}{\partial g} + \frac{\partial g}{\partial w}$

$\frac{\partial g}{\partial w} = 2 \cdot w$

C. For a composite function $f(g_1(w), \ldots, g_d(w))$, $\frac{\partial f}{\partial w} = \left(\frac{\partial f}{\partial g_1} \cdot \frac{\partial g_1}{\partial w}, \ldots, \frac{\partial f}{\partial g_d} \cdot \frac{\partial g_d}{\partial w}\right)$

D. ✓ For a composite function $f(g_1(w), \ldots, g_d(w))$, $\frac{\partial f}{\partial w} = \sum_{i=1}^{d} \frac{\partial f}{\partial g_i} \cdot \frac{\partial g_i}{\partial w}$

$\frac{\partial f}{\partial g_1} \cdot \frac{\partial g_1}{\partial w} = 2w$

$1 \quad 2w$

$-\frac{1}{g_2^2(w)} \cdot 3 \cdot w^2$

$d = 2$

$\frac{\partial f}{\partial g_2} \cdot \frac{\partial g_2}{\partial w} = -\frac{3}{w^4}$

$g_1(w) = w^2$

$g_2(w) = w^3$

$f(x, y) = x + 1/y$

$\frac{\partial f}{\partial w} = \partial\left(w^2 + \frac{1}{w^3}\right)$

$= 2w - \frac{3}{w^4}$

Exercise

$(Ax)_i = \sum_{j=1}^{n} A_{ij} x_j$

$\sum_{i=1}^{n} x_i \sum_{j=1}^{n} A_{ij} x_j$

- A function $f: \mathbb{R}^{n\times 1} \mapsto \mathbb{R}$ is defined as $f(x) = x^\top A x + b^\top x$ for some $b \in \mathbb{R}^{n\times 1}$ and $A \in \mathbb{R}^{n\times n}$. What is the derivative $\nabla f(x)$?

A. ✓ $(A + A^\top)x + b$

B. $2A^\top x + b$

C. $2Ax + b$

D. $2Ax + x$

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

$\frac{\partial f}{\partial x_i} = ((A+A^T)x)_i + b_i$

$\sum_{i=1}^{n} x_i^2 A_{ii} + \sum_{(i,j),\, j\neq i} (A_{ij} + A_{ji}) x_j x_i$

$2A_{ii}$

$\partial \downarrow$

$2 x_i A_{ii} + \sum_{j\neq i} (A_{ij} + A_{ji}) x_j$

$= \sum_{j=1}^{n} (A_{ij} + A_{ji}) x_j = (Ax)_i + (A^T x)_i$

$2 x_i A_{ii} + \sum_{j=1,\, j\neq i}^{n} (A_{ij} + A_{ji}) x_j$

$(A_{ii} + A_{ii}) x_i$

Gradient of a Quadratic Function

- For $x \in \mathbb{R}^n$, let $f(x) = x^\top Ax$ for some known matrix $A \in \mathbb{R}^{n\times n}$. Then,

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

- Using previous slides, product rule for $f(x) = g(x)^\top x$, with $g(x) = A^\top x$:

$$\begin{aligned} \nabla_x f(x) &= \nabla_x^T g(x) x + \nabla_x x^T g(x) \\ &= (A^T)^T x + I^T A^T x \\ &= (A + A^T) x \end{aligned}$$

- This gives the Hessian:

$$\nabla_x^2 f(x) = A + A^T$$

Exercise

- A function $f: \mathbb{R}^{n\times 1} \mapsto \mathbb{R}$ is defined as $f(w) = \ln(1 + e^{-w^\top x})$ for some $x \in \mathbb{R}^{n\times 1}$. What is the derivative $\nabla_w f(w)$?

$g(w) = e^{-h(w)}$ $\quad h(w) = w^T x$

A. $-\frac{w}{1+e^{w^\top x}}$

B. $-\frac{x}{1+e^{w^\top x}}$

C. $-\frac{w}{1+e^{-w^\top x}}$

D. $-\frac{x}{1+e^{-w^\top x}}$

$f(w) = \ln(1 + g(w))$

$$\frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial h} \cdot \frac{\partial h}{\partial w} \cdot$$

$$\frac{1}{1+g(w)} \cdot -e^{-h(w)} \cdot x$$

$$= \frac{1}{1+e^{-w^T x}} \cdot -e^{-w^T x} \cdot x = \frac{-x}{1+e^{w^T x}}$$

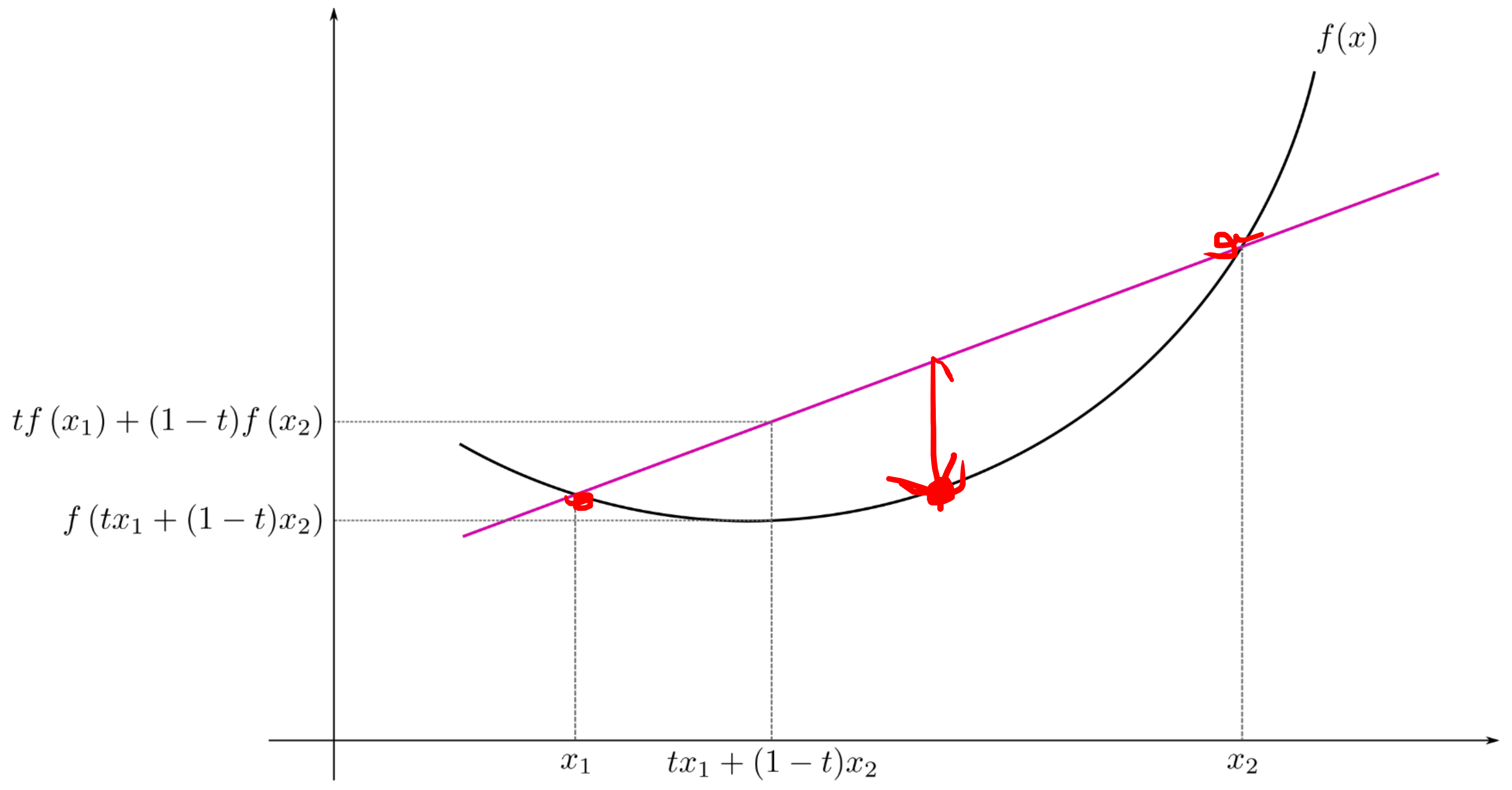
Convexity

- A function $f: \mathbb{R}^n \mapsto \mathbb{R}$, a scaler function that takes as input a vector $x \in \mathbb{R}^n$, is convex if for any $x_1, x_2 \in \mathbb{R}^n$ and any $t \in [0,1]$

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

- Feasible domain $\mathbb{R}^n$ can be replaced by any subset of $\mathbb{R}^n$

- Decide whether $f(x)$ is convex:
  - By definition
  - If $f(x)$ is second-order differentiable,
    - $\nabla_x^2 f(x)$ is positive semi-definite $\Leftrightarrow f(x)$ is convex



- A function f is concave if $-f$ is convex.

Exercise

- Suppose $a \in \mathbb{R}^{n\times 1}$ is an arbitrary vector. Which one of the following functions in NOT convex:

$x \quad y \quad t \in [0,1]$

$|a|+|b| \geq |a+b|$

A. $f(x) = \sum_{i=1}^{n} |x_i|$

$t f(x) + (1-t) f(y) = \sum_{i=1}^{n} t|x_i| + (1-t)|y_i|$

$\geq \sum_{i=1}^{n} |t x_i + (1-t) y_i|$

$= f(tx + (1-t)y).$

B. $f(x) = \sum_{i=1}^{n} a_i x_i$

$x = (1,0)$

$y = (0,1)$

$t = 1/2$

C. $f(x) = \min_{i\in\{1,2,\ldots,n\}} a_i x_i$

$\underbrace{\frac{1}{2}f(x) + \frac{1}{2}f(y)}_{0} \leq f(\frac{1}{2}(x+y))$ $\quad 1/2$

D. $f(x) = \sum_{i=1}^{n} \exp(x_i)$

$H(f(x)) = \Big[\quad \mathrm{diag}(e^{x_1}, \ldots, e^{x_n}) \Big]$

Exercise

- For a differential function $f: \mathbb{R}^n \mapsto \mathbb{R}$, which of the following statements are correct

A. If $x^\star$ is a minimizer of f, then $\nabla f(x^\star) = 0$

B. If $x^\star$ is a maximizer of f, then $\nabla f(x^\star) = 0$

C. If $\nabla f(x^\star) = 0$, then $x^\star$ is a minimizer of f.

D. If $\nabla f(x^\star) = 0$, then $x^\star$ is a maximizer of f.

Questions?