Lecture 1. Vectors, matrices, and derivatives.

8/21/23

1. Notations

scalar: x (single number). lower-case name.

$x \in \mathbb{R}$ (x is a real value $\in (-\infty, +\infty)$)

$x \in \mathbb{N}$ (x is a natural value)

$N = \{1, 2, 3, \cdots\}$.

vectors: array of numbers

$\mathbf{x}$ (lower-case, bold typeface)

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

: all vectors are column vector in default.

$\mathbf{x}^T = (x_1, x_2, \cdots, x_n)$

row vector.

matrix: 2D array of numbers.

$\mathbf{A}$ (upper case, bold typeface)

$$A = \begin{bmatrix} A_{11} & A_{12}, & \cdots, & A_{1n} \\ A_{21} & A_{22}, & \cdots, & A_{2n} \\ \vdots & & & \\ A_{m1} & A_{m2}, & \cdots, & A_{mn} \end{bmatrix}_{m \times n}$$

A_{ij}: element at the ith row and jth column.

$A_{i:}$ $\Rightarrow$ All elements in the ith row.

$A_{:j}$ $\Rightarrow$ All elements in the jth column.

2. Operations.

1) inner and outter product.

$\mathbf{x} \in R^n$ ($\mathbf{x}$ is a vector of n values) $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

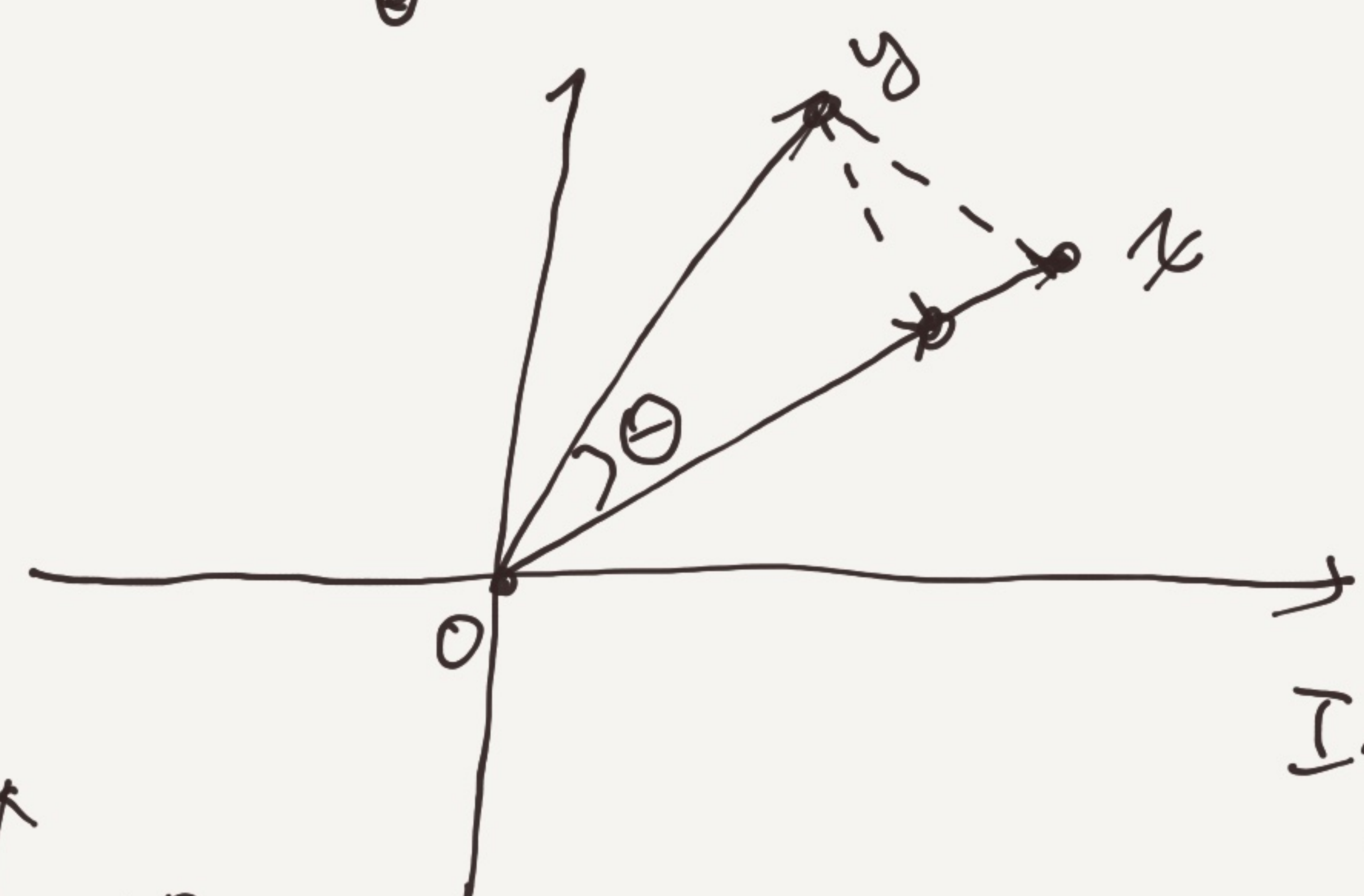
$\mathbf{y} \in R^n$ $\quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

Inner product between $\mathbf{x}$ and $\mathbf{y}$:

$$\mathbf{x}^T \cdot \mathbf{y} = [x_1, x_2, \cdots x_n]_{1\times n} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = x_1 y_1 + x_2 \cdot y_2 + \cdots + x_n y_n$$

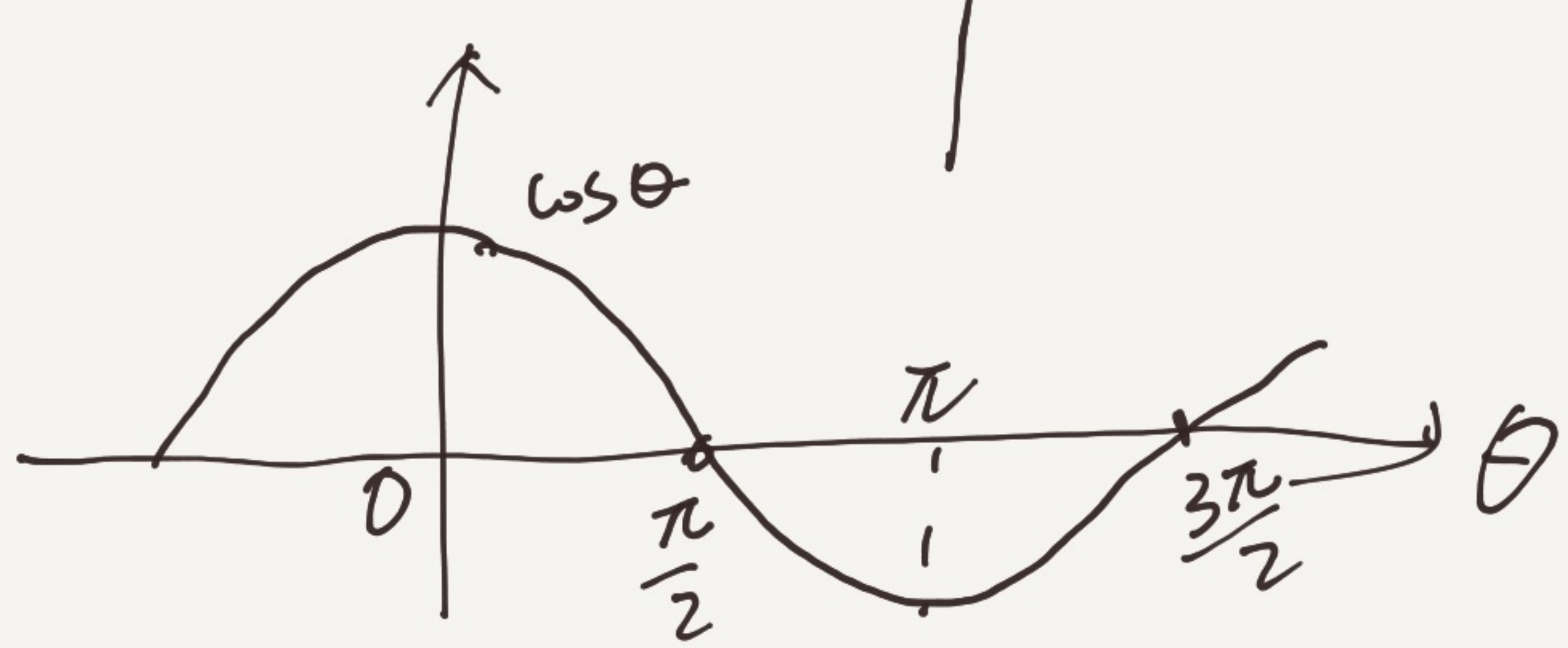
$$= \sum_{i=1}^{n} x_i \cdot y_i$$

$$= \|\mathbf{x}\|_2 \cdot \left[\|\mathbf{y}\|_2 \cdot \cos\theta \right]$$



If $\theta = 0$, $\quad \mathbf{x}^T \mathbf{y} = \underbrace{\|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2 \cdot 1}$

$\theta = \frac{\pi}{2}$, $\quad \mathbf{x}^T \cdot \mathbf{y} = 0$



outter product.

$$x \otimes y = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \otimes \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1y_1, & x_1y_2, & \cdots, & x_1y_n \\ & & & \vdots \\ x_ny_1, & x_ny_2, & \cdots, & x_ny_n \end{bmatrix}$$

3. Calculus

1) Derivatives: D of a function (f) of an input variable.

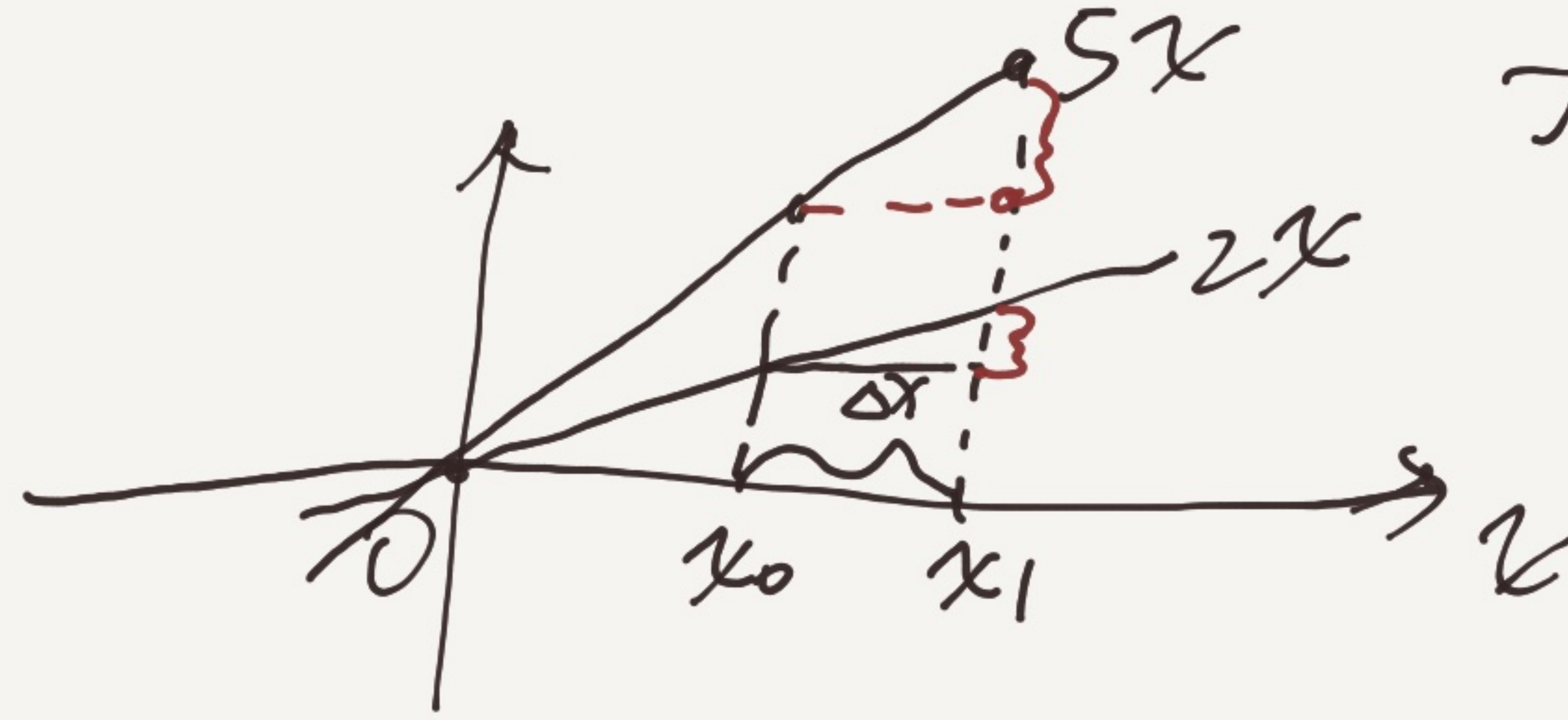
measures the sensitivity of function values to the changes of its input varible.

① $f(x): R \to R. \Rightarrow x \in R, f(x) \in R$

$$Df(x) = \frac{df(x)}{dx}$$

if $f(x) = 5x, \Rightarrow Df(x) = 5$

$f(x) = 2x \Rightarrow Df(x) = 2$



$f(x) = nx$

$$Df(x) = n = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{5\Delta x}{\Delta x} = 5$$

2 $f: R^n \to R \quad \boldsymbol{x} \in \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \; y \in R.$

$$Df(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots \frac{\partial f}{\partial x_n} \right)$$ row vector.

$$f(x) = \boldsymbol{x}^T \boldsymbol{x} = x_1^2 + x_2^2 + \cdots + x_n^2$$

$$Df(x) = \left(\frac{\partial f}{\partial x_1} = 2x_1, 2x_2, \cdots 2x_n \right) = 2\boldsymbol{x}^T$$

Gradient: $\nabla \qquad \nabla f(x) = (Df(x))^T$

$$\nabla f(x) = 2\boldsymbol{x}$$

③ $f: R^n \to R^m$, $x \in \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $y \in \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}, & \frac{\partial f_1}{\partial x_2}, & \cdots & \frac{\partial f_1}{\partial n} \\ \vdots & & & \\ \frac{\partial f_m}{\partial x_1}, & \frac{\partial f_m}{\partial x_2}, & \cdots, & \frac{\partial f_m}{\partial x_n} \end{pmatrix}_{m \times n}$$

④ Examples: $f(x)$. $x \in R$

Function	Derivatives
$f(x) = a$ (constant)	$Df(x) = 0$
$f(x) = ax$	$Df(x) = a$
$f(x) = x^n$	$Df(x) = n \cdot x^{n-1}$
* $f(x) = e^x$	$Df(x) = e^x$
* $f(x) = \log x$	$Df(x) = 1/x$