



# Tensor Derivatives

---

## Linear Algebra

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee [rabiee@sharif.edu](mailto:rabiee@sharif.edu)

Maryam Ramezani [maryam.ramezani@sharif.edu](mailto:maryam.ramezani@sharif.edu)

# Introduction

---

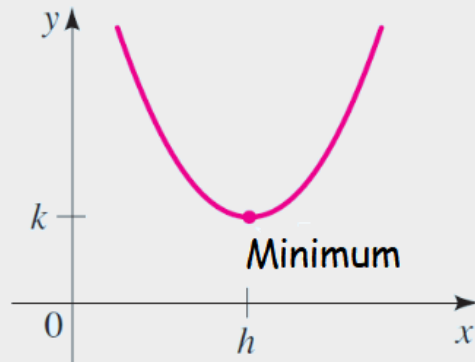
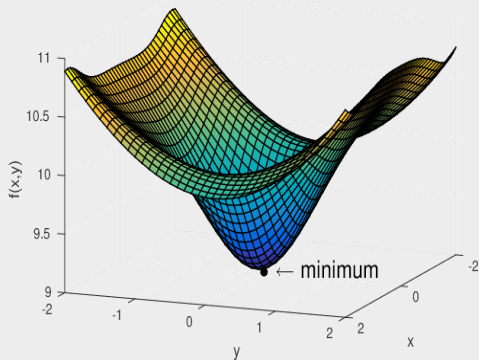


## Types of matrix derivative

Types	Scalar	Vector	Matrix
Scalar	$\frac{\partial y}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{Y}}{\partial x}$
Vector	$\frac{\partial y}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	Tensor! (Optional part of this course)
Matrix	$\frac{\partial y}{\partial \mathbf{X}}$		



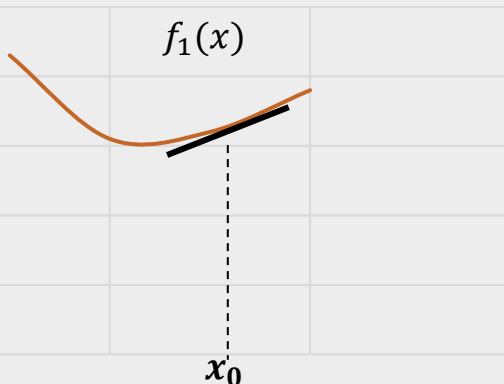
- ❑ Machine Learning training requires one to evaluate how one vector changes with respect to another
- ❑ How output changes with respect to parameters
- ❑ How do we find minimum of a scalar function?
- ❑ How do we find minimum of two variables?



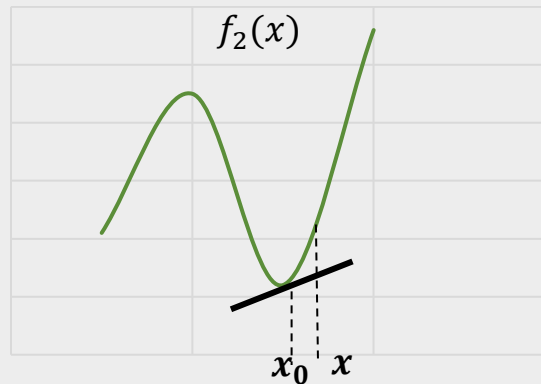
□ Derivative of a vector-valued function  $f: \mathbb{R} \rightarrow \mathbb{R}^n$  with respect to scalar  $x \in \mathbb{R}$ :

$$\frac{\partial f(x)}{\partial x} \triangleq \begin{bmatrix} \frac{\partial f_1(x)}{\partial x} \\ \frac{\partial f_2(x)}{\partial x} \\ \vdots \\ \frac{\partial f_n(x)}{\partial x} \end{bmatrix}$$

$$f(x) \approx f(x_0) + m(x - x_0) \quad m = \begin{bmatrix} f'_1(x_0) \\ \vdots \\ f'_n(x_0) \end{bmatrix}$$



CE282: Linear Algebra



Hamid R. Rabiee & Maryam Ramezani

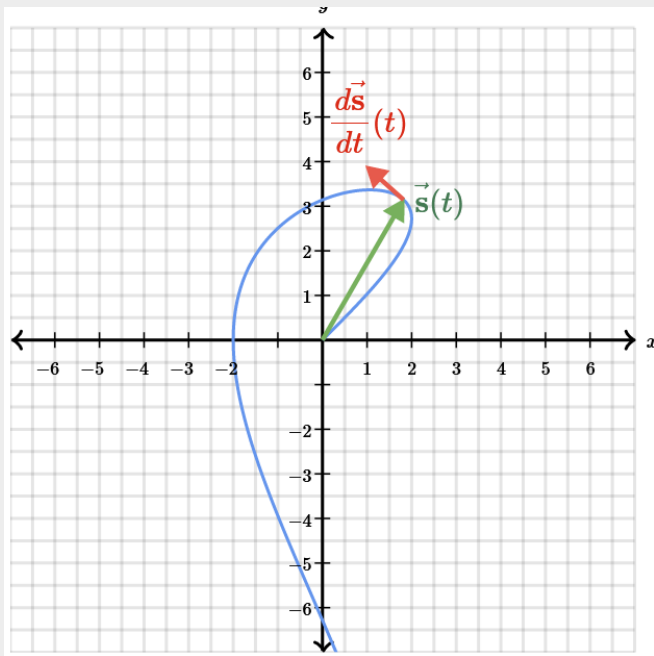
Example

$$f(x) = \begin{bmatrix} \sin(x) \\ \cos(x) \end{bmatrix}$$

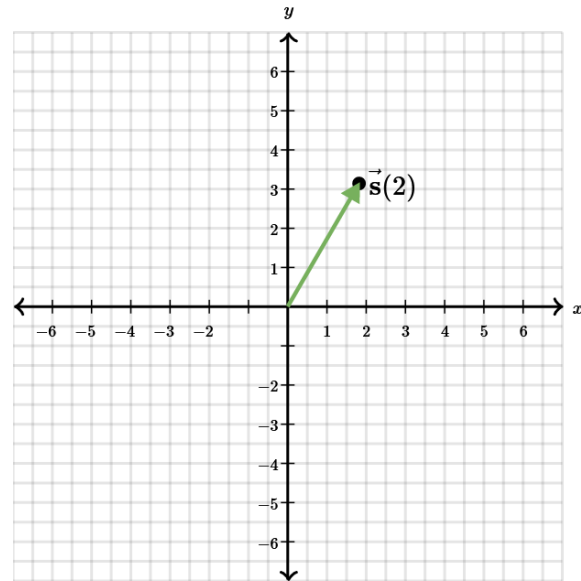
# Vector-Valued Function



$$\vec{s}(t) = \begin{bmatrix} 2 \sin(t) \\ 2 \cos(t/3)t \end{bmatrix}$$



$$\vec{s}(2) = \begin{bmatrix} 2 \sin(2) \\ 2 \cos(2/3) \cdot 2 \end{bmatrix} \approx \begin{bmatrix} 1.819 \\ 3.144 \end{bmatrix}$$



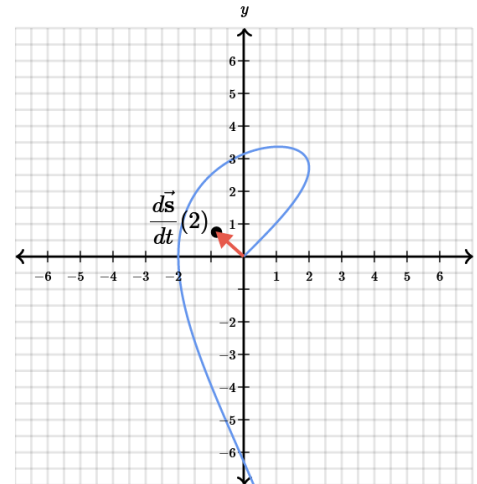


$$\frac{d\vec{s}}{dt}(t) = \begin{bmatrix} \frac{d}{dt}(2 \sin(t)) \\ \frac{d}{dt}(2 \cos(t/3))t \end{bmatrix}$$

$$= \begin{bmatrix} 2 \cos(t) \\ 2 \cos(t/3) - \frac{2}{3} \sin(t/3)t \end{bmatrix}$$

$$\begin{aligned} \frac{d\vec{s}}{dt}(2) &= \begin{bmatrix} 2 \cos(2) \\ 2 \cos(2/3) - \frac{2}{3} \sin(2/3) \cdot 2 \end{bmatrix} \\ &\approx \begin{bmatrix} -0.832 \\ 0.747 \end{bmatrix} \end{aligned}$$

This is also some two-dimensional vector.





□ Derivative of a matrix-valued function  $f: \mathbb{R} \rightarrow \mathbb{R}^{m \times n}$  with respect to scalar  $x \in \mathbb{R}$ :

$$\frac{\partial f(x)}{\partial x} \triangleq \begin{bmatrix} \frac{\partial f_{11}(x)}{\partial x} & \frac{\partial f_{12}(x)}{\partial x} & \dots & \frac{\partial f_{1n}(x)}{\partial x} \\ \frac{\partial f_{21}(x)}{\partial x} & \frac{\partial f_{22}(x)}{\partial x} & \dots & \frac{\partial f_{2n}(x)}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{m1}(x)}{\partial x} & \frac{\partial f_{m2}(x)}{\partial x} & \dots & \frac{\partial f_{mn}(x)}{\partial x} \end{bmatrix}$$

Example

□ Rotation Matrix

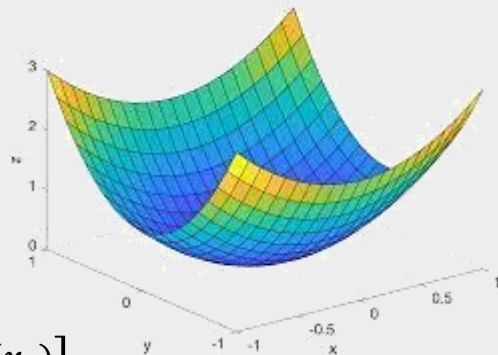




□ Derivative of a real-valued function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  with respect to vector  $\mathbf{x} \in \mathbb{R}^n$ :

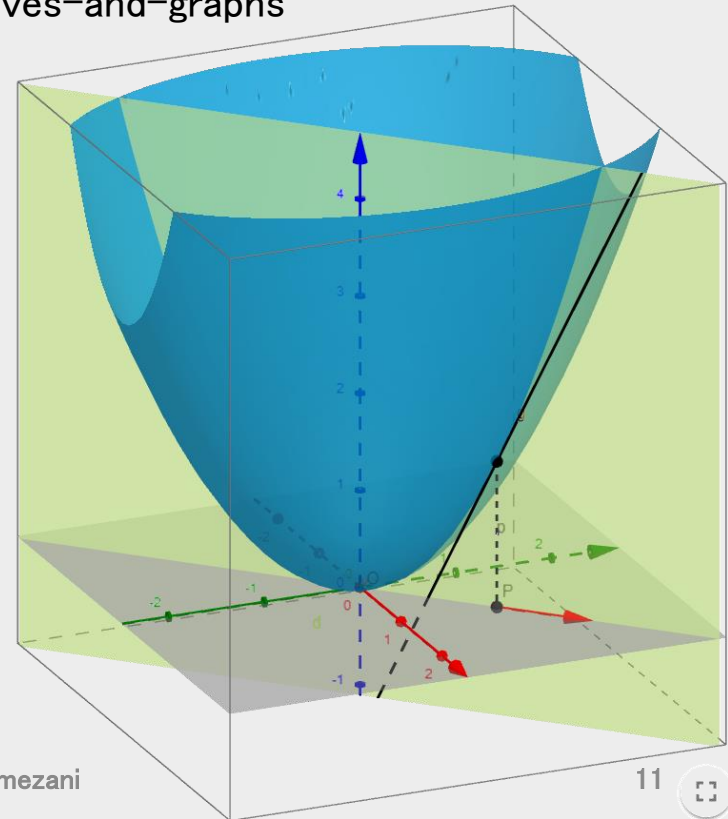
$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

$$f(\mathbf{x}) - f(\mathbf{x}_0) = \mathbf{m}^T (\mathbf{x} - \mathbf{x}_0) \quad \mathbf{m} = \begin{bmatrix} f'_1(\mathbf{x}_0) \\ \vdots \\ f'_n(\mathbf{x}_0) \end{bmatrix}$$



□ Gradient

- ❑ <https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivatives/v/partial-derivatives-and-graphs>
- ❑ <https://www.geogebra.org/m/bxhwxr2x>





- $\frac{dY}{dx} = \frac{dY}{du} \frac{du}{dx}$   $x, u$ : scalars  $Y$ : matrix
- $\frac{dy}{dX} = \frac{dy}{du} \frac{du}{dX}$   $y, u$ : scalars  $X$ : matrix
- $\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$   $x, y, u$ : vectors



- $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$
- $\frac{\partial(AB)}{\partial \alpha} = A \frac{\partial(B)}{\partial \alpha} + \frac{\partial(A)}{\partial \alpha} B$  if A and B be matrices which elements are function of scalar  $\alpha$
- $\frac{\partial(x^T y)}{\partial z} = x^T \frac{\partial(y)}{\partial z} + y^T \frac{\partial(x)}{\partial z}$  if x and y be vectors which elements are function of vector z

## Example

- $f, g: \mathbb{R} \rightarrow \mathbb{R}^n$        $h(x) = f(x)^T g(x)$        $h'(x) = ?$
- $f: \mathbb{R}^n \rightarrow \mathbb{R}$     $g: \mathbb{R} \rightarrow \mathbb{R}$     $h(x) = f(x)g(x)$        $h'(x) = ?$
- $H: \mathbb{R} \rightarrow \mathbb{R}^{m \times n}, F: \mathbb{R} \rightarrow \mathbb{R}^{m \times p}, G: \mathbb{R} \rightarrow \mathbb{R}^{p \times n}$        $H(x) = F(x)G(x)$



□ Derivative of a scalar function  $f: \mathbb{R}^N \rightarrow \mathbb{R}$  with respect to vector  $\mathbf{x} \in \mathbb{R}^N$ :

$$\square \quad \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} & \frac{\partial f(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_N} \end{bmatrix}$$

□ Derivative of a vector function  $f: \mathbb{R}^N \rightarrow \mathbb{R}^M$  with respect to vector  $\mathbf{x} \in \mathbb{R}^N$ :

$$\square \quad \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_N} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\mathbf{x})}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_M(\mathbf{x})}{\partial x_1} & \frac{\partial f_M(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_M(\mathbf{x})}{\partial x_N} \end{bmatrix}$$



## Definition

□ Derivative of a scalar function  $f: \mathbb{R}^{M \times N} \rightarrow \mathbb{R}$  with respect to matrix  $\mathbf{X} \in \mathbb{R}^{M \times N}$ :

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial X_{1,1}} & \frac{\partial f(\mathbf{X})}{\partial X_{2,1}} & \dots & \frac{\partial f(\mathbf{X})}{\partial X_{M,1}} \\ \frac{\partial f(\mathbf{X})}{\partial X_{1,2}} & \frac{\partial f(\mathbf{X})}{\partial X_{2,2}} & \dots & \frac{\partial f(\mathbf{X})}{\partial X_{M,2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial X_{1,N}} & \frac{\partial f(\mathbf{X})}{\partial X_{2,N}} & \dots & \frac{\partial f(\mathbf{X})}{\partial X_{M,N}} \end{bmatrix}$$

□ Using the above definitions, we can generalize the chain rule, Given  $\mathbf{u} = \mathbf{h}(x)$  (i.e.  $\mathbf{u}$  is a function of  $x$ ) and  $\mathbf{g}$  is a vector function of  $\mathbf{u}$ , the vector-by-vector chain rule states:

$$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial x} = \frac{\partial \mathbf{u}}{\partial x} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$$



□ Board 😊



□ Board 😊





## DEFINITION

Suppose  $z = f(x, y)$  is a function of two variables with a domain of  $D$ . Let  $(a, b) \in D$  and define  $\mathbf{u} = \cos \theta \mathbf{i} + \sin \theta \mathbf{j}$ . Then the **directional derivative** of  $f$  in the direction of  $\mathbf{u}$  is given by

$$D_{\mathbf{u}}f(a, b) = \lim_{h \rightarrow 0} \frac{f(a + h \cos \theta, b + h \sin \theta) - f(a, b)}{h},$$

provided the limit exists.

$$\nabla_{\vec{\mathbf{v}}}f(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\vec{\mathbf{v}}) - f(\mathbf{x})}{h\|\vec{\mathbf{v}}\|}$$



Try to prove the followings:

$$\square \frac{\partial(u(x)+v(x))}{\partial x} = \frac{\partial u(x)}{\partial x} + \frac{\partial v(x)}{\partial x}$$

$$\square \frac{\partial(Ax)}{\partial x} = A$$

$$\square \frac{\partial(x^T a)}{\partial x} = a^T$$

$$\square \frac{\partial(x^T Ax)}{\partial x} = x^T (A + A^T)$$

$$\square \frac{\partial(x^T Ax)}{\partial x} = 2x^T A \text{ if } A \text{ is symmetric}$$



$$A\vec{x} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_1x_1 + a_2x_2 \\ a_3x_1 + a_4x_2 \end{bmatrix}$$

$$\begin{aligned} \frac{dA\vec{x}}{dx} &= \begin{bmatrix} \frac{\partial(a_1x_1 + a_2x_2)}{\partial x_1} & \frac{\partial(a_1x_1 + a_2x_2)}{\partial x_2} \\ \frac{\partial(a_3x_1 + a_4x_2)}{\partial x_1} & \frac{\partial(a_3x_1 + a_4x_2)}{\partial x_2} \end{bmatrix} \\ &= \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = A \end{aligned}$$



## Important

1. Derivative of a linear function:

$$\frac{\partial}{\partial \vec{x}} \vec{a} \cdot \vec{x} = \frac{\partial}{\partial \vec{x}} \vec{a}^T \vec{x} = \frac{\partial}{\partial \vec{x}} \vec{x}^T \vec{a} = \vec{a}^T$$

(If you think back to calculus, this is just like  $\frac{d}{dx} ax = a$ ).

2. Derivative of a quadratic function:

$$\frac{\partial}{\partial \vec{x}} \vec{x}^T A \vec{x} = 2A\vec{x}$$

(Again, if you think back to calculus, this is just like  $\frac{d}{dx} ax^2 = 2ax$ ).

If you ever need it, the more general rule (for non-symmetric A) is:

$$\frac{\partial}{\partial \vec{x}} \vec{x}^T A \vec{x} = \vec{x}^T (A + A^T)$$

which of course is the same thing as  $2A\vec{x}$  when A is symmetric.



Given  $A = [a_{ij}]$ , the  $(i, j)$ -cofactor of  $A$  is the number  $C_{ij}$  given by

$$C_{ij} = (-1)^{i+j} \det(A_{ij})$$

Then

$$\det(A) = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

Which is a **cofactor expansion across the first row** of  $A$ .

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} = A^{-1} = \frac{1}{|A|} \text{adj } A$$

$$\text{Adj } A = C^T$$

The matrix of cofactors is called the **adjugate** (or **classical adjoint**) of  $A$ , denoted by  $\text{adj } A$ .



## Proof the followings:

$$\square \frac{\partial (A(t))^{-1}}{\partial t} = -A(t)^{-1} \frac{\partial (A(t))}{\partial t} A(t)^{-1}$$

$$\square \frac{\partial \det(A)}{\partial A} = \det(A) A^{-1}$$

$$\square \frac{\partial \ln(\det(A))}{\partial A} = (A^{-1})^T$$

$$\square \frac{\partial \det(A(t))}{\partial t} = \det(A) \operatorname{trace}\left(A^{-1} \frac{\partial (A(t))}{\partial t}\right)$$

$$\square \frac{\partial \operatorname{trace}(BA^{-1})}{\partial A} = -A^{-1}BA^{-1}$$

$$\square \frac{\partial (y^T Ax)}{\partial A} = yx^T$$

$$\square \frac{\partial (x^T Ax)}{\partial A} = xx^T$$

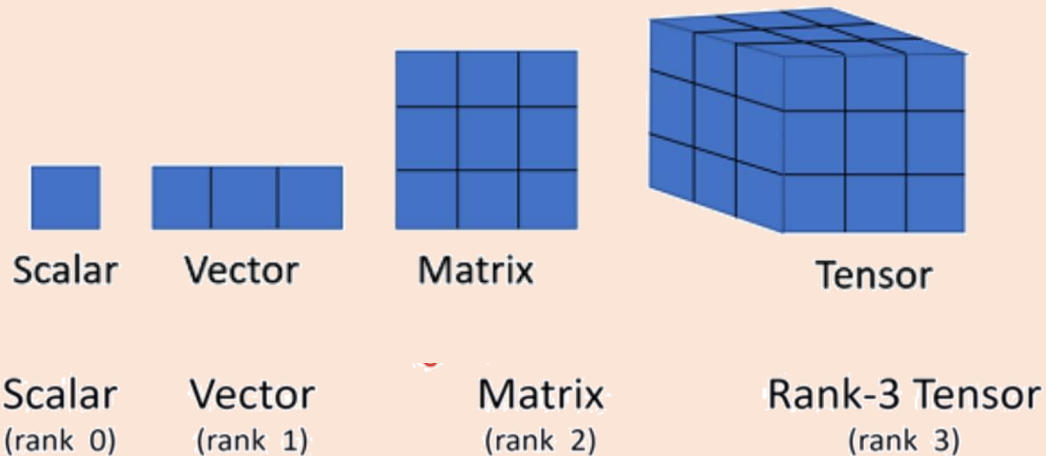
# Tensor (Optional)

---

## Definition

- Multi-dimensional array of numbers

```
w = torch.empty(3)
x = torch.empty(2, 3)
y = torch.empty(2, 3, 4)
z = torch.empty(2, 3, 2, 4)
```

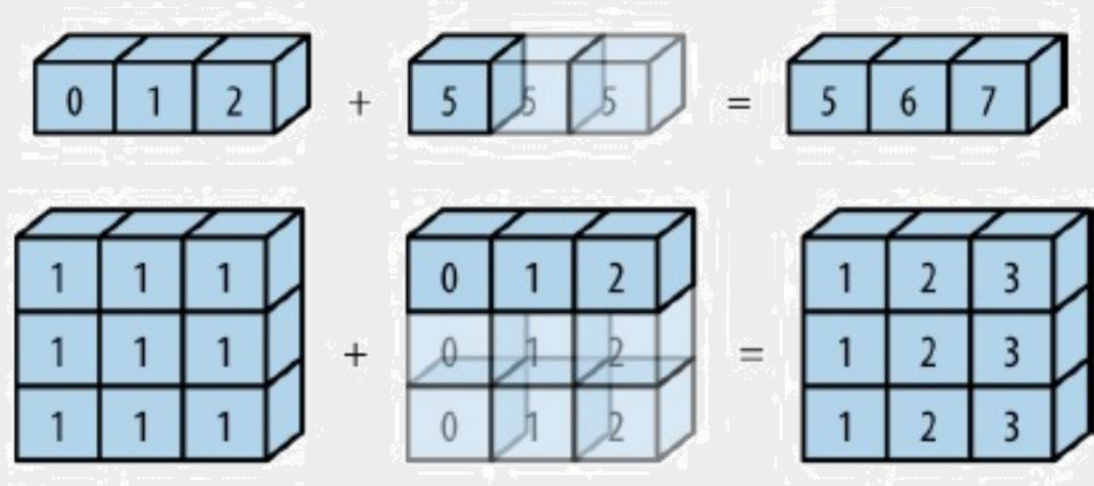




# Tensors Addition



- ❑ Adding tensors with same size
- ❑ Adding scalar to tensor
- ❑ Adding tensors with different size: if **broadcastable**





- ❑ Two tensors are “broadcastable” if the following rules hold:
  - Each tensor has at least one dimension.
  - When iterating over the dimension sizes, starting at the trailing dimension, the dimension sizes must either be equal, one of them is 1, or one of them does not exist.
- ❑ Example
  - T1: (5,7,3) T2:(5,7,3)
  - T1: (5,3,4,1) T2:(3,1,1)



## □ Matrix Product on tensors

$$(m \times n) \cdot (n \times k) = (m \times k)$$

product is defined







- ❑ Linear Algebra and Its Applications, David C. Lay
- ❑ Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares
- ❑ [https://en.Wikipedia.org/wiki/matrix\\_calculus](https://en.Wikipedia.org/wiki/matrix_calculus)
- ❑ <https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf>
- ❑ [https://www.kamperh.com/notes/kamper\\_matrixcalculus13.pdf](https://www.kamperh.com/notes/kamper_matrixcalculus13.pdf)