

Tensor Derivatives

Linear Algebra

Department of Computer Engineering

Sharif University of Technology

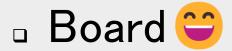
Hamid R. Rabiee <u>rabiee@sharif.edu</u>

Maryam Ramezani <u>maryam.ramezani@sharif.edu</u>

Introduction

Linearization

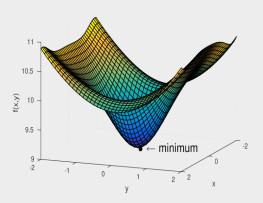


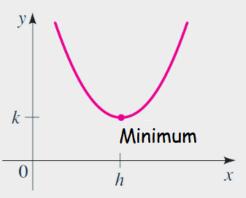


Motivation



- 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?





Vector-Valued Function



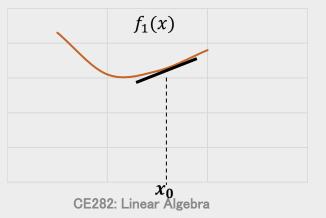
□ Derivative of a vector-valued function $f: \mathbb{R} \to \mathbb{R}^n$ with respect to scalar $x \in \mathbb{R}$:

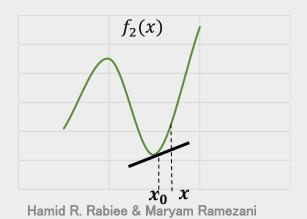
$$\frac{\partial f(x)}{\partial x} \triangleq \begin{bmatrix} \frac{\partial f_1(x)}{\partial x} \\ \frac{\partial f_2(x)}{\partial x} \\ \vdots \\ \frac{\partial f_n(x)}{\partial x} \end{bmatrix}$$

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

$$m = \begin{bmatrix} f_1'(x_0) \\ \dots \\ f_n'(x_0) \end{bmatrix}$$

Example





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

Matrix-Valued Function



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

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

Example

■ Rotation Matrix

Real-Valued Multivariant Function



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

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

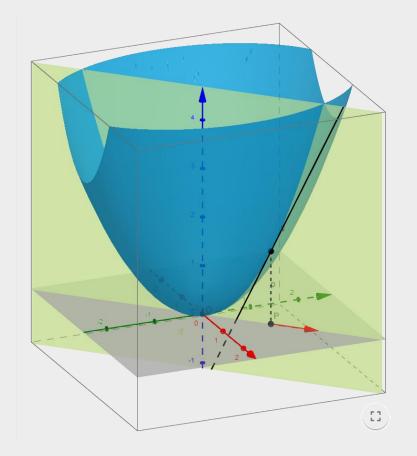
$$f(x) - f(x_0) = m^T(x - x_0) \qquad m = \begin{bmatrix} f'_1(x_0) \\ \vdots \\ f'_n(x_0) \end{bmatrix}$$

□ Gradient

Directional Derivative



□ https://www.geogebra.org/m/bxhwxr2x



Product Rule



$$\Box (f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$$

Example

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

Chain Rule





Motivation



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

$$\Box \quad \frac{\partial f(x)}{\partial x} \triangleq \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \frac{\partial f(x)}{\partial x_2} & \dots & \frac{\partial f(x)}{\partial x_N} \end{bmatrix}$$

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

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

Definitions



Definition

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

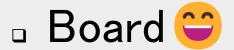
$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \triangleq \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial X_{1,1}} & \frac{\partial f(\mathbf{X})}{\partial X_{2,1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{M,1}} \\ \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} & \frac{\partial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

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

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

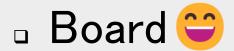
Scalar & Vectors





Vectors & Vectors





Tensor



Definition

☐ Multi-dimensional array of numbers

w = torch.empty(3)

x = torch.empty(2, 3)

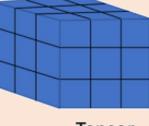
y = torch.empty(2, 3, 4)

z = torch.empty(2, 3, 2, 4)









Tensor

Scalar (rank 0)

Vector (rank 1)

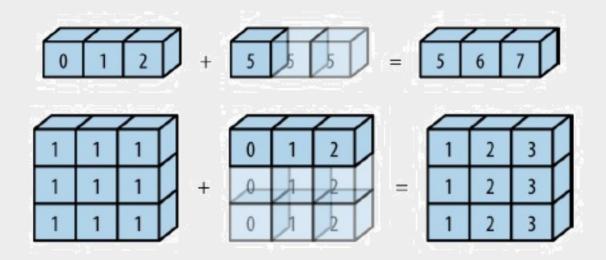
Matrix (rank 2)

Rank-3 Tensor

Tensors Addition



- Adding tensors with same size
- Adding scalar to tensor
- □ Adding tensors with different size: if broacastable



Tensors Addition



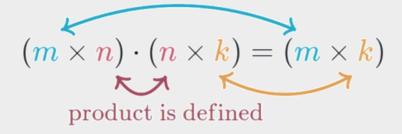
- ☐ Two tensors are "broadcastable" if the following rules hold:
 - Each tensor has at least one dimension.
 - When iterating over the dimension sizes, starting at the trailing dimension, the dimension sizes must either be equal, one of them is 1, or one of them does not exist.

Example

- o T1: (5,7,3) T2:(5,7,3)
- o T1: (5,3,4,1) T2:(3,1,1)



Matrix Product on tensors



Conclusion



Conclusion

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

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

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

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

$$\Box \frac{\partial (\ln(|X|))}{\partial X} = (X^{-1})^T$$

Hint!



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

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

$$= \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = A$$

Derivative of a vector with respect to a matrix



Derivative of a matrix with respect to a matrix



Conclusion



Important

1. Derivative of a linear function:

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

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

2. Derivative of a quadratic function:

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

(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:

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

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

Good Examples!!!



Proof the followings:

$$\Box \frac{\partial (y^T A X)}{\partial A} = y X^T$$

$$\square \frac{\partial (X^T A X)}{\partial A} = X X^T$$

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

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

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

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

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

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

References



- □ 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://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf
- https://www.kamperh.com/notes/kamper_matrixcalculus13.pdf