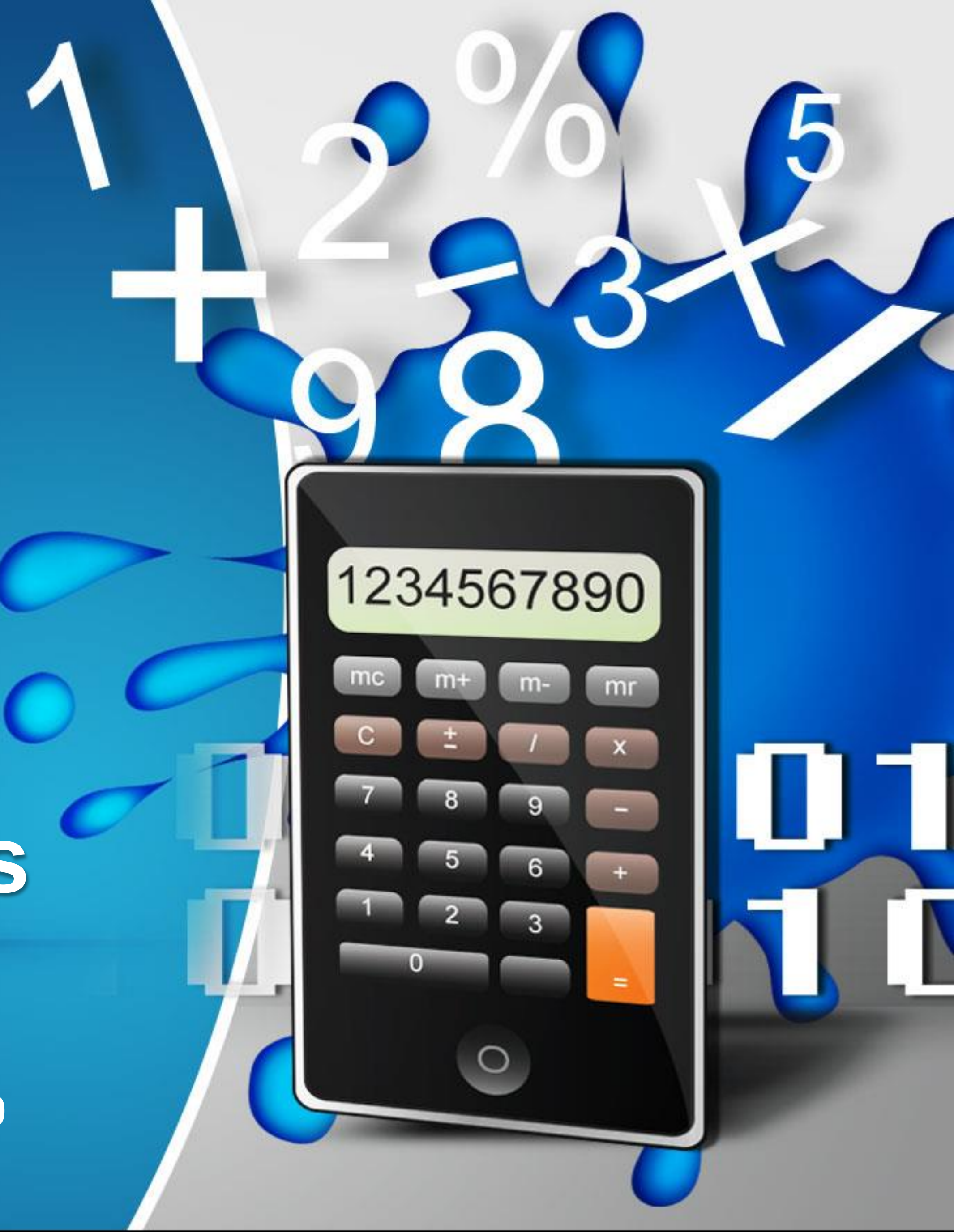
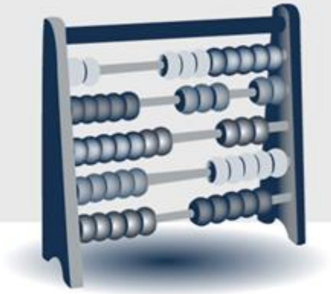
Matrix Calculus

Po-Chen Wu

Media IC and System Lab
GIEE, NTU



Outline

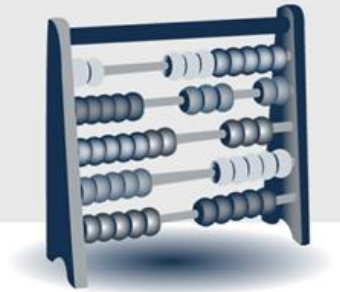


- Introduction
- Derivatives in Numerator-Layout Notation
- Derivatives in Denominator-Layout Notation
- Identities

Outline

- Introduction
  - Notation
  - Matrix Operations
- Derivatives in Numerator-Layout Notation
- Derivatives in Denominator-Layout Notation
- Identities

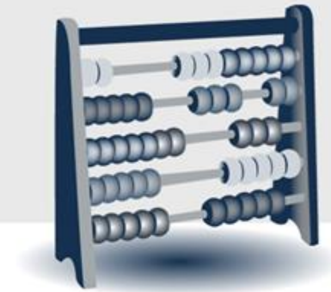
Matrix Calculus



- Matrix calculus is a specialized notation for doing multivariable calculus, especially over spaces of matrices.
- Two competing notational conventions split the field of matrix calculus into two separate groups.

Numerator-Layout Notation	**Denominator-Layout Notation**
$\frac{\partial \mathbf{y}}{\partial \mathrm{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}, \frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}$	$\frac{\partial \mathbf{y}}{\partial \mathrm{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix}, \frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$

Notation



- Matrix: $\mathbf{A}, \mathbf{X}, \mathbf{Y}$
  - bold capital letter
- Vector: $\mathbf{a}, \mathbf{x}, \mathbf{y}$ (**column**)
  - boldface lowercase letter
- Scalar: a, x, y
  - lowercase itatilc typeface
- Transpose: $\mathbf{A}^T$, $\mathbf{a}^T$
- Trace: $\mathrm{tr}(\mathbf{A})$
  - $\mathrm{tr}(\mathbf{A}) = A_{11} + A_{22} + \cdots + A_{nn} = \sum_{i=1}^{n} A_{ii}$
- Determinant: $\det(\mathbf{A})$

n columns

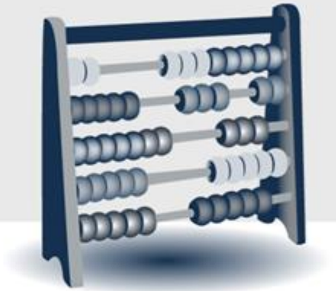
m rows

$$\underset{m \times n \text{ matrix}}{\mathbf{A}} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

$m \times 1$ vector

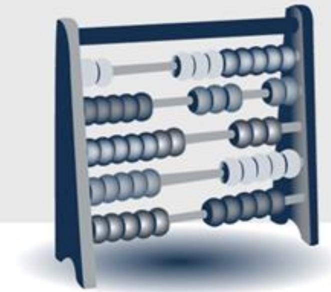
1×1 scalar

Properties of Transpose



- $(\mathbf{A}^T)^T = \mathbf{A}$
- $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
- $(\mathbf{A} + \mathbf{B} + \mathbf{C})^T = \mathbf{A}^T + \mathbf{B}^T + \mathbf{C}^T$
- $(r\mathbf{A})^T = r\mathbf{A}^T$
- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
- $(\mathbf{ABC})^T = \mathbf{C}^T\mathbf{B}^T\mathbf{A}^T$

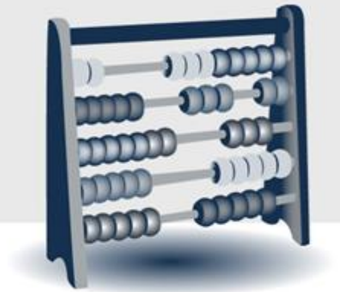
Trace & Determinant



- If $\mathbf{A}$ is a square n−by−n matrix and if $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of $\mathbf{A}$, then
  - $\mathrm{tr}(\mathbf{A}) = A_{11} + A_{22} + \cdots + A_{nn} = \sum_{i=1}^{n} A_{ii} = \sum_{i=1}^{n} \lambda_i$
  - $\det(\mathbf{A}) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} M_{i,j} = \prod_{i=1}^{n} \lambda_i$
    - Minor $M_{i,j}$: the determinant of the $(n-1) \times (n-1)$-matrix that results from $\mathbf{A}$ by removing the ith row and the jth column.
    - Cofactor $C_{i,j}$: $(-1)^{i+j} M_{i,j}$
    - Cofactor matrix: $\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}$

$$\mathbf{A}^{-1} = \frac{(\mathbf{C})^T}{\det(\mathbf{A})}$$

Hadamard Product



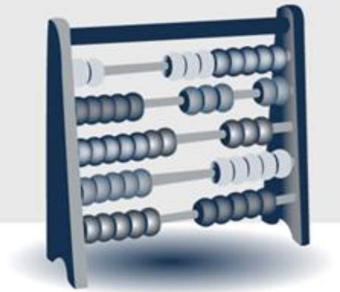
- For two matrices, $\mathbf{A}$, $\mathbf{B}$, of the same dimension, $m \times n$ the Hadamard product, $\mathbf{A} \circ \mathbf{B}$, is a matrix, of the same dimension as the operands, with elements given by

$$(\mathbf{A} \circ \mathbf{B})_{i,j} = (\mathbf{A})_{i,j} \cdot (\mathbf{B})_{i,j}$$

  - For example the Hadamard product for a 3×3 matrix $\mathbf{A}$ with a 3×3 matrix $\mathbf{B}$ is:

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \circ \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} & A_{12}B_{12} & A_{13}B_{13} \\ A_{21}B_{21} & A_{22}B_{22} & A_{23}B_{23} \\ A_{31}B_{31} & A_{32}B_{32} & A_{33}B_{33} \end{bmatrix}$$

Kronecker Product



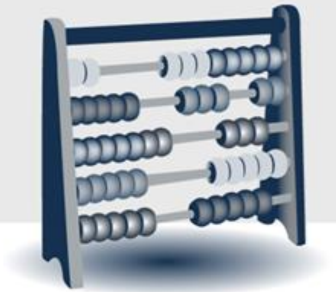
- If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is a $p \times q$ matrix, then the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $mp \times nq$ block matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & \cdots & A_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ A_{m1}\mathbf{B} & \cdots & A_{mn}\mathbf{B} \end{bmatrix}$$

  – For example, the Kronecker product for a 2×2 matrix $\mathbf{A}$ with a 2×3 matrix $\mathbf{B}$ is:

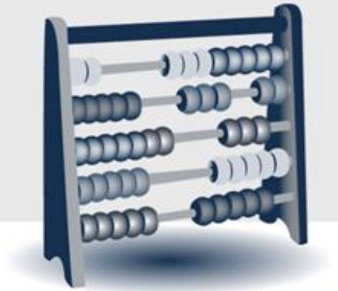
$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{11}B_{11} & A_{11}B_{12} & A_{11}B_{13} & A_{12}B_{11} & A_{12}B_{12} & A_{12}B_{13} \\ A_{11}B_{21} & A_{11}B_{22} & A_{11}B_{23} & A_{12}B_{21} & A_{12}B_{22} & A_{12}B_{23} \\ A_{21}B_{11} & A_{21}B_{12} & A_{21}B_{13} & A_{22}B_{11} & A_{22}B_{12} & A_{22}B_{13} \\ A_{21}B_{21} & A_{21}B_{22} & A_{21}B_{23} & A_{22}B_{21} & A_{22}B_{22} & A_{22}B_{23} \end{bmatrix}$$

Outline



- Introduction
- Derivatives in Numerator-Layout Notation
  - List of Differentiation
  - Derivative Formulas
  - Chain rule
  - The Matrix Differential
- Derivatives in Denominator-Layout Notation
- Identities

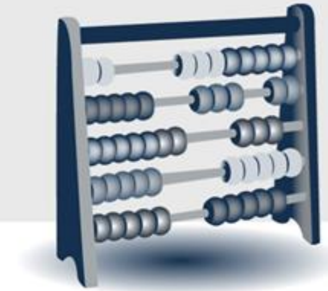
Vector-by-Scalar



- The derivative of $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$, by x is written as:

$$\frac{\partial \mathbf{y}}{\partial \mathrm{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}$$

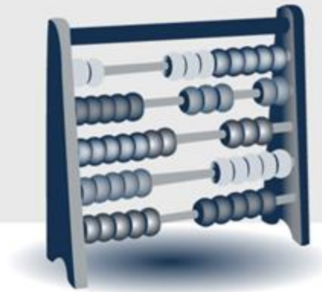
Scalar-by-Vector



- The derivative of y by $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is written as:

$$\frac{\partial \mathrm{y}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}$$

Vector-by-Vector

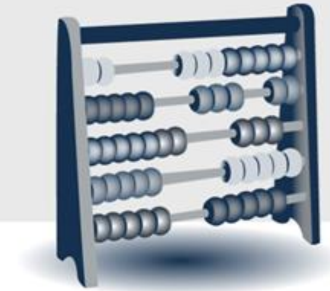


- The derivative of $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$ with respect to $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \overset{\text{def}}{=} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & & \frac{\partial y_2}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

  – Also known as the Jacobian matrix

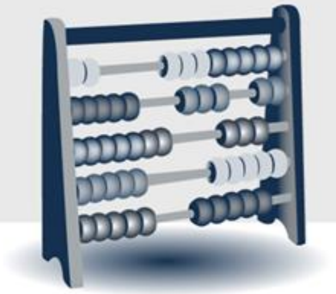
Example 1



- Given $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, and $y_1 = x_1^2 - 2x_2$, $y_2 = x_3^2 - 4x_2$, the Jacobian matrix $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ is:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \frac{\partial y_1}{\partial x_3} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 & -2 & 0 \\ 0 & -4 & 2x_3 \end{bmatrix}$$

Example 2



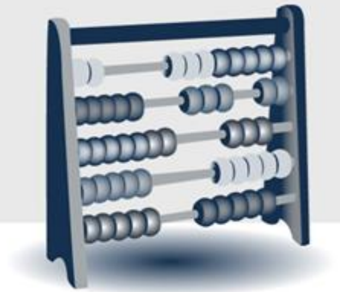
- The transformation from spherical to Cartesian coordinates is defined by

$$x = r\sin\theta\cos\phi\,, y = r\sin\theta\sin\phi\,, z = r\cos\theta$$

  – Let $\mathbf{y} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} r \\ \theta \\ \phi \end{bmatrix}$, the Jacobian matrix $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ is:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ \cos\theta & -r\sin\theta & 0 \end{bmatrix}$$

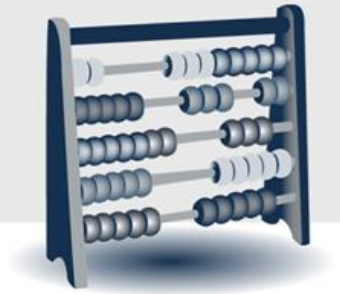
Matrix-by-Scalar



- The derivative of a matrix function $\mathbf{Y}$ by a scalar x is known as the tangent matrix and is given by

$$\frac{\partial \mathbf{Y}}{\partial x} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial Y_{11}}{\partial x} & \frac{\partial Y_{12}}{\partial x} & \cdots & \frac{\partial Y_{1n}}{\partial x} \\ \frac{\partial Y_{21}}{\partial x} & \frac{\partial Y_{22}}{\partial x} & & \frac{\partial Y_{2n}}{\partial x} \\ & \vdots & \ddots & \vdots \\ \frac{\partial Y_{m1}}{\partial x} & \frac{\partial Y_{m2}}{\partial x} & \cdots & \frac{\partial Y_{mn}}{\partial x} \end{bmatrix}$$

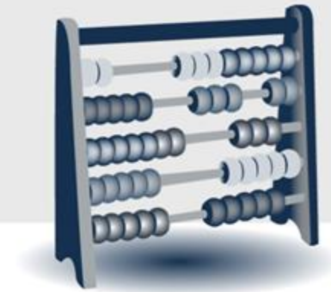
Scalar-by-Matrix



- The derivative of a scalar y function by a matrix $\mathbf{X}$ is known as the gradient matrix and is given by

$$\frac{\partial y}{\partial \mathbf{X}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y}{\partial X_{11}} & \frac{\partial y}{\partial X_{21}} & \cdots & \frac{\partial y}{\partial X_{m1}} \\ \frac{\partial y}{\partial X_{12}} & \frac{\partial y}{\partial X_{22}} & \cdots & \frac{\partial y}{\partial X_{m2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial X_{1n}} & \frac{\partial y}{\partial X_{2n}} & \cdots & \frac{\partial y}{\partial X_{mn}} \end{bmatrix}$$

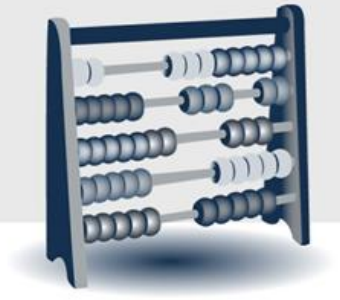
List of Differentiation



- Result of differentiating various kinds of aggregates with other kinds of aggregates.

	Scalar y		Vector $\mathbf{y}$ (size m)		Matrix $\mathbf{Y}$ (size $m \times n$)	
	Notation	Type	Notation	Type	Notation	Type
Scalar x	$\frac{\partial y}{\partial x}$	scalar	$\frac{\partial \mathbf{y}}{\partial x}$	size-m column vector	$\frac{\partial \mathbf{Y}}{\partial x}$	$m \times n$ matrix
Vector $\mathbf{x}$ (size n)	$\frac{\partial y}{\partial \mathbf{x}}$	size-n row vector	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	$m \times n$ matrix	$\frac{\partial \mathbf{Y}}{\partial \mathbf{x}}$	–
Matrix $\mathbf{X}$ (size $p \times q$)	$\frac{\partial y}{\partial \mathbf{X}}$	$q \times p$ matrix	$\frac{\partial \mathbf{y}}{\partial \mathbf{X}}$	–	$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$	–

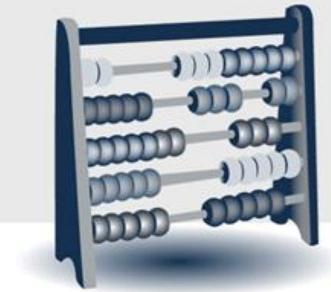
Derivative Formulas



$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$
$\mathbf{Ax}$	$\mathbf{A}$
$\mathbf{x}^T\mathbf{A}$	$\mathbf{A}^T$
$\mathbf{x}^T\mathbf{x}$	$2\mathbf{x}^T$
$\mathbf{x}^T\mathbf{Ax}$	$\mathbf{x}^T\mathbf{A} + \mathbf{x}^T\mathbf{A}^T$

- Hint: Derive $\mathbf{x}$
  - If you have to differentiate $\mathbf{x}^T$, transpose the rest.
  - If you have two $\mathbf{x}$-terms, differentiate them separately in turn and then sum up the two derivatives.

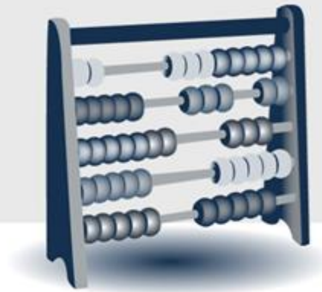
Chain Rule (1/2)



- Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_r \end{bmatrix}$, where $\mathbf{z}$ is a function of $\mathbf{y}$, which is in turn a function of $\mathbf{x}$. Then

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \cdots & \frac{\partial z_1}{\partial x_n} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & & \frac{\partial z_2}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial z_r}{\partial x_1} & \frac{\partial z_r}{\partial x_2} & \cdots & \frac{\partial z_r}{\partial x_n} \end{bmatrix}, \qquad \text{where } \frac{\partial z_i}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial z_i}{\partial y_k} \frac{\partial y_k}{\partial x_j} \quad \begin{cases} i = 1,2,\cdots,r \\ j = 1,2,\cdots,n \end{cases}$$

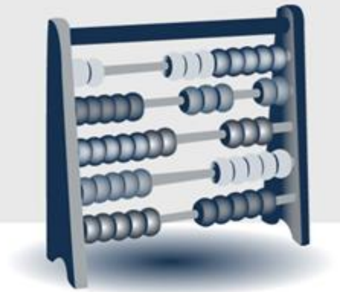
Chain Rule (2/2)



$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \begin{bmatrix} \sum \frac{\partial z_1}{\partial y_k}\frac{\partial y_k}{\partial x_1} & \sum \frac{\partial z_1}{\partial y_k}\frac{\partial y_k}{\partial x_2} & \cdots & \sum \frac{\partial z_1}{\partial y_k}\frac{\partial y_k}{\partial x_n} \\ \sum \frac{\partial z_2}{\partial y_k}\frac{\partial y_k}{\partial x_1} & \sum \frac{\partial z_2}{\partial y_k}\frac{\partial y_k}{\partial x_2} & & \sum \frac{\partial z_2}{\partial y_k}\frac{\partial y_k}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \sum \frac{\partial z_r}{\partial y_k}\frac{\partial y_k}{\partial x_1} & \sum \frac{\partial z_r}{\partial y_k}\frac{\partial y_k}{\partial x_2} & \cdots & \sum \frac{\partial z_r}{\partial y_k}\frac{\partial y_k}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_m} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & & \frac{\partial z_2}{\partial y_m} \\ \vdots & & \ddots & \vdots \\ \frac{\partial z_r}{\partial y_1} & \frac{\partial z_r}{\partial y_2} & \cdots & \frac{\partial z_r}{\partial y_m} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & & \frac{\partial y_2}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} = \frac{\partial \mathbf{z}}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$$

Exercise 1 (Numerator-Layout)



- Find $\mathbf{w}^*$ to minimize $E(\mathbf{w})$, where

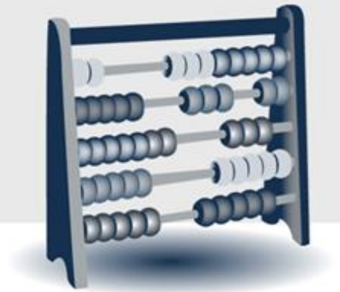
$$E(\mathbf{w}) = \frac{1}{N}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2 = \frac{1}{N}(\mathbf{X}\mathbf{w} - \mathbf{y})^T(\mathbf{X}\mathbf{w} - \mathbf{y})$$

$$= \frac{1}{N}(\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} - 2\mathbf{w}^T\mathbf{X}^T\mathbf{y} + \mathbf{y}^T\mathbf{y})$$

  – Assume $\mathbf{X}^T\mathbf{X}$ is invertible.

$$\mathbf{w}^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

The Matrix Differential (1/2)



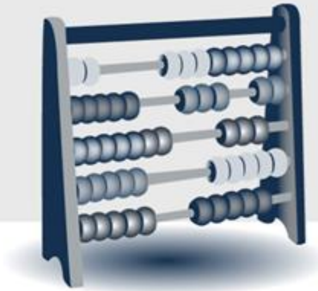
- For a scalar function $f(\mathbf{x})$, where $\mathbf{x}$ is an n-vector, the ordinary differential of multivariate calculus is defined as

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$$

- In harmony with this formula, we define the differential of an $m \times n$ matrix $\mathbf{X} = [X_{ij}]$ to be

$$d\mathbf{X} \stackrel{\text{def}}{=} \begin{bmatrix} dX_{11} & dX_{12} & \cdots & dX_{1n} \\ dX_{21} & dX_{22} & & dX_{2n} \\ \vdots & & \ddots & \vdots \\ dX_{m1} & dX_{m2} & \cdots & dX_{mn} \end{bmatrix}$$

The Matrix Differential (2/2)



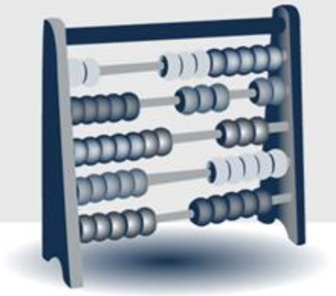
- This definition complies with the multiplicative and associative rules

$$d(\alpha\mathbf{X}) = \alpha d\mathbf{X} \qquad\qquad d(\mathbf{X}+\mathbf{Y}) = d\mathbf{X} + d\mathbf{Y}$$

- If $\mathbf{X}$ and $\mathbf{Y}$ are product-conforming matrices, it can be verified that the differential of their product is

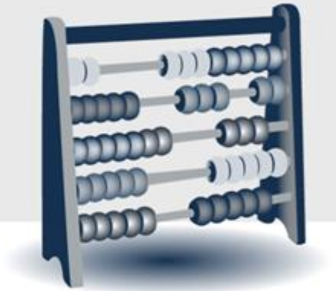
$$d(\mathbf{XY}) = (d\mathbf{X})\mathbf{Y} + \mathbf{X}(d\mathbf{Y})$$

Summary of Numerator-Layout



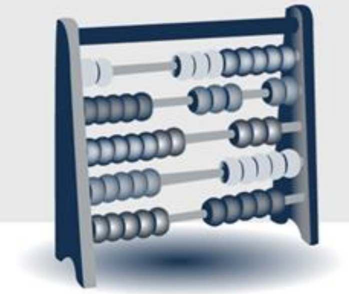
- Three straightforward key points:
  1. Ice cream
     - ➢ Derivatives
  2. Derive **x**
     - ➢ Derivative Formulas
  3. Chain rule
     - ➢ From left to right

Outline



- Introduction
- Derivatives in Numerator-Layout Notation
- Derivatives in Denominator-Layout Notation
  - List of Differentiation
  - Derivative Formulas
  - Chain Rule
- Identities

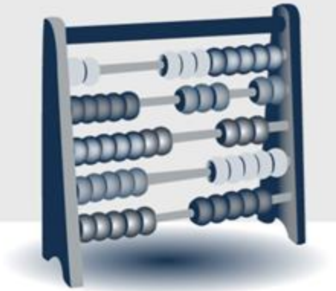
List of Differentiation



- The results of operations will be transposed when switching from numerator-layout notation to denominator-layout notation.

	Scalar y		Vector $\mathbf{y}$ (size m)		Matrix $\mathbf{Y}$ (size $m \times n$)	
	Notation	Type	Notation	Type	Notation	Type
Scalar x	$\frac{\partial y}{\partial x}$	scalar	$\frac{\partial \mathbf{y}}{\partial x}$	size-m row vector	$\frac{\partial \mathbf{Y}}{\partial x}$	–
Vector $\mathbf{x}$ (size n)	$\frac{\partial y}{\partial \mathbf{x}}$	size-n column rector	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	$n \times m$ matrix	$\frac{\partial \mathbf{Y}}{\partial \mathbf{x}}$	–
Matrix $\mathbf{X}$ (size $p \times q$)	$\frac{\partial y}{\partial \mathbf{X}}$	$p \times q$ matrix	$\frac{\partial \mathbf{y}}{\partial \mathbf{X}}$	–	$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$	–

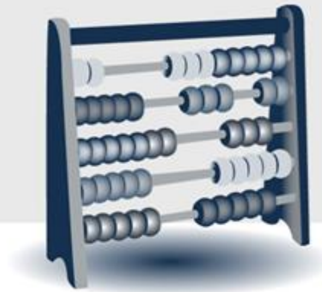
Derivative Formulas



$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$
$\mathbf{Ax}$	$\mathbf{A}^T$
$\mathbf{x}^T\mathbf{A}$	$\mathbf{A}$
$\mathbf{x}^T\mathbf{x}$	$2\mathbf{x}$
$\mathbf{x}^T\mathbf{Ax}$	$\mathbf{Ax} + \mathbf{A}^T\mathbf{x}$

- Hint: Derive $\mathbf{x}^T$
  - If you have to differentiate $\mathbf{x}$, transpose the rest.
  - If you have two $\mathbf{x}$-terms, differentiate them separately in turn and then sum up the two derivatives.

Chain Rule

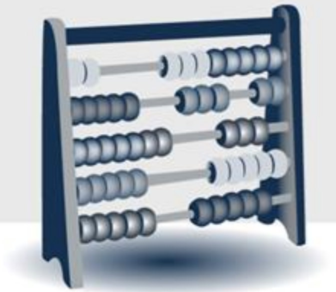


- Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_r \end{bmatrix}$, where $\mathbf{z}$ is a function of $\mathbf{y}$, which is in turn a function of $\mathbf{x}$. Then

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \frac{\partial \mathbf{z}}{\partial \mathbf{y}}$$

  – the chain of matrices builds “toward the left.”

Exercise 2 (Denominator-Layout)



- Find $\mathbf{w}^*$ to minimize $E(\mathbf{w})$, where

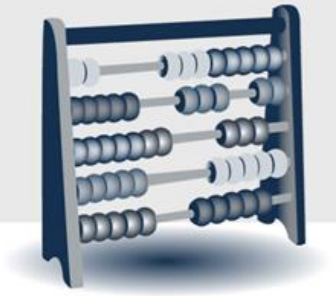
$$E(\mathbf{w}) = \frac{1}{N}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2 = \frac{1}{N}(\mathbf{X}\mathbf{w} - \mathbf{y})^T(\mathbf{X}\mathbf{w} - \mathbf{y})$$

$$= \frac{1}{N}(\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} - 2\mathbf{w}^T\mathbf{X}^T\mathbf{y} + \mathbf{y}^T\mathbf{y})$$

  - Assume $\mathbf{X}^T\mathbf{X}$ is invertible.

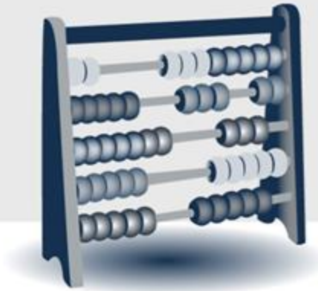
$$\mathbf{w}^* = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Outline



- Introduction
- Derivatives in Numerator-Layout Notation
- Derivatives in Denominator-Layout Notation
- Identities
  - Vector-by-Vector
  - Scalar-by-Vector
  - Vector-by-Scalar
  - Scalar-by-Matrix
  - Matrix-by-Scalar

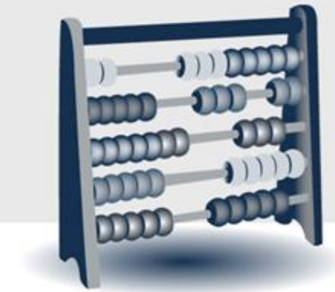
Vector-by-Vector



- Identities: vector-by-vector $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

Condition	Expression	Numerator layout	Denominator layout
$a = a(\mathbf{x})$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\frac{\partial a}{\partial \mathbf{x}}$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial a}{\partial \mathbf{x}}\mathbf{u}^T$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial(\mathbf{u} + \mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

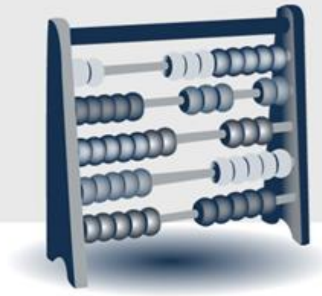
Scalar-by-Vector



- Identities: scalar-by-vector $\frac{\partial y}{\partial \mathbf{x}}$

Condition	Expression	Numerator layout	Denominator layout
$u = u(\mathbf{x})$	$\frac{\partial f(g(u))}{\partial \mathbf{x}} =$	$\frac{\partial f(g)}{\partial g} \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial \mathbf{u}^T \mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^T \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^T \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \mathbf{u}$
$\mathbf{A} \neq \mathbf{A}(\mathbf{x})$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial \mathbf{u}^T \mathbf{A} \mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^T \mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^T \mathbf{A}^T \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{A} \mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \mathbf{A}^T \mathbf{u}$
$\mathbf{a} \neq \mathbf{a}(\mathbf{x})$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{a}^T \mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{a}^T \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{a}$

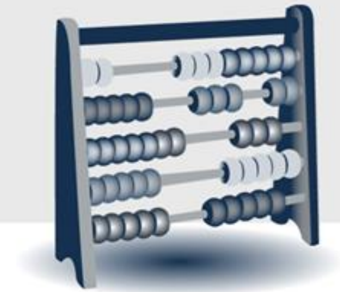
Vector-by-Scalar



- Identities: vector-by-scalar $\frac{\partial \mathbf{y}}{\partial x}$

Condition	Expression	Numerator layout	Denominator layout
$\mathbf{u} = \mathbf{u}(x)$, $\mathbf{v} = \mathbf{v}(x)$	$\frac{\partial(\mathbf{u}+\mathbf{v})}{\partial x} =$	$\frac{\partial \mathbf{u}}{\partial x} + \frac{\partial \mathbf{v}}{\partial x}$	
$\mathbf{u} = \mathbf{u}(x)$, $\mathbf{v} = \mathbf{v}(x)$	$\frac{\partial(\mathbf{u}\times\mathbf{v})}{\partial x} =$	$\mathbf{u} \times \frac{\partial \mathbf{v}}{\partial x} + \frac{\partial \mathbf{u}}{\partial x} \times \mathbf{v}$	
$\mathbf{u} = \mathbf{u}(x)$	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial x} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial x}$	$\frac{\partial \mathbf{u}}{\partial x} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

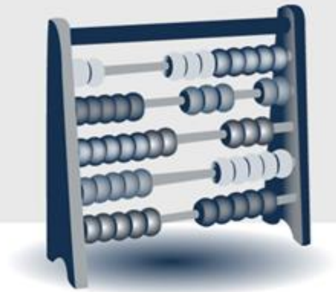
Scalar-by-Matrix



- Identities: scalar-by-matrix $\frac{\partial y}{\partial \mathbf{X}}$

Condition	Expression	Numerator layout	Denominator layout
$u = u(\mathbf{X})$, $v = v(\mathbf{X})$	$\frac{\partial(u+v)}{\partial \mathbf{X}} =$	$\frac{\partial u}{\partial \mathbf{X}} + \frac{\partial v}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$, $v = v(\mathbf{X})$	$\frac{\partial uv}{\partial \mathbf{X}}$	$u\frac{\partial v}{\partial \mathbf{X}} + v\frac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$	$\frac{\partial f(g(u))}{\partial \mathbf{X}} =$	$\frac{\partial f(g)}{\partial g}\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{X}}$	

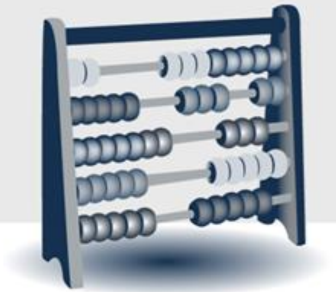
Matrix-by-Scalar



- Identities: scalar-by-matrix $\frac{\partial \mathbf{Y}}{\partial x}$

Condition	Expression	Numerator layout
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\frac{\partial(\mathbf{U}+\mathbf{V})}{\partial x} =$	$\frac{\partial \mathbf{U}}{\partial x} + \frac{\partial \mathbf{V}}{\partial x}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\frac{\partial(\mathbf{U}\mathbf{V})}{\partial x} =$	$\mathbf{U}\frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x}\mathbf{V}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\frac{\partial(\mathbf{U}\circ\mathbf{V})}{\partial x} =$	$\mathbf{U}\circ\frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x}\circ\mathbf{V}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\frac{\partial(\mathbf{U}\otimes\mathbf{V})}{\partial x} =$	$\mathbf{U}\otimes\frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x}\otimes\mathbf{V}$

Reference



- Matrix calculus, *Wiki*

  http://en.wikipedia.org/wiki/Matrix_calculus

- The Matrix Cookbook

  http://www.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf