Deep Learning

Lecture 3: Calculus

Dr. Mehrdad Maleki

A function $f : X \to Y$ is given.

A function $f : X \to Y$ is given. Then the derivative of f at $x_0 \in X$ means that if x_0 changes a little bit, how $f(x_0)$ changes accordingly.

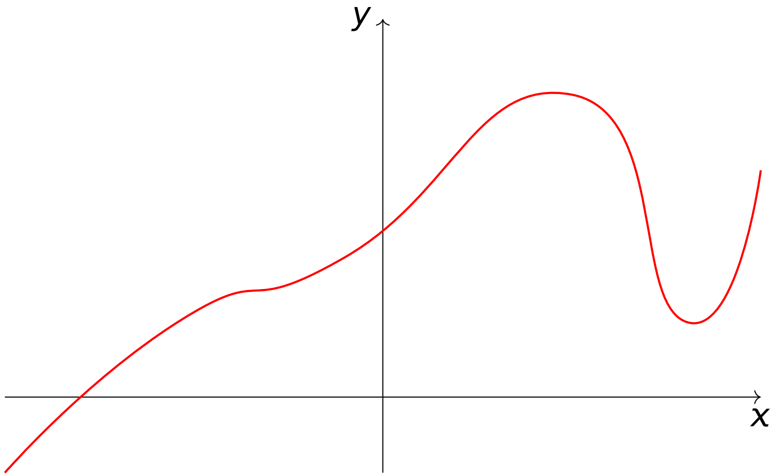
A function $f : X \to Y$ is given. Then the derivative of f at $x_0 \in X$ means that if x_0 changes a little bit, how $f(x_0)$ changes accordingly. For example if the radius of a balloon of radius 50 cm change to 55 cm, how much change we have in the volume of the baloon?

A function $f : X \to Y$ is given. Then the derivative of f at $x_0 \in X$ means that if x_0 changes a little bit, how $f(x_0)$ changes accordingly. For example if the radius of a balloon of radius 50 cm change to 55 cm, how much change we have in the volume of the baloon? Since $V = \frac{4}{3}\pi R^3$ then $\frac{dV}{dR} = 4\pi R^2$.

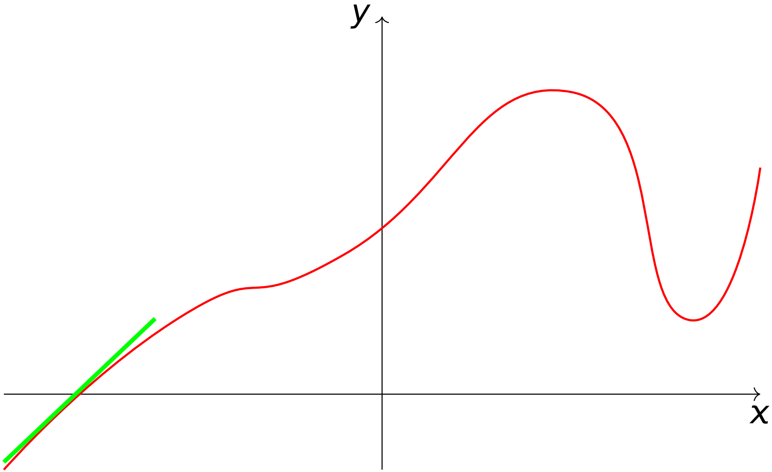
A function $f : X \to Y$ is given. Then the derivative of f at $x_0 \in X$ means that if x_0 changes a little bit, how $f(x_0)$ changes accordingly. For example if the radius of a balloon of radius 50 cm change to 55 cm, how much change we have in the volume of the baloon? Since $V = \frac{4}{3}\pi R^3$ then $\frac{dV}{dR} = 4\pi R^2$. So $V_2 - V_1 \approx \frac{dV}{dR}(R_2 - R_1)$.

A function $f : X \to Y$ is given. Then the derivative of f at $x_0 \in X$ means that if x_0 changes a little bit, how $f(x_0)$ changes accordingly. For example if the radius of a balloon of radius 50 cm change to 55 cm, how much change we have in the volume of the baloon? Since $V = \frac{4}{3}\pi R^3$ then $\frac{dV}{dR} = 4\pi R^2$. So $V_2 - V_1 \approx \frac{dV}{dR}(R_2 - R_1)$. Since $R_2 - R_1 = 5$ and $\frac{dV}{dR}|_{R=50} = 4\pi \times 50^2 \approx 31400$. So $V_2 - V_1 \approx 31400 * 5 = 157000\, cm^3$ but $V_1 \approx 52333$ and $V_2 \approx 157000 + 52333$ thus $\frac{V_2}{V_1} = \frac{209333}{52333} \approx 4$. Hence, the volume of the sphere increases by a factor of 4 if the radius of the sphere is increased by 5 cm.

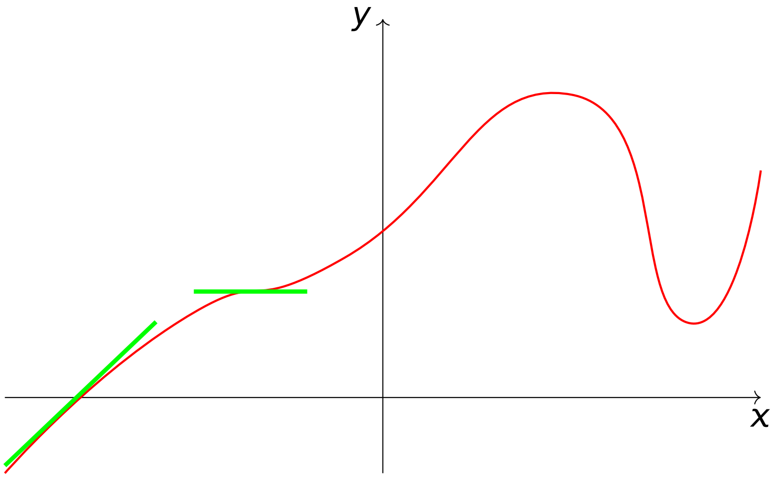
In one dimensional case when we have a scalar function of one variable $y = f(x)$ we can plot this on the $x - y$ coordinate and the slop of to the graph of the function at point x_0 is the derivative of f at x_0 and it is denoted by $f'(x_0)$ or $\frac{dy}{dx}\Big|_{x=x_0}$.



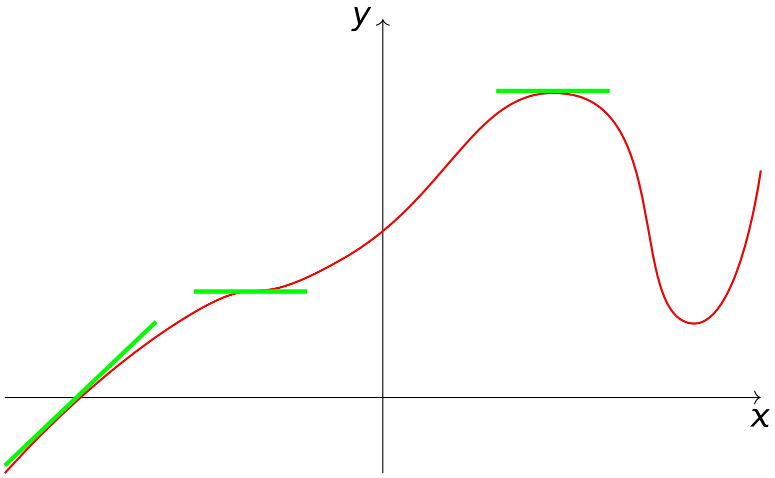
In one dimensional case when we have a scalar function of one variable $y = f(x)$ we can plot this on the $x - y$ coordinate and the slop of to the graph of the function at point x_0 is the derivative of f at x_0 and it is denoted by $f'(x_0)$ or $\frac{dy}{dx}\Big|_{x=x_0}$.



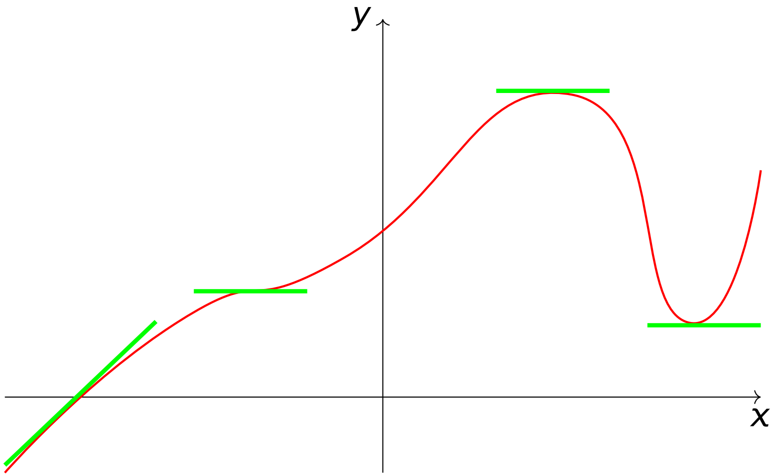
In one dimensional case when we have a scalar function of one variable $y = f(x)$ we can plot this on the $x-y$ coordinate and the slop of to the graph of the function at point x_0 is the derivative of f at x_0 and it is denoted by $f'(x_0)$ or $\frac{dy}{dx}\Big|_{x=x_0}$.



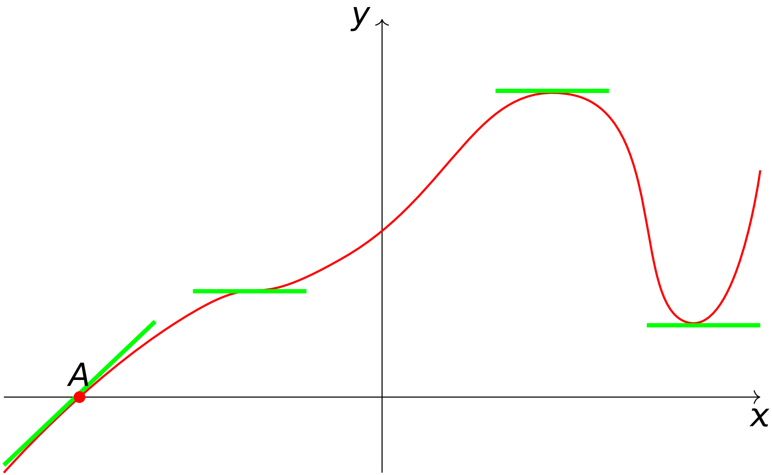
In one dimensional case when we have a scalar function of one variable $y = f(x)$ we can plot this on the $x-y$ coordinate and the slop of to the graph of the function at point x_0 is the derivative of f at x_0 and it is denoted by $f'(x_0)$ or $\left.\frac{dy}{dx}\right|_{x=x_0}$.



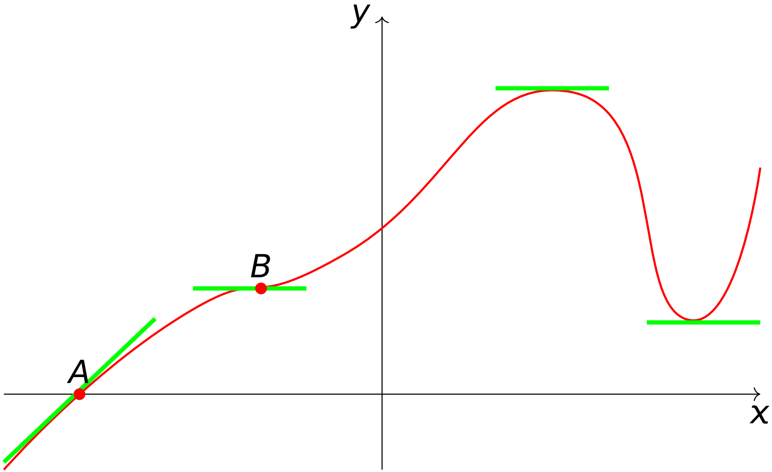
In one dimensional case when we have a scalar function of one variable $y = f(x)$ we can plot this on the $x - y$ coordinate and the slop of to the graph of the function at point x_0 is the derivative of f at x_0 and it is denoted by $f'(x_0)$ or $\frac{dy}{dx}\Big|_{x=x_0}$.



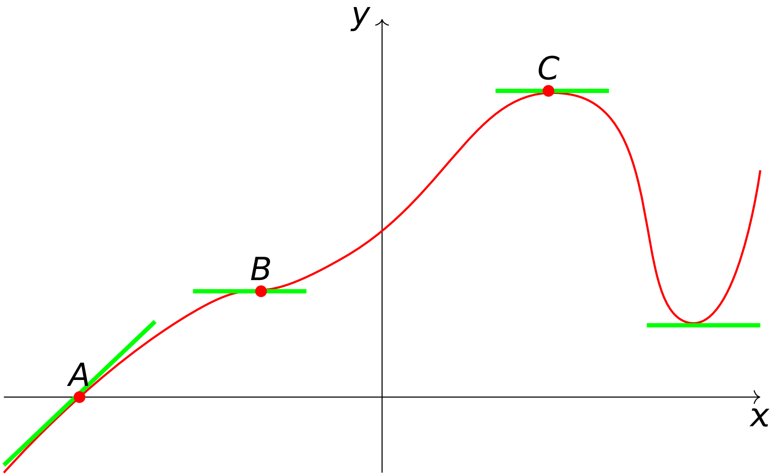
In one dimensional case when we have a scalar function of one variable $y = f(x)$ we can plot this on the $x - y$ coordinate and the slop of to the graph of the function at point x_0 is the derivative of f at x_0 and it is denoted by $f'(x_0)$ or $\frac{dy}{dx}\Big|_{x=x_0}$.



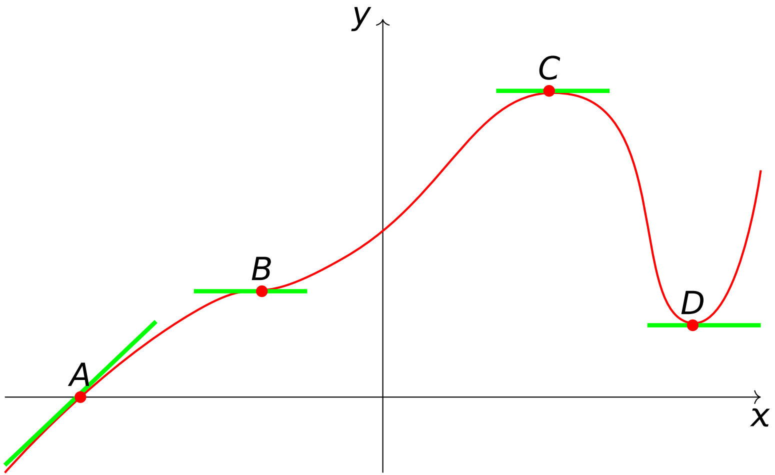
In one dimensional case when we have a scalar function of one variable $y = f(x)$ we can plot this on the $x - y$ coordinate and the slop of to the graph of the function at point x_0 is the derivative of f at x_0 and it is denoted by $f'(x_0)$ or $\frac{dy}{dx}\Big|_{x=x_0}$.



In one dimensional case when we have a scalar function of one variable $y = f(x)$ we can plot this on the $x - y$ coordinate and the slop of to the graph of the function at point x_0 is the derivative of f at x_0 and it is denoted by $f'(x_0)$ or $\frac{dy}{dx}\Big|_{x=x_0}$.



In one dimensional case when we have a scalar function of one variable $y = f(x)$ we can plot this on the $x - y$ coordinate and the slop of to the graph of the function at point x_0 is the derivative of f at x_0 and it is denoted by $f'(x_0)$ or $\left.\frac{dy}{dx}\right|_{x=x_0}$.



Chain Rule

If $y = f(z)$ and $z = g(x)$ then $y = h(x)$. The derivative of y with respect to x is,

Chain Rule

If $y = f(z)$ and $z = g(x)$ then $y = h(x)$. The derivative of y with respect to x is,

$$\frac{dy}{dx} =$$

Chain Rule

If $y = f(z)$ and $z = g(x)$ then $y = h(x)$. The derivative of y with respect to x is,

$$\frac{dy}{dx} = \frac{dy}{dz}$$

Chain Rule

If $y = f(z)$ and $z = g(x)$ then $y = h(x)$. The derivative of y with respect to x is,

$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$$

Chain Rule

If $y = f(z)$ and $z = g(x)$ then $y = h(x)$. The derivative of y with respect to x is,

$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$$

or equivalently,

$$h'(x) =$$

Chain Rule

If $y = f(z)$ and $z = g(x)$ then $y = h(x)$. The derivative of y with respect to x is,

$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$$

or equivalently,

$$h'(x) = f'(z)$$

Chain Rule

If $y = f(z)$ and $z = g(x)$ then $y = h(x)$. The derivative of y with respect to x is,

$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$$

or equivalently,

$$h'(x) = f'(z)g'(x)$$

Chain Rule

If $y = f(z)$ and $z = g(x)$ then $y = h(x)$. The derivative of y with respect to x is,

$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$$

or equivalently,

$$\begin{aligned} h'(x) &= f'(z)g'(x) \\ &= f'(g(x)) \end{aligned}$$

Chain Rule

If $y = f(z)$ and $z = g(x)$ then $y = h(x)$. The derivative of y with respect to x is,

$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}$$

or equivalently,

$$\begin{aligned} h'(x) &= f'(z)g'(x) \\ &= f'(g(x))g'(x) \end{aligned}$$

Important Functions

$$f(x) = x^n$$

Important Functions

$$f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$$

Important Functions

$$f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$$

$$f(x) = \sin(x)$$

Important Functions

$$f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$$

$$f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$$

Important Functions

$$f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$$
$$f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$$
$$f(x) = \cos(x)$$

Important Functions

$$f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$$

$$f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$$

$$f(x) = \cos(x) \Rightarrow f'(x) = -\sin(x)$$

Important Functions

$$f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$$

$$f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$$

$$f(x) = \cos(x) \Rightarrow f'(x) = -\sin(x)$$

$$f(x) = e^x$$

Important Functions

$$f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$$
$$f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$$
$$f(x) = \cos(x) \Rightarrow f'(x) = -\sin(x)$$
$$f(x) = e^x \Rightarrow f'(x) = e^x$$

Important Functions

$$f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$$
$$f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$$
$$f(x) = \cos(x) \Rightarrow f'(x) = -\sin(x)$$
$$f(x) = e^x \Rightarrow f'(x) = e^x$$

$$f(x) = \log(x)$$

Important Functions

$$f(x) = x^n \Rightarrow f'(x) = nx^{n-1}$$
$$f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$$
$$f(x) = \cos(x) \Rightarrow f'(x) = -\sin(x)$$
$$f(x) = e^x \Rightarrow f'(x) = e^x$$
$$f(x) = \log(x) \Rightarrow f'(x) = \frac{1}{x}$$

Derivative Ruels

$$(u(x) + v(x))'$$

Derivative Ruels

$$(u(x) + v(x))' = u'(x) + v'(x)$$

Derivative Ruels

$$(u(x) + v(x))' = u'(x) + v'(x)$$

$$(u(x)v(x))'$$

Derivative Ruels

$$(u(x) + v(x))' = u'(x) + v'(x)$$
$$(u(x)v(x))' = u'(x)v(x) + u(x)v'(x)$$

Derivative Ruels

$$(u(x)+v(x))' = u'(x)+v'(x)$$
$$(u(x)v(x))' = u'(x)v(x)+u(x)v'(x)$$
$$(cu(x))'$$

Derivative Ruels

$$(u(x) + v(x))' = u'(x) + v'(x)$$
$$(u(x)v(x))' = u'(x)v(x) + u(x)v'(x)$$
$$(cu(x))' = cu'(x)$$

Chain Rule of multi-variables

If $y = f(z_1, \ldots, z_n)$ and $z_i = g_i(x)$ then $y = h(x)$ and we have,

Chain Rule of multi-variables

If $y = f(z_1, \ldots, z_n)$ and $z_i = g_i(x)$ then $y = h(x)$ and we have,

$$\frac{dy}{dx} =$$

Chain Rule of multi-variables

If $y = f(z_1, \ldots, z_n)$ and $z_i = g_i(x)$ then $y = h(x)$ and we have,

$$\frac{dy}{dx} = \frac{\partial y}{\partial z_1}\frac{dz_1}{dx}$$

Chain Rule of multi-variables

If $y = f(z_1, \ldots, z_n)$ and $z_i = g_i(x)$ then $y = h(x)$ and we have,

$$\frac{dy}{dx} = \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \cdots +$$

Chain Rule of multi-variables

If $y = f(z_1, \ldots, z_n)$ and $z_i = g_i(x)$ then $y = h(x)$ and we have,

$$\frac{dy}{dx} = \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \cdots + \frac{\partial y}{\partial z_n}\frac{dz_n}{dx}$$

Chain Rule of multi-variables

If $y = f(z_1, \ldots, z_n)$ and $z_i = g_i(x)$ then $y = h(x)$ and we have,

$$\begin{aligned} \frac{dy}{dx} &= \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \cdots + \frac{\partial y}{\partial z_n}\frac{dz_n}{dx} \\ &= \frac{\partial y}{\partial z_1} g_1'(x) \end{aligned}$$

Chain Rule of multi-variables

If $y = f(z_1, \dots, z_n)$ and $z_i = g_i(x)$ then $y = h(x)$ and we have,

$$\begin{aligned}\frac{dy}{dx} &= \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \cdots + \frac{\partial y}{\partial z_n}\frac{dz_n}{dx} \\ &= \frac{\partial y}{\partial z_1} g_1'(x) + \cdots +\end{aligned}$$

Chain Rule of multi-variables

If $y = f(z_1, \ldots, z_n)$ and $z_i = g_i(x)$ then $y = h(x)$ and we have,

$$\begin{aligned}\frac{dy}{dx} &= \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \cdots + \frac{\partial y}{\partial z_n}\frac{dz_n}{dx} \\ &= \frac{\partial y}{\partial z_1} g_1'(x) + \cdots + \frac{\partial y}{\partial z_n} g_n'(x)\end{aligned}$$

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x)$, $z_2 = sin(x)$ then $\left.\frac{dy}{dx}\right|_{x=1}$ is

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x)$, $z_2 = sin(x)$ then $\frac{dy}{dx}\Big|_{x=1}$ is

$$\frac{dy}{dx} =$$

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x)$, $z_2 = sin(x)$ then $\left.\frac{dy}{dx}\right|_{x=1}$ is

$$\frac{dy}{dx} = \frac{\partial y}{\partial z_1}\frac{dz_1}{dx}$$

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x), z_2 = sin(x)$ then $\frac{dy}{dx}\Big|_{x=1}$ is

$$\frac{dy}{dx} = \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \frac{\partial y}{\partial z_2}\frac{dz_2}{dx}$$

$$=$$

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x)$, $z_2 = sin(x)$ then $\frac{dy}{dx}\Big|_{x=1}$ is

$$\frac{dy}{dx} = \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \frac{\partial y}{\partial z_2}\frac{dz_2}{dx}$$

$$= (2z_1)(\frac{1}{x})$$

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x)$, $z_2 = sin(x)$ then $\frac{dy}{dx}\Big|_{x=1}$ is

$$
\begin{aligned}
\frac{dy}{dx} &= \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \frac{\partial y}{\partial z_2}\frac{dz_2}{dx} \\
&= (2z_1)(\frac{1}{x}) + (2z_2)(cos(x)) \\
&=
\end{aligned}
$$

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x)$, $z_2 = sin(x)$ then $\frac{dy}{dx}\Big|_{x=1}$ is

$$\begin{aligned}
\frac{dy}{dx} &= \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \frac{\partial y}{\partial z_2}\frac{dz_2}{dx} \\
&= (2z_1)(\frac{1}{x}) + (2z_2)(cos(x)) \\
&= (2 \times 0)(\frac{1}{1})
\end{aligned}$$

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x), z_2 = sin(x)$ then $\frac{dy}{dx}\Big|_{x=1}$ is

$$\begin{aligned}
\frac{dy}{dx} &= \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \frac{\partial y}{\partial z_2}\frac{dz_2}{dx} \\
&= (2z_1)(\frac{1}{x}) + (2z_2)(cos(x)) \\
&= (2 \times 0)(\frac{1}{1}) + (2 \times \sin 1)(\cos 1) \\
&=
\end{aligned}$$

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x), z_2 = sin(x)$ then $\frac{dy}{dx}\Big|_{x=1}$ is

$$\frac{dy}{dx} = \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \frac{\partial y}{\partial z_2}\frac{dz_2}{dx}$$

$$= (2z_1)(\frac{1}{x}) + (2z_2)(cos(x))$$

$$= (2 \times 0)(\frac{1}{1}) + (2 \times \sin 1)(\cos 1)$$

$$= sin(2)$$

$$\approx$$

Example

Let $y = z_1^2 + z_2^2$ and $z_1 = log(x)$, $z_2 = sin(x)$ then $\frac{dy}{dx}\Big|_{x=1}$ is

$$\begin{aligned}
\frac{dy}{dx} &= \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \frac{\partial y}{\partial z_2}\frac{dz_2}{dx} \\
&= (2z_1)(\frac{1}{x}) + (2z_2)(cos(x)) \\
&= (2 \times 0)(\frac{1}{1}) + (2 \times \sin 1)(\cos 1) \\
&= sin(2) \\
&\approx 0.034899497
\end{aligned}$$

If $f : \mathbb{R}^n \rightarrow \mathbb{R}$, i.e., $y = f(x_1, \ldots, x_n)$, i.e., scalar y depends on reals $x_1, \ldots, x_n$, then the **Gradient** of f at $\mathbf{x_0}$ is

If $f : \mathbb{R}^n \to \mathbb{R}$, i.e., $y = f(x_1, \ldots, x_n)$, i.e., scalar y depends on reals $x_1, \ldots, x_n$, then the **Gradient** of f at $\mathbf{x_0}$ is

$$\nabla f(\mathbf{x_0}) = \begin{bmatrix} \left.\frac{\partial y}{\partial x_1}\right|_{\mathbf{x_0}} \\ \vdots \\ \left.\frac{\partial y}{\partial x_n}\right|_{\mathbf{x_0}} \end{bmatrix}$$

$y = x_1 + x_2$, then if $y = sum(x_1, x_2)$ then the Gradient of sum at $\mathbf{x_0} = (2, 3)$ is,

$y = x_1 + x_2$, then if $y = sum(x_1, x_2)$ then the Gradient of sum at $\mathbf{x_0} = (2, 3)$ is,

$$\nabla sum(2, 3) =$$

$y = x_1 + x_2$, then if $y = sum(x_1, x_2)$ then the Gradient of sum at $\mathbf{x_0} = (2, 3)$ is,

$$\nabla sum(2,3) = \begin{bmatrix} \frac{\partial sum}{\partial x_1} \\ \frac{\partial sum}{\partial x_2} \end{bmatrix}$$

$y = x_1 + x_2$, then if $y = sum(x_1, x_2)$ then the Gradient of sum at $\mathbf{x_0} = (2, 3)$ is,

$$\nabla sum(2,3) = \begin{bmatrix} \frac{\partial sum}{\partial x_1} \\ \frac{\partial sum}{\partial x_2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Directional Derivative

Let $\mathbf{u} \in \mathbb{R}^n$ and $\|\mathbf{u}\| = 1$, then the **Directional Derivative** of f in the direction of $\|\mathbf{u}\|$ in $\mathbf{x_0}$ is denoted by $\mathcal{D}_{\mathbf{u}} f(\mathbf{x_0})$ and is define by,

Directional Derivative

Let $\mathbf{u} \in \mathbb{R}^n$ and $\|\mathbf{u}\| = 1$, then the **Directional Derivative** of f in the direction of $\|\mathbf{u}\|$ in $\mathbf{x_0}$ is denoted by $\mathcal{D}_{\mathbf{u}} f(\mathbf{x_0})$ and is define by,

$$\mathcal{D}_{\mathbf{u}} f(\mathbf{x_0}) = \nabla f(\mathbf{x_0}) \cdot \mathbf{u}$$

In which direction, we have maximum Directional Derivative?

In which direction, we have maximum Directional Derivative? If we choose $\mathbf{u} = \frac{\nabla f(\mathbf{x_0})}{\|\nabla f(\mathbf{x_0})\|}$ then we have maximum directional derivative and

In which direction, we have maximum Directional Derivative? If we choose $\mathbf{u} = \frac{\nabla f(\mathbf{x_0})}{\|\nabla f(\mathbf{x_0})\|}$ then we have maximum directional derivative and in the direction of $-\mathbf{u}$ we have minimum directional derivative.

What is the directional derivative of the funcion $f(x_1, x_2) = x_1^2 + x_2^2$ in the direction of $\mathbf{u} = \frac{1}{25}(3, 4)$ in $\mathbf{x_0} = (1, 1)$?

What is the directional derivative of the funcion $f(x_1, x_2) = x_1^2 + x_2^2$ in the direction of $\mathbf{u} = \frac{1}{25}(3, 4)$ in $\mathbf{x_0} = (1, 1)$?

Since $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$ then,

What is the directional derivative of the funcion $f(x_1, x_2) = x_1^2 + x_2^2$ in the direction of $\mathbf{u} = \frac{1}{25}(3, 4)$ in $\mathbf{x_0} = (1, 1)$?

Since $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$ then,

$$\mathcal{D}_{\mathbf{u}} f(x_1, x_2) = \nabla f(x_1, x_2) \cdot \mathbf{u}$$

What is the directional derivative of the funcion $f(x_1, x_2) = x_1^2 + x_2^2$ in the direction of $\mathbf{u} = \frac{1}{25}(3, 4)$ in $\mathbf{x_0} = (1, 1)$?

Since $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$ then,

$$\begin{aligned} \mathcal{D}_{\mathbf{u}} f(x_1, x_2) &= \nabla f(x_1, x_2) \cdot \mathbf{u} \\ &= \begin{bmatrix} 2 \\ 2 \end{bmatrix} \cdot \frac{1}{25} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \end{aligned}$$

What is the directional derivative of the funcion $f(x_1, x_2) = x_1^2 + x_2^2$ in the direction of $\mathbf{u} = \frac{1}{25}(3, 4)$ in $\mathbf{x_0} = (1, 1)$?

Since $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$ then,

$$\begin{aligned} \mathcal{D}_{\mathbf{u}} f(x_1, x_2) &= \nabla f(x_1, x_2) \cdot \mathbf{u} \\ &= \begin{bmatrix} 2 \\ 2 \end{bmatrix} \cdot \frac{1}{25} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\ &= \frac{14}{25} \end{aligned}$$

What is the directional derivative of the funcion $f(x_1, x_2) = x_1^2 + x_2^2$ in the direction of $\mathbf{u} = \frac{1}{25}(3, 4)$ in $\mathbf{x_0} = (1, 1)$?

Since $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$ then,

$$\begin{aligned} \mathcal{D}_{\mathbf{u}} f(x_1, x_2) &= \nabla f(x_1, x_2) \cdot \mathbf{u} \\ &= \begin{bmatrix} 2 \\ 2 \end{bmatrix} \cdot \frac{1}{25} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\ &= \frac{14}{25} \\ &= 0.56 \end{aligned}$$

What is the directional derivative of the funcion $f(x_1, x_2) = x_1^2 + x_2^2$ in the direction of $\mathbf{u} = \frac{1}{25}(3, 4)$ in $\mathbf{x_0} = (1, 1)$?

Since $\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$ then,

$$
\begin{aligned}
\mathcal{D}_{\mathbf{u}} f(x_1, x_2) &= \nabla f(x_1, x_2) \cdot \mathbf{u} \\
&= \begin{bmatrix} 2 \\ 2 \end{bmatrix} \cdot \frac{1}{25} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\
&= \frac{14}{25} \\
&= 0.56
\end{aligned}
$$

while $\mathcal{D}_{\frac{\nabla f(\mathbf{x_0})}{\|\nabla f(\mathbf{x_0})\|}} f(x_1, x_2) = \|\nabla f(\mathbf{x_0})\| \approx 2.82$

Jacobian

For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, such that $(y_1, \ldots, y_m) = f(x_1, \ldots, x_n)$, the **Jacobian** of f at $\mathbf{x_0}$ is a $m \times n$ matrix and denoted by $\mathbf{J}_f(\mathbf{x_0})$ and is defined as follow:

Jacobian

For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, such that $(y_1, \ldots, y_m) = f(x_1, \ldots, x_n)$, the **Jacobian** of f at $\mathbf{x_0}$ is a $m \times n$ matrix and denoted by $\mathbf{J}_f(\mathbf{x_0})$ and is defined as follow:

$$\mathbf{J}_f(\mathbf{x_0}) =$$

Jacobian

For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, such that $(y_1, \ldots, y_m) = f(x_1, \ldots, x_n)$, the **Jacobian** of f at $\mathbf{x_0}$ is a $m \times n$ matrix and denoted by $\mathbf{J}_f(\mathbf{x_0})$ and is defined as follow:

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1}\Big|_{\mathbf{x_0}} & \cdots & \frac{\partial y_1}{\partial x_n}\Big|_{\mathbf{x_0}} \\ & & \\ & & \end{bmatrix}$$

Jacobian

For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, such that $(y_1, \ldots, y_m) = f(x_1, \ldots, x_n)$, the **Jacobian** of f at $\mathbf{x_0}$ is a $m \times n$ matrix and denoted by $\mathbf{J}_f(\mathbf{x_0})$ and is defined as follow:

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1}\Big|_{\mathbf{x_0}} & \cdots & \frac{\partial y_1}{\partial x_n}\Big|_{\mathbf{x_0}} \\ \vdots & \ddots & \vdots \\ & & \end{bmatrix}$$

Jacobian

For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, such that $(y_1, \dots, y_m) = f(x_1, \dots, x_n)$, the **Jacobian** of f at $\mathbf{x_0}$ is a $m \times n$ matrix and denoted by $\mathbf{J}_f(\mathbf{x_0})$ and is defined as follow:

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1}\Big|_{\mathbf{x_0}} & \cdots & \frac{\partial y_1}{\partial x_n}\Big|_{\mathbf{x_0}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1}\Big|_{\mathbf{x_0}} & \cdots & \frac{\partial y_m}{\partial x_n}\Big|_{\mathbf{x_0}} \end{bmatrix}$$

Jacobian

For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, such that $(y_1, \ldots, y_m) = f(x_1, \ldots, x_n)$, the **Jacobian** of f at $\mathbf{x_0}$ is a $m \times n$ matrix and denoted by $\mathbf{J}_f(\mathbf{x_0})$ and is defined as follow:

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \left.\frac{\partial y_1}{\partial x_1}\right|_{\mathbf{x_0}} & \cdots & \left.\frac{\partial y_1}{\partial x_n}\right|_{\mathbf{x_0}} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial y_m}{\partial x_1}\right|_{\mathbf{x_0}} & \cdots & \left.\frac{\partial y_m}{\partial x_n}\right|_{\mathbf{x_0}} \end{bmatrix}$$

another notation,

Jacobian

For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, such that $(y_1, \ldots, y_m) = f(x_1, \ldots, x_n)$, the **Jacobian** of f at $\mathbf{x_0}$ is a $m \times n$ matrix and denoted by $\mathbf{J}_f(\mathbf{x_0})$ and is defined as follow:

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1}\Big|_{\mathbf{x_0}} & \cdots & \frac{\partial y_1}{\partial x_n}\Big|_{\mathbf{x_0}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1}\Big|_{\mathbf{x_0}} & \cdots & \frac{\partial y_m}{\partial x_n}\Big|_{\mathbf{x_0}} \end{bmatrix}$$

another notation,

$$\mathbf{J}_f(\mathbf{x_0}) =$$

Jacobian

For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, such that $(y_1, \ldots, y_m) = f(x_1, \ldots, x_n)$, the **Jacobian** of f at $\mathbf{x_0}$ is a $m \times n$ matrix and denoted by $\mathbf{J}_f(\mathbf{x_0})$ and is defined as follow:

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1}\Big|_{\mathbf{x_0}} & \cdots & \frac{\partial y_1}{\partial x_n}\Big|_{\mathbf{x_0}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1}\Big|_{\mathbf{x_0}} & \cdots & \frac{\partial y_m}{\partial x_n}\Big|_{\mathbf{x_0}} \end{bmatrix}$$

another notation,

$$\mathbf{J}_f(\mathbf{x_0}) = \frac{\partial(y_1, \ldots, y_m)}{\partial(x_1, \ldots, x_n)}\Big|_{\mathbf{x_0}}$$

When $m = 1$ then the Jacobian of $f : \mathbb{R}^n \to \mathbb{R}$ is a row vector and its transpos is the gradient of f, i.e., When $m = 1$ then the Jacobian of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a row vector and its transpos is the gradient of f, i.e.,

$$\mathbf{J}_f(\mathbf{x}) = (\nabla f(\mathbf{x}))^T$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2, 3)$?

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2, 3)$?

$$\mathbf{J}_f(2, 3) =$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2, 3)$?

$$\mathbf{J}_f(2,3) = \begin{bmatrix} \frac{\partial(x_1 x_2)}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(x_1 x_2)}{\partial x_2}\Big|_{(2,3)} \\ & \\ & \end{bmatrix}$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2, 3)$?

$$\mathbf{J}_f(2,3) = \begin{bmatrix} \left.\frac{\partial(x_1x_2)}{\partial x_1}\right|_{(2,3)} & \left.\frac{\partial(x_1x_2)}{\partial x_2}\right|_{(2,3)} \\ \left.\frac{\partial(sin(x_1))}{\partial x_1}\right|_{(2,3)} & \left.\frac{\partial(sin(x_1))}{\partial x_2}\right|_{(2,3)} \\ & \end{bmatrix}$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2, 3)$?

$$\mathbf{J}_f(2, 3) = \begin{bmatrix} \frac{\partial(x_1x_2)}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(x_1x_2)}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(sin(x_1))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(sin(x_1))}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(cos(x_2))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(cos(x_2))}{\partial x_2}\Big|_{(2,3)} \end{bmatrix}$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2,3)$?

$$\mathbf{J}_f(2,3) = \begin{bmatrix} \frac{\partial(x_1x_2)}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(x_1x_2)}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(sin(x_1))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(sin(x_1))}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(cos(x_2))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(cos(x_2))}{\partial x_2}\Big|_{(2,3)} \end{bmatrix}$$

$$= \Bigg[\quad x_2\Big|_{(2,3)} \qquad x_1\Big|_{(2,3)}$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2, 3)$?

$$\mathbf{J}_f(2,3) = \begin{bmatrix} \frac{\partial(x_1 x_2)}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(x_1 x_2)}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(sin(x_1))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(sin(x_1))}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(cos(x_2))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(cos(x_2))}{\partial x_2}\Big|_{(2,3)} \end{bmatrix}$$

$$= \left[\begin{matrix} x_2\Big|_{(2,3)} & x_1\Big|_{(2,3)} \\ cos(x_1)\Big|_{(2,3)} & 0\Big|_{(2,3)} \\ & \end{matrix} \right.$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2, 3)$?

$$\mathbf{J}_f(2,3) = \begin{bmatrix} \frac{\partial(x_1 x_2)}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(x_1 x_2)}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(sin(x_1))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(sin(x_1))}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(cos(x_2))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(cos(x_2))}{\partial x_2}\Big|_{(2,3)} \end{bmatrix}$$

$$= \begin{bmatrix} x_2\Big|_{(2,3)} & x_1\Big|_{(2,3)} \\ cos(x_1)\Big|_{(2,3)} & 0\Big|_{(2,3)} \\ 0\Big|_{(2,3)} & -sin(x_2)\Big|_{(2,3)} \end{bmatrix}$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2,3)$?

$$\mathbf{J}_f(2,3) = \begin{bmatrix} \frac{\partial(x_1x_2)}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(x_1x_2)}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(sin(x_1))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(sin(x_1))}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(cos(x_2))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(cos(x_2))}{\partial x_2}\Big|_{(2,3)} \end{bmatrix}$$

$$= \begin{bmatrix} x_2\Big|_{(2,3)} & x_1\Big|_{(2,3)} \\ cos(x_1)\Big|_{(2,3)} & 0\Big|_{(2,3)} \\ 0\Big|_{(2,3)} & -sin(x_2)\Big|_{(2,3)} \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 2 \\ & \\ & \end{bmatrix}$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2, 3)$?

$$\mathbf{J}_f(2,3) = \begin{bmatrix} \frac{\partial(x_1x_2)}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(x_1x_2)}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(sin(x_1))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(sin(x_1))}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(cos(x_2))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(cos(x_2))}{\partial x_2}\Big|_{(2,3)} \end{bmatrix}$$

$$= \begin{bmatrix} x_2\Big|_{(2,3)} & x_1\Big|_{(2,3)} \\ cos(x_1)\Big|_{(2,3)} & 0\Big|_{(2,3)} \\ 0\Big|_{(2,3)} & -sin(x_2)\Big|_{(2,3)} \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 2 \\ 0.99 & 0 \\ & \end{bmatrix}$$

Let $f(x_1, x_2) = (x1x2, sin(x_1), cos(x_2))$. What is $\mathbf{J}_f(2,3)$?

$$\mathbf{J}_f(2,3) = \begin{bmatrix} \frac{\partial(x_1 x_2)}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(x_1 x_2)}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(sin(x_1))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(sin(x_1))}{\partial x_2}\Big|_{(2,3)} \\ \frac{\partial(cos(x_2))}{\partial x_1}\Big|_{(2,3)} & \frac{\partial(cos(x_2))}{\partial x_2}\Big|_{(2,3)} \end{bmatrix}$$

$$= \begin{bmatrix} x_2\Big|_{(2,3)} & x_1\Big|_{(2,3)} \\ cos(x_1)\Big|_{(2,3)} & 0\Big|_{(2,3)} \\ 0\Big|_{(2,3)} & -sin(x_2)\Big|_{(2,3)} \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 2 \\ 0.99 & 0 \\ 0 & -0.05 \end{bmatrix}$$

Let define $f(\mathbf{x}) = \mathbf{x}^T \mathbf{a}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \rightarrow \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

Let define $f(\mathbf{x}) = \mathbf{x}^T\mathbf{a}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \rightarrow \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x_0}) =$$

Let define $f(\mathbf{x}) = \mathbf{x}^T\mathbf{a}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \rightarrow \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x_0}) = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}$$

Let define $f(\mathbf{x}) = \mathbf{x}^T \mathbf{a}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \rightarrow \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x_0}) = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Let define $f(\mathbf{x}) = \mathbf{x}^T\mathbf{a}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \to \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

$$
\begin{aligned}
f(\mathbf{x_0}) &= \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \\
&= x_1 a_1 + \cdots + x_n a_n
\end{aligned}
$$

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_1} & \ldots & \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_n} \end{bmatrix}$$

$$
\begin{aligned}
\mathbf{J}_f(\mathbf{x_0}) &= \begin{bmatrix} \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_1} & \dots & \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{J}_f(\mathbf{x_0}) &= \begin{bmatrix} \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_1} & \dots & \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} \\
&= \mathbf{a}^T
\end{aligned}
$$

Let define $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \to \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

Let define $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \to \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x_0}) =$$

Let define $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \rightarrow \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x_0}) = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix}$$

Let define $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \to \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x_0}) = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Let define $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$, where $\mathbf{a} : \mathbb{R}^n$ is a constant vector. Then $f : \mathbb{R}^n \to \mathbb{R}$. What is the Jacobian of f at $\mathbf{x_0}$?

$$
\begin{aligned}
f(\mathbf{x_0}) &= \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= x_1 a_1 + \cdots + x_n a_n
\end{aligned}
$$

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_1} & \ldots & \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_n} \end{bmatrix}$$

$$
\begin{aligned}
\mathbf{J}_f(\mathbf{x_0}) &= \begin{bmatrix} \frac{\partial(x_1a_1+\cdots+x_na_n)}{\partial x_1} & \dots & \frac{\partial(x_1a_1+\cdots+x_na_n)}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{J}_f(\mathbf{x_0}) &= \begin{bmatrix} \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_1} & \ldots & \frac{\partial(x_1 a_1 + \cdots + x_n a_n)}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} a_1 & \ldots & a_n \end{bmatrix} \\
&= \mathbf{a}^T
\end{aligned}
$$

Let $\mathbf{W}_{m \times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \to \mathbb{R}^m$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x}) =$$

Let $\mathbf{W}_{m \times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \to \mathbb{R}^m$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ & & \\ & & \end{bmatrix}$$

Let $\mathbf{W}_{m\times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \to \mathbb{R}^m$.
What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ & & \end{bmatrix}$$

Let $\mathbf{W}_{m \times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \to \mathbb{R}^m$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix}$$

Let $\mathbf{W}_{m \times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \\ \\ \end{bmatrix}$$

Let $\mathbf{W}_{m \times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \to \mathbb{R}^m$. What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \left[\begin{matrix} x_1 \\ \vdots \\ \end{matrix}\right.$$

Let $\mathbf{W}_{m \times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \to \mathbb{R}^m$.
What is the Jacobian of f at $\mathbf{x_0}$?

$$f(\mathbf{x}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Let $\mathbf{W}_{m \times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \to \mathbb{R}^m$. What is the Jacobian of f at $\mathbf{x_0}$?

$$
\begin{aligned}
f(\mathbf{x}) &= \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} w_{11}x_1 + \cdots + w_{1n}x_n \\ \\ \\ \end{bmatrix}
\end{aligned}
$$

Let $\mathbf{W}_{m\times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \to \mathbb{R}^m$. What is the Jacobian of f at $\mathbf{x_0}$?

$$
\begin{aligned}
f(\mathbf{x}) &= \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \left[\begin{matrix} w_{11}x_1 + \cdots + w_{1n}x_n \\ \vdots \\ \\ \end{matrix} \right.
\end{aligned}
$$

Let $\mathbf{W}_{m \times n}$ be a constant matrix and let define $f(\mathbf{x}) = \mathbf{W}\mathbf{x}$. Then $f : \mathbb{R}^n \to \mathbb{R}^m$. What is the Jacobian of f at $\mathbf{x_0}$?

$$
\begin{aligned}
f(\mathbf{x}) &= \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} w_{11}x_1 + \cdots + w_{1n}x_n \\ \vdots \\ w_{m1}x_1 + \cdots + w_{mn}x_n \end{bmatrix}
\end{aligned}
$$

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_1} & \cdots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_n} \\ & & \\ & & \end{bmatrix}$$

$$\mathbf{J}_f(\mathbf{x_0}) = \left[\begin{array}{ccc} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_1} & \cdots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ & & \end{array} \right.$$

$$\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_1} & \cdots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial x_1} & \cdots & \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial x_n} \end{bmatrix}$$

$$
\begin{aligned}
\mathbf{J}_f(\mathbf{x_0}) &= \begin{bmatrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_1} & \dots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial x_1} & \dots & \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial x_n} \end{bmatrix} \\
&= \left[\begin{matrix} w_{11} & \dots & w_{1n} \\ & & \\ & & \end{matrix} \right.
\end{aligned}
$$

$$
\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial(w_{11}x_1 + \cdots + w_{1n}x_n)}{\partial x_1} & \cdots & \frac{\partial(w_{11}x_1 + \cdots + w_{1n}x_n)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1 + \cdots + w_{mn}x_n)}{\partial x_1} & \cdots & \frac{\partial(w_{m1}x_1 + \cdots + w_{mn}x_n)}{\partial x_n} \end{bmatrix}
$$

$$
= \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ & & \end{bmatrix}
$$

$$
\mathbf{J}_f(\mathbf{x_0}) = \begin{bmatrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_1} & \dots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial x_1} & \dots & \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial x_n} \end{bmatrix}
$$

$$
= \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix}
$$

$$
\begin{aligned}
\mathbf{J}_f(\mathbf{x_0}) &= \begin{bmatrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_1} & \cdots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial x_1} & \cdots & \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \\
&= \mathbf{W}
\end{aligned}
$$

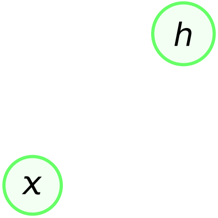
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic map defined by $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x}$, where A is a symmetric matrix of dimension $(n \times n)$.

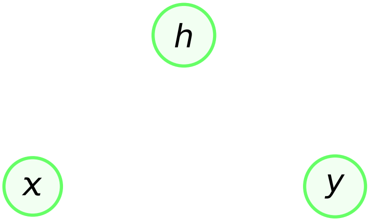
Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a quadratic map defined by $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$, where A is a symmetric matrix of dimension $(n \times n)$. Since matrix multiplication is associative we can see f as the multiplication of two matrix $\mathbf{x}^T$ and $\mathbf{y}(\mathbf{x}) = \mathbf{A}\mathbf{x}$.

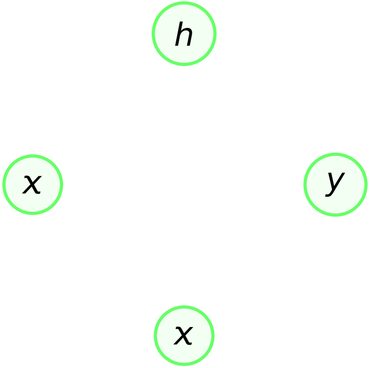
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic map defined by $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$, where A is a symmetric matrix of dimension $(n \times n)$. Since matrix multiplication is associative we can see f as the multiplication of two matrix $\mathbf{x}^T$ and $\mathbf{y}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Note that the matrix multiplication is exactly the function composition!

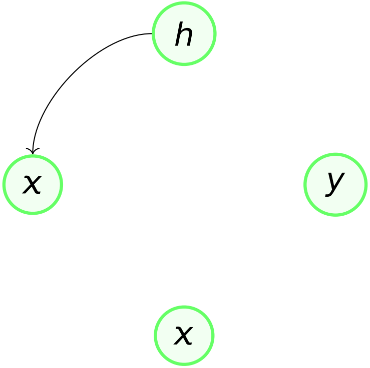
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic map defined by $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x}$, where A is a symmetric matrix of dimension $(n \times n)$. Since matrix multiplication is associative we can see f as the multiplication of two matrix $\mathbf{x}^T$ and $\mathbf{y}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Note that the matrix multiplication is exactly the function composition! So $f(\mathbf{x}) = h(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$ and if we apply the chaine rule to h we have:

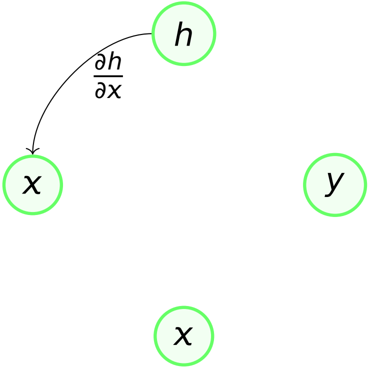

h

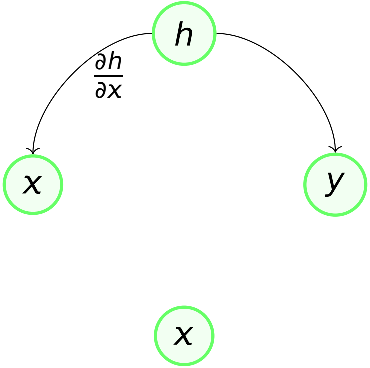

h
x


h
x
y

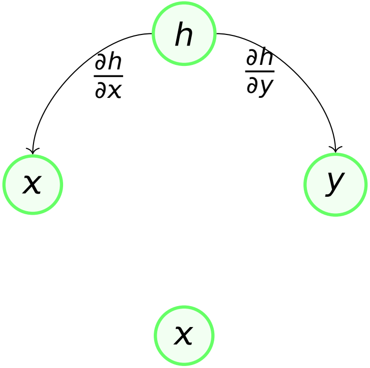

h
x
y
x

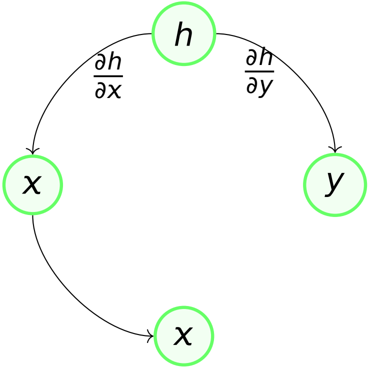

h
x
y
x


h
$\frac{\partial h}{\partial x}$
x
y
x

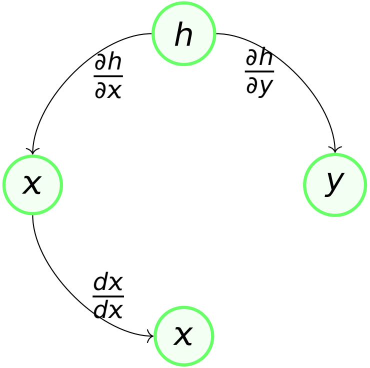


h
∂h/∂x
x
y
x

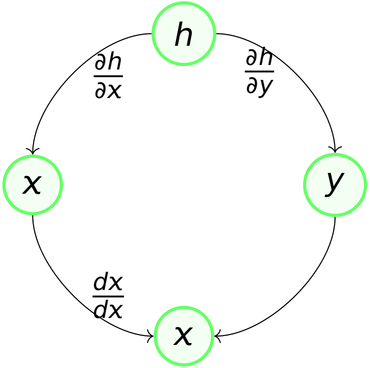

h
∂h/∂x
∂h/∂y
x
y
x



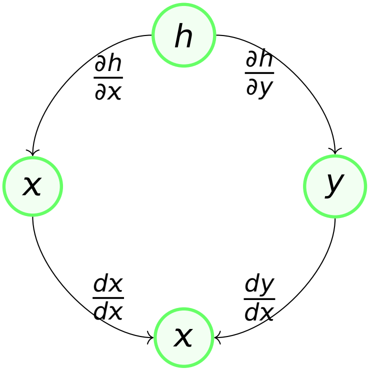
h
$\frac{\partial h}{\partial x}$
$\frac{\partial h}{\partial y}$
x
y
x



h
x
y
x
∂h/∂x
∂h/∂y
dx/dx



h
$\frac{\partial h}{\partial x}$
$\frac{\partial h}{\partial y}$
x
y
$\frac{dx}{dx}$
x


h
x
y
x
$\frac{\partial h}{\partial x}$
$\frac{\partial h}{\partial y}$
$\frac{dx}{dx}$
$\frac{dy}{dx}$

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial h(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}}$$

$$
\begin{aligned}
\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \frac{\partial h(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \\
&= \frac{\partial h}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{x}} + \frac{\partial h}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}}
\end{aligned}
$$

$$\begin{aligned}\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \frac{\partial h(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \\ &= \frac{\partial h}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{x}} + \frac{\partial h}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \\ &= \mathbf{y}^T + \mathbf{x}^T \mathbf{A}\end{aligned}$$

$$
\begin{aligned}
\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \frac{\partial h(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \\
&= \frac{\partial h}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{x}} + \frac{\partial h}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \\
&= \mathbf{y}^T + \mathbf{x}^T \mathbf{A} \\
&= (\mathbf{A}\mathbf{x})^T + \mathbf{x}^T \mathbf{A}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \frac{\partial h(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \\
&= \frac{\partial h}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{x}} + \frac{\partial h}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \\
&= \mathbf{y}^T + \mathbf{x}^T \mathbf{A} \\
&= (\mathbf{A}\mathbf{x})^T + \mathbf{x}^T \mathbf{A} \\
&= \mathbf{x}^T \mathbf{A}^T + \mathbf{x}^T \mathbf{A}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \frac{\partial h(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \\
&= \frac{\partial h}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{x}} + \frac{\partial h}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \\
&= \mathbf{y}^T + \mathbf{x}^T \mathbf{A} \\
&= (\mathbf{A}\mathbf{x})^T + \mathbf{x}^T \mathbf{A} \\
&= \mathbf{x}^T \mathbf{A}^T + \mathbf{x}^T \mathbf{A} \\
&= \mathbf{x}^T (\mathbf{A}^T + \mathbf{A})
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &= \frac{\partial h(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} \\
&= \frac{\partial h}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{x}} + \frac{\partial h}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \\
&= \mathbf{y}^T + \mathbf{x}^T \mathbf{A} \\
&= (\mathbf{A}\mathbf{x})^T + \mathbf{x}^T \mathbf{A} \\
&= \mathbf{x}^T \mathbf{A}^T + \mathbf{x}^T \mathbf{A} \\
&= \mathbf{x}^T (\mathbf{A}^T + \mathbf{A}) \\
&= 2\mathbf{x}^T \mathbf{A}
\end{aligned}
$$

The last equality holds because $\mathbf{A}$ is symmetric, i.e., $\mathbf{A}^T = \mathbf{A}$

Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{W}\mathbf{x}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$f(\mathbf{W}) =$$

Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{W}\mathbf{x}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$f(\mathbf{W}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ & & \\ & & \end{bmatrix}$$

Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{W}\mathbf{x}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$f(\mathbf{W}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ & & \end{bmatrix}$$

Let $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{Wx}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$f(\mathbf{W}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix}$$

Let $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{W}\mathbf{x}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$f(\mathbf{W}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \\ \\ \end{bmatrix}$$

Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{W}\mathbf{x}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$f(\mathbf{W}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \left[\begin{matrix} x_1 \\ \vdots \\ \end{matrix} \right.$$

Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{W}\mathbf{x}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$f(\mathbf{W}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{W}\mathbf{x}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$
\begin{aligned}
f(\mathbf{W}) &= \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \left[\begin{matrix} w_{11}x_1 + \cdots + w_{1n}x_n \\ \\ \\ \end{matrix} \right.
\end{aligned}
$$

Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{Wx}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$f(\mathbf{W}) = \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \left[\begin{matrix} w_{11}x_1 + \dots + w_{1n}x_n \\ \vdots \\ \end{matrix} \right.$$

Let $f : \mathbb{R}^{m \times n} \to \mathbb{R}^n$ with $f(\mathbf{W}) = \mathbf{W}\mathbf{x}$ for a fixed $\mathbf{x} : \mathbb{R}^n$. Then,

$$
\begin{aligned}
f(\mathbf{W}) &= \begin{bmatrix} w_{11} & \dots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \dots & w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} w_{11}x_1 + \cdots + w_{1n}x_n \\ \vdots \\ w_{m1}x_1 + \cdots + w_{mn}x_n \end{bmatrix}
\end{aligned}
$$

$$\mathbf{J}_f(\mathbf{W_0}) = \left[\begin{matrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{11}} & \cdots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{mn}} \\ & & \\ & & \end{matrix} \right.$$

$$\mathbf{J}_f(\mathbf{W_0}) = \left[\begin{matrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{11}} & \cdots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{mn}} \\ \vdots & \ddots & \vdots \\ & & \end{matrix} \right.$$

$$\mathbf{J}_f(\mathbf{W_0}) = \begin{bmatrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{11}} & \cdots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{mn}} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial w_{11}} & \cdots & \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial w_{mn}} \end{bmatrix}$$

$$\mathbf{J}_f(\mathbf{W_0}) = \begin{bmatrix} \frac{\partial(w_{11}x_1+\dots+w_{1n}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{11}x_1+\dots+w_{1n}x_n)}{\partial w_{mn}} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\dots+w_{mn}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{m1}x_1+\dots+w_{mn}x_n)}{\partial w_{mn}} \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & \dots & x_n & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ & & & & & & & & & \\ & & & & & & & & & \end{bmatrix}$$

$$\mathbf{J}_f(\mathbf{W_0}) = \begin{bmatrix} \frac{\partial(w_{11}x_1+\dots+w_{1n}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{11}x_1+\dots+w_{1n}x_n)}{\partial w_{mn}} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\dots+w_{mn}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{m1}x_1+\dots+w_{mn}x_n)}{\partial w_{mn}} \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & \dots & x_n & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & x_1 & \dots & x_n & \dots & 0 & \dots & 0 \\ & & & & & & & & & \\ & & & & & & & & & \end{bmatrix}$$

$$
\mathbf{J}_f(\mathbf{W_0}) = \begin{bmatrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{mn}} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial w_{mn}} \end{bmatrix}
$$

$$
= \begin{bmatrix} x_1 & \dots & x_n & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & x_1 & \dots & x_n & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & x_1 & \dots & x_n \end{bmatrix}
$$

$$
\begin{aligned}
\mathbf{J}_f(\mathbf{W_0}) &= \begin{bmatrix} \frac{\partial(w_{11}x_1+\dots+w_{1n}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{11}x_1+\dots+w_{1n}x_n)}{\partial w_{mn}} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\dots+w_{mn}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{m1}x_1+\dots+w_{mn}x_n)}{\partial w_{mn}} \end{bmatrix} \\
&= \begin{bmatrix} x_1 & \dots & x_n & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & x_1 & \dots & x_n & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & x_1 & \dots & x_n \end{bmatrix} \\
&= \begin{bmatrix} x_1 & \dots x_n \end{bmatrix} \otimes \mathbf{I}_m
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{J}_f(\mathbf{W_0}) &= \begin{bmatrix} \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{11}x_1+\cdots+w_{1n}x_n)}{\partial w_{mn}} \\ \vdots & \ddots & \vdots \\ \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial w_{11}} & \dots & \frac{\partial(w_{m1}x_1+\cdots+w_{mn}x_n)}{\partial w_{mn}} \end{bmatrix} \\
&= \begin{bmatrix} x_1 & \dots & x_n & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & x_1 & \dots & x_n & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & x_1 & \dots & x_n \end{bmatrix} \\
&= \begin{bmatrix} x_1 & \dots x_n \end{bmatrix} \otimes \mathbf{I}_m \\
&= \mathbf{x} \otimes \mathbf{I}_m
\end{aligned}
$$

Thank You