Linear Algebra & Convex Optimization – Lecture 11

Functions & Derivatives

Outline

- Functions & Derivatives
- Matrix Derivatives

Functions

$f : \mathbf{R}^n \to \mathbf{R}^m$

f maps (some) n-vectors into m-vectors

it does not mean that $f(x)$ is defined for every $x \in \mathbf{R}^n$

Example:

$f : \mathbf{S}^n \to \mathbf{R}$, given by

$f(X) = \log\det X$ with $\mathbf{dom}\, f = \mathbf{S}^n_{++}$

$f : \mathbf{S}^n \to \mathbf{R}$ specifies the *syntax* of f:

$f : A \to B \quad \Rightarrow \quad f$ is a function on the set $\mathbf{dom}\, f \subseteq A$ into the set B;

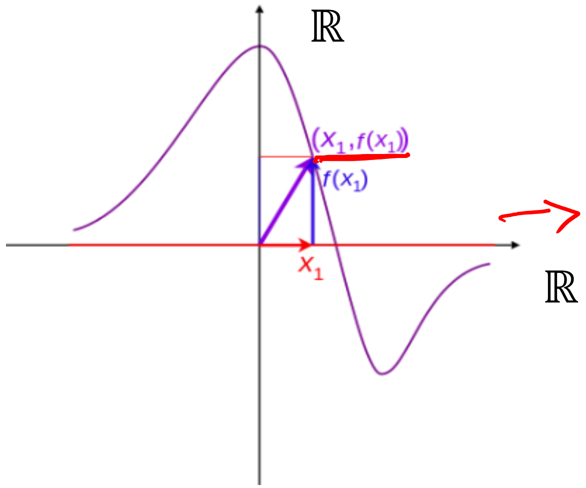
- Finding minimum/maximum of a function is critical in many scenarios

Graph of Functions

A real valued function $f(x_1, x_2, \ldots . x_n)$ is a mapping $f: \mathbb{R}^n \to \mathbb{R}$ such that the **dom** f is a subset of $\mathbb{R}^n$

$f(x_1, x_2, \ldots . x_n)$ can be plotted as a graph in $\mathbb{R}^{n+1}$

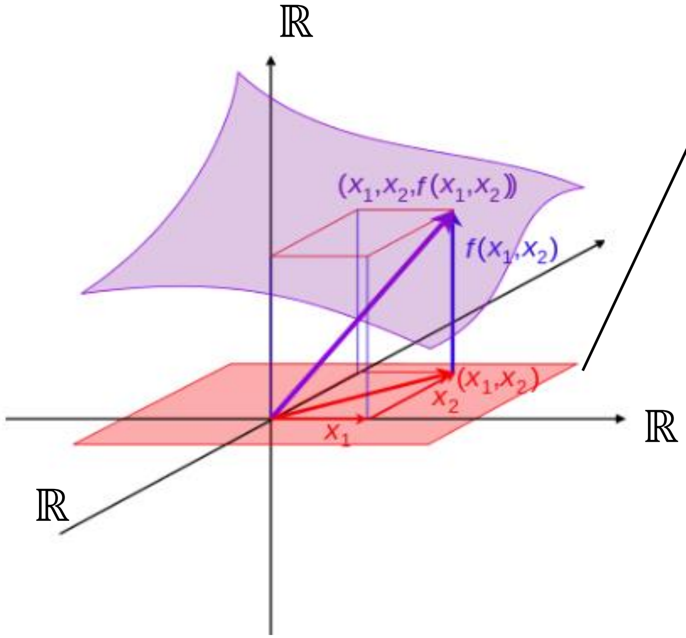
$f: \mathbb{R} \to \mathbb{R}$



$n = 1$

$f: \mathbb{R}^2 \to \mathbb{R}$

dom f



$n = 2$

First Derivative

$$z + te_i = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

For $f : \mathbf{R} \to \mathbf{R}$

$$f'(z) = \lim_{t \to 0} \frac{f(z+t) - f(z)}{t}$$ is called the *derivative* of the function f at the point z.

We can think of f' as a scalar-valued function of a scalar variable

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} \qquad z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

Partial Derivatives:

Suppose $f : \mathbf{R}^n \to \mathbf{R}$ → Scalar Function

$f(x) = f(x_1, \ldots, x_n)$

At $x = z$

$$\frac{\partial f}{\partial x_i}(z) = \lim_{t \to 0} \frac{f(z_1, \ldots, z_{i-1}, z_i + t, z_{i+1}, \ldots, z_n) - f(z)}{t}$$

$$= \lim_{t \to 0} \frac{f(z + te_i) - f(z)}{t}$$

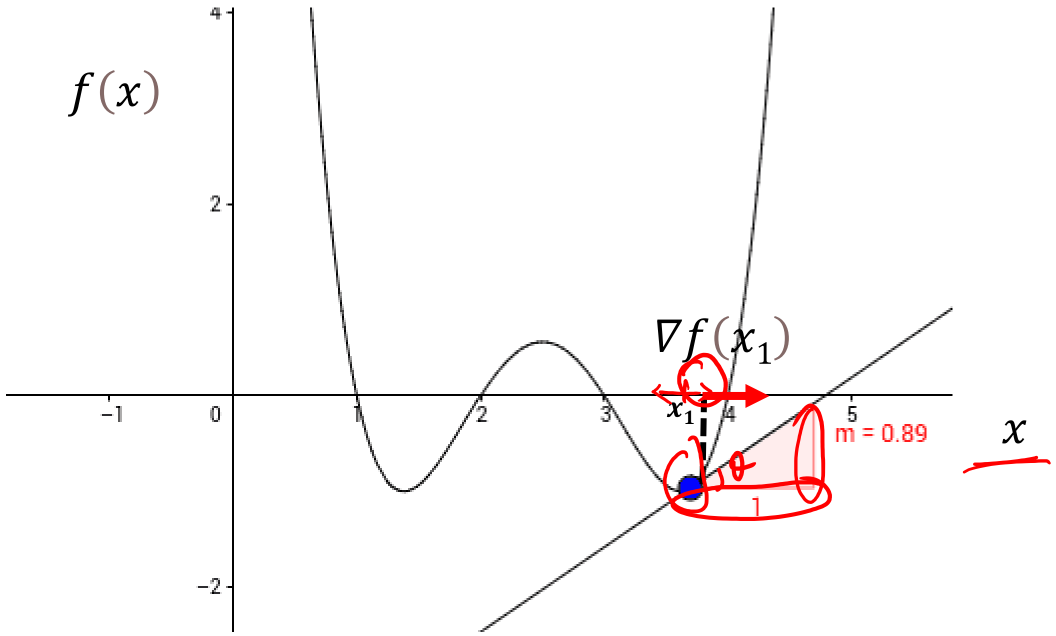
Gradient Function of f:

$\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ → vector function

$$\nabla f(z) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(z) \\ \vdots \\ \frac{\partial f}{\partial x_n}(z) \end{bmatrix} \in \mathbb{R}^n$$

Gradient ($f: \mathbb{R} \to \mathbb{R}$)

$f(x) = (x-1)(x-2)(x-3)(x-4)$



Gradient vector ∇f lies on the domain of the function $f(x)$

$\nabla f(x_1)$ is the 1D Gradient Vector at $x = x_1$

Value of Gradient $\nabla f(x_1)$ = Slope of the tangent

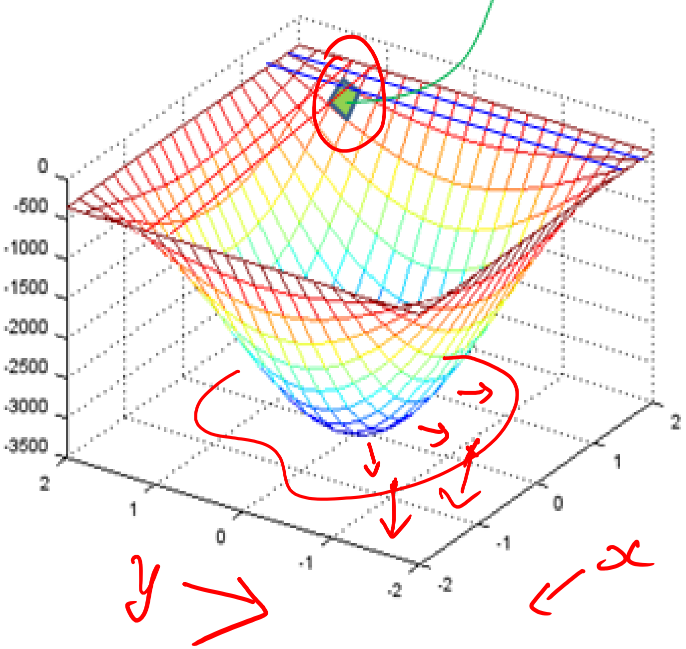
Direction of Gradient $\nabla f(x_1)$ = sign of the Slope of the tangent

Gradient ($f: \mathbb{R}^2 \to \mathbb{R}$)

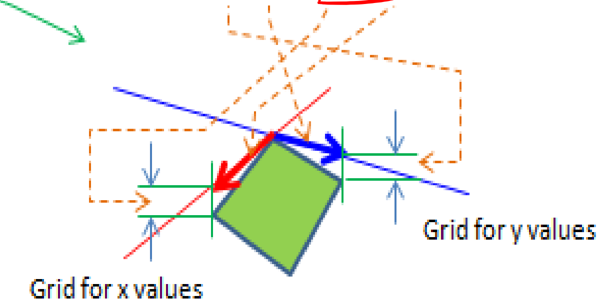
$\nabla f : \mathbb{R}^2 \to \mathbb{R}^2$

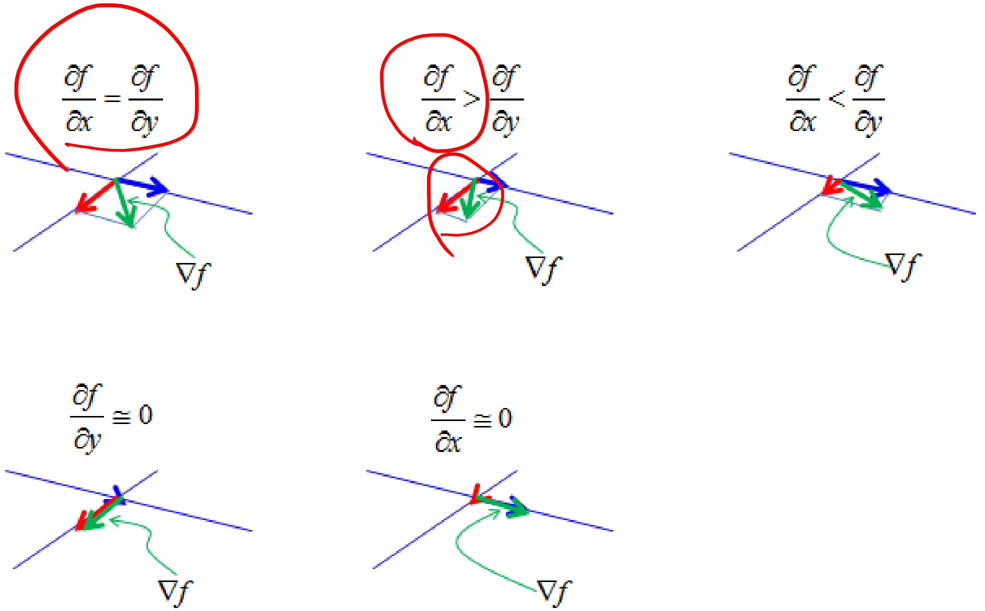
$$\begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

$f: \mathbb{R}^2 \to \mathbb{R}$



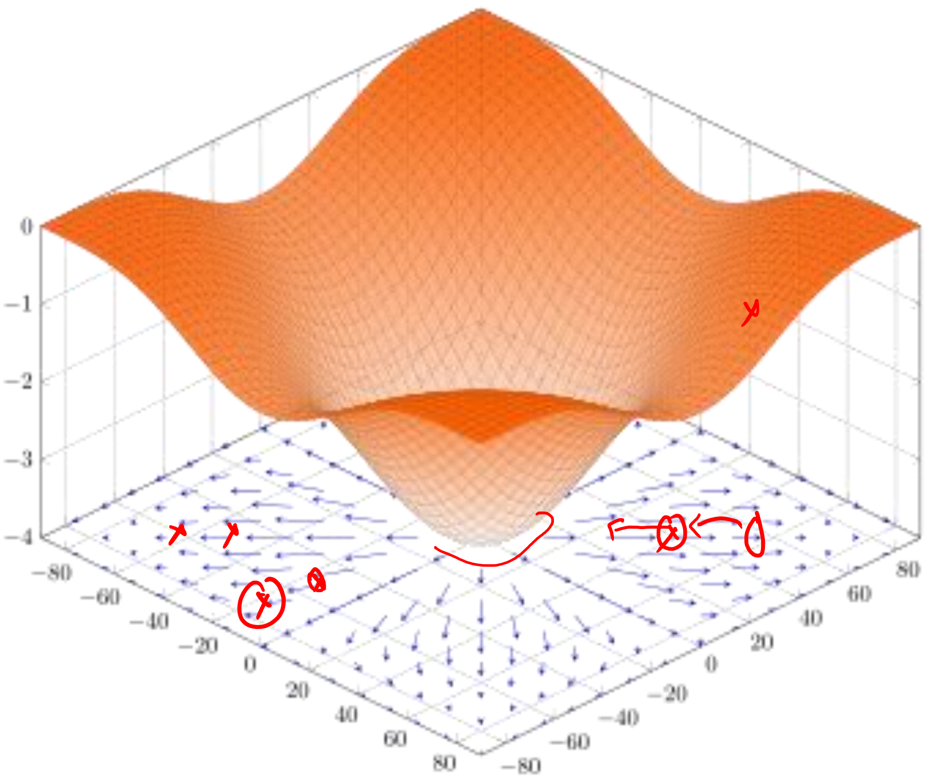
$$\nabla f = \frac{\partial f}{\partial y}\hat{j} + \frac{\partial f}{\partial x}\hat{i}$$





Gradient ($f: \mathbb{R}^2 \to \mathbb{R}$)

$R^{10^6} \to \mathbb{R}$



$$f(x_1, x_2) = -\left(\cos^2 x_1 + \cos^2 x_2\right)^2$$

Gradient for $f: \mathbb{R}^n \to \mathbb{R}$

$$\nabla f = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix}$$

First Order Taylor Approximation

The (first-order) Taylor approximation of f at the point z

$\hat{f} : \mathbf{R}^n \to \mathbf{R}$

$$\hat{f}(x) = f(z) + \frac{\partial f}{\partial x_1}(z)(x_1 - z_1) + \cdots + \frac{\partial f}{\partial x_n}(z)(x_n - z_n)$$

interpret $x_i - z_i$ as the deviation of x_i from z_i

$\frac{\partial f}{\partial x_i}(z)(x_i - z_i)$: approximation of the change in f due to the deviation of x_i from z_i.

First Order Taylor Approximation in Compact Form:

$$\hat{f}(x) = f(z) + \nabla f(z)^T (x - z).$$

$$\hat{f}(z) = f(z)$$

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \qquad \frac{\partial f(z)}{\partial x_1} = \frac{f(z') - f(z)}{(x_1 - z_1)}$$

$f: \mathbb{R}^n \to \mathbb{R}$

$x = [x_1, x_2, \dots . x_n]$ is a point in the neighborhood of z

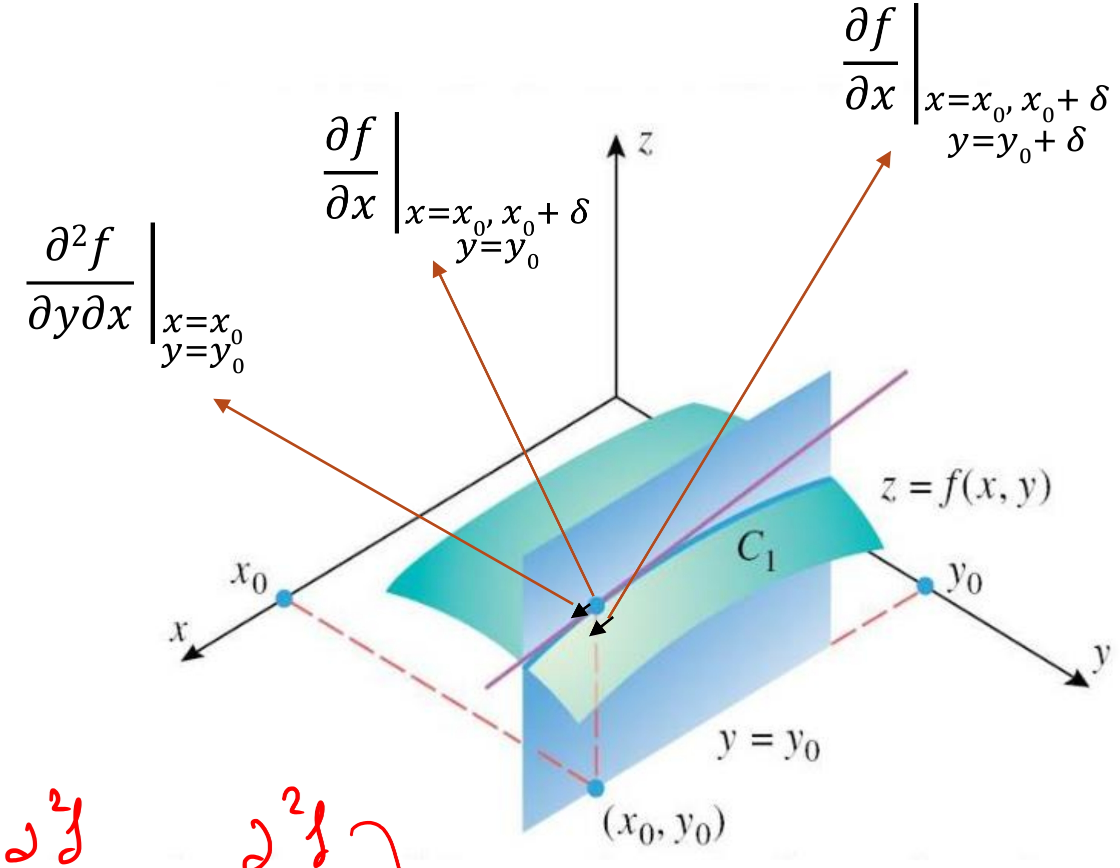
$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}}_{\nabla f^T} \begin{bmatrix} x_1 - z_1 \\ x_2 - z_2 \\ \vdots \\ x_n - z_n \end{bmatrix} \quad x - z$$

Hessian ($f: \mathbb{R}^n \to \mathbb{R}$)

$f: \mathbb{R}^n \to \mathbb{R}$ $\quad \nabla f: \mathbb{R}^n \to \mathbb{R}^n$ $\quad \nabla^2 f: \mathbb{R}^n \to \mathbb{R}^{n \times n}$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$$



For $f(x_1, x_2, \ldots . x_n)$ depending on n variables:

$$\nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

$$\nabla^2 f\ (i, j) = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

$$\left.\begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}\right|_{x_0, y_0}$$

Hessian is symmetric if f is continuous at $x_1, x_2, \ldots . x_n$.

For $f: \mathbb{R}^n \to \mathbb{R}$, $\quad \nabla^2 f: \mathbb{R}^n \to \mathbb{R}^{n \times n}$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = \frac{\partial^2 f}{\partial x_1 \partial x_2}$$

$$\frac{\partial f}{\partial x_1} = \frac{f(x_2, x_1 + \delta) - f(x_2, x_1)}{\delta}$$

$$\frac{\partial f}{\partial x_2 \partial x_1} = \left[\frac{f(x_2 + \delta, x_1 + \delta) - f(x_2 + \delta, x)}{\delta} \right] - \left[\frac{f(x_2, x_1 + \delta) - f(x_2, x_1)}{\delta} \right] / \delta$$

$$\left[\frac{f(x_2 + \delta, x_1 + \delta) - f(x_2, x_1 + \delta)}{\delta} \right] - \left[\frac{f(x_2 + \delta, x) - f(x_2, x_1)}{\delta} \right] / \delta = \frac{\partial f}{\partial x_1 \partial x_2}$$

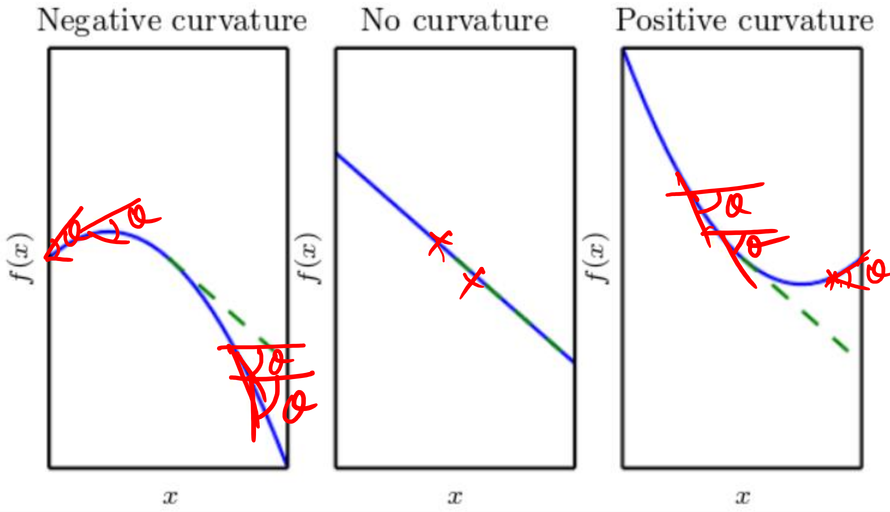
Properties of Hessian Matrix

Hessian Matrix evaluated at location x captures
the nature of curvature of a function

$(\nabla f)_{(1,1)}$ is large positive means function displays positive curvature in first dimension

$(\nabla f)_{(2,2)}$ is negative means function displays negative curvature in second dimension

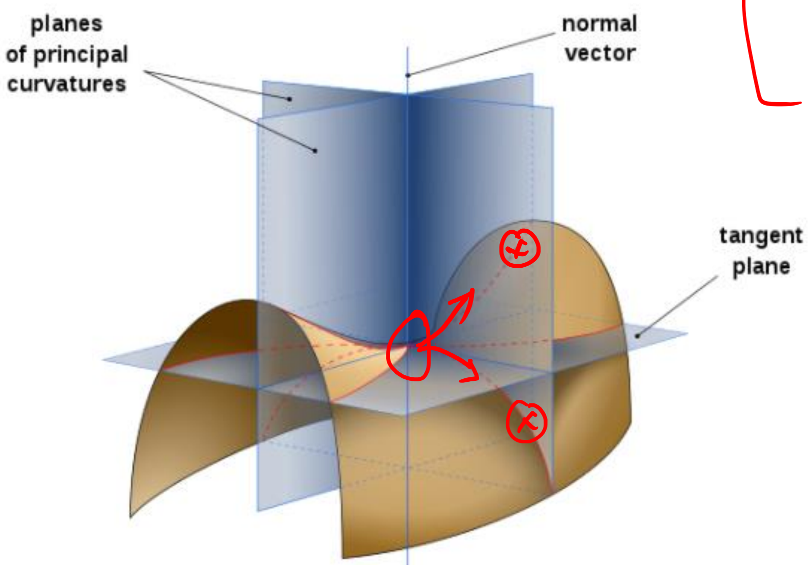
$(\nabla f)_{(1,2)}$ is negative means function displays curvature in opposite directions along dimension 1 and 2.



$\mathbb{R} \to \mathbb{R}$

$\mathbb{R}^2 \to \mathbb{R}$

$2\times$



Second Order Taylor Approximation

$f : \mathbf{R}^n \to \mathbf{R}.$

Hessian matrix of f at $x{=}z$, denoted $\nabla^2 f(z)$

$$\nabla^2 f(z)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad \text{at } x{=}z$$

$$i = 1, \ldots, n, \quad j = 1, \ldots, n,$$

$$\hat{f}(x) = f(z) + \nabla f(z)^T(x - z) + (1/2)(x - z)^T \nabla^2 f(z)(x - z).$$

$$\underbrace{x^T A x}_{y} = \sum_{i,j=1}^{N} a_{ij} x_i x_j$$

$$[x_1 \cdots x_n] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\sum_{i\ j} \frac{\partial^2 f}{\partial x_i \partial x_j} (x_i - z_i)(x_j - z_j)$$

Jacobian ($f: \mathbb{R}^n \to \mathbb{R}^m$)

$\mathbb{R}^n \to \mathbb{R}$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \Rightarrow \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}$$

Defined for functions of the form $f: \mathbb{R}^n \to \mathbb{R}^m$

$f: \mathbb{R} \to \mathbb{R}$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Jacobian of $f: \mathbb{R}^n \to \mathbb{R}$ is equivalent to transpose of gradient of f, ∇f^T

Matrix Calculus – (First Order Derivatives)

o/p →

i/p ↓

Types of matrix derivative

Types	Scalar	Vector	Matrix
Scalar	$\frac{\partial y}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{Y}}{\partial x}$
Vector	$\frac{\partial y}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$	
Matrix	$\frac{\partial y}{\partial \mathbf{X}}$		

Derivative of Scalar function $(y : \mathbb{R}^m \to \mathbb{R})$ **by vector:**

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$$

The derivative is column vector (**Gradient**) by convention (*denominator layout or Gradient convention*)

Derivative of Vector function $(y : \mathbb{R} \to \mathbb{R}^m)$ **by scalar:**

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

The derivative is row vector by convention (*denominator layout or Gradient convention*)

Derivative of vector function ($\mathrm{y}: \mathbb{R}^n \to \mathbb{R}^m$ **) by vector:**

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

Transpose of Jacobian Matrix

Matrix Calculus

Some Common Derivatives

(in denominator layout convention)

u & **v** are functions of vector **x**

A & **b** are independent of vector **x**

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top$$

$$\frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{u}$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}$$

$$\frac{\partial \mathbf{b}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top \mathbf{b}$$

$$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^\top$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^\top)\,\mathbf{x}$$

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \\ a_{n1} & \cdots & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_m \end{array} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & & a_{mn} \end{bmatrix}$$

$$\sum_{i,j} a_{ij}\, x_i x_j$$

$$x_i \Rightarrow \sum_j \left(a_{ij} x_j + a_{ji} x_j\right)$$