Lecture 8.1: Functions of n variables

Optimization and Computational Linear Algebra for Data Science

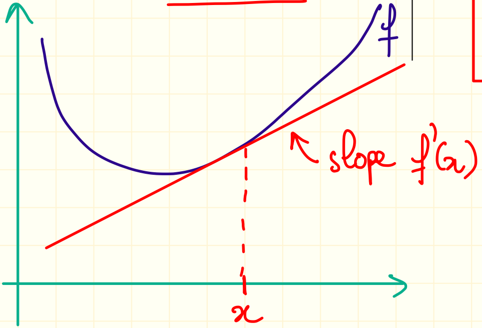
Léo Miolane

Derivative / Gradient

$$f : \mathbb{R} \to \mathbb{R}$$
$$x \mapsto f(x)$$

Derivative at $x \in \mathbb{R}$:

$$\underline{f'(x) \in \mathbb{R}}$$



$$f : \mathbb{R}^n \to \mathbb{R}$$
$$x \mapsto f(x) = f(x_1, \ldots, x_n)$$

Gradient at $x \in \mathbb{R}^n$:

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

<u>Interpretation</u>: $\nabla f(x)$ indicates the direction of largest increase of f at x.

Gradient and contour lines

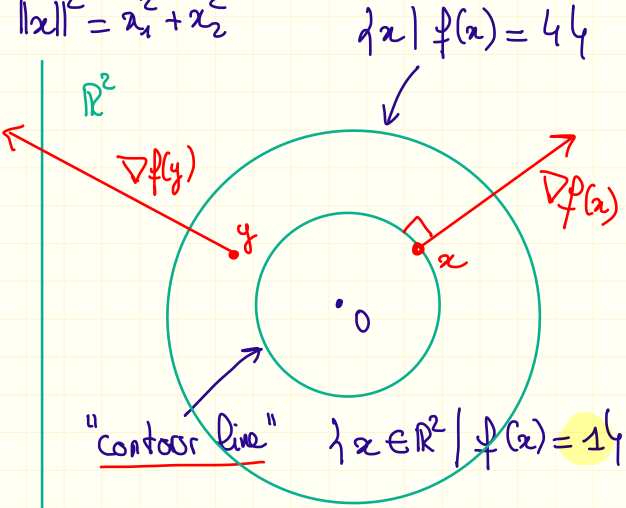
$$\underline{Ex}: \quad f: \mathbb{R}^2 \longrightarrow \mathbb{R}$$
$$x \longmapsto \|x\|^2 = x_1^2 + x_2^2$$

$$\frac{\partial f}{\partial x_1}(x) = 2x_1$$

$$\frac{\partial f}{\partial x_2}(x) = 2x_2$$

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

$$= 2x$$



Hessian matrix

What is the equivalent of the second derivative for function of n variables ?

$$\left(H_f(x)\right)_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

$$\begin{array}{rccl} f: & \mathbb{R}^n & \to & \mathbb{R} \\ & x & \mapsto & f(x) = f(x_1, \ldots, x_n) \end{array}$$

Hessian at $x \in \mathbb{R}^n$:

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1\,\partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1}(x) & \frac{\partial^2 f}{\partial x_n\,\partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Example

$$f: \mathbb{R}^n \longrightarrow \mathbb{R}$$
$$x \longmapsto x_1^3 + \dots + x_n^3$$

$$\left(Hf(x)\right)_{i,j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}(x)\right)$$
$$= \frac{\partial}{\partial x_i}\left(3x_j^2\right) = \begin{cases} 6x_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$Hf(x) = \begin{pmatrix} 6x_1 & & & (0) \\ & 6x_2 & & \\ & & \ddots & \\ (0) & & & 6x_n \end{pmatrix}$$

Schwarz's Theorem

technical assumption

Theorem

If $f : \mathbb{R}^n \to \mathbb{R}$ is «twice differentiable», then for all $x \in \mathbb{R}^n$ and all $i, j \in \{1, \ldots, n\}$ we have:

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(x) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x).$$

Consequence : $Hf(x)$ is a $\boxed{\text{symmetric matrix}}$!