$L : E \to F$, linear with E and F nvs

$$L(\lambda \vec{u} + \mu \vec{v}) = \lambda L(\vec{u}) + \mu L(\vec{v})$$

$$L(u+h) = L(u) + \boxed{DL(u)}(h) + o(\|h\|)$$

$$L(u+h) = L(u) + L(h) + \cancel{o(\|h\|)}$$

$$DL(u) = L$$

$A : E \to F$, affine: $A(x) = L(x) + b$ with linear L and $b \in F$

$$A(u + h) = A(u) + DA(u)(h) + o(\|h\|)$$

$$DA(u) = L$$

$f(X) = \|X\|_2^2$, with $X \in \mathbb{R}^n$

$$\|x+h\|_2^2 = \|x\|_2^2 + D_{\|\cdot\|}(x)(h) + o(\|h\|)$$

$$\|x+h\|^2 = \|x\|^2 + \underbrace{2\langle x, h\rangle}_{} + \underbrace{\|h\|^2}_{o(\|h\|)}$$

$$D_{\|.\|^2}(x)(h) = 2\langle x, h\rangle$$

$\mathbb{R}^n \longleftrightarrow \mathcal{L}(\mathbb{R}^n)$ isomorphism Riesz theorem

$\forall L \in \mathcal{L}(\mathbb{R}^n)\ \exists \vec{v}$ s.t.

$\vec{v} \longleftrightarrow L, \ \forall x \ L(\vec{x}) = \langle \vec{v}, \vec{x} \rangle$

$D_{\|\ \|^2}(x)(h) = \langle 2x, h\rangle$ therefore

one can identify $D_{\|\ \|^2}(x)$ to $2x$

$$\nabla f(x) = 2x$$

Example 5

$\mathbb{R}^2 \to \mathbb{R}$ $\quad f(x_1, x_2) = \begin{cases} 0 & \text{on } (0,0) \\ \dfrac{x_1 x_2}{x_1^2 + x_2^2} \end{cases}$

f is not differentiable because f is not continuous

$f(h,h) = \dfrac{h^2}{h^2 + h^2}$ for $h \neq 0$

$= \dfrac{1}{2} \;\not\xrightarrow[h \to 0]{}\; 0$

Example 6 exercise

Example 7

exercise

Example $f(x) = \|Ax + b\|^2$ with $A \in M_{m\times n}(\mathbb{R})$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$

$$\|A(x+h)+b\|^2 = \|Ax+b\|^2 + Df(x)(h) + o(\|h\|)$$

$$\|A(x+h)+b\|^2 = \|(Ax+b) + Ah\|^2 = \|Ax+b\|^2 + 2\langle Ax+b, Ah\rangle + \|Ah\|^2$$

$$\langle A\vec{u}, \vec{v}\rangle = \langle \vec{u}, A^T\vec{v}\rangle$$

$$= \|Ax+b\|^2 + 2\langle A^T(Ax+b), h\rangle + \|Ah\|^2 = o(\|h\|)$$

$$\|Ah\|^2 \le \|A\|^2\|h\|^2$$

$$\frac{\|Ah\|^2}{\|h\|} \le \|A\|^2\|h\| \to 0, \quad \|h\| \to 0$$

$$Df(x)(h) = 2\langle A^T(Ax+b), h\rangle$$

$$\nabla f(x) = 2A^T(Ax+b)$$

$$Df(x)(h) = \langle \nabla f(x), h\rangle$$

Example $f(x) = \frac{\langle Ax, x\rangle}{\|x\|_2^2}$ with $x \in \mathbb{R}^n$ and $A \in S^n$, $x \neq 0_{\mathbb{R}^n}$

Symetric matrix on $\mathbb{R}^n$

$$\frac{\langle A(x+h), x+h\rangle}{\|x+h\|^2} = \frac{\langle Ax, x\rangle}{\|x\|^2} + Df(x)(h) + o(\|h\|)$$

$$\langle Ax + Ah, x+h\rangle = \langle Ax, x\rangle + \langle Ax, h\rangle + \langle Ah, x\rangle + \langle Ah, h\rangle$$

$$= \langle Ax, x\rangle + 2\langle Ax, h\rangle + \langle Ah, h\rangle$$

$$\|x+h\|^2 = \|x\|^2 + 2\langle x, h\rangle + \|h\|^2$$

$$\frac{1}{1+\varepsilon} = 1 - \varepsilon + o(\varepsilon)$$

$$f(x+h) = \frac{\langle Ax, x\rangle + 2\langle Ax, h\rangle + \langle Ah, h\rangle}{\|x\|^2\left(1 + \frac{2\langle x, h\rangle}{\|x\|^2} + \frac{\|h\|^2}{\|x\|^2}\right)}$$

$$= \left(\frac{\langle Ax, x\rangle}{\|x\|^2} + \boxed{\frac{2\langle Ax, h\rangle}{\|x\|^2}} + \frac{\langle Ah, \varepsilon h\rangle}{\|x\|^2}\right)\left(1 - \boxed{\frac{2\langle x, h\rangle}{\|x\|^2}} - \frac{\|h\|^2}{\|x\|^2} + o(\|h\|)\right)$$

$$= \frac{\langle Ax, x\rangle}{\|x\|^2} + \boxed{\frac{2\langle Ax, h\rangle}{\|x\|^2} - 2\frac{\langle Ax, x\rangle}{\|x\|^2}\frac{\langle x, h\rangle}{\|x\|^2}} = Df(x)(h)$$

$$+ \frac{\langle Ah, h\rangle}{\|x\|^2} - \boxed{4\frac{\langle Ax, h\rangle}{\|x\|^2}\frac{\langle x, h\rangle}{\|x\|^2}} + o(\|h\|) - 2\frac{\langle Ah, h\rangle}{\|x\|^2}\frac{\langle x, h\rangle}{\|x\|^2}$$

$$Df(x)(h) = \frac{2\langle Ax - \langle Ax, x\rangle x / \|x\|^2, h\rangle}{\|x\|^2}$$

Definition of the gradient for a real function

$\mathcal{L}(\mathbb{R}^n) \longleftrightarrow \mathbb{R}^n$

- If V is a Hibert space, if f is differentiable, the Riesz representation theorem leads to the definition of the gradient of f: $\nabla f(x) \in V$

$$\langle \nabla f(x), y \rangle = Df(x)y.$$

$\nabla f(x)$ is the unique vector associated with $Df(x)$ by the Riesz theorem

the gradient is also the vector of the partial derivatives

Definition of the gradient for a real function

- If V is a Hibert space, if f is differentiable, the Riesz representation theorem leads to the definition of the gradient of f: $\nabla f(x) \in V$

$$\langle \nabla f(x), y \rangle = Df(x)y.$$

- Directional derivative of f in the direction $d \in \mathbb{R}^n$

$$\lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha} = \langle \nabla f(x), d \rangle.$$

$$f(x+\alpha d) = f(x) + Df(x)(\alpha d) + o(\|\alpha d\|)$$
$$= f(x) + \langle \nabla f(x), \alpha d \rangle + o(\|\alpha d\|)$$
$$\frac{f(x+\alpha d) - f(x)}{\alpha} = \langle \nabla f(x), d \rangle + \underbrace{\frac{o(\alpha\|d\|)}{\alpha}}_{\xrightarrow[\alpha \to 0]{} 0}$$

Definition of the gradient for a real function

- If V is a Hibert space, if f is differentiable, the Riesz representation theorem leads to the definition of the gradient of f: $\nabla f(x) \in V$

$$\langle \nabla f(x), y \rangle = Df(x)y.$$

- Directional derivative of f in the direction $d \in \mathbb{R}^n$

$$\lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha} = \langle \nabla f(x), d \rangle.$$

- Direction of descent $d \in \mathbb{R}^n$ at x

$$\langle \nabla f(x), d \rangle < 0.$$

Why is it called a direction of descent ?

$$f(x+\alpha d) = f(x) + \alpha \langle \nabla f(x), d \rangle + o(\alpha)$$

$$\frac{f(x+\alpha d) - f(x)}{\alpha} = \underbrace{\langle \nabla f(x), d \rangle}_{<0} + \frac{o(\alpha)}{\alpha}$$

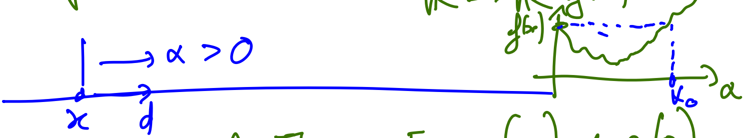
$\alpha \to 0 \quad 0$

$\exists\, \alpha_0 >$ s.t. $\forall \alpha < \alpha_0 \quad \langle \nabla f(x), d \rangle + \frac{o(\alpha)}{\alpha} < 0$

$$f(x+\alpha d) < f(x)$$

$g(\alpha) = f(x+\alpha d)$ is a scalar function

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



on the interval $]0, \alpha_0[\quad g(\alpha) < g(0)$

Definition of the gradient for a function of $\mathbb{R}^n$ in $\mathbb{R}$

- $\frac{\partial f(x)}{\partial x_j} = Df(x)(e_i)$

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

definition 1) f is montonous on an interval I iff f is decreasing on I or f is increasing on I

2) f is strictly decreasing on I iff

$\forall$ $x < y$ $x, y \in I$

$f(x) > f(y)$

Definition of the gradient for a function of $\mathbb{R}^n$ in $\mathbb{R}$

- $\frac{\partial f(x)}{\partial x_j} = Df(x)(e_i)$
- If $V = \mathbb{R}^n$, the gradient is the vector of partial derivatives $\left(\frac{\partial f(x)}{\partial x_j}\right)_{j=1,\ldots n}$.

Definition of the gradient for a function of $\mathbb{R}^n$ in $\mathbb{R}$

- $\frac{\partial f(x)}{\partial x_j} = Df(x)(e_i)$
- If $V = \mathbb{R}^n$, the gradient is the vector of partial derivatives $\left(\dfrac{\partial f(x)}{\partial x_j}\right)_{j=1,\dots n}$.
- A function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is said to be of class C^k if all its partial derivatives up to order k exist and are continuous on U

Example 1: $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \langle Ax, x\rangle$ $\quad A \in \mathcal{M}_{n\times n}(\mathbb{R})$

$$\langle A(x+h), x+h\rangle = \langle Ax, x\rangle + \langle Ax, h\rangle + \langle x, Ah\rangle + \langle h, Ah\rangle$$

$$= \langle Ax, x\rangle + \underbrace{\langle (A+A^T)x, h\rangle}_{\langle \nabla f(x), h\rangle} + \underbrace{\langle h, Ah\rangle}_{o(\|h\|)}$$

$$= f(x) + \langle \nabla f(x), h\rangle + o(\|h\|)$$

$$\nabla f(x) = (A + A^T)x$$

if $A \in S_n(\mathbb{R})$ (symmetric)

then $\nabla f(x) = 2Ax$

Example 2: $f : \mathbb{R}^n \to \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$,

$f(x) = \sum_{i=2}^{n} (x_i - x_{i-1})^2 = (x_2 - x_1)^2 + \dots + (x_i - x_{i-1})^2 + (x_{i+1} - x_i)^2 + \dots + (x_n - x_{n-1})^2$

expand the sum

$i = 1$

$$\frac{\partial f}{\partial x_1} = 2(x_1 - x_2)$$

for $i = 2, \dots, n-1$

$$\frac{\partial f}{\partial x_i} = 2(x_i - x_{i-1}) + 2(x_i - x_{i+1}) = 2(2x_i - x_{i-1} - x_{i+1})$$

$i = n$

$$\frac{\partial f}{\partial x_n} = 2(x_n - x_{n-1})$$

$$\nabla f(x) = 2\begin{bmatrix} x_1 - x_2 \\ 2x_2 - x_1 - x_3 \\ \vdots \\ 2x_i - x_{i-1} - x_{i+1} \\ \vdots \\ 2x_{n-1} - x_n - x_{n-2} \\ x_n - x_{n-1} \end{bmatrix}$$

Example 3: $f(x,y) = \langle Ax, x\rangle + \langle y, Bx\rangle$

$A \in \mathcal{M}_{n\times n}(\mathbb{R})$, $B \in \mathcal{M}_{m\times n}(\mathbb{R})$, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$

Calculate $\nabla_x f$ and $\nabla_y f$

$$f(x+h, y) = f(x,y) + \langle \nabla_x f(x,y), h\rangle + o(\|h\|)$$

$$\langle A(x+h), x+h\rangle + \langle y, B(x+h)\rangle = \langle Ax, x\rangle + \langle (A+A^T)x, h\rangle + \langle Ah, h\rangle + \langle y, Bx\rangle + \langle y, Bh\rangle$$

$$= \langle Ax, x\rangle + \langle y, Bx\rangle + \langle (A+A^T)x + B^T y, h\rangle + \langle Ah, h\rangle$$

$$\nabla_x f(x,y) = (A+A^T)x + B^T y$$

$$f(x, y+h) = \langle Ax, x\rangle + \langle y+h, Bx\rangle = \langle Ax, x\rangle + \langle y, Bx\rangle + \langle h, Bx\rangle$$

$$= f(x,y) + \langle Bx, h\rangle$$

$$\nabla_y f(x,y) = Bx$$

apply the rule on affine function since $f(x,y)$ is affine in y

Jacobian matrix definition

E and F nvs, dimensions n and m, bases $\mathcal{B}$ and $\mathcal{B}'$.
Let $U \subset E$ be open and $f : U \to F$, $f(x) = (f_1(x), \ldots, f_m(x))$ differentiable in $a \in U$.
$Df(a) \in \mathcal{L}(E, F)$ so there is a unique Jacobian matrix $Jf(a)$, $m \times n$, which represents $Df(a)$ in bases $\mathcal{B}$ and $\mathcal{B}'$.
Let $h = (h_1, \ldots, h_n) \in E$ we have $Df(a).h = Jf(a)h$.

$$Df(a) = \begin{pmatrix} Df_1(a) \\ \vdots \\ Df_m(a) \end{pmatrix}, \text{ i.e. } Df(a)h = \begin{pmatrix} Df_1(a)h \\ \vdots \\ Df_m(a)h \end{pmatrix} = J_f(a) \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}.$$

The jth column vector of $J_f(a)$ is the vector $\begin{pmatrix} Df_1(a)e_j \\ \vdots \\ Df_m(a)e_j \end{pmatrix}$

(with $(e_j)_m = \delta_{jm}$), and $Df_i(a)e_j = \dfrac{\partial f_i}{\partial x_j}(a)$,

$Jf(a) \in \mathcal{M}_{m \times n}(\mathbb{R})$

or $Jf(a) = (Jf(a)_{i,j})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} = (\dfrac{\partial f_i}{\partial x_j})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$

Relation between gradient and Jacobian matrix

$f_i : \mathbb{R}^n \to \mathbb{R} \Rightarrow \nabla f_i \in \mathbb{R}^n$

E nvs, dimensions n, base $\mathcal{B}$.
$U \subset E$ open and $f : U \to \mathbb{R}^m$, $f(x) = (f_1(x), \ldots, f_m(x))$
differentiable in $a \in U$.

$$Jf(a) = (Jf(a)_{i,j})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} = (\frac{\partial f_i}{\partial x_j})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} = \begin{pmatrix} \nabla f_1(x)^t \\ \vdots \\ \nabla f_m(x)^t \end{pmatrix}$$

m rows
n columns

Taylor formulas at 1st order

$f(x+d) = f(x) + \langle \nabla f(x), d \rangle + o(\|d\|)$

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ differentiable on S centered at x. For all $d \in \mathbb{R}^n$ t.q. $x + d \in S$ there is $\alpha \in [0, 1]$ $]0,1[$ t. q.

$$f(x+d) = f(x) + \langle \nabla f(x + \alpha d), d \rangle.$$

the value of α is not known and depends on x and d

$f(x-d) = f(x) + \langle \nabla f(x - \beta d), d \rangle$

with $\beta \in]0,1[$

Second-order Taylor formula

Definition : Let f be a differentiable function on V. f is twice differentiable at a if there exists a linear map $L(a) : V \to V'$ such that

$$Df(a+h) = Df(a) + L(a)h + o(\|h\|_V) \in V',$$

where V' denotes the topological dual of V. The second differential of f, denoted $D^2 f(a)$, is the map $L(a) : V \to V'$

Second-order Taylor formula in $\mathbb{R}^n$

Schwarz theorem.

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^n$. If ∇f is differentiable we identify $d^2 f(x)$ with the Hessian matrix $Hf(x)$, defined by $Hf : \mathbb{R}^n \to \mathbb{R}^{n\times n}$

$$Hf(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{i,j=1,\dots n}.$$

Schwarz's theorem: If the function f is twice differentiable, its Hessian matrix is symmetric.

Relation between Hessian and Jacobian matrices

The Hessian matrix of $f(x)$ is the Jacobian matrix of $Df(x)$