Relation between gradient and Jacobian matrix

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 $a \in U$.

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

Quiz

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

Which of the following expressions is correct?

1. $\nabla f(x) = 4Ax$
2. $\nabla f(x) = 4AA^t b$
3. $\nabla f(x) = 4A^t b$

Taylor formulas at 1st order

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 $\alpha \in [0, 1]$ t. q.

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

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 $\mathbb{R}^n$

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 $Df(x)$

Second order Taylor formula

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ twice differentiable on $U \subset \mathbb{R}^n$. If the segment $[a, a+h]$ is contained in U, so

- there is $\alpha \in [0,1]$t. q.

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} \langle Hf(a+\alpha h)h, h \rangle .$$

Second order Taylor formula

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ twice differentiable on $U \subset \mathbb{R}^n$. If the segment $[a, a+h]$ is contained in U, so

- there is $\alpha \in [0, 1]$t. q.

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2}\langle Hf(a+\alpha h)h, h \rangle.$$

- Taylor formula with integral remainder.

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \int_0^1 (1-t)\langle Hf(a+th)h, h \rangle dt$$

Taylor formula with Lagrange remainder

Let f be twice differentiable in an open set $U \subset \mathbb{R}^n$ with values in $\mathbb{R}$.
If $[a, a+h] \subset U$ and $\exists C > 0$ t.q.

$$\forall h, \quad \forall\, t \in [0,1], \quad |\langle Hf(a+th)h, h\rangle| \leq C\|h\|^2$$

So,

$$|f(a+h) - f(a) - \langle \nabla f(a), h\rangle| \leq \frac{C}{2}\|h\|^2$$

Taylor Mac-Laurin Formula

Let $f : [a, b] \longrightarrow \mathbb{R}$, p times differentiable. Then $\exists \gamma \in]a, b[$ such that

$$f(b) = f(a) + \sum_{k=1}^{p-1} \frac{(ba)^k}{k!} f^{(k)}(a) + \frac{(ba)^p}{p!} f^{(p)}(\gamma).$$

Exercises

We consider a function $v : \mathbb{R}^n \to \mathbb{R}^p$ twice differentiable on $\mathbb{R}^n$, a positive definite symmetric matrix $A \in \mathbb{R}^{p\times p}$, and the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) = \langle Av(x), v(x)\rangle.$$

- Show that f is twice differentiable

Exercises

We consider a function $v : \mathbb{R}^n \to \mathbb{R}^p$ twice differentiable on $\mathbb{R}^n$, a positive definite symmetric matrix $A \in \mathbb{R}^{p \times p}$, and the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) = \langle Av(x), v(x) \rangle.$$

- Show that f is twice differentiable
- Calculate its differential

Exercises

We consider a function $v : \mathbb{R}^n \to \mathbb{R}^p$ twice differentiable on $\mathbb{R}^n$, a positive definite symmetric matrix $A \in \mathbb{R}^{p \times p}$, and the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(x) = \langle Av(x), v(x) \rangle.$$

- Show that f is twice differentiable
- Calculate its differential
- Calculate its differential second

Exercises

Reminders for the functions of $\mathbb{R}^2$ in $\mathbb{R}$

Let $f : \mathbb{R}^2 \to \mathbb{R}$ in $\mathcal{C}^2$

- The critical points of f are the points such that $\dfrac{\partial f}{\partial x}(x, y) = \dfrac{\partial f}{\partial y}(x, y) = 0$

Reminders for the functions of $\mathbb{R}^2$ in $\mathbb{R}$

Let $f : \mathbb{R}^2 \to \mathbb{R}$ in $\mathcal{C}^2$

- The critical points of f are the points such that
  $\frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial y}(x, y) = 0$
- Monge notation: let (x^*, y^*) be a critical point of f

$$r = \frac{\partial^2 f}{\partial x^2}(x^*, y^*), \quad s = \frac{\partial^2 f}{\partial x \partial y}(x^*, y^*), \quad t = \frac{\partial^2 f}{\partial y^2}(x^*, y^*)$$

Reminders for the functions of $\mathbb{R}^2$ in $\mathbb{R}$

Let $f : \mathbb{R}^2 \to \mathbb{R}$ in $\mathcal{C}^2$

- The critical points of f are the points such that $\frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial y}(x, y) = 0$
- Monge notation: let (x^*, y^*) be a critical point of f

$$r = \frac{\partial^2 f}{\partial x^2}(x^*, y^*), \quad s = \frac{\partial^2 f}{\partial x \partial y}(x^*, y^*), \quad t = \frac{\partial^2 f}{\partial y^2}(x^*, y^*)$$

- Extrema: i.e. $\Delta = rt - s^2$

Reminders for the functions of $\mathbb{R}^2$ in $\mathbb{R}$

Let $f : \mathbb{R}^2 \to \mathbb{R}$ in $\mathcal{C}^2$

- The critical points of f are the points such that $\dfrac{\partial f}{\partial x}(x, y) = \dfrac{\partial f}{\partial y}(x, y) = 0$
- Monge notation: let (x^*, y^*) be a critical point of f

$$r = \frac{\partial^2 f}{\partial x^2}(x^*, y^*), \quad s = \frac{\partial^2 f}{\partial x \partial y}(x^*, y^*), \quad t = \frac{\partial^2 f}{\partial y^2}(x^*, y^*)$$

- Extrema: i.e. $\Delta = rt - s^2$
  - If $\Delta > 0$ then (x^*, y^*) is $\begin{cases} \text{a maximum if } r < 0 \\ \text{a minimum if } r > 0 \end{cases}$

Reminders for the functions of $\mathbb{R}^2$ in $\mathbb{R}$

Let $f : \mathbb{R}^2 \to \mathbb{R}$ in $\mathcal{C}^2$

- The critical points of f are the points such that
  $\frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial y}(x, y) = 0$
- Monge notation: let (x^*, y^*) be a critical point of f

$$r = \frac{\partial^2 f}{\partial x^2}(x^*, y^*), \quad s = \frac{\partial^2 f}{\partial x \partial y}(x^*, y^*), \quad t = \frac{\partial^2 f}{\partial y^2}(x^*, y^*)$$

- Extrema: i.e. $\Delta = rt - s^2$
  - If $\Delta > 0$ then (x^*, y^*) is $\begin{cases} \text{a maximum if } r < 0 \\ \text{a minimum if } r > 0 \end{cases}$
  - If $\Delta < 0$ then (x^*, y^*) is not an extremum

Reminders for the functions of $\mathbb{R}^2$ in $\mathbb{R}$

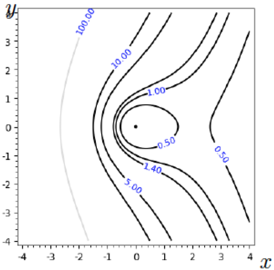
Let $f : \mathbb{R}^2 \to \mathbb{R}$ in $\mathcal{C}^2$

- The critical points of f are the points such that
  $\frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial y}(x, y) = 0$
- Monge notation: let (x^*, y^*) be a critical point of f

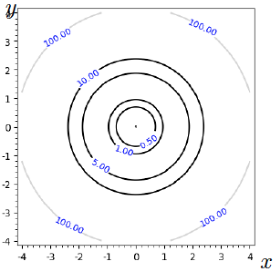
$$r = \frac{\partial^2 f}{\partial x^2}(x^*, y^*), \quad s = \frac{\partial^2 f}{\partial x \partial y}(x^*, y^*), \quad t = \frac{\partial^2 f}{\partial y^2}(x^*, y^*)$$

- Extrema: i.e. $\Delta = rt - s^2$
  - If $\Delta > 0$ then (x^*, y^*) is $\begin{cases} \text{a maximum if } r < 0 \\ \text{a minimum if } r > 0 \end{cases}$
  - If $\Delta < 0$ then (x^*, y^*) is not an extremum
  - If $\Delta = 0$ nothing can be said

Example: $f(x, y) = (x^2 + y^2)e^{-x}$



a)



b)

Example: $f(x, y) = (x^2 + y^2)e^{-x}$

Numerical approximation of derivatives and gradient

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is not known explicitly with a formula. Computing its gradient can be achieved numerically using Taylor formula.

For $f : \mathbb{R} \to \mathbb{R}$, C^2

$$f(x+h) = f(x) + f'(x)h + o(h)$$

$$\frac{f(x+h) - f(x)}{h} = f'(x) + o(1).$$

Or, for h small enough

$$\frac{f(x+h) - f(x)}{h} \approx f'(x).$$

Numerical approximation of derivatives

Better approximation using

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + o(h^2),$$

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) + o(h^2).$$

Substracting (2) from (1) provides

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + o(h)$$

Or, for h small enough

$$\frac{f(x+h) - f(x-h)}{2h} \approx f'(x).$$

Why is it a better approximation than $\dfrac{f(x+h) - f(x)}{h}$?

Numerical approximation of derivatives

Can we do better using higher order Taylor expansions ?

Find a $4^t h$ order approximation of $f'(x)$ using expansions like

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + \frac{h^3}{6} f^{(3)}(x) + \frac{h^4}{24} f^{(4)}(x) + o(h^4)$$

Expand...

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f^{(3)}(x) + \frac{h^4}{24}f^{(4)}(x) + o(h^4)$$

$$f(x-h) =$$

$$f(x+2h) =$$

$$f(x-2h) =$$

Choose correct coefficients before summing

Numerical approximation of gradient

And if $f : \mathbb{R}^n \to \mathbb{R}$?

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}$$

with e_i the canonical basis vector $(e_i)_j = \delta_{i,j} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ otherwise} \end{cases}$

Therefore $\frac{\partial f}{\partial x_i}(x) \approx \frac{f(x + he_i) - f(x)}{h}$ or

$\frac{\partial f}{\partial x_i}(x) \approx \frac{f(x + he_i) - f(x - he_i)}{2h}$ for h small enough

Experiment numerically with notebook 1

Outline

Course goals and terms

Introduction to Optimization

Reminders : Differential calculus

Convexity
- Convex sets
- Convex functions

Unconstrained optimisation
- Optimality conditions in the unconstrained case
- Solving systems of non linear equations
- Descent methods

Optimisation with constraints
- Duality
- Algorithms for constrained optimization

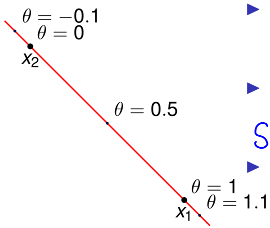
Convex sets

- Affine and convex sets
- Important Examples
- Convexity Preserving Operations

Affine set

Line passing through two points x_1 and x_2 in $\mathbb{R}^n$

$$x = \theta x_1 + (1-\theta)x_2, \quad \theta \in \mathbb{R}.$$



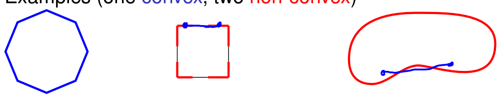
- *Definition :* a set A is affine if it contains all the lines passing through any two points of A
- Example: the set of solutions of a linear system
  $S = \{x, Ax = b\}$ S is affine
- (Conversely any affine set can be expressed as the set of solutions of a linear system

Convex set

- Line segment connecting two points x_1 and x_2

$$x = \theta x_1 + (1 - \theta) x_2, \quad 0 \leq \theta \leq 1.$$

- *Definition :* a set C is convex if it contains all the line segments connecting any two points of C
- Examples (one convex, two non-convex)



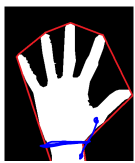
Combination convex and convex hull

Definition : a convex combination of k points of $\mathbb{R}^n$, $x_1, \ldots, x_k$ is a point $x \in \mathbb{R}^n$ such that

$$x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k = \sum_{i=1}^{k} \theta_i x_i$$

with $\theta_1 + \cdots + \theta_k = 1$ and $\theta_i \geq 0 \ \forall i = 1, \ldots, k$

Definition : the convex hull of a set E, denoted conv E, is the set of all convex combinations of points of E



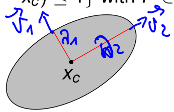
Euclidean balls and ellipsoids

Definition : ball (Euclidean) with center x_c and radius r

$B(x_c, r) = \{x, \|x - x_c\|_2 \le r\} = \{x_c + ru, \|u\|_2 \le 1\}$

Definition : a ellipsoid is a set of the form

$E = \{x, (x - x_c)^t P^{-1}(x - x_c) \le 1\}$ with $P \in S^n_{++}$



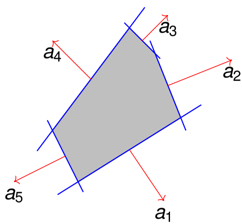
Property : $E = \{x_c + Au, \|u\|_2 \le 1\}$ with A invertible square

Polyhedrons

$Ax \preceq b$ $(Ax)_i \leq b_i$

$\sum A_{ij} x_j = b_i$ hyperplan of $\mathbb{R}^n$ $\rightarrow \sum A_{ij} x_j \leq b_i$

Definition : a polyhedron is a set of the form
$P := \{x \in \mathbb{R}^n, Ax \preceq b\}$ with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ fixed.
Property : a polyhedron is a finite intersection of closed half spaces of $\mathbb{R}^n$.



Convexity criterion of a set C

- Check definition
  $\forall x_1, x_2 \in C, \quad \forall 0 \le \theta \le 1, \quad \theta x_1 + (1-\theta) x_2 \in C$
- Show that C can be obtained from simple convex sets (hyperplanes, half spaces, balls,...) by convexity-preserving operations
  - intersection — $\bigcap_i C_i$ is convex
  - affine functions

f affine: $f(C)$ is convex, $f^{-1}(C)$ is convex

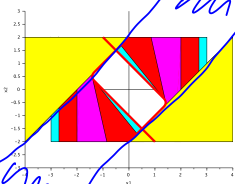
Intersection of convexes

Property : The intersection of any number of convexes is a convexe

Example: $S = \{x \in \mathbb{R}^m, |p(t)| \leq 1, \forall |t| \leq \pi/3\}$

with $p(t) = x_1 \cos t + x_2 \cos 2t + \ldots + x_m \cos mt$

For $m = 2$



114

Affine function   exercise

Consider an affine function on $\mathbb{R}^n$ $f(x) = Ax + b$ with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$

Property : The image of a convex by f is a convex

$$C \subset \mathbb{R}^n \text{ convex } \Rightarrow f(C) = \{f(x), x \in C\} \text{ convex}$$

Property : The inverse image of a convex by f is a convex

$$C \subset \mathbb{R}^m \text{ convex } \Rightarrow f^{-1}(C) = \{x \in \mathbb{R}^n,\ f(x) \in C\} \text{ convex}$$

Proof

Convex function

Definition : A function $f : C \longrightarrow \mathbb{R}$ is

- convex iff for all $x, y \in C$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in [0,1].$$

- strictly convex iff for all $x, y \in C$, $x \neq y$

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \quad \forall \lambda \in]0,1[.$$

Examples of convex/concave functions on $\mathbb{R}^n$

i) Affine $f(x) = a^t x + b$ $\Leftrightarrow$ both convex and concave

Norms $\|x\|_p = (\sum_{i=1}^n x_i^p)^{1/p}$ for $p \geq 1$, $\|x\|_\infty = \max_k |x_k|$

$$a^t(\lambda x + (1-\lambda) y) + b = \lambda a^t x + (1-\lambda) a^t y + (\lambda + 1 - \lambda) b$$
$$= \lambda (a^t x + b) + (1-\lambda)(a^t y + b)$$

$f : x \longmapsto \|x\|$ $\quad$ f concave or convex?

$\|x + y\| \leq \|x\| + \|y\|$ $\quad$ therefore convex

$\|\lambda x + (1-\lambda) y\| \leq \lambda \|x\| + (1-\lambda)\|y\|$ $\quad$ $\lambda \in [0,1]$

Examples of convex/concave functions on $\mathcal{M}_{m\times n}(\mathbb{R})$

$\|v\|_2^2 = \sum v_i^2$ $\quad$ $|||X|||_2 = (\lambda_{max}(X^TX))^{1/2}$

Affine functions $\quad$ both convex and concave

$$f(X) = tr(A^TX) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b$$

Spectral Norm: Maximum Singular Value

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^TX))^{1/2}$$ convex

$|||X|||_2$ euclidian matrix norm associated with the $\|v\|_2$ by

$$|||X|||_2 = \sup_{v\neq 0} \frac{\|Xv\|_2}{\|v\|_2} = \sup_{\|v\|_2=1} \|Xv\|_2$$

$$|||X|||_1 = \sup_{\|v\|_1=1} \|Xv\|_1 \quad \text{with } \|v\|_1 = \sum_i |v_i|$$

$$|||X|||_1 = \max_j \sum_{i=1}^{n} |X_{ij}|$$

$$|||X|||_\infty = \sup_{\|v\|_\infty=1} \|Xv\|_\infty = \max_i \sum_j |X_{ij}|$$

Restriction of a convex function on lines

$X + tM = X^{\frac{1}{2}}(I + t X^{-\frac{1}{2}} M X^{-\frac{1}{2}}) X^{\frac{1}{2}}$

$X = P D P^{-1}$ with $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$ $P = (\text{eigenvectors})$

$X^{\frac{1}{2}} = P D^{\frac{1}{2}} P^{-1}$ with $D^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{pmatrix}$

(could be any $r \in \mathbb{R}$)

$f : \mathbb{R}^n \to \mathbb{R}$
$g_{x,v}(t) = f(x + tv)$ with $x, v \in \mathbb{R}^n$, $t \in \mathbb{R}$
f is convex iff $g_{x,v}(t)$ convex for all $x, v \in \mathbb{R}^n$

Example : $f(X) = \log\det X$, dom $f = S^n_{++}$ Concave

$X, M \in S^n_{++}$, $\lambda_i > 0$

$g_{X,M}(t) = \log\det(X + tM)$

Epigraph

$X + tM = X^{\frac{1}{2}}(I + t X^{-\frac{1}{2}} M X^{-\frac{1}{2}}) X^{\frac{1}{2}}$

$\det(X + tM) = \det X^{\frac{1}{2}} \det X^{\frac{1}{2}} \det(I + t X^{-\frac{1}{2}} M X^{-\frac{1}{2}})$, $X^{-\frac{1}{2}} M X^{-\frac{1}{2}} \in S^n_{++}$

$\log\det(X + tM) = C + \log\det(I + t X^{-\frac{1}{2}} M X^{-\frac{1}{2}})$, $\lambda_i > 0$

$= C + \log \prod_i (1 + \lambda_i t)$ → $(1 + \lambda_i t)$

1. Definition of $C_\alpha = \{x \in \text{dom } f, f(x) \leq \alpha\}$
   the level sets of a convex function are convex

$= C + \sum \log(1 + \lambda_i t)$ where it is defined $\log(1 + \lambda_i t)$?

a sum of concave functions it is concave

Epigraph

1. Definition of $C_\alpha = \{x \in \text{dom } f, f(x) \leq \alpha\}$
   the level sets of a convex function are convex
2. Defining the epigraph of $f : \mathbb{R}^n \to \mathbb{R}$

$$epi\ f = \{(x, t) \in \mathbb{R}^{n+1}, x \in \text{dom } f,\ f(x) \leq t\}$$

f convex if and only if epi f is convex