# 2 Optimality conditions

**Elementary remark:**

$$\min_{[a,b]\subset\mathbb{R}} J$$

if $x_0$ is a local minimum of $J$, then (necessary condition):

- $J'(x_0) \geq 0$ if $x_0 = a$,
- $J'(x_0) = 0$ if $x_0 \in ]a, b[$,
- $J'(x_0) \leq 0$ if $x_0 = b$.

Indeed, if $x_0 \in [a, b[$, we can consider $x = x_0 + h$ for some $h > 0$ small enough.

$$J(x) = J(x_0) + hJ'(x_0) + o(h) \geq J(x_0) \quad \rightarrow \quad J'(x_0) \geq 0.$$

If $x_0 \in ]a, b]$, $x = x_0 - h$ and then $J'(x_0) \leq 0$. Note that if $x_0 \in ]a, b[$,

$$J(x_0) + \frac{h^2}{2} J''(x_0) + o(h^2) \geq J(x_0)$$

and then $J''(x_0) \geq 0$.
One must take into account the constraints ($x \in [a, b]$) in order to test the optimality: there are **admissible directions**.

## 2.1 Fréchet and Gâteaux differentiability

For clarity purpose, we may denote by $dJ$ the Fréchet derivative, and by $J'$ the Gâteaux derivative.

**Definition 2.1** *$J : V \rightarrow H$ has a directional derivative at point $u \in V$ along the direction $v \in V$ if*

$$\lim_{\theta \to 0^+} \frac{J(u + \theta v) - J(u)}{\theta}$$

*exists, and we denote by $J'(u; v)$ the limit.*

Example: $V = \mathbb{R}^n$ and $H = \mathbb{R}$, partial derivatives of $J$:

$$\frac{J(u + \theta e_i) - J(u)}{\theta} \rightarrow \frac{\partial J}{\partial x_i}(u).$$

**Definition 2.2** *$J : V \rightarrow H$ is Gâteaux differentiable at point $u$ if*

- *$J'(u; v)$ exists, $\forall v \in V$,*
- *and $v \mapsto J'(u; v)$ is a linear continuous function:*

$$J'(u; v) = (J'(u), v), \qquad J'(u) \in \mathcal{L}(V, H)$$

*and we denote by $J'(u)$ the Gâteaux derivative of $J$ at $u$.*

**Definition 2.3** *$J : V \rightarrow H$ is Fréchet differentiable at point $u$ if there exists $J'(u) \in \mathcal{L}(V, H)$ such that*

$$J(u + v) = J(u) + (J'(u), v) + \|v\|\varepsilon(v),$$

*where $\varepsilon(v) \rightarrow 0$ when $v \rightarrow 0$.*

**Proposition 2.1** *If $J$ is Fréchet differentiable at $u$, then $J$ is Gâteaux differentiable, and the two derivatives are equal.*

**Proof**
$J(u+\theta v) = J(u)+(J'(u), \theta v)+\|\theta v\|\varepsilon(\theta v)$, and then $\dfrac{J(u+\theta v)-J(u)}{\theta} = (J'(u), v)+\|v\|\varepsilon(\theta v)$, and $\varepsilon(\theta v) \to 0$ when $\theta \to 0$. Then the directional derivative exists everywhere, and $J'(u)$ is the Gâteaux derivative. □

**Remark:** the converse proposition is not true. Consider the following function in $\mathbb{R}^2$:

$$J(x, y) = 1 \;\text{ if }\; y = x^2 \;\text{ and }\; x \neq 0; \qquad J(x, y) = 0 \;\text{ otherwise.}$$

This function is Gâteaux differentiable. But it is not continuous, and hence cannot have a Fréchet derivative at 0.

**Finite-dimensional case:** $V = \mathbb{R}^n$ and $H = \mathbb{R}$. If all partial derivatives of $J$ exist at $u$, then

$$(J'(u), v) = \sum_{i=1}^{n} \frac{\partial J}{\partial x_i}(u).v_i = \langle \nabla J(u), v \rangle.$$

**Example 1:** $J(v) = Av$ where $A \in \mathcal{L}(V, H)$. Then $J'(u) = A$, $\forall u$.

**Example 2:** $J(v) = \frac{1}{2}a(v, v) - L(v)$, where $a$ is a continuous symmetric bilinear form, and $L$ is a continuous linear form.

$$J(u + v) = J(u) + (J'(u), v) + \frac{1}{2}a(v, v).$$

As $a$ is continuous, $a(v, v) \leq M\|v\|^2$ and then

$$(J'(u), v) = a(u, v) - L(v).$$

**Second (Gâteaux) derivative:** If $\dfrac{J'(u+\theta v; w) - J'(u; w)}{\theta}$ has a finite limit $\forall u, w, v \in V$ when $\theta \to 0^+$, then we denote this limit by $J''(u; v, w)$, and it is the second directional derivative at point $u$ in the directions $v$ and $w$.
If $J''(u; v, w)$ is a continuous bilinear form, then $J''(u)$ is called the second derivative of $J$, or **hessian** of $J$.

**Finite increments and Taylor formula:**

**Proposition 2.2** *If $J : V \to \mathbb{R}$ is Gâteaux differentiable $\forall u + \theta v$, where $\theta \in [0, 1]$, then $\exists \theta_0 \in ]0, 1[$ such that*

$$J(u + v) = J(u) + J'(u + \theta_0 v, v).$$

**Proof**
Let $f(\theta) = J(u + \theta v)$. $f : \mathbb{R} \to \mathbb{R}$.

$$f'(\theta) = \lim \frac{f(\theta+\varepsilon) - f(\theta)}{\varepsilon} = \lim \frac{J(u + \theta v + \varepsilon v) - J(u + \theta v)}{\varepsilon} = (J'(u + \theta v), v).$$

Moreover, $\exists \theta_0 \in ]0, 1[$ such that $f(1) = f(0) + f'(\theta_0)$. □

**Proposition 2.3** *If $J : V \to \mathbb{R}$ is twice Gâteaux differentiable $\forall u + \theta v$, $\theta \in [0,1]$, then $\exists \theta_0 \in ]0,1[$ such that*

$$J(u+v) = J(u) + (J'(u), v) + \frac{1}{2} J''(u + \theta_0 v; v, v).$$

**Proof**
Same as before, with $f''(\theta) = (J''(u+\theta v); v, v)$. □

## 2.2 Convexity and Gâteaux differentiability

Let $V$ be a Banach space on $\mathbb{R}$, and $K \subset V$ a convex subset.

**Definition 2.4** *$J : K \to \mathbb{R}$ is a convex function if*

$$J\left((1-\theta)u + \theta v\right) \leq (1-\theta)J(u) + \theta J(v), \quad \forall \theta \in [0,1], \ \forall u, v \in K.$$

*$J$ is called strictly convex if*

$$J\left((1-\theta)u + \theta v\right) < (1-\theta)J(u) + \theta J(v), \quad \forall \theta \in ]0,1[, \ \forall u \neq v \in K.$$

*$J$ is $\alpha$-convex ($\alpha > 0$), or strongly convex, if*

$$J\left((1-\theta)u + \theta v\right) \leq (1-\theta)J(u) + \theta J(v) - \frac{\alpha}{2}\theta(1-\theta)\|u-v\|^2, \quad \forall \theta \in [0,1], \ \forall u, v \in K.$$

**Proposition 2.4** *$\alpha$-convex $\Rightarrow$ strictly convex $\Rightarrow$ convex.*

**Proposition 2.5** *Let $J : K \to \mathbb{R}$ be a convex function, then every local minimum of $J$ is a global minimum.*

**Proof**
Let $u$ be a local minimum of $J$. Then $J(u) \leq J(v)$, $\forall v \in \mathcal{V}(u) \cap K$. If $u$ is not a global minimum, $\exists w \in K$ such that $J(w) < J(u)$. Then, for $\theta > 0$, by convexity,

$$J((1-\theta)u + \theta w) < J(u).$$

For $\theta > 0$ small enough, $(1-\theta)u + \theta w \in \mathcal{V}(u) \cap K$, and there is a contradiction. □

**Proposition 2.6** *If $J$ is strictly convex, then $J$ has at most one minimum.*

**Proof**
Let $u_1$ and $u_2$ be two (local $\Rightarrow$ global) minima of $J$. Then $J(u_1) = J(u_2)$. By strict convexity, $J(w) < J(u_1)$ for all $w \in ]u_1, u_2[$. □

**Proposition 2.7** *If $J : V \to \mathbb{R}$ is Gâteaux-differentiable in $V$, then:*

*i* *$J$ is convex in $V \Leftrightarrow J(v) \geq J(u) + (J'(u), v-u)$, $\forall u, v$*

*ii* *$J$ is strictly convex in $V \Leftrightarrow J(v) > J(u) + (J'(u), v-u)$, $\forall u \neq v$*

*iii* *$J$ is $\alpha$-convex in $V \Leftrightarrow J(v) \geq J(u) + (J'(u), v-u) + \frac{\alpha}{2}\|v-u\|^2$, $\forall u, v$*

**Proof**

i) If $J$ is convex:

$$J(u+\theta(v-u)) \leq J(u)+\theta(J(v)-J(u))$$
$$\frac{J(u+\theta(v-u))-J(u)}{\theta} \leq J(v)-J(u)$$

and consider the limit when $\theta \to 0^+$. Conversely,

$$J(u) \geq J(u+\theta(v-u)) - \theta\left(J'(u+\theta(v-u)), v-u\right)$$

and

$$J(v) \geq J(u+\theta(v-u)) + (1-\theta)\left(J'(u+\theta(v-u)), v-u\right)$$

Let us multiply the first inequality by $(1-\theta)$ and the second by $\theta$, and then

$$(1-\theta)J(u)+\theta J(v) \geq J(u+\theta(v-u))$$

ii) If $J$ is strictly convex: then $J$ is convex, and by (i) between $u$ and $u+\theta(v-u)$,

$$(J'(u), \theta(v-u)) \leq J(u+\theta(v-u)) - J(u)$$
$$(J'(u), v-u) \leq \frac{J(u+\theta(v-u))-J(u)}{\theta} < J(v)-J(u)$$

by strict convexity. Conversely, see (i).

iii) Same as (ii). □

**Geometric interpretation:** $J$ is convex if the function lies above its tangent.

**Proposition 2.8** *If $J$ is Gâteaux-differentiable, then:*

*i) $J$ is convex iff $J'$ is monotonic, i.e.*

$$(J'(u)-J'(v), u-v) \geq 0$$

*ii) $J$ is strictly convex iff $J'$ is strictly monotonic, i.e.*

$$(J'(u)-J'(v), u-v) > 0, \ u \neq v$$

*iii) $J$ is $\alpha$-convex iff*

$$(J'(u)-J'(v), u-v) \geq \alpha\|u-v\|^2$$

**Proof**

i) If $J$ is convex, $J(v) \geq J(u)+(J'(u), v-u)$ and $J(u) \geq J(v)+(J'(v), u-v)$. Conversely, let $\phi(\theta) = J(u+\theta(v-u))$. Then $\phi'(u) = (J'(u+\theta(v-u)), v-u)$. We have $\phi'(\theta)-\phi'(0) \geq 0$, and by integrating between 0 and $\theta$, $\phi(\theta)-\theta\phi'(0)-\phi(0) \geq 0$. $\theta = 1 \Rightarrow J$ is convex.

ii) Same as (i).

iii) Add $\frac{\alpha}{2}\theta^2\|v-u\|^2$. □

**Proposition 2.9** *If $J : V \to \mathbb{R}$ is twice Gâteaux-differentiable, then*

*i) $J$ is convex in $V$ iff $(J''(u)w, w) \geq 0, \ \forall u, w$.*

*ii) $J$ is $\alpha$-convex in $V$ iff $(J''(u)w, w) \geq \alpha\|w\|^2, \ \forall u, w$.*

**Proof**
If $J$ is convex, $(J'(u+\theta w) - J'(u), \theta w) \geq 0$. Let $\theta \to 0^+$, and then $(J''(u)w, w) \geq 0$. Conversely: Taylor formula. □

**Geometric interpretation:** if $J$ is $\alpha$-convex, the function $J$ has a curvature larger than (or equal to) the curvature of $w \mapsto \frac{1}{2}w^2$.

**Remark:** $J$ strictly convex $\Leftarrow (J''(u)w, w) > 0$, this is a sufficient condition, but not a necessary condition!

## 2.3 Optimality conditions

### 2.3.1 First-order necessary optimality conditions

**Proposition 2.10** *Let $V$ be a Banach space, and $J : \Omega \to \mathbb{R}$, where $\Omega$ is an open subset of $V$. Let $u \in \Omega$ be a local extremum of $J$, and assume that $J$ is Gâteaux-differentiable at $u$. Then,*

$$J'(u) = 0.$$

*This is the Euler's equation.*

**Proof**
For all $v$, for $\theta > 0$ small enough, $u + \theta v \in \mathcal{V}(u) \subset \Omega$. Then $J(u) \leq J(u + \theta v)$. Then $(J'(u), v) \geq 0$. But $(J'(u), -v) \geq 0$, and then $(J'(u), v) = 0$, $\forall v$. Then $J'(u) = 0$. □

**Remark:** This proposition is not true if $\Omega$ is not an open subset! For instance, consider $J(x) = x$ on $\Omega = [0, 1]$. 0 is a minimum of $J$, but $J'(0) = 1$. The proposition is usually applied to $\Omega = V$, or to an inner point of a subset.

**Notation:** The points $u$ where $J'(u) = 0$ are called the **critical points** of $J$.

**Proposition 2.11** *Let $V$ be a Banach space, and $K$ a convex subset of $V$. If $u$ is a local minimum of $J$ on $K$, and if $J$ is Gâteaux-differentiable at $u$, then*

$$(J'(u), v - u) \geq 0, \ \forall v \in K.$$

*This is the Euler's inequality.*

**Proof**
For all $v \in K$, for all $\theta > 0$ small enough, $u + \theta(v - u) \in \mathcal{V}(u) \cap K$ as long as $\theta \leq 1$. Then $J(u) \leq J(u + \theta(v - u))$, and with $\theta \to 0^+$, $(J'(u), v - u) \geq 0$. □

**Particular cases:**

i) $K$ is a sub-vector space of $V$: then $(J'(u), w) = 0$, $\forall w \in K$.

ii) $K$ is a sub-affine space of $V$: $v = a + v_0$, $v_0 \in$ sub-vector space $V_0$ of $V$. Then $(J'(u), w) = 0$, $\forall w \in V_0$.

iii) **Projection on a closed convex set:**

Let $K$ be a closed convex subset of a Hilbert space $V$. Let $J(v) = \|v-f\|^2$, for a given $f \in V\backslash K$. Let $u$ be the projection of $f$ on $K$. Then $(J'(u), v) = 2\langle u-f, v\rangle$. Then $(J'(u), v-u) = 2\langle u-f, v-u\rangle \geq 0$. Then

$$\langle f-u, v-u\rangle \leq 0, \ \forall v \in K.$$

This property characterizes the projection on a closed convex set.

iv) **Least-square estimation:**

$$y(x) = \sum_{j=1}^{p} v_j w_j(x)$$

The goal is to approximate the $b_i$'s: $y(x_i) \approx b_i$. Define

$$J(v) = \sum_{i=1}^{n}[y(x_i)-b_i]^2 = \sum_{i=1}^{n}\left[\sum_{j=1}^{p} v_j w_j(x_i) - b_i\right]^2$$

$$= \|Av-b\|^2.$$

Then $(J'(u), v) = \langle Au-b, Av\rangle = \langle {}^tA(Au-b), v\rangle$ and then

$${}^tAAu = {}^tAb$$

This equality is called the **normal equation** of the least-square estimation problem. The same idea can be used for estimating the solution of an overdetermined system of linear equations.

### 2.3.2 Convex case: sufficient optimality conditions

Let $V$ be a Banach space, and $K$ a convex subset of $V$.

**Proposition 2.12** *Let $J : V \to \mathbb{R}$ Gâteaux-differentiable and convex on $K$, where $K$ is an open convex subset of $V$. If $J'(u) = 0$, where $u \in K$, then $u$ is a global minimum of $J$ on $K$.*

**Proof**
$J(v) \geq J(u) + (J'(u), v-u) = J(u)$. □

**Proposition 2.13** *If $J : K \to \mathbb{R}$ is convex, Gâteaux-differentiable, and if $(J'(u), v-u) \geq 0$, $\forall v \in K$, then $u$ is a global minimum of $J$ on $K$.*

These are necessary conditions in a general case, and necessary and sufficient conditions in the convex case.

**Particular case:** $K = V$: $J'(u) = 0$ is the Euler's equation.

**Application to the minimization of quadratic cost functions** Let $a$ be a continuous symmetric bilinear coercive form: $a(v,v) \geq \alpha\|v\|^2$, $\alpha > 0$, $\forall v \in V$, and $L$ be a continuous linear form on $V$. We define

$$J(v) = \frac{1}{2}a(v,v) - L(v).$$

**Proposition 2.14** *$J$ is Fréchet-differentiable, and $\alpha$-convex.*

**Proof**
$J(u+h) = J(u) + (J'(u),h) + \frac{1}{2}a(h,h)$ and $(J'(u),h) = a(u,h) - L(h)$.
$(J'(v) - J'(u), v-u) = a(v-u, v-u) \geq \alpha\|v-u\|^2$, then $J$ is $\alpha$-convex. □

Euler's inequality for the optimization problem $\inf\limits_{v\in K} J(v)$ is

$$a(u, v-u) \geq L(v-u),\ \forall v \in K.$$

If $K = V$, then $a(u,v) = L(v),\ \forall v \in V$.

In the particular case where $V = \mathbb{R}^n$, then $a(v,v) = (Av, v)$ where $A$ is a symmetric positive definite matrix ($\alpha = \inf \lambda_i$). And $L(v) = (b,v)$, with $b \in \mathbb{R}^n$. In this case, the Euler's equation is:

$$Au = b.$$

**Non-differentiable convex cost functions** Let $J(v) = j_0(v) + j_1(v)$, where $j_0$ is convex and Gâteaux-differentiable, and $j_1$ is convex. If $J(u) = \inf\limits_{v\in K} J(v)$, with $u \in K$ a nonempty closed convex subset of $V$:

**Proposition 2.15** *$u$ is characterized by the following inequality:*

$$(j_0'(u), v-u) + j_1(v) - j_1(u) \geq 0,\ \forall v \in K.$$

**Proof**
Sufficient condition:

$$j_0(v) + j_1(v) - j_0(u) - j_1(u) = J(v) - J(u) \geq (j_0'(u), v-u) + j_1(v) - j_1(u) \geq 0.$$

Necessary condition:

$$j_0(u)+j_1(u) \leq j_0(u+\theta(v-u))+j_1(u+\theta(v-u)) \leq j_1(u)+\theta(j_1(v)-j_1(u))+j_0(u+\theta(v-u))$$

and then $(j_0'(u), v-u) + j_1(v) - j_1(u) \geq 0$. □

### 2.3.3 Second-order optimality conditions

**Proposition 2.16** *Necessary condition for a local minimum: let $\Omega$ be an open subset of a normed vector space $V$, and $J : \Omega \to \mathbb{R}$ a differentiable function in $\Omega$, twice differentiable at $u \in \Omega$. If $u$ is a local minimum of $J$, then*

$$J''(u)(w,w) \geq 0,\ \forall w \in V.$$

**Proof**
Taylor-Young:

$$J(u+\theta w) = J(u) + \theta(J'(u), w) + \frac{\theta^2}{2}[J''(u)(w,w) + \varepsilon(\theta)]$$

where $\varepsilon(\theta) \to 0$ when $\theta \to 0$. As $u$ is a local minimum of $J$, $J'(u) = 0$ and $J(u+\theta w) \geq J(u)$ for all $w$, for $\theta \geq 0$ small enough. Then $J''(u)(w,w) \geq 0$. □

**Definition 2.5** *$u$ is a strict local minimum of $J$ if there exists a neighborhood $\mathcal{V}(u)$ such that*

$$J(u) < J(v), \ \forall v \in \mathcal{V}(u)\backslash\{u\}.$$

**Proposition 2.17** *Sufficient condition for a strict local minimum: let $\Omega$ be an open subset of $V$, and $J : \Omega \to \mathbb{R}$ a differentiable function such that $J'(u) = 0$. If $J$ is twice differentiable at $u$, and if there exists $\alpha > 0$ such that $J''(u)(w,w) \geq \alpha\|w\|^2$, $\forall w \in V$, then $u$ is a strict local minimum of $J$.*

**Proof**

$$J(u+w) - J(u) = \frac{1}{2}[J''(u)(w,w) + \|w\|^2\varepsilon(w)] \geq \frac{\alpha - \varepsilon(w)}{2}\|w\|^2 > 0$$

in an open ball centered at $u$, of radius $r$ (small enough) such that $\|\varepsilon(w)\| < \alpha$ if $\|w\| \leq r$. □

**Proposition 2.18** *Let $\Omega$ be an open subset of $V$, $J$ a differentiable function such that $J'(u) = 0$. If $J$ is twice differentiable in $\Omega$, and if there exists a ball $\mathcal{B}$ centered at $u$ in $\Omega$ such that*

$$J''(v)(w,w) \geq 0, \ \forall v \in \mathcal{B}, \ \forall w \in V,$$

*then $u$ is a local minimum of $J$.*

**Proof**

Taylor Mac Laurin: $\exists v \in ]u, u+w[$ such that

$$J(u+w) = J(u) + \frac{1}{2}J''(v)(w,w) \geq J(u)$$

for all $u + w \in \mathcal{B}$. □

**Remark:** Straightforward application to a convex function.