

Continuous (convex) optimisation

M2 - PSL / Dauphine / S.U.

Antonin Chambolle, CNRS, CEREMADE

Université Paris Dauphine PSL

Sep.-Nov. 2024

Lecture 3: Subgradients, Monotone operators.

Contents

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

- 1 Monotone operators
 - Subgradients of convex functions
 - Elements of monotone operators theory

Generalized gradients: Subgradients of convex functions

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Consider f convex, proper (the definition also is valid for a non-convex function but conflicts with more reasonable, local definitions).

Definition: subgradient

The subgradient of f at $x \in \text{dom } f$ is the set:

$$\partial f(x) := \{p \in \mathcal{X} : f(y) \geq f(x) + \langle p, y - x \rangle \ \forall y \in \mathcal{X}\}.$$

This is clearly a **closed, convex** set.

Subgradient: fundamental property

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Theorem (?)

Let $f : \mathcal{X} \rightarrow (-\infty, +\infty]$ be convex, proper. Then $x \in \mathcal{X}$ is a minimizer of f if and only if $0 \in \partial f(x)$.

Proof: actually this is the definition of the subgradient.

Subgradient

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

If f (convex) is Gateaux-differentiable at x , that is if there exists $\nabla f(x) \in \mathcal{X}$ such that for any h ,

$$\lim_{t \rightarrow 0} \frac{f(x + th) - f(x)}{t} = \langle \nabla f(x), h \rangle$$

then $\partial f = \{\nabla f(x)\}$.

Subgradient

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

If f (convex) is Gateaux-differentiable at x , that is if there exists $\nabla f(x) \in \mathcal{X}$ such that for any h ,

$$\lim_{t \rightarrow 0} \frac{f(x + th) - f(x)}{t} = \langle \nabla f(x), h \rangle$$

then $\partial f = \{\nabla f(x)\}$.

- Indeed since f is convex then, for any h , $\phi : t \mapsto f(x + th)$ is convex and using $\phi(1) \geq \phi(0) + \phi'(0)$, that is:

$$f(x + h) \geq f(x) + \langle \nabla f(x), h \rangle,$$

which shows that $\nabla f(x) \in \partial f(x)$.

- On the other hand, for $p \in \partial f(x)$, $t > 0$ small, then $f(x + th) - f(x) \geq t \langle p, h \rangle$. Dividing by t and letting $t \rightarrow 0$ we deduce $\langle \nabla f(x) - p, h \rangle \geq 0$. Since this is true for any h , $p = \nabla f(x)$.

Subgradient

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

If f (convex) is Gateaux-differentiable at x , that is if there exists $\nabla f(x) \in \mathcal{X}$ such that for any h ,

$$\lim_{t \rightarrow 0} \frac{f(x + th) - f(x)}{t} = \langle \nabla f(x), h \rangle$$

then $\partial f = \{\nabla f(x)\}$.

- Indeed since f is convex then, for any h , $\phi : t \mapsto f(x + th)$ is convex and using $\phi(1) \geq \phi(0) + \phi'(0)$, that is:

$$f(x + h) \geq f(x) + \langle \nabla f(x), h \rangle,$$

which shows that $\nabla f(x) \in \partial f(x)$.

- On the other hand, for $p \in \partial f(x)$, $t > 0$ small, then $f(x + th) - f(x) \geq t \langle p, h \rangle$. Dividing by t and letting $t \rightarrow 0$ we deduce $\langle \nabla f(x) - p, h \rangle \geq 0$. Since this is true for any h , $p = \nabla f(x)$.

Subgradient

Existence of subgradients

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

If f is convex, $x \in \text{dom } f$, $v \in \mathcal{X}$, $t > s > 0$:

$$f(x + sv) = f((s/t)(x + tv) + (1 - s/t)x) \leq \frac{s}{t}f(x + tv) + (1 - \frac{s}{t})f(x)$$

so that

$$\frac{f(x + sv) - f(x)}{s} \leq \frac{f(x + tv) - f(x)}{t}.$$

It follows that

$$f'(x; v) := \lim_{t \downarrow 0^+} \frac{f(x + tv) - f(x)}{t} = \inf_{t > 0} \frac{f(x + tv) - f(x)}{t}$$

is well defined (in $[-\infty, \infty]$), and $< +\infty$ as soon as $\{x + tv : t > 0\} \cap \text{dom } f \neq \emptyset$.

Subgradient

Existence of subgradients

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

If f is convex, $x \in \text{dom } f$, $v \in \mathcal{X}$, $t > s > 0$:

$$f(x + sv) = f((s/t)(x + tv) + (1 - s/t)x) \leq \frac{s}{t}f(x + tv) + (1 - \frac{s}{t})f(x)$$

so that

$$\frac{f(x + sv) - f(x)}{s} \leq \frac{f(x + tv) - f(x)}{t}.$$

It follows that

$$f'(x; v) := \lim_{t \downarrow 0^+} \frac{f(x + tv) - f(x)}{t} = \inf_{t > 0} \frac{f(x + tv) - f(x)}{t}$$

is well defined (in $[-\infty, \infty]$), and $< +\infty$ as soon as $\{x + tv : t > 0\} \cap \text{dom } f \neq \emptyset$.

Hence: if $x \in \overbrace{\text{dom } f}^{\circ}$, then $f'(x; v) < \infty$ for all v . In addition
 $f'(x; 0) = 0 \leq f'(x; v) + f'(x; -v)$ hence $f'(x; v) > -\infty$.

Subgradient

Existence of subgradients

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

One has: $f'(x; \cdot)$ is a limit of convex functions, and hence convex, moreover, it is clearly positively 1-homogeneous: $f'(x; \lambda v) = \lambda f'(x; v)$ for all $\lambda \geq 0$ and all v .

Let $C = \{p : \langle p, v \rangle \leq f'(x; v) \forall v\} = \partial f'(x, \cdot)(0)$. It is a subgradient and hence a closed, convex set. In fact, $C = \partial f(x)$:

if $p \in C$ then $f(x + v) - f(x) \geq f'(x; v) \geq \langle p, v \rangle$ for all v and $p \in \partial f(x)$. Conversely if $p \in \partial f(x)$ then for any v , $(f(x + tv) - f(x))/t \geq \langle p, v \rangle$ and in the limit $f'(x; v) \geq \langle p, v \rangle$.

Subgradient

Existence of subgradients

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

One has: $f'(x; \cdot)$ is a limit of convex functions, and hence convex, moreover, it is clearly positively 1-homogeneous: $f'(x; \lambda v) = \lambda f'(x; v)$ for all $\lambda \geq 0$ and all v .

Let $C = \{p : \langle p, v \rangle \leq f'(x; v) \forall v\} = \partial f'(x, \cdot)(0)$. It is a subgradient and hence a closed, convex set. In fact, $C = \partial f(x)$:

if $p \in C$ then $f(x + v) - f(x) \geq f'(x; v) \geq \langle p, v \rangle$ for all v and $p \in \partial f(x)$. Conversely if $p \in \partial f(x)$ then for any v , $(f(x + tv) - f(x))/t \geq \langle p, v \rangle$ and in the limit $f'(x; v) \geq \langle p, v \rangle$.

By Fenchel duality the convex, lower-semicontinuous envelope of $v \mapsto f'(x; v)$ is the support function of C (which could be empty).

Subgradient

Existence of subgradients

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

One has: $f'(x; \cdot)$ is a limit of convex functions, and hence convex, moreover, it is clearly positively 1-homogeneous: $f'(x; \lambda v) = \lambda f'(x; v)$ for all $\lambda \geq 0$ and all v .

Let $C = \{p : \langle p, v \rangle \leq f'(x; v) \forall v\} = \partial f'(x, \cdot)(0)$. It is a subgradient and hence a closed, convex set. In fact, $C = \partial f(x)$:

if $p \in C$ then $f(x + v) - f(x) \geq f'(x; v) \geq \langle p, v \rangle$ for all v and $p \in \partial f(x)$. Conversely if $p \in \partial f(x)$ then for any v , $(f(x + tv) - f(x))/t \geq \langle p, v \rangle$ and in the limit $f'(x; v) \geq \langle p, v \rangle$.

By Fenchel duality the convex, lower-semicontinuous envelope of $v \mapsto f'(x; v)$ is the support function of C (which could be empty).

In finite dimension we deduce that the subgradient $\partial f(x)$ is not empty for any x in the interior of the domain (actually in $\text{ri dom } f$, also).

Subgradient

Existence of subgradients

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

In ∞ dimension it remains true if we assume in addition f *lower semicontinuous*:

Then we have seen that f is bounded in the interior of its domain and therefore locally Lipschitz. Hence for v in the unit ball and t small enough,

$(f(x + tv) - f(x))/t$ is also Lipschitz therefore also $v \mapsto f'(x; v)$ is.

Since then

$$f'(x; v) = \sup_{p \in C} \langle p, v \rangle$$

it shows that $C = \partial f(x)$ is not empty, and bounded.

We will show later on that in general, for a convex lsc function, $\text{dom } \partial f$ is dense in $\text{dom } f$ (even when this set has empty interior).

Subgradient

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Additionally, for x in the interior of $\text{dom } f$, in case $\partial f(x) = \{p\}$, then $f'(x; v) = \langle p, v \rangle$ for any v , that is: f is Gateaux differentiable in x .

Lemma

Let f be convex lsc and $x \in \overbrace{\text{dom } f}$. Then f is (Gateaux) differentiable at x if and only if ∂f has exactly one element.

Remark: One could assume $x \in \text{ri dom } f$ in the finite-dimensional case yet we could just deduce that the restriction of f to its domain D is Gateaux at x differentiable if and only if $\partial f(x) = D^\perp + \{p\}$

Remark 1: subgradient of strongly convex function

Lemma

Let f be strongly convex with parameter $\mu > 0$. Then for any $x \in \text{dom } \partial f$, $y \in \text{dom } f$ and $p \in \partial f(x)$,

$$f(y) \geq f(x) + \langle p, y - x \rangle + \frac{\mu}{2} |x - y|^2$$

Proof: Let $p \in \partial f(x)$. For all $y, t \in (0, 1)$,

$$f(ty + (1 - t)x) \leq tf(y) + (1 - t)f(x) - \mu \frac{t(1 - t)}{2} |x - y|^2$$

and

$$f(ty + (1 - t)x) \geq f(x) + \langle p, ty + (1 - t)x - x \rangle = f(x) + t \langle p, y - x \rangle.$$

Combining the two, we find:

$$tf(y) + (1 - t)f(x) - \mu \frac{t(1 - t)}{2} |x - y|^2 \geq f(x) + t \langle p, y - x \rangle$$

Remark 1: subgradient of strongly convex function

Lemma

Let f be strongly convex with parameter $\mu > 0$. Then for any $x \in \text{dom } \partial f$, $y \in \text{dom } f$ and $p \in \partial f(x)$,

$$f(y) \geq f(x) + \langle p, y - x \rangle + \frac{\mu}{2} |x - y|^2$$

Proof: Let $p \in \partial f(x)$. For all $y, t \in (0, 1)$,

$$f(ty + (1 - t)x) \leq tf(y) + (1 - t)f(x) - \mu \frac{t(1 - t)}{2} |x - y|^2$$

and

$$f(ty + (1 - t)x) \geq f(x) + \langle p, ty + (1 - t)x - x \rangle = f(x) + t \langle p, y - x \rangle.$$

Combining the two, we find:

$$tf(y) + (1 - t)f(x) - \mu \frac{t(1 - t)}{2} |x - y|^2 \geq f(x) + t \langle p, y - x \rangle$$

We recover the claim after dividing by t and sending $t \rightarrow 0$.

Remark 2: subgradient of functions with Lipschitz gradient

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Lemma

Let f be convex with L -Lipschitz gradient. Then for any $x, y \in \mathcal{X}$,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} |\nabla f(x) - \nabla f(y)|^2$$

[proved later on from the previous result, by duality.]

Legendre-Fenchel Identity

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

If x realizes the sup in $f^*(y) = \sup_x \langle y, x \rangle - f(x)$ then for all z ,

$$\langle y, x \rangle - f(x) \geq \langle y, z \rangle - f(z) \Leftrightarrow f(z) \geq f(x) + \langle y, z - x \rangle$$

which means that $y \in \partial f(x)$. (And conversely, $y \in \partial f(x)$ implies x realizes the sup...)

In addition, using, $\langle y, x \rangle - f(x) = f^*(y)$ and $f(x) \geq f^{**}(x) \geq \langle y, x \rangle - f^*(y)$, one sees that $f(x) = f^{**}(x)$, y realizes the sup in the definition of $f^{**}(y)$, so that $x \in \partial f^*(y)$, and $y \in \partial f^{**}(x)$ (and f is lsc at x).

In particular we see that $\partial f^{**}(x) \supseteq \partial f(x)$ for all x . Precisely we have:

Legendre-Fenchel identity

$$y \in \partial f(x) \Leftrightarrow \langle x, y \rangle = f(x) + f^*(y) \Rightarrow x \in \partial f^*(y),$$

the latter being also an equivalence if f is lsc, convex (if $f = f^{**}$).

“Subdifferential calculus”

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

A first simple example: minimizing $f + g$ with g smooth.

Lemma

Assume $x \in \mathcal{X}$ is a minimizer of $f + g$, where f is convex and g is C^1 . Then for all $y \in \mathcal{X}$,

$$f(y) \geq f(x) - \langle \nabla g(x), y - x \rangle$$

that is, $-\nabla g(x) \in \partial f(x) \Leftrightarrow \partial f(x) + \nabla g(x) \ni 0$.

Proof: For $t > 0$ small enough,

$$f(x) + g(x) \leq f(x + t(y - x)) + g(x + t(y - x)) \leq f(x) + t(f(y) - f(x)) + g(x + t(y - x))$$

so that

$$\frac{g(x) - g(x + t(y - x))}{t} \leq f(y) - f(x)$$

and we recover the claim in the limit $t \rightarrow 0$.

Remark: density of subgradients

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Corollary

Let f be convex, lsc: then $\text{dom } \partial f$ is dense in $\text{dom } f$.

Proof: Let $\bar{x} \in \text{dom } f$, $\tau > 0$ and let x_τ be the minimizer of $|x - \bar{x}|^2/(2\tau) + f(x)$. Then by the previous result,

$$\frac{\bar{x} - x_\tau}{\tau} \in \partial f(x_\tau)$$

so that $x_\tau \in \text{dom } \partial f$. In addition, $|x_\tau - \bar{x}|^2 \leq 2\tau f(\bar{x}) \rightarrow 0$ as $\tau \rightarrow 0$ since $f(\bar{x}) < +\infty$. □

Subdifferential calculus

The subgradient of a sum

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Theorem

Let f, g be convex, proper.

- For all x , $\partial f(x) + \partial g(x) \subset \partial(f + g)(x)$.
- If there exists $\bar{x} \in \text{dom } f$ where g is continuous, then $\partial f(x) + \partial g(x) = \partial(f + g)(x)$. (In finite dimension, a relevant, weaker condition is $\text{ri dom } g \cap \text{ri dom } f \neq \emptyset$.)

Proof: the inclusion is obvious from the definition. For the reverse inclusion, we assume $p \in \partial(f + g)(x)$ and want to show that it can be decomposed as $q + r$ with $q \in \partial f(x)$ and $r \in \partial g(x)$.

By definition, we have that $f(y) + g(y) \geq f(x) + g(x) + \langle p, y - x \rangle$.

Subdifferential calculus

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Thanks to the assumption that g is continuous at \bar{x} , $\text{epi}(g(\cdot) - \langle p, \cdot \rangle)$ contains a ball B centered at $(\bar{x}, g(\bar{x}) - \langle p, \bar{x} \rangle + 1)$ and has non empty interior. Denote E this interior, and F the following translation/flip of $\text{epi } f$:

$$F = \{(y, t) : -t \geq f(y) - [f(x) + g(x) - \langle p, x \rangle]\},$$

which is convex.

For $(y, t) \in F$, one has $-t \geq f(y) - [f(x) + g(x) - \langle p, x \rangle] \geq -[g(y) - \langle p, y \rangle]$, that is $t \leq [g(y) - \langle p, y \rangle]$ so that $(y, t) \notin E$.

Hence by the separation theorem there exists $(q, \lambda) \neq (0, 0)$, such that for all $(y, t) \in E$, $(y', t') \in F$,

$$\langle q, y \rangle + \lambda t \geq \langle q, y' \rangle + \lambda t'.$$

As t' can be sent to $-\infty$ (or t to $+\infty$), $\lambda \geq 0$. Moreover since \bar{x} is in $\text{dom } f$, if $\lambda = 0$ one finds that $\langle q, y - \bar{x} \rangle \leq 0$ for all $y \in \text{dom } g$ which contains a ball centered in \bar{x} , so that $q = 0$, which is a contradiction.

Subdifferential calculus

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Hence $\lambda > 0$ so that without loss of generality we can assume $\lambda = 1$.

In particular choosing $t' = f(x) + g(x) - \langle p, x \rangle - f(y')$,

$$\langle q, y \rangle + t \geq \langle q, y' \rangle + f(x) + g(x) - \langle p, x \rangle - f(y').$$

for all $(y, t) \in E$. The closure of E contains $\text{epi}(g(\cdot) - \langle p, \cdot \rangle)$: indeed any $(y, t) \in \text{epi}(g(\cdot) - \langle p, \cdot \rangle)$ is on the boundary of the set $\{ty + (1-t)B : 0 < t < 1\} \subset \text{epi}(g(\cdot) - \langle p, \cdot \rangle)$.

Hence it follows that for all y, y' ,

$$\begin{aligned} \langle q, y \rangle + g(y) - \langle p, y \rangle &\geq \langle q, y' \rangle + f(x) + g(x) - \langle p, x \rangle - f(y') \\ &\Leftrightarrow f(y') + g(y) \geq f(x) + g(x) + \langle p, y - x \rangle + \langle q, y' - y \rangle \\ &= f(x) + g(x) + \langle p - q, y - x \rangle + \langle q, y' - x \rangle \end{aligned}$$

showing that $q \in \partial f(x)$ and $r = p - q \in \partial g(x)$, as requested. \square

Remark: For f, g convex, proper, lsc. the result is also deduced from the theorem on inf-convolutions...

Subdifferential calculus

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Theorem

Let $A : \mathcal{X} \rightarrow \mathcal{Y}$ be a continuous operator between two Hilbert spaces and f a proper, convex function on \mathcal{Y} . Let $g = f(Ax)$, then if there is \bar{x} such that f is continuous at $A\bar{x}$, $\partial g(x) = A^* \partial f(Ax)$. In finite dimension, one can just require that $A\bar{x} \in \text{ri dom } f$.

Proof is similar (again, one inclusion is easy).

Application: Karush-Kuhn-Tucker's theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

KKT's Theorem

Let $f, g_i, i = 1, \dots, m$ be C^1 , convex and assume

$$\exists \bar{x}, (g_i(\bar{x}) < 0 \forall i = 1, \dots, m) \quad (\text{Slater's condition})$$

Then x^* is a solution of

$$\min_{g_i(x) \leq 0, i=1, \dots, m} f(x)$$

if and only if there exists $(\lambda_i)_{i=1}^m, \lambda_i \geq 0$ such that

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) = 0,$$

$$\sum_{i=1}^m \lambda_i g_i(x^*) = 0 \quad (\text{complementary slackness condition})$$

KKT's Theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Proof: Observe that since $g_i(x^*) \leq 0$ and $\lambda_i \geq 0$ the complementary condition is also equivalent to:
 $\forall i, g_i(x^*) = 0$ or $\lambda_i = 0$.

If the last statements are true, then x^* is is a minimizer of the convex function $f + \sum_i \lambda_i g_i$. Then obviously for any x with $g_i(x) \leq 0$ for all i ,

$$f(x) \geq f(x) + \sum_i \lambda_i g_i(x) \geq f(x^*) + \sum_i \lambda_i g_i(x^*) = f(x^*).$$

Conversely, consider for all i the function

$$\delta_i(x) = \begin{cases} 0 & \text{if } g_i(x) \leq 0, \\ +\infty & \text{else.}, \end{cases}$$

then the problem is equivalent to $\min_x f(x) + \sum_i \delta_i(x)$. By Slater's condition, we know that there exists \bar{x} where all functions f, δ_i are continuous. Hence by the previous theorems:

$$0 \in \partial(f + \sum_i \delta_i)(x^*) = \nabla f(x^*) + \sum_{i=1}^m \partial \delta_i(x^*).$$

KKT's Theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

It remains to characterize $\partial\delta_i(x^*)$.

If $g_i(x^*) < 0$ then it is negative in a neighborhood of x^* and $\partial\delta_i(x^*) = \{0\}$.

If $g_i(x^*) = 0$, then we need to characterize the vectors p such that for all y with $g_i(y) \leq 0$,

$$0 \geq \langle p, y - x^* \rangle.$$

Let $v \perp \nabla g_i(x^*)$, and consider $y = x^* - t(\nabla g_i(x^*) + v)$: then

$$g_i(y) = -t \langle \nabla g_i(x^*), \nabla g_i(x^*) + v \rangle + o(t) = -t |\nabla g_i(x^*)|^2 + o(t) < 0$$

if $t > 0$ is small enough, hence

$$0 \leq \langle p, \nabla g_i(x^*) + v \rangle.$$

We easily deduce that we must have $p = \lambda_i \nabla g_i(x^*)$, for some $\lambda_i \geq 0$ (in other words, $\partial\delta_i(x^*) = \mathbb{R}_+ \nabla g_i(x^*)$). The theorem follows. □

KKT's Theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Remark: in case g_i is affine it is enough to assume $g_i(\bar{x}) = 0$, this allows in particular to treat also the case of affine equality constraints ($g(x) = 0 \Leftrightarrow (g(x) \leq 0 \text{ and } -g(x) \leq 0)$).

Monotone operators in Hilbert spaces

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

A fundamental property of subgradients is the *monotonicity*: Using that for all $p \in \partial f(x)$, $p' \in \partial f(x')$:

$$f(x') \geq f(x) + \langle p, x' - x \rangle, \quad f(x) \geq f(x') + \langle p', x - x' \rangle,$$

and summing both inequalities, we find

$$0 \geq \langle p - p', x' - x \rangle.$$

In [1D](#), this is equivalent to saying that ∂f is non-decreasing (if $x' > x$, p' must be $\geq p$). In general one says that ∂f is a “monotone operator”:

Definition

The operator $A : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X})$ is monotone if and only if $\forall x, x' \in \mathcal{X}$, $\forall p \in Ax$ and $p' \in Ax'$, one has

$$\langle p' - p, x' - x \rangle \geq 0.$$

Monotone operators in Hilbert spaces

More definitions

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Definition

The operator $A : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X})$ is (μ) -strongly monotone if and only if $\forall x, x' \in \mathcal{X}$, $\forall p \in Ax$ and $p' \in Ax'$, one has

$$\langle p - p', x - x' \rangle \geq \mu |x - x'|^2.$$

It is (μ) -co-coercive if

$$\langle p - p', x - x' \rangle \geq \mu |p - p'|^2.$$

It is *maximal* if the graph $\{(x, p) : p \in Ax\} \subset \mathcal{X} \times \mathcal{X}$ is maximal with respect to inclusion, among all the graphs of monotone operators.

In dimension 1: graphs of nondecreasing functions / (sub)gradients of convex functions.
In higher dimension, not true anymore (example: an antisymmetric linear mapping in \mathbb{R}^d , $d \geq 2$).

The subgradient of a convex function f is monotone, strongly monotone if f is strongly convex, co-coercive if ∇f is Lipschitz ("Baillon-Haddad").

Monotone operators in Hilbert spaces

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Lemma

Let f be convex. Then ∂f is a maximal-monotone operator if and only if it is the subgradient of a lower-semicontinuous function.

Proof: (cf Rockafellar): if f is lsc, to show that ∂f is maximal we must show that if $x \in \mathcal{X}$ and $p \notin \partial f(x)$ then one can find y and $q \in \partial f(y)$ with $\langle p - q, x - y \rangle < 0$.

Replacing f with $f(x) - \langle p, x \rangle$ we can assume that $p = 0$, that is, $0 \notin \partial f(x)$.

Consider now the minimizer of $f(y) + |y - x|^2/2$ which exists as this function is strongly convex and lsc. It is characterized by $\partial f(y) + (y - x) \ni 0$ that is, $q = x - y \in \partial f(y)$. Then, necessarily $q \neq 0$ otherwise this means $0 \in \partial f(x)$. Then,

$$\langle p - q, x - y \rangle = \langle -q, x - y \rangle = -|x - y|^2 = -|q|^2 < 0.$$

This shows that ∂f is maximal.

Conversely if ∂f is maximal, since $\partial f^{**} \supset \partial f$, then this operator is also the subgradient of the convex, lsc function f^{**} . We are *not* proving here that $f = f^{**}$, only that ∂f is *also* the subgradient of the convex, lsc function f^{**} . f and f^{**} could differ at some point where $\partial f(x) = \emptyset$.

Monotone operators in Hilbert spaces

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Definition

Given A a monotone operator, with graph $\{(x, p) : p \in Ax\}$, its *inverse* is $A^{-1} : p \mapsto \{x : Ax \ni p\}$, with graph $\{(p, x) : p \in Ax\}$.

Therefore, it is maximal if and only if A is maximal, co-coercive if and only if A is strongly monotone.

Remark: For f convex lsc.*, $(\partial f)^{-1} = \partial f^*$ (by Legendre-Fenchel's identity).

Monotone operators in Hilbert spaces

Sum of Maximal-Monotone operators

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Lemma

Let A, B be maximal monotone operators. if $\overset{\circ}{\text{dom } A} \cap \text{dom } B \neq \emptyset$, then $A + B$ (which is always monotone) is maximal monotone.

(Cor 2.7 in H. Brézis: *Opérateurs maximaux-monotones et semi-groupes de contraction dans les espaces de Hilbert*).

Monotone operators in Hilbert spaces

Minty's theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Theorem (Minty 62)

The *resolvent* of a maximal-monotone operator A , defined by

$$x \mapsto y = (I + A)^{-1}x =: J_A x \Leftrightarrow y + Ay \ni x$$

is a well (everywhere) defined single-valued nonexpansive mapping. (Conversely, for a monotone operator A if $(I + A)$ is surjective then A is maximal.)

Minty's theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Proof: We introduce the graph $G = \{(y + x, y - x) : x \in \mathcal{X}, y \in Ax\}$. If $(a, b), (a', b') \in G$, with $a = y + x, b = y - x$ and $a' = y' + x', b' = y' - x'$, then

$$|b - b'|^2 = |y - y'|^2 - 2 \langle y - y', x - x' \rangle + |x - x'|^2 = |a - a'|^2 - 4 \langle y - y', x - x' \rangle \leq |a - a'|^2$$

that is G is the graph of a 1-Lipschitz function. [Conversely, G 1-Lipschitz implies A monotone.]

Moreover, if $G' \supseteq G$ is also the graph of a 1-Lipschitz function, then defining

$A' = \{((a - b)/2, (a + b)/2) : (a, b) \in G'\}$ the same computation shows that $A' \supseteq A$ is the graph of a monotone operator, hence if A is maximal: $A' = A$ and $G' = G$.

In particular, if G is defined for all a then clearly G and therefore A are maximal (Remark: being 1-Lipschitz, G is necessarily single-valued).

So the theorem is equivalent to the question whether a 1-Lipschitz function which is not defined in the whole of \mathcal{X} can be extended.

This result (which is true only in Hilbert spaces) is known as Kirszbraun-Valentine's theorem (1935), we give a quick proof derived from Federer (*Geometric measure theory*, 2.10.43).

Minty's / Kirschbraun-Valentine's theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

The basic brick is the following extension from n to $n + 1$ points:

Lemma

If $(x_i)_{i=1}^n, (y_i)_{i=1}^n$ are points in Hilbert spaces respectively \mathcal{X}, \mathcal{Y} such that $\forall i, j, |y_i - y_j| \leq |x_i - x_j|$, then for any $x \in \mathcal{X}$ there exists $y \in \mathcal{Y}$ with $|y_i - y| \leq |x_i - x|$ for all $i = 1, \dots, n$.

Proof: It is enough to prove this for $x = 0$: we need to find a common point to $\bar{B}(y_i, |x_i|)$. There is nothing to prove if $x = x_i$ for some i , so we assume $x_i \neq 0, i = 1, \dots, n$.

We define

$$\bar{c} = \min \left\{ c \geq 0 : \bigcap_{i=1}^n \bar{B}(y_i, c|x_i|) \neq \emptyset \right\} > 0$$

(if the y_i are distinct, which we may also assume). This is a min because the closed balls are weakly compact, and we can consider y such that $|y - y_i| \leq \bar{c}|x_i|, i = 1, \dots, n$.

We must show that $\bar{c} \leq 1$.

Kirszbraun-Valentine's theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Then: y must be a convex combination of the points $(y_i)_{i \in I}$ such that $|y - y_i| = \bar{c}|x_i|$.
Indeed, if not, let y' be the projection of y onto $\overline{\text{co}} \{y_i : i \in I\}$. As for any $i \in I$, $\langle y_i - y', y - y' \rangle \leq 0$ one has, letting $y_t = (1 - t)y + ty'$, that for any $i \in I$:

$$\begin{aligned} |y_i - y_t|^2 &= |y_i - y + t(y - y')|^2 = |y_i - y|^2 + 2t \langle y_i - y, y - y' \rangle + t^2 |y - y'|^2 \\ &= |y_i - y|^2 + 2t \langle y_i - y', y - y' \rangle - 2t |y - y'|^2 + t^2 |y - y'|^2 \\ &\leq |y_i - y|^2 - t(2 - t) |y - y'|^2 < |y_i - y|^2 \end{aligned}$$

if $t \in (0, 2)$.

Hence if $t > 0$ is small enough, one sees that $|y_i - y_t| < |y_i - y| = \bar{c}|x_i|$ for $i \in I$, while since for $i \notin I$, $|y_i - y| < \bar{c}|x_i|$, one can still guarantee the same strict inequality for y_t if t is small enough. But this contradicts the definition of \bar{c} , since then there would exist $c < \bar{c}$ such that $y_t \in \bigcap_{i=1}^n \bar{B}(y_i, c|x_i|)$.

Kirschbraun-Valentine's theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Hence we can write $y = \sum_{i \in I} \theta_i y_i$ as a convex combination ($\theta_i \in [0, 1], \sum_{i \in I} \theta_i = 1$). Then since $2 \langle a, b \rangle = |a|^2 + |b|^2 - |a - b|^2$,

$$\begin{aligned} 0 &= \left| \sum_{i \in I} \theta_i y_i - y \right|^2 = \sum_{i, j \in I} \theta_i \theta_j \langle y_i - y, y_j - y \rangle \\ &= \frac{1}{2} \sum_{i, j \in I} \theta_i \theta_j (|y_i - y|^2 + |y_j - y|^2 - |y_i - y_j|^2) \\ &\geq \frac{1}{2} \sum_{i, j \in I} \theta_i \theta_j (\bar{c}^2 |x_i|^2 + \bar{c}^2 |x_j|^2 - |x_i - x_j|^2) \\ &= \bar{c}^2 \sum_{i, j \in I} \theta_i \theta_j \langle x_i, x_j \rangle - \frac{1 - \bar{c}^2}{2} |x_i - x_j|^2 \end{aligned}$$

which shows that

$$(1 - \bar{c}^2) \sum_{i, j \in I} \theta_i \theta_j |x_i - x_j|^2 \geq 2\bar{c}^2 \left| \sum_{i \in I} \theta_i x_i \right|^2$$

so that $\bar{c} \leq 1$. Hence, y satisfies $|y - y_i| \leq |x_i|$, as requested, which shows the Lemma.

Minty's theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

We finish the proof of Minty's Theorem: if there exists $x \in \mathcal{X}$ such that $\{x\} \times \mathcal{X} \cap G = \emptyset$, consider the set

$$K = \bigcap_{(a,b) \in G} \bar{B}(b, |x - a|)$$

which is an intersection of weakly compact sets.

We show that because the compact sets defining K have the "*finite intersection property*", K can not be empty: Choosing $(a_0, b_0) \in G$, if $\bar{B}_0 = \bar{B}(b_0, |x - a_0|)$, we see that

$$K = \bar{B}_0 \cap \left(\bigcap_{(a,b) \in G} \bar{B}(b, |x - a|) \right)$$

hence $\bar{B}_0 \setminus K = \bar{B}_0 \cap \bigcup_{(a,b) \in G} \bar{B}(b, |x - a|)^c$.

If this is \bar{B}_0 , by compactness one can extract a finite covering $\bigcup_{i=1}^n \bar{B}(b_i, |x - a_i|)^c$ for $(a_i, b_i) \in G$, $i = 1, \dots, n$. We find that

$$\bar{B}_0 \cap \bigcup_{i=1}^n \bar{B}(b_i, |x - a_i|)^c = \bar{B}_0$$

or equivalently that

$$\bar{B}_0 \cap \bigcap_{i=1}^n \bar{B}(b_i, |x - a_i|) = \emptyset$$

which contradicts The Lemma.

Minty's theorem

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Hence, $\bar{B}_0 \setminus K \neq \bar{B}_0$ which means that $K \neq \emptyset$. Choosing $y \in K$, we find that $G \cup \{(x, y)\}$ is the graph of a 1-Lipschitz function and is strictly larger than G , which contradicts the maximality of A .

The non-expansiveness of $(I + A)^{-1}$ follows from, if $y + Ay \ni x$, $y' + Ay' \ni x'$, $p = x - y \in Ay$, $p' = x' - y' \in Ay'$:

$$|x - x'|^2 = |y - y'|^2 + 2 \langle p - p', y - y' \rangle + |p - p'|^2 \geq |y - y'|^2 + |p - p'|^2,$$

that is, for $T = (I + A)^{-1}$:

$$|Tx - Tx'|^2 + |(I - T)x - (I - T)x'|^2 \leq |x - x'|^2.$$

An operator which satisfies this is said firmly non-expansive (\Leftrightarrow 1-co-coercive).

Reflexion operator

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Given A maximal monotone, we define the *Reflexion* of A :

$$R_A = 2J_A - I = 2(I + A)^{-1} - I$$

Lemma

R_A is nonexpansive, and in particular, $J_A = I/2 + R_A/2$ is $(1/2)$ -averaged.

In fact one has even:

Proposition

For an operator $T : \mathcal{X} \rightarrow \mathcal{X}$, the following are equivalent:

- ① T is the resolvent of a maximal-monotone operator.
- ② T is firmly non-expansive;
- ③ T is $1/2$ -averaged, that is, $R = 2T - I$ is non-expansive;

Reflexion operator

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Proof of the lemma: We prove $(2) \Leftrightarrow (3)$ in the theorem. It follows in an obvious way from the parallelogram identity: for any x, x' ,

$$\begin{aligned} |Rx - Rx'|^2 &= |(Tx - x) - (Tx' - x') + Tx - Tx'|^2 \\ &= 2|(I - T)x - (I - T)x'|^2 + 2|Tx - Tx'|^2 - |x - x'|^2 \leq |x - x'|^2 \\ &\Leftrightarrow |(I - T)(x) - (I - T)(x')|^2 + |Tx - Tx'|^2 \leq |x - x'|^2. \end{aligned}$$

Remark: more generally, the parallelogram identity/strong convexity of $|\cdot|^2/2$ shows that: T_θ is θ -averaged for some $0 < \theta \leq 1$ (that is $T_\theta = (1 - \theta)I + \theta T$, T 1-Lipschitz) if and only if for all x, x' :

$$|T_\theta x - T_\theta x'|^2 + \frac{1 - \theta}{\theta} |(I - T_\theta)x - (I - T_\theta)x'|^2 \leq |x - x'|^2$$

To finish the proof of the theorem, we have to prove that if an operator $T = I/2 + R/2$ is $(1/2)$ -averaged (R is non-expansive), then there exists a maximal monotone operator A such that $T = J_A$.

Reflexion operator

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

The proof follows by the same (or reverse) construction as in the beginning of the proof of Minty's theorem: we consider the graph

$$G = \{((x + y)/2, (x - y)/2) : x \in \mathcal{X}, y = Rx\} = \{(Tx, (I - T)x) : x \in \mathcal{X}\}$$

and denote by A the corresponding operator ($y \in Ax \Leftrightarrow (x, y) \in G$). Then A is monotone: if $(\xi, \eta), (\xi', \eta') \in G$, then for some $x, x' \in \mathcal{X}$, $\xi = (x + Rx)/2$, $\eta = (x - Rx)/2$, etc., and we find:

$$\begin{aligned}\langle \xi - \xi', \eta - \eta' \rangle &= \frac{1}{4} \langle x + Rx - x' - Rx', x - Rx - x' + Rx' \rangle \\ &= \frac{1}{4} (|x - x'|^2 - |Rx - Rx'|^2) \geq 0.\end{aligned}$$

Moreover, A is maximal, if not, one could build as before from $A' \supset A$ a non-expansive graph $\{(\xi + \eta, \xi - \eta) : \eta \in A'\xi\}$ strictly larger than the graph $\{(x, Rx) : x \in \mathcal{X}\}$, which is of course impossible. By construction, $ATx \ni (I - T)x$ for all x , hence $(I + A)Tx \ni x \Leftrightarrow Tx = (I + A)^{-1}x$.

A practical consequence: proximal point algorithm

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

If $x^0 \in \mathcal{X}$ and $x^{k+1} = (I + A)^{-1}x^k$, $k \geq 0$, and there exists \bar{x} with $A\bar{x} \ni 0 \Leftrightarrow (I + A)^{-1}\bar{x} = \bar{x}$, then $x^k \rightarrow \bar{x}$ where $Ax \ni 0$ (KM's theorem).

In particular if $A = \tau \partial g$ for g convex, lsc and $\tau > 0$,

$$x^{k+1} = (I + A)^{-1}(x^k) \Leftrightarrow x^{k+1} \in x^k - \tau \partial g(x^{k+1}) \Leftrightarrow x^{k+1} = \arg \min_x g(x) + \frac{1}{2\tau} |x - x^k|^2$$

we see that the implicit gradient descent converges, as the iterations of a $1/2$ -averaged operator.

Definition

The resolvent of the subgradient ∂g of a convex, lsc function is called the “proximity operator” (or “proximal”) of g :

$$\text{prox}_g(x) = (I + \partial g)^{-1}(x) = \arg \min_{x'} g(x') + \frac{1}{2} |x' - x|^2.$$

Moreau's identity

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Lemma

Let A be a maximal-monotone operator. Then for any $x \in \mathcal{X}$,

$$x = (I + A)^{-1}(x) + (I + A^{-1})^{-1}x.$$

Proof: one has $y = (I + A)^{-1}x \Leftrightarrow y + Ay \ni x \Leftrightarrow y \in A^{-1}(x - y)$, letting then $z = x - y$, this is $x \in z + A^{-1}z \Leftrightarrow z = (I + A^{-1})^{-1}x$. □

This is often written, for $\tau > 0$:

$$x = (I + \tau A)^{-1}(x) + \tau(I + \frac{1}{\tau}A^{-1})^{-1}(\frac{x}{\tau}),$$

or for $A = \partial g$, g convex lsc,

$$x = (I + \tau \partial g)^{-1}(x) + \tau(I + \frac{1}{\tau} \partial g^*)^{-1}(\frac{x}{\tau}) = \text{prox}_{\tau g}(x) + \tau \text{prox}_{g^*/\tau}(\frac{x}{\tau}).$$

Remark: Yosida regularization and gradient flows

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Given A a maximal monotone operator, the maximal monotone operator $A_\tau = [x - (I + \tau A)^{-1}x]/\tau$ is called a *Yosida* approximation of A : it is a $(1/\tau)$ -Lipschitz-continuous mapping, with full domain. In case $A = \partial f$, $A_\tau = \nabla f_\tau$ where

$$f_\tau(x) = \min_{x'} f(x') + \frac{1}{2\tau}|x - x'|^2.$$

The operator τA_τ is firmly non-expansive, since $I - \tau A_\tau$ is. It is a key tool for establishing the existence of solutions to:

$$\dot{x} + Ax \ni 0$$

(cf H. Brézis, *Opérateurs maximaux-monotones et semi-groupes de contraction dans les espaces de Hilbert*).

Back to Fenchel-Rockafellar duality

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

Consider again:

$$\min_{x \in \mathcal{X}} f(Kx) + g(x)$$

with $K : \mathcal{X} \rightarrow \mathcal{Y}$ is continuous linear map and f, g convex, lsc. Then we have seen that a solution can be found as a saddle-point of

$$\mathcal{L}(x, y) = \langle y, Kx \rangle - f^*(y) + g(x),$$

that is (x^*, y^*) such that:

$$\mathcal{L}(x^*, y) \leq \mathcal{L}(x^*, y^*) \leq \mathcal{L}(x, y^*) \tag{S}$$

for all $x \in \mathcal{X}, y \in \mathcal{Y}$. Then:

Fenchel-Rockafellar duality: saddle point

Continuous
(convex)
optimisation

A. Chambolle

Monotone
operators

Subgradients of
convex functions

Elements of
monotone operators
theory

By optimality in the saddle-point problem: $Kx^* - \partial f^*(y^*) \ni 0$,
 $K^*y^* + \partial g(x^*) \ni 0$, that is:

$$0 \in \begin{pmatrix} \partial g(x) \\ \partial f^*(y) \end{pmatrix} + \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

meaning the solution can be found by finding the “zero” of the sum of two monotone operators. So a solution can be computed if we have an algorithm for solving $Ax + Bx \ni 0$, A, B maximal monotone.

This can be solved by a class or methods called (operator) “splitting algorithms”.