

Université Paris-Saclay – M2 Optimisation

Convex analysis and optimization

Part II.1: Fixed point algorithms

jean-christophe@pesquet.eu, emilie.chouzenoux@inria.fr
Center for Visual Computing – OPIS Inria group
CentraleSupélec

A first answer

Fixed point theorem (E. Picard, 1856-1941)

Let $\mathcal{H}$ be a Hilbert space. If $T: \mathcal{H} \rightarrow \mathcal{H}$ is such that

T is a strict contraction, i.e. there exists $\rho \in [0, 1[$ such that

$$(\forall (x, x') \in \mathcal{H}^2) \quad \|Tx - Tx'\| \leq \rho \|x - x'\|.$$

Then T has a unique fixed point $\hat{x}$.

The sequence $(x_n)_{n \in \mathbb{N}}$ defined as $(\forall n \in \mathbb{N}) \ x_{n+1} = Tx_n$ with $x_0 \in \mathcal{H}$, converges to $\hat{x}$.



Proof: For every $n \in \mathbb{N}$ and $m \in \mathbb{N} \setminus \{0\}$,

$$\|x_{n+m} - x_n\| = \left\| \sum_{\ell=1}^m x_{n+\ell} - x_{n+\ell-1} \right\| \leq \sum_{\ell=1}^m \|x_{n+\ell} - x_{n+\ell-1}\|.$$

For every $\ell \in \{1, \dots, m\}$,

$$\begin{aligned} \|x_{n+\ell} - x_{n+\ell-1}\| &= \|Tx_{n+\ell-1} - Tx_{n+\ell-2}\| \\ &\leq \rho \|x_{n+\ell-1} - x_{n+\ell-2}\| \leq \dots \leq \rho^{n+\ell-1} \|x_1 - x_0\|. \end{aligned}$$

Then $\|x_{n+m} - x_n\| \leq (1 - \rho)^{-1} (1 - \rho^m) \rho^n \|x_1 - x_0\|$.

Therefore $(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence and it converges to some $\hat{x} \in \mathcal{H}$.

→ prove that $\forall \epsilon > 0, (x_n)$ is a Cauchy sequence :
 $\forall \epsilon > 0, \forall n, m \exists N > 0 / \forall n, m \geq N, \|x_n - x_m\| \leq \epsilon.$

A first answer

Fixed point theorem (E. Picard, 1856-1941)

Let $\mathcal{H}$ be a Hilbert space. If $T: \mathcal{H} \rightarrow \mathcal{H}$ is such that

T is a strict contraction, i.e. there exists $\rho \in [0, 1[$ such that

$$(\forall (x, x') \in \mathcal{H}^2) \quad \|Tx - Tx'\| \leq \rho \|x - x'\|.$$

Then T has a unique fixed point $\hat{x}$.

The sequence $(x_n)_{n \in \mathbb{N}}$ defined as $(\forall n \in \mathbb{N}) \ x_{n+1} = Tx_n$ with $x_0 \in \mathcal{H}$, converges to $\hat{x}$.



Proof: Since T is continuous, $x_n \rightarrow \hat{x} \Rightarrow T\hat{x} = \hat{x}$. This shows that there exists a fixed point of T .

In addition, let $\bar{x}$ be a fixed point of T , for every $n \in \mathbb{N}$,

$$\begin{aligned} \|x_{n+1} - \bar{x}\| &= \|Tx_n - T\bar{x}\| \\ &\leq \rho \|x_n - \bar{x}\|. \end{aligned}$$

Consequently, $\|x_n - \bar{x}\| \leq \rho^n \|x_0 - \bar{x}\|$. Hence, we have proved that $(x_n)_{n \in \mathbb{N}}$ converges linearly to $\bar{x}$ and there exists a unique fixed point.

Objective of this course

- ▶ Extend this theorem to more general operators
 - ▶ not necessarily *strictly* contractive
 - ▶ possibly dependent on the iteration number n
 - ▶ built from **composition of simpler operators** (*splitting techniques*).
- ▶ Apply this to solve minimization problems.
 - ↪ How to relate T to the objective function f ?

Fixed point algorithms: convergence

Let $\mathcal{H}$ be a Hilbert space.

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$ and $\hat{x} \in \mathcal{H}$.

- ▶ $(x_n)_{n \in \mathbb{N}}$ converges strongly to $\hat{x}$ if

$$\lim_{n \rightarrow +\infty} \|x_n - \hat{x}\| = 0.$$

It is denoted by $x_n \rightarrow \hat{x}$.

- ▶ $(x_n)_{n \in \mathbb{N}}$ converges weakly to $\hat{x}$ if

$$(\forall y \in \mathcal{H}) \quad \lim_{n \rightarrow +\infty} \langle y | x_n - \hat{x} \rangle = 0.$$

It is denoted by $x_n \rightharpoonup \hat{x}$.

Remark: In a finite dimensional Hilbert space, strong and weak convergences are equivalent.

Fixed point algorithms: convergence

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of $\mathcal{H}$.

$(x_n)_{n \in \mathbb{N}}$ converges weakly if and only if

- ▶ $(x_n)_{n \in \mathbb{N}}$ is bounded $\forall n, \exists M \mid |x_n| \leq M$

and

- ▶ $(x_n)_{n \in \mathbb{N}}$ possesses at most one sequential cluster point in the weak topology. $\exists \hat{x} \in \mathcal{H}, \text{ and } \varphi: \mathbb{N} \rightarrow \mathbb{N} \mid x_{\varphi(m)} \rightarrow \hat{x}$

- ▶ $\hat{x}$ is a sequential cluster point of $(x_n)_{n \in \mathbb{N}}$ in the weak topology if there exists a subsequence $(x_{n_k})_{k \in \mathbb{N}}$ of $(x_n)_{n \in \mathbb{N}}$ that converges weakly to $\hat{x}$.

Fixed point algorithms: convergence

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of $\mathcal{H}$.

$(x_n)_{n \in \mathbb{N}}$ converges weakly if and only if

- ▶ $(x_n)_{n \in \mathbb{N}}$ is bounded
and
- ▶ $(x_n)_{n \in \mathbb{N}}$ possesses at most one sequential cluster point in the weak topology.

Illustration:

x_0	x_1	x_2	x_3	x_4	x_5	$\dots$
1	-1	1	-1	1	-1	$\dots$

→ $(x_n)_{n \in \mathbb{N}}$ is bounded but it has 2 sequential cluster points: -1 and 1 .

→ $(x_n)_{n \in \mathbb{N}}$ does not converge.

Fixed point algorithms: convergence

Opial's lemma

Let D be a nonempty subset of $\mathcal{H}$.

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$.

$(x_n)_{n \in \mathbb{N}}$ **weakly converges** to a point in D if

► for every $x \in D$, $(\|x_n - x\|)_{n \in \mathbb{N}}$ converges and

► every weak sequential cluster point of $(x_n)_{n \in \mathbb{N}}$ lies in D .

$$\begin{aligned} & \cdot \forall x \in D, (\|x_n - x\|)_{n \in \mathbb{N}} \text{ conv.} \\ & \cdot S\omega((x_n)_{n \in \mathbb{N}}) \subset D \\ & \left\{ \hat{x} \in \mathcal{H} \mid \exists (q_p)_{p \in \mathbb{N}} / q_p(m) \rightarrow \hat{x} \right\} \end{aligned}$$

$$\begin{aligned} & \text{let } \hat{x}, \hat{x}' \in \mathcal{H} \mid \exists (q_p)_{p \in \mathbb{N}} / q_p(m) \rightarrow \hat{x} \quad q_p(m) \rightarrow \hat{x}' \\ & \forall y \in D, \langle y | x_m - x \rangle \rightarrow 0 \quad 2\langle y | x_m - x \rangle = \|x_m - x - y\|^2 - \|x_m - x\|^2 - \|y\|^2 \\ & 2\langle x_m | \hat{x} - \hat{x}' \rangle = \|x_m - \hat{x}\|^2 - \|x_m - \hat{x}'\|^2 + \|\hat{x}\|^2 - \|\hat{x}'\|^2 \rightarrow 2\alpha \\ & \Rightarrow \langle x_m | \hat{x} - \hat{x}' \rangle \rightarrow \alpha \quad \text{also } \langle q_p(m) | \hat{x} - \hat{x}' \rangle \rightarrow \langle \hat{x} | \hat{x} - \hat{x}' \rangle \text{ and} \\ & \quad \quad \quad \|\hat{x} - \hat{x}'\|^2 = 0 \quad \text{so } \hat{x} = \hat{x}' \end{aligned}$$

Fixed point algorithms: convergence

Proof:

Let $x \in D$. Since $(\|x_n - x\|)_{n \in \mathbb{N}}$ converges, then $(\|x_n - x\|)_{n \in \mathbb{N}}$ and thus $(x_n)_{n \in \mathbb{N}}$ are bounded.

We assume that $(x_{n_k})_{k \in \mathbb{N}}$ and $(x_{n_\ell})_{\ell \in \mathbb{N}}$ are such that $x_{n_k} \rightharpoonup \hat{x}$ and $x_{n_\ell} \rightharpoonup \hat{x}'$ where $(\hat{x}, \hat{x}') \in D^2$. For every $n \in \mathbb{N}$,

$$2 \langle x_n | \hat{x}' - \hat{x} \rangle = \|x_n - \hat{x}\|^2 - \|x_n - \hat{x}'\|^2 - \|\hat{x}\|^2 + \|\hat{x}'\|^2.$$

Because $(\|x_n - \hat{x}\|)_{n \in \mathbb{N}}$ and $(\|x_n - \hat{x}'\|)_{n \in \mathbb{N}}$ converge, there exists $\alpha \in \mathbb{R}$ such that $\langle x_n | \hat{x}' - \hat{x} \rangle \rightarrow \alpha$ and thus

$\langle x_{n_k} | \hat{x}' - \hat{x} \rangle \rightarrow \langle \hat{x} | \hat{x}' - \hat{x} \rangle = \alpha$. Similarly, $\langle \hat{x}' | \hat{x}' - \hat{x} \rangle = \alpha$.

Consequently, $\|\hat{x}' - \hat{x}\|^2 = 0 \Rightarrow \hat{x} = \hat{x}'$.

Fixed point algorithms: Fejér-monotone sequence

Let D be a nonempty subset of a Hilbert space $\mathcal{H}$.

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$.

$(x_n)_{n \in \mathbb{N}}$ is **Fejér-monotone** with respect to D if

$$(\forall x \in D)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \leq \|x_n - x\|.$$

Let $D \subset \mathcal{H}$.

Let $(x_n)_{n \in \mathbb{N}}$ be Fejér-monotone with respect to D then

▶ for every $x \in D$, $(\|x_n - x\|)_{n \in \mathbb{N}}$ converges,

▶ $(x_n)_{n \in \mathbb{N}}$ is bounded.

$$\forall M > 0 / \|x_{n_0}\| \geq M \quad \|x_{n_0+1}\| \leq \|x_{n_0}\|$$

$$\|x_{n_0+1}\| \leq M \leq \|x_{n_0}\|$$

since $u_n = \|x_n - x\| \searrow$
 $u_n \geq 0 \Rightarrow$
 $u_n \rightarrow l$
 $\forall x \in D, \quad x_n \rightarrow x$
 $\Rightarrow$ la suite est
 cauchy.

Fixed point algorithms: Fejér-monotone sequence

Fejér-monotone convergence

Let D be a nonempty subset of a Hilbert space $\mathcal{H}$.

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$.

$(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in D if

- ▶ $(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone with respect to D
and
- ▶ every weak sequential cluster point of $(x_n)_{n \in \mathbb{N}}$ lies in D .

Lemma

Let C be a nonempty closed convex subset of $\mathcal{H}$.

If $(x_n)_{n \in \mathbb{N}}$ denotes a sequence in C that weakly converges to $\hat{x}$ then $\hat{x} \in C$.

Fixed point algorithms: Fejér-monotone sequence

Lemma

Let C be a nonempty closed convex subset of $\mathcal{H}$.

If $(x_n)_{n \in \mathbb{N}}$ denotes a sequence in C that weakly converges to $\hat{x}$ then $\hat{x} \in C$.

Proof:

Let P_C be the projection onto C .

Since $(\forall n \in \mathbb{N}) x_n \in C$, we have

$$\begin{aligned}
 & x_n \rightharpoonup \hat{x} \quad \forall n, x_n \in C \\
 & \forall y, \langle x_n - \hat{x}, y \rangle \rightarrow 0 \\
 & \langle x_n - \hat{x}, \hat{x} - P_C(\hat{x}) \rangle \\
 & \leq \|x_n - P_C(\hat{x})\| \|\hat{x} - P_C(\hat{x})\|
 \end{aligned}$$

$$\langle x_n - P_C(\hat{x}) | \hat{x} - P_C(\hat{x}) \rangle \leq 0.$$

By using $x_n \rightharpoonup \hat{x}$, it results that $\|\hat{x} - P_C(\hat{x})\|^2 = 0$, and thus $\hat{x} = P_C(\hat{x}) \in C$.

Nonexpansive operator: definition

Let C be a nonempty set of a Hilbert space $\mathcal{H}$. Let $T: C \rightarrow \mathcal{H}$.

The set of fixed points of T is

$$\text{Fix}T = \{x \in C \mid x = Tx\}.$$

Let $C \subset \mathcal{H}$ be a nonempty set.

Let $T: C \rightarrow \mathcal{H}$.

T is a nonexpansive operator if $(\forall (x, y) \in C^2) \quad \|Tx - Ty\| \leq \|x - y\|$.

Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$.

Let $T: C \rightarrow \mathcal{H}$ be nonexpansive.

Then $\text{Fix}T$ is a closed convex set.

Nonexpansive operator: definition

Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$.

Let $T: C \rightarrow \mathcal{H}$ be nonexpansive.

Then $\text{Fix } T$ is a closed convex set.

Proof: Closedness:

Let $(x_n)_{n \in \mathbb{N}}$ is a sequence of $\text{Fix } T \subset C$ converging to $x \in \mathcal{H}$. Since T is continuous, $x \in C$ and $(Tx_n)_{n \in \mathbb{N}} = (x_n)_{n \in \mathbb{N}}$ converges to $Tx = x$.

Thus $x \in \text{Fix } T$.

Nonexpansive operator: definition

Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$.

Let $T: C \rightarrow \mathcal{H}$ be nonexpansive.

Then $\text{Fix } T$ is a closed convex set.

Proof: Convexity:

Let $(x, y) \in (\text{Fix } T)^2$ and let $\alpha \in]0, 1[$. We have $\alpha x + (1 - \alpha)y \in C$ and

$$\begin{aligned}
 & \|T(\alpha x + (1 - \alpha)y) - \alpha x - (1 - \alpha)y\|^2 \\
 = & \|\alpha(T(\alpha x + (1 - \alpha)y) - x) + (1 - \alpha)(T(\alpha x + (1 - \alpha)y) - y)\|^2 \\
 = & \alpha^2\|T(\alpha x + (1 - \alpha)y) - x\|^2 + (1 - \alpha)^2\|T(\alpha x + (1 - \alpha)y) - y\|^2 \\
 & + 2\alpha(1 - \alpha)\langle T(\alpha x + (1 - \alpha)y) - x \mid T(\alpha x + (1 - \alpha)y) - y \rangle \\
 = & \alpha^2\|T(\alpha x + (1 - \alpha)y) - x\|^2 + (1 - \alpha)^2\|T(\alpha x + (1 - \alpha)y) - y\|^2 \\
 & - \alpha(1 - \alpha)(\|T(\alpha x + (1 - \alpha)y) - x - T(\alpha x + (1 - \alpha)y) + y\|^2 \\
 & - \|T(\alpha x + (1 - \alpha)y) - x\|^2 - \|T(\alpha x + (1 - \alpha)y) - y\|^2) \\
 = & \alpha\|T(\alpha x + (1 - \alpha)y) - x\|^2 + (1 - \alpha)\|T(\alpha x + (1 - \alpha)y) - y\|^2 - \alpha(1 - \alpha)\|x - y\|^2.
 \end{aligned}$$

Nonexpansive operator: definition

Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$.

Let $T: C \rightarrow \mathcal{H}$ be nonexpansive.

Then $\text{Fix } T$ is a closed convex set.

Proof: Convexity:

Let $(x, y) \in (\text{Fix } T)^2$ and let $\alpha \in]0, 1[$. We have

$$\begin{aligned} & \|T(\alpha x + (1 - \alpha)y) - \alpha x - (1 - \alpha)y\|^2 \\ &= \alpha \|T(\alpha x + (1 - \alpha)y) - x\|^2 + (1 - \alpha) \|T(\alpha x + (1 - \alpha)y) - y\|^2 - \alpha(1 - \alpha) \|x - y\|^2. \end{aligned}$$

Since $Tx = x$, $Ty = y$ and T is nonexpansive,

$$\begin{aligned} & \|T(\alpha x + (1 - \alpha)y) - \alpha x - (1 - \alpha)y\|^2 \\ &= \alpha \|T(\alpha x + (1 - \alpha)y) - Tx\|^2 + (1 - \alpha) \|T(\alpha x + (1 - \alpha)y) - Ty\|^2 - \alpha(1 - \alpha) \|x - y\|^2 \\ &\leq \alpha \|\alpha x + (1 - \alpha)y - x\|^2 + (1 - \alpha) \|\alpha x + (1 - \alpha)y - y\|^2 - \alpha(1 - \alpha) \|x - y\|^2 \\ &= (\alpha(1 - \alpha)^2 + (1 - \alpha)\alpha^2 - \alpha(1 - \alpha)) \|x - y\|^2 = 0. \end{aligned}$$

Hence $T(\alpha x + (1 - \alpha)y) = \alpha x + (1 - \alpha)y \Leftrightarrow \alpha x + (1 - \alpha)y \in \text{Fix } T$.

Nonexpansive operator: fixed point algorithm

Demiclosedness principle

Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$.

Let $T: C \rightarrow \mathcal{H}$ be a nonexpansive operator.

If $(x_n)_{n \in \mathbb{N}}$ is a sequence in C that converges weakly to $\hat{x}$ and if $Tx_n - x_n \rightarrow 0$ then $\hat{x} \in \text{Fix } T$.

Proof:

$x_n \rightharpoonup \hat{x} \Rightarrow \hat{x} \in C$ and $T\hat{x}$ defined. For every $n \in \mathbb{N}$,

$$\|x_n - T\hat{x}\|^2 = \|x_n - \hat{x}\|^2 + \|\hat{x} - T\hat{x}\|^2 + 2 \langle x_n - \hat{x} | \hat{x} - T\hat{x} \rangle$$

$$\|x_n - T\hat{x}\|^2 = \|x_n - Tx_n\|^2 + \|Tx_n - T\hat{x}\|^2 + 2 \langle x_n - Tx_n | Tx_n - T\hat{x} \rangle$$

$$\begin{aligned} \Rightarrow \|\hat{x} - T\hat{x}\|^2 &= \|x_n - Tx_n\|^2 + \|Tx_n - T\hat{x}\|^2 - \|x_n - \hat{x}\|^2 \\ &\quad + 2 \langle x_n - Tx_n | Tx_n - T\hat{x} \rangle - 2 \langle x_n - \hat{x} | \hat{x} - T\hat{x} \rangle \end{aligned}$$

Nonexpansive operator: fixed point algorithm

Demiclosedness principle

Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$.

Let $T: C \rightarrow \mathcal{H}$ be a nonexpansive operator.

If $(x_n)_{n \in \mathbb{N}}$ is a sequence in C that converges weakly to $\hat{x}$ and if $Tx_n - x_n \rightarrow 0$ then $\hat{x} \in \text{Fix } T$.

Proof:

$$\begin{aligned} \|\hat{x} - T\hat{x}\|^2 &= \|x_n - Tx_n\|^2 + \|Tx_n - T\hat{x}\|^2 - \|x_n - \hat{x}\|^2 \\ &\quad + 2 \langle x_n - Tx_n \mid Tx_n - T\hat{x} \rangle - 2 \langle x_n - \hat{x} \mid \hat{x} - T\hat{x} \rangle. \end{aligned}$$

Since T is nonexpansive, by using the Cauchy-Schwarz inequality,

$$\begin{aligned} \|\hat{x} - T\hat{x}\|^2 &\leq \|x_n - Tx_n\|^2 + 2\|x_n - Tx_n\|\|Tx_n - T\hat{x}\| - 2 \langle x_n - \hat{x} \mid \hat{x} - T\hat{x} \rangle \\ &\leq \|x_n - Tx_n\|^2 + 2\|x_n - Tx_n\|\|x_n - \hat{x}\| - 2 \langle x_n - \hat{x} \mid \hat{x} - T\hat{x} \rangle. \end{aligned}$$

$x_n \rightharpoonup \hat{x} \Rightarrow (x_n)_{n \in \mathbb{N}}$ bounded. The result follows by taking the limit.

Nonexpansive operator: fixed point algorithm

Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$.

Let $T: C \rightarrow C$ be a nonexpansive operator such that $\text{Fix } T \neq \emptyset$.

Let $x_0 \in C$,

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n.$$

If $x_n - Tx_n \rightarrow 0$, then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{Fix } T$.

Exercise :

Prove this result by showing that $(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone with respect to $\text{Fix } T$.

Nonexpansive operator: fixed point algorithm

Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$.

Let $T: C \rightarrow C$ be a nonexpansive operator such that $\text{Fix } T \neq \emptyset$.

Let $x_0 \in C$,

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n.$$

If $x_n - Tx_n \rightarrow 0$, then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{Fix } T$.

Answer : For every $n \in \mathbb{N}$ and $y \in \text{Fix } T$,

$$\|x_{n+1} - y\| = \|Tx_n - Ty\| \leq \|x_n - y\|.$$

$(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone with respect to $\text{Fix } T$.

Let $(x_{n_k})_{k \in \mathbb{N}}$ be a subsequence of $(x_n)_{n \in \mathbb{N}}$ such that $x_{n_k} \rightharpoonup \hat{x}$ where $\hat{x} \in \mathcal{H}$.

By assumption $x_{n_k} - Tx_{n_k} \rightarrow 0$ and thus, according to the demiclosedness principle, $\hat{x} \in \text{Fix } T$.

This shows the weak convergence of $(x_n)_{n \in \mathbb{N}}$.

Fixed point algorithms: Fejér-monotone sequence

Красносельский-Mann algorithm

Let C be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$.

Let $T: C \rightarrow C$ be a nonexpansive operator such that $\text{Fix } T \neq \emptyset$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 1]$ such that

$$\sum_{n \in \mathbb{N}} \lambda_n (1 - \lambda_n) = +\infty.$$

Let $x_0 \in C$ and $(\forall n \in \mathbb{N}) \ x_{n+1} = x_n + \lambda_n (Tx_n - x_n)$. Then,

- ▶ $(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone with respect to $\text{Fix } T$.
- ▶ $(Tx_n - x_n)_{n \in \mathbb{N}}$ converges strongly to 0.
- ▶ $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{Fix } T$.

Typical choice: $(\forall n \in \mathbb{N}) \ \lambda_n = \lambda \in]0, 1[$.

Fixed point algorithms: Fejér-monotone sequence

Proof :

For every $n \in \mathbb{N}$, by convex combination, $x_n \in C$.

Fejér-monotonicity with respect to $\text{Fix } T : (\forall x \in \text{Fix } T)(\forall n \in \mathbb{N})$

$$\begin{aligned}
 & \|x_{n+1} - x\|^2 \\
 &= \|x_n + \lambda_n(Tx_n - x_n) - x\|^2 \\
 &= \|(1 - \lambda_n)(x_n - x) + \lambda_n(Tx_n - x)\|^2 \\
 &= (1 - \lambda_n)^2 \|x_n - x\|^2 + \lambda_n^2 \|Tx_n - x\|^2 - 2\lambda_n(1 - \lambda_n) \langle x - x_n \mid Tx_n - x \rangle \\
 &= (1 - \lambda_n) \|x_n - x\|^2 + \lambda_n \|Tx_n - x\|^2 - \lambda_n(1 - \lambda_n) \|Tx_n - x + x - x_n\|^2 \\
 &= (1 - \lambda_n) \|x_n - x\|^2 + \lambda_n \|Tx_n - Tx\|^2 - \lambda_n(1 - \lambda_n) \|Tx_n - x_n\|^2 \\
 &\leq (1 - \lambda_n) \|x_n - x\|^2 + \lambda_n \|x_n - x\|^2 - \lambda_n(1 - \lambda_n) \|Tx_n - x_n\|^2 \\
 &\leq \|x_n - x\|^2.
 \end{aligned}$$

Fixed point algorithms: Fejér-monotone sequence

Proof :

We want to prove that $\|Tx_n - x_n\| \rightarrow 0$.

We deduce from $\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - \lambda_n(1 - \lambda_n)\|Tx_n - x_n\|^2$ that

$$\begin{aligned}
 (\forall N \in \mathbb{N}) \quad & \sum_{n=0}^N \lambda_n(1 - \lambda_n)\|Tx_n - x_n\|^2 \leq \|x_0 - x\|^2 - \|x_{N+1} - x\|^2 \\
 \Rightarrow \sum_{n \in \mathbb{N}} & \lambda_n(1 - \lambda_n)\|Tx_n - x_n\|^2 \leq \|x_0 - x\|^2 \\
 \Rightarrow \sum_{k=n}^{+\infty} & \lambda_k(1 - \lambda_k)\|Tx_k - x_k\|^2 \rightarrow 0 \\
 \Rightarrow \inf_{k \geq n} & \|Tx_k - x_k\|^2 \sum_{k=n}^{+\infty} \lambda_k(1 - \lambda_k) \rightarrow 0.
 \end{aligned}$$

The assumptions on the sequence $(\lambda_n)_{n \in \mathbb{N}}$ lead to $\liminf_{n \rightarrow +\infty} \|Tx_n - x_n\| = 0$.

Fixed point algorithms: Fejér-monotone sequence

Proof :

Let $(x_{n_k})_{k \in \mathbb{N}}$ be a subsequence of $(x_n)_{n \in \mathbb{N}}$ such that $x_{n_k} \rightharpoonup \hat{x}$.

According to the demiclosedness principle, $Tx_{n_k} - x_{n_k} \rightarrow 0$ yields $\hat{x} \in \text{Fix } T$. The weak convergence of $(x_n)_{n \in \mathbb{N}}$ to $\hat{x}$ results from the Fejér-monotonicity of $(x_n)_{n \in \mathbb{N}}$ with respect to $\text{Fix } T$.

α -averaged operator: definition

Let $C \subset \mathcal{H}$ be a nonempty set of a Hilbert space $\mathcal{H}$.

Let $A : C \rightarrow \mathcal{H}$ and let $\alpha \in]0, 1[$.

A is an α -averaged operator if there exists a nonexpansive operator $R : C \rightarrow \mathcal{H}$ such that

$$A = (1 - \alpha)\text{Id} + \alpha R.$$

Remark : When $\alpha = 1/2$, we say that A is firmly nonexpansive .

Fixed point algorithms: α -averaged operator

Let $T: \mathcal{H} \rightarrow \mathcal{H}$ be an α -averaged operator with $\alpha \in]0, 1[$ such that $\text{Fix } T \neq \emptyset$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 1/\alpha]$ such that

$$\sum_{n \in \mathbb{N}} \lambda_n (1 - \alpha \lambda_n) = +\infty.$$

Let $x_0 \in \mathcal{H}$ and $(\forall n \in \mathbb{N}) \ x_{n+1} = x_n + \lambda_n (Tx_n - x_n)$.

The following properties are satisfied:

- ▶ $(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone with respect to $\text{Fix } T$.
- ▶ $(Tx_n - x_n)_{n \in \mathbb{N}}$ converges strongly to 0.
- ▶ $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{Fix } T$.

Remark: since $\alpha < 1$, one can choose $(\forall n \in \mathbb{N}) \ \lambda_n = 1$, that is

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n.$$

Fixed point algorithms: α -averaged operator

Proof :

Since T is α -averaged, there exists a nonexpansive operator R such that $T = (1 - \alpha)\text{Id} + \alpha R$.

Let $(\forall n \in \mathbb{N}) \mu_n = \alpha \lambda_n \in [0, 1]$.

$$\sum_{n \in \mathbb{N}} \lambda_n (1 - \alpha \lambda_n) = +\infty \quad \Leftrightarrow \quad \sum_{n \in \mathbb{N}} \mu_n (1 - \mu_n) = +\infty.$$

The iterations can be written as

$$\begin{aligned} (\forall n \in \mathbb{N}) \quad x_{n+1} &= x_n + \lambda_n (Tx_n - x_n) \\ &= x_n + \mu_n (Rx_n - x_n). \end{aligned}$$

Moreover, for every $x \in \mathcal{H}$,

$$x \in \text{Fix} T \quad \Leftrightarrow \quad x = (1 - \alpha)x + \alpha Rx \quad \Leftrightarrow \quad x \in \text{Fix} R,$$

that is $\text{Fix} R = \text{Fix} T$.

+ Krasnosel'skii-Mann algorithm.

Exercise

Let $\mathcal{H}$ be a Hilbert space. Let $m \in \mathbb{N}$, $m \geq 2$. For every $i \in \{1, \dots, m\}$, let $T_i: \mathcal{H} \rightarrow \mathcal{H}$ be a nonexpansive operator. Assume that $\bigcap_{i=1}^m \text{Fix } T_i \neq \emptyset$. Let $(\omega_i)_{1 \leq i \leq m} \in]0, +\infty[^m$ be such that $\sum_{i=1}^m \omega_i = 1$, and let $T = \sum_{i=1}^m \omega_i T_i$.

1. Show that, for every $i \in \{1, \dots, m\}$ and for every $(x, y) \in \mathcal{H} \times \text{Fix } T_i$,

$$\langle T_i x - x \mid x - y \rangle \leq -\frac{1}{2} \|T_i x - x\|^2.$$

2. Deduce that $\text{Fix } T = \bigcap_{i=1}^m \text{Fix } T_i$.
3. In the following, we assume that, for every $i \in \{1, \dots, m\}$, T_i is α_i -averaged with $\alpha_i \in]0, 1[$. Show that T is α -averaged where α is a constant to be precised.
4. Study the convergence of the iteration

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (1 - \lambda_n)x_n + \lambda_n \sum_{i=1}^m \omega_i T_i x_n$$

where $x_0 \in \mathcal{H}$ and $\lambda_n \in]0, +\infty[$.

Fixed point algorithms: Fejér-monotone sequence

Proof :

The assumptions on $(\lambda_n)_{n \in \mathbb{N}}$ lead to $\liminf_{n \rightarrow +\infty} \|Tx_n - x_n\| = 0$.

Moreover, as T is a nonexpansive operator

$$\begin{aligned}\|Tx_{n+1} - x_{n+1}\| &= \|Tx_{n+1} - Tx_n + (1 - \lambda_n)(Tx_n - x_n)\| \\ &\leq \|Tx_{n+1} - Tx_n\| + (1 - \lambda_n)\|Tx_n - x_n\| \\ &\leq \|x_{n+1} - x_n\| + (1 - \lambda_n)\|Tx_n - x_n\| \\ &= \lambda_n\|Tx_n - x_n\| + (1 - \lambda_n)\|Tx_n - x_n\| \\ &= \|Tx_n - x_n\|.\end{aligned}$$

Consequently, $(\|Tx_n - x_n\|)_{n \in \mathbb{N}}$ converges and

$$Tx_n - x_n \rightarrow 0.$$

Université Paris-Saclay – M2 Optimisation

Convex analysis and optimization

Part II.2: Nonexpansive operators

jean-christophe@pesquet.eu

Center for Visual Computing – OPIS Inria group
CentraleSupélec

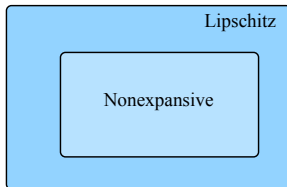
Nonexpansive operator: definition

Let $\mathcal{H}$ be a Hilbert space and let C be a nonempty subset of $\mathcal{H}$.

Let $A: C \rightarrow \mathcal{H}$ and $\nu \in]0, +\infty[$

$\nu^{-1}A$ is nonexpansive if $(\forall (x, y) \in C^2) \quad \|Ax - Ay\| \leq \nu \|x - y\|$.

$\nu^{-1}A$ is nonexpansive $\Leftrightarrow A$ is ν -Lipschitzian.



Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space.

Let $A: C \rightarrow \mathcal{H}$.

A is firmly nonexpansive if

$$(\forall x \in C)(\forall y \in C) \quad \|Ax - Ay\|^2 \leq \langle Ax - Ay \mid x - y \rangle .$$

Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.

Let $A: C \rightarrow \mathcal{H}$.

A is **firmly nonexpansive** if

$$(\forall (x, y) \in C^2) \quad \|Ax - Ay\|^2 + \|(\text{Id} - A)x - (\text{Id} - A)y\|^2 \leq \|x - y\|^2.$$

Proof: For every $(x, y) \in C^2$,

$$\begin{aligned} & \|Ax - Ay\|^2 + \|(\text{Id} - A)x - (\text{Id} - A)y\|^2 \leq \|x - y\|^2 \\ \Leftrightarrow & \|Ax - Ay\|^2 + \|x - y\|^2 - 2 \langle x - y \mid Ax - Ay \rangle + \|Ax - Ay\|^2 \leq \|x - y\|^2 \\ \Leftrightarrow & \|Ax - Ay\|^2 \leq \langle x - y \mid Ax - Ay \rangle. \end{aligned}$$

Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.

Let $A: C \rightarrow \mathcal{H}$.

A is **firmly nonexpansive** if

$$(\forall (x, y) \in C^2) \quad \|Ax - Ay\|^2 + \|(\text{Id} - A)x - (\text{Id} - A)y\|^2 \leq \|x - y\|^2.$$

- ▶ A is firmly nonexpansive $\Leftrightarrow \text{Id} - A$ is firmly nonexpansive.
- ▶ A is firmly nonexpansive $\Leftrightarrow \underbrace{2A - \text{Id}}_{\text{Reflection of } A}$ is nonexpansive.

Reflection of A

Proof: For every $(x, y) \in C^2$,

$$\begin{aligned} & \| (2A - \text{Id})x - (2A - \text{Id})y \|^2 \leq \|x - y\|^2 \\ \Leftrightarrow & 4\|Ax - Ay\|^2 - 4\langle Ax - Ay \mid x - y \rangle + \|x - y\|^2 \leq \|x - y\|^2 \\ \Leftrightarrow & \|Ax - Ay\|^2 \leq \langle Ax - Ay \mid x - y \rangle. \end{aligned}$$

Nonexpansive operator: definition

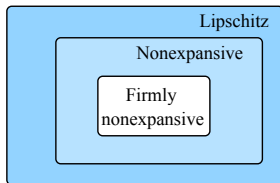
Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.

Let $A: C \rightarrow \mathcal{H}$.

A is **firmly nonexpansive** if

$$(\forall (x, y) \in C^2) \quad \|Ax - Ay\|^2 + \|(\text{Id} - A)x - (\text{Id} - A)y\|^2 \leq \|x - y\|^2.$$

A is firmly nonexpansive $\Rightarrow A$ is nonexpansive.



Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.
Let $A: C \rightarrow \mathcal{H}$ and $\beta \in]0, +\infty[$.

A is **β -cocoercive** if βA is firmly nonexpansive, i.e.,

$$(\forall x \in C)(\forall y \in C) \quad \beta \|Ax - Ay\|^2 \leq \langle x - y \mid Ax - Ay \rangle .$$

- ▶ Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces, $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ nonzero, and $A: \mathcal{G} \rightarrow \mathcal{G}$. A is β -cocoercive $\Rightarrow L^*AL$ is $\|L\|^{-2}\beta$ -cocoercive.

Proof: For every $(x, y) \in \mathcal{H}^2$,

$$\begin{aligned} \langle L^*ALx - L^*ALy \mid x - y \rangle &= \langle ALx - ALy \mid Lx - Ly \rangle \\ &\geq \beta \|ALx - ALy\|^2 \end{aligned}$$

Furthermore, $\|L^*ALx - L^*ALy\|^2 \leq \|L\|^2 \|ALx - ALy\|^2$.

Then $\langle L^*ALx - L^*ALy \mid x - y \rangle \geq \beta \|L^*ALx - L^*ALy\|^2 / \|L\|^2$.

Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.
Let $A: C \rightarrow \mathcal{H}$ and $\beta \in]0, +\infty[$.

A is β -cocoercive if βA is firmly nonexpansive, i.e.,

$$(\forall x \in C)(\forall y \in C) \quad \beta \|Ax - Ay\|^2 \leq \langle x - y \mid Ax - Ay \rangle .$$

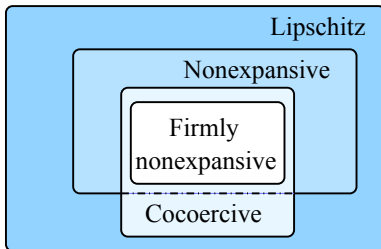
- ▶ Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces, $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ nonzero, and $A: \mathcal{G} \rightarrow \mathcal{G}$. A is β -cocoercive $\Rightarrow L^*AL$ is $\|L\|^{-2}\beta$ -cocoercive.
- ▶ A is β -cocoercive $\Rightarrow A$ is β^{-1} -Lipschitzian.

Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.
Let $A: C \rightarrow \mathcal{H}$ and $\beta \in]0, +\infty[$.

A is β -cocoercive if βA is firmly nonexpansive, i.e.,

$$(\forall x \in C)(\forall y \in C) \quad \beta \|Ax - Ay\|^2 \leq \langle x - y \mid Ax - Ay \rangle .$$



Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.

Let $A : C \rightarrow \mathcal{H}$ and let $\alpha \in]0, 1[$.

A is α -averaged if there exists a nonexpansive operator $R : C \rightarrow \mathcal{H}$ such that

$$A = (1 - \alpha)\text{Id} + \alpha R .$$

Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.

Let $A : C \rightarrow \mathcal{H}$ and let $\alpha \in]0, 1[$.

A is α -averaged if

$$(\forall (x, y) \in C^2) \quad \|Ax - Ay\|^2 + \frac{1 - \alpha}{\alpha} \|(\text{Id} - A)x - (\text{Id} - A)y\|^2 \leq \|x - y\|^2.$$

Proof: For every $(x, y) \in C^2$,

$$\|Ax - Ay\|^2 + \frac{1 - \alpha}{\alpha} \|(\text{Id} - A)x - (\text{Id} - A)y\|^2 \leq \|x - y\|^2$$

$$\Leftrightarrow \alpha \|Ax - Ay\|^2 + (1 - \alpha)(\|Ax - Ay\|^2 - 2 \langle x - y \mid Ax - Ay \rangle + \|x - y\|^2) \leq \alpha \|x - y\|^2$$

$$\Leftrightarrow \|Ax - Ay\|^2 - 2(1 - \alpha) \langle x - y \mid Ax - Ay \rangle + (1 - 2\alpha) \|x - y\|^2 \leq 0$$

$$\Leftrightarrow \|Ax - Ay - (1 - \alpha)(x - y)\|^2 \leq \alpha^2 \|x - y\|^2$$

$$\Leftrightarrow R = \frac{A - (1 - \alpha)\text{Id}}{\alpha} \text{ nonexpansive.}$$

Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.

Let $A : C \rightarrow \mathcal{H}$ and let $\alpha \in]0, 1[$.

A is α -averaged if

$$(\forall (x, y) \in C^2) \quad \|Ax - Ay\|^2 + \frac{1 - \alpha}{\alpha} \|(\text{Id} - A)x - (\text{Id} - A)y\|^2 \leq \|x - y\|^2.$$

- ▶ A is α -averaged $\Rightarrow A$ is nonexpansive.
 - ▶ A is $\frac{1}{2}$ -averaged $\Leftrightarrow A$ is firmly nonexpansive.
 - ▶ A is α -averaged $\Rightarrow A$ is α' -averaged for every $\alpha' \in [\alpha, 1[$.
 - ▶ Let $\lambda \in]0, 1/\alpha[$. A is α -averaged $\Rightarrow (1 - \lambda)\text{Id} + \lambda A$ is $\lambda\alpha$ -averaged.
- Proof: If A is α -averaged, there exists a nonexpansive operator R such that $A = (1 - \alpha)\text{Id} + \alpha R$. We have thus

$$\begin{aligned} (1 - \lambda)\text{Id} + \lambda A &= (1 - \lambda)\text{Id} + \lambda((1 - \alpha)\text{Id} + \alpha R) \\ &= (1 - \lambda\alpha)\text{Id} + \lambda\alpha R. \end{aligned}$$

Nonexpansive operator: definition

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.
Let $A : C \rightarrow \mathcal{H}$ and let $\alpha \in]0, 1[$.

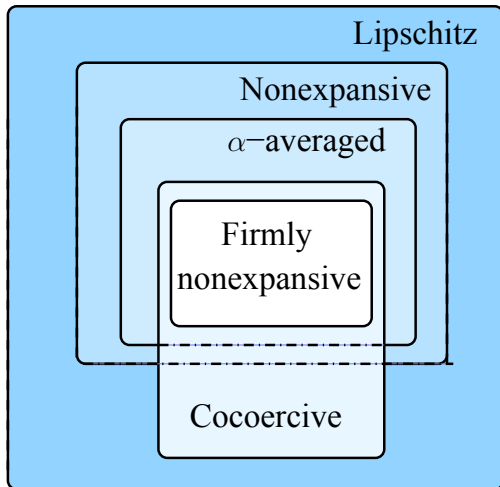
A is **α -averaged** if

$$(\forall (x, y) \in C^2) \quad \|Ax - Ay\|^2 + \frac{1 - \alpha}{\alpha} \|(\text{Id} - A)x - (\text{Id} - A)y\|^2 \leq \|x - y\|^2.$$

- ▶ Let $(\omega_i)_{1 \leq i \leq n} \in]0, 1]^n$ be such that $\sum_{i=1}^n \omega_i = 1$ and let $(\alpha_i)_{1 \leq i \leq n} \in]0, 1[^n$. If, for every $i \in \{1, \dots, n\}$, $A_i : C \rightarrow \mathcal{H}$ is α_i -averaged, then **$\sum_{i=1}^n \omega_i A_i$ is α -averaged** with $\alpha = \sum_{i=1}^n \omega_i \alpha_i$.
- ▶ Let $(\alpha_i)_{1 \leq i \leq n} \in]0, 1[^n$. If, for every $i \in \{1, \dots, n\}$, $A_i : C \rightarrow C$ is α_i -averaged, then **$A_1 \cdots A_n$ is α -averaged** with

$$\alpha = \frac{1}{1 + 1/(\sum_{i=1}^n \frac{\alpha_i}{1 - \alpha_i})}.$$

Nonexpansive operator: recap



Nonexpansive operator: main property

Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty subset of $\mathcal{H}$.

Let $A: C \rightarrow \mathcal{H}$.

Let $\beta \in]0, +\infty[$ and $\gamma \in]0, 2\beta[$.

If A is β -cocoercive, then $\text{Id} - \gamma A$ is $\gamma/(2\beta)$ -averaged.

Proof :

A β -cocoercive $\Leftrightarrow \beta A$ firmly nonexpansive.

There exists a nonexpansive operator $R: C \rightarrow \mathcal{H}$ such that

$$\beta A = (\text{Id} + R)/2.$$

Thus

$$\text{Id} - \gamma A = \left(1 - \frac{\gamma}{2\beta}\right)\text{Id} + \frac{\gamma}{2\beta}(-R).$$

$(-R)$ being nonexpansive, $\text{Id} - \gamma A$ is $\gamma/(2\beta)$ -averaged.

Nonexpansive operators

What is their use ?



Descent lemma

Let $\mathcal{H}$ be a real Hilbert space, $f: \mathcal{H} \rightarrow \mathbb{R}$ and $\nu \in]0, +\infty[$.

If f is Fréchet differentiable and its gradient is ν -Lipschitzian, then

$$(\forall (x, y) \in \mathcal{H}^2) \quad f(y) \leq f(x) + \langle y - x \mid \nabla f(x) \rangle + \frac{\nu}{2} \|y - x\|^2.$$

Proof : For every $(x, y) \in \mathcal{H}^2$ and $t \in \mathbb{R}$, let $\varphi(t) = f(x + t(y - x))$. φ is differentiable and $\varphi'(t) = \langle y - x \mid \nabla f(x + t(y - x)) \rangle$. We have then

$$\begin{aligned} \varphi(1) - \varphi(0) &= \int_0^1 \varphi'(t) dt \\ \Leftrightarrow f(y) - f(x) - \langle y - x \mid \nabla f(x) \rangle &= \int_0^1 \langle y - x \mid \nabla f(x + t(y - x)) - \nabla f(x) \rangle dt. \end{aligned}$$

In addition, according to the Cauchy-Schwarz inequality,

$$\begin{aligned} &\langle y - x \mid \nabla f(x + t(y - x)) - \nabla f(x) \rangle \\ &\leq \|y - x\| \|\nabla f(x + t(y - x)) - \nabla f(x)\| \leq t\nu \|y - x\|^2. \end{aligned}$$

This leads to

$$(\forall (x, y) \in \mathcal{H}^2) \quad f(y) \leq f(x) + \langle y - x \mid \nabla f(x) \rangle + \frac{\nu}{2} \|y - x\|^2.$$

Reminders on conjugation

Let $\mathcal{H}$ be a Hilbert space and $f: \mathcal{H} \rightarrow]-\infty, +\infty]$.

The **conjugate** of f is the function $f^*: \mathcal{H} \rightarrow [-\infty, +\infty]$ such that

$$(\forall u \in \mathcal{H}) \quad f^*(u) = \sup_{x \in \mathcal{H}} (\langle x | u \rangle - f(x)) .$$

- Property: Let $\Gamma_0(\mathcal{H})$ be the class of proper lower-semicontinuous convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$.

If $f \in \Gamma_0(\mathcal{H})$, then $f^* \in \Gamma_0(\mathcal{H})$.

Reminders on conjugation

Let $\mathcal{H}$ be a Hilbert space and $f: \mathcal{H} \rightarrow]-\infty, +\infty]$.

The **conjugate** of f is the function $f^*: \mathcal{H} \rightarrow [-\infty, +\infty]$ such that

$$(\forall u \in \mathcal{H}) \quad f^*(u) = \sup_{x \in \mathcal{H}} (\langle x | u \rangle - f(x)) .$$

Fenchel-Young inequality : If f is proper, then

1. $(\forall (x, u) \in \mathcal{H}^2) \quad f(x) + f^*(u) \geq \langle x | u \rangle$
2. $(\forall (x, u) \in \mathcal{H}^2) \quad u \in \partial f(x) \Leftrightarrow f(x) + f^*(u) = \langle x | u \rangle .$

Baillon-Haddad theorem

Let $\mathcal{H}$ be a real Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\nu \in]0, +\infty[$.
 If f is Fréchet differentiable, then ∇f ν -Lipschitzian $\Leftrightarrow \nabla f$ ν^{-1} -cocoercive.

Proof : Assume that ∇f is ν -Lipschitzian. According to Fenchel-Young inequality, for every $(x, y, z) \in \mathcal{H}^3$,

$$f^*(\nabla f(y)) \geq \langle z \mid \nabla f(y) \rangle - f(z).$$

From the descent lemma,

$$f^*(\nabla f(y)) \geq \langle z \mid \nabla f(y) - \nabla f(x) \rangle + \langle x \mid \nabla f(x) \rangle - f(x) - \frac{\nu}{2} \|z - x\|^2.$$

Moreover, using again the Fenchel-Young result,

$$\langle x \mid \nabla f(x) \rangle - f(x) = f^*(\nabla f(x)).$$

Thus,

$$f^*(\nabla f(y)) \geq \langle z \mid \nabla f(y) - \nabla f(x) \rangle + f^*(\nabla f(x)) - \frac{\nu}{2} \|z - x\|^2$$

Baillon-Haddad theorem

Let $\mathcal{H}$ be a real Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\nu \in]0, +\infty[$.
 If f is Fréchet differentiable, then ∇f ν -Lipschitzian $\Leftrightarrow \nabla f$ ν^{-1} -cocoercive.

Proof : Thus,

$$\begin{aligned} f^*(\nabla f(y)) &\geq \langle z \mid \nabla f(y) - \nabla f(x) \rangle + f^*(\nabla f(x)) - \frac{\nu}{2} \|z - x\|^2 \\ &= f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x) \rangle + \langle z - x \mid \nabla f(y) - \nabla f(x) \rangle - \frac{\nu}{2} \|z - x\|^2. \end{aligned}$$

Taking the supremum with respect to z yields

$$\begin{aligned} f^*(\nabla f(y)) &\geq f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x) \rangle \\ &\quad + (\nu \|\cdot\|^2/2)^*(\nabla f(y) - \nabla f(x)) \\ &= f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x) \rangle + \frac{1}{2\nu} \|\nabla f(y) - \nabla f(x)\|^2. \end{aligned}$$

Consequently,

$$f^*(\nabla f(y)) \geq f^*(\nabla f(x)) + \langle x \mid \nabla f(y) - \nabla f(x) \rangle + \frac{1}{2\nu} \|\nabla f(y) - \nabla f(x)\|^2.$$

Baillon-Haddad theorem

Let $\mathcal{H}$ be a real Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\nu \in]0, +\infty[$.
 If f is Fréchet differentiable, then ∇f ν -Lipschitzian $\Leftrightarrow \nabla f$ ν^{-1} -cocoercive.

Proof : For every $(x, y) \in \mathcal{H}^2$,

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

and symmetrically

$$f^*(\nabla f(x)) \geq f^*(\nabla f(y)) + \langle y \mid \nabla f(x) - \nabla f(y) \rangle + \frac{1}{2\nu} \|\nabla f(x) - \nabla f(y)\|^2.$$

By summing, we finally obtain

$$-\langle y - x \mid \nabla f(y) - \nabla f(x) \rangle + \frac{1}{\nu} \|\nabla f(x) - \nabla f(y)\|^2 \leq 0,$$

which shows that ∇f is $1/\nu$ -cocoercive.

Nonexpansive operator: example

Baillon-Haddad theorem

Let $\mathcal{H}$ be a real Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\nu \in]0, +\infty[$.

If f is Fréchet differentiable, then ∇f ν -Lipschitzian $\Leftrightarrow \nabla f$ ν^{-1} -cocoercive.

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$, $\nu \in]0, +\infty[$ and $\gamma \in]0, 2/\nu[$.

f Fréchet differentiable and ∇f ν -Lipschitzian $\Rightarrow$ $\underbrace{\text{Id} - \gamma \nabla f}_{\text{gradient descent operator}}$ is $\gamma\nu/2$ -

averaged.

Resolvent: definition

Let $\mathcal{H}$ be a Hilbert space.

Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$.

The **revolvent** of A is

$$J_A = (\text{Id} + A)^{-1}.$$

Let $\mathcal{H}$ be a Hilbert space.

Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ and $\gamma \in]0, +\infty[$.

The **Yosida approximation** of A of index γ is

$$\gamma A = \frac{1}{\gamma}(\text{Id} - J_{\gamma A}).$$

Proximity operator: definition

Let $\mathcal{H}$ be a Hilbert space. Let $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in]0, +\infty[$. For every $x \in \mathcal{H}$, there exists a unique $p \in \mathcal{H}$ such that

$$f(p) + \frac{1}{2\gamma} \|p - x\|^2 = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2.$$

Let $\mathcal{H}$ be a Hilbert space. Let $f \in \Gamma_0(\mathcal{H})$.

- ▶ The **Moreau envelope** of f of parameter $\gamma \in]0, +\infty[$ is

$$\gamma f: \mathcal{H} \rightarrow \mathbb{R}: x \mapsto \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2.$$

- ▶ The **proximity operator** of f is

$$\text{prox}_f: \mathcal{H} \rightarrow \mathcal{H}: x \mapsto \operatorname{argmin}_{y \in \mathcal{H}} f(y) + \frac{1}{2} \|y - x\|^2.$$

Proximity operator: definition

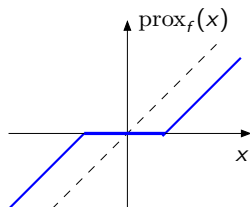
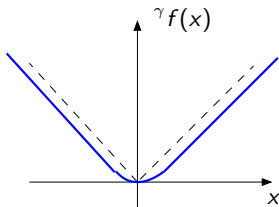
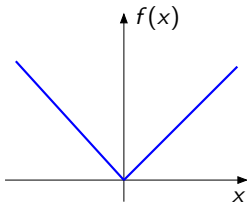
Let $\mathcal{H}$ be a Hilbert space. Let $f \in \Gamma_0(\mathcal{H})$.

- The **Moreau envelope** of f of parameter $\gamma \in]0, +\infty[$ is

$$\gamma f: \mathcal{H} \rightarrow \mathbb{R}: x \mapsto \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2.$$

- The **proximity operator** of f is

$$\text{prox}_f: \mathcal{H} \rightarrow \mathcal{H}: x \mapsto \operatorname{argmin}_{y \in \mathcal{H}} f(y) + \frac{1}{2} \|y - x\|^2.$$



Proximity operator: definition

Let $\mathcal{H}$ be a Hilbert space. Let $f \in \Gamma_0(\mathcal{H})$.

- ▶ The **Moreau envelope** of f of parameter $\gamma \in]0, +\infty[$ is

$$\gamma f: \mathcal{H} \rightarrow \mathbb{R}: x \mapsto \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma} \|y - x\|^2.$$

- ▶ The **proximity operator** of f is

$$\text{prox}_f: \mathcal{H} \rightarrow \mathcal{H}: x \mapsto \operatorname{argmin}_{y \in \mathcal{H}} f(y) + \frac{1}{2} \|y - x\|^2.$$

Remark: For every $x \in \mathcal{H}$,

$$\gamma f(x) \leq f(x).$$

Proximity operator: definition

Let $\mathcal{H}$ be a Hilbert space and $f \in \Gamma_0(\mathcal{H})$.

$$\text{prox}_f = J_{\partial f}.$$

Proof: By using Fermat's rule, for every $x \in \mathcal{H}$,

$$\begin{aligned} p = \arg \min \quad f + \frac{1}{2} \|\cdot - x\|^2 &\Leftrightarrow 0 \in \partial \left(f + \frac{1}{2} \|\cdot - x\|^2 \right) (p) \\ &\Leftrightarrow 0 \in \partial f(p) + p - x \\ &\Leftrightarrow x \in (\text{Id} + \partial f)(p) \\ &\Leftrightarrow p = (\text{Id} + \partial f)^{-1}(x). \end{aligned}$$

Proximity operator: properties

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $(x, p) \in \mathcal{H}^2$.

$$p = \text{prox}_f x \iff (\forall y \in \mathcal{H}) \quad \langle y - p \mid x - p \rangle + f(p) \leq f(y).$$

Let $\mathcal{H}$ be a Hilbert space and $f \in \Gamma_0(\mathcal{H})$.

Then prox_f is **firmly nonexpansive**.

Proximity operator: properties

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in]0, +\infty[$.

γf is differentiable and $\nabla \gamma f$ is γ^{-1} -Lipschitzian

$$(\forall x \in \mathcal{H}) \quad \nabla \underbrace{\gamma f}_{\text{Moreau envelope}} = \gamma^{-1}(\text{Id} - \text{prox}_{\gamma f}) = \underbrace{\gamma \partial f}_{\text{Yosida approximation}}.$$

Proximity operator: properties

Lemma

Let $\mathcal{H}$ be a real Hilbert space and let $g: \mathcal{H} \rightarrow \mathbb{R}$.

Assume that there exists $G: \mathcal{H} \rightarrow \mathcal{H}$ continuous such that, for every $x \in \mathcal{H}$, $G(x) \in \partial g(x)$. Then g is Fréchet differentiable on $\mathcal{H}$ and $\nabla g = G$.

Proof: For every $(x, y) \in \mathcal{H}^2$,

$$g(x+y) \geq g(x) + \langle G(x) | y \rangle$$

$$g(x) \geq g(x+y) - \langle G(x+y) | y \rangle$$

$\forall x \in \mathcal{H}, G(x) \in \partial g(x)$
 ie $\bigcap_{x \in \mathcal{H}} \partial g(x) \ni G(x)$

Then

$$0 \leq g(x+y) - g(x) - \langle G(x) | y \rangle$$

$$\leq \langle G(x+y) - G(x) | y \rangle \leq \|G(x+y) - G(x)\| \|y\|.$$

We deduce that

$$\lim_{\substack{y \rightarrow 0 \\ y \neq 0}} \frac{g(x+y) - g(x) - \langle G(x) | y \rangle}{\|y\|} = 0.$$

Proximity operator: properties

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in]0, +\infty[$.

γf is differentiable and $\nabla \gamma f$ is γ^{-1} -Lipschitzian

$$(\forall x \in \mathcal{H}) \quad \underbrace{\nabla \gamma f}_{\text{Moreau envelope}} = \gamma^{-1}(\text{Id} - \text{prox}_{\gamma f}) = \underbrace{\gamma \partial f}_{\text{Yosida approximation}}.$$

Proof: Since $\text{prox}_{\gamma f}$ is firmly nonexpansive, $\text{Id} - \text{prox}_{\gamma f}$ is also firmly nonexpansive.

It is thus nonexpansive and $\gamma^{-1}(\text{Id} - \text{prox}_{\gamma f})$ is γ^{-1} -Lipschitzian.

Proximity operator: properties

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in]0, +\infty[$.

γf is differentiable and $\nabla \gamma f$ is γ^{-1} -Lipschitzian

$$(\forall x \in \mathcal{H}) \quad \underbrace{\nabla \gamma f}_{\text{Moreau envelope}} = \gamma^{-1}(\text{Id} - \text{prox}_{\gamma f}) = \underbrace{\gamma \partial f}_{\text{Yosida approximation}}.$$

Proof: Let $(x, y, q) \in \mathcal{H}^3$ and let $p_x = \text{prox}_{\gamma f} x$. We have

$$\begin{aligned} \gamma f(q) &\geq \gamma f(p_x) + \langle q - p_x \mid x - p_x \rangle \\ &= \gamma f(p_x) + \langle q - y + y - x + x - p_x \mid x - p_x \rangle \\ &= \gamma f(p_x) + \|x - p_x\|^2 + \langle y - x \mid x - p_x \rangle + \langle q - y \mid x - p_x \rangle. \end{aligned}$$

We know that $\gamma f(x) = \inf_{p \in \mathcal{H}} f(p) + \frac{1}{2\gamma} \|p - x\|^2 = f(p_x) + \frac{1}{2\gamma} \|p_x - x\|^2$.

Hence

$$\begin{aligned} \gamma f(q) &\geq \gamma \gamma f(x) + \langle y - x \mid x - p_x \rangle + \frac{1}{2} \|x - p_x\|^2 + \langle q - y \mid x - p_x \rangle. \end{aligned}$$

Proximity operator: properties

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in]0, +\infty[$.

γf is differentiable and $\nabla \gamma f$ is γ^{-1} -Lipschitzian

$$(\forall x \in \mathcal{H}) \quad \underbrace{\nabla \gamma f}_{\text{Moreau envelope}} = \gamma^{-1}(\text{Id} - \text{prox}_{\gamma f}) = \underbrace{\gamma \partial f}_{\text{Yosida approximation}}.$$

Proof: Let $(x, y, q) \in \mathcal{H}^3$ and let $p_x = \text{prox}_{\gamma f} x$. We have

$$\begin{aligned} \gamma f(q) &\geq \gamma f(p_x) + \langle q - p_x \mid x - p_x \rangle \\ &= \gamma f(p_x) + \langle q - y + y - x + x - p_x \mid x - p_x \rangle \\ &= \gamma f(p_x) + \|x - p_x\|^2 + \langle y - x \mid x - p_x \rangle + \langle q - y \mid x - p_x \rangle. \end{aligned}$$

We know that $\gamma f(x) = \inf_{p \in \mathcal{H}} f(p) + \frac{1}{2\gamma} \|p - x\|^2 = f(p_x) + \frac{1}{2\gamma} \|p_x - x\|^2$.

Hence

$$\begin{aligned} &\gamma f(q) + \frac{1}{2} \|q - y\|^2 \\ &\geq \gamma f(x) + \langle y - x \mid x - p_x \rangle + \frac{1}{2} \|x - p_x\|^2 + \langle q - y \mid x - p_x \rangle + \frac{1}{2} \|q - y\|^2. \end{aligned}$$

Proximity operator: properties

Proof: Let $(x, y, q) \in \mathcal{H}^3$ and let $p_x = \text{prox}_{\gamma f} x$. We have

$$\begin{aligned}\gamma f(q) &\geq \gamma f(p_x) + \langle q - p_x \mid x - p_x \rangle \\ &= \gamma f(p_x) + \langle q - y + y - x + x - p_x \mid x - p_x \rangle \\ &= \gamma f(p_x) + \|x - p_x\|^2 + \langle y - x \mid x - p_x \rangle + \langle q - y \mid x - p_x \rangle.\end{aligned}$$

We know that $\gamma f(x) = \inf_{p \in \mathcal{H}} f(p) + \frac{1}{2\gamma} \|p - x\|^2 = f(p_x) + \frac{1}{2\gamma} \|p_x - x\|^2$.

Hence

$$\begin{aligned}\gamma f(q) + \frac{1}{2} \|q - y\|^2 \\ \geq \gamma \gamma f(x) + \langle y - x \mid x - p_x \rangle + \frac{1}{2} \|x - p_x\|^2 + \langle q - y \mid x - p_x \rangle + \frac{1}{2} \|q - y\|^2.\end{aligned}$$

Since $\gamma f(y) = \inf_{q \in \mathcal{H}} f(q) + \frac{1}{2\gamma} \|q - y\|^2$,

$$\begin{aligned}\gamma \gamma f(y) &\geq \gamma \gamma f(x) + \langle y - x \mid x - p_x \rangle + \frac{1}{2} \|x - p_x\|^2 \\ &\quad + \underbrace{\inf_{q \in \mathcal{H}} \langle q - y \mid x - p_x \rangle + \frac{1}{2} \|q - y\|^2}_{-\frac{1}{2} \|x - p_x\|^2}.\end{aligned}$$

Proximity operator: properties

Proof: Since $\gamma f(y) = \inf_{q \in \mathcal{H}} f(q) + \frac{1}{2\gamma} \|q - y\|^2$,

$$\begin{aligned} \gamma \gamma f(y) &\geq \gamma \gamma f(x) + \langle y - x \mid x - p_x \rangle + \frac{1}{2} \|x - p_x\|^2 \\ &\quad + \underbrace{\inf_{q \in \mathcal{H}} \langle q - y \mid x - p_x \rangle + \frac{1}{2} \|q - y\|^2}_{-\frac{1}{2} \|x - p_x\|^2}. \end{aligned}$$

Consequently

$$\gamma f(y) \geq \gamma f(x) + \frac{1}{\gamma} \langle x - p_x \mid y - x \rangle,$$

which shows that $\gamma^{-1}(x - p_x) \in \partial \gamma f(x)$.

Since $\gamma^{-1}(\text{Id} - \text{prox}_{\gamma f})$ is continuous, we deduce that γf is Fréchet differentiable and $\nabla \gamma f = \gamma^{-1}(\text{Id} - \text{prox}_{\gamma f})$.

Proximity operator: properties

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in]0, +\infty[$.

γf is differentiable and $\nabla \gamma f$ is γ^{-1} -Lipschitzian

$$(\forall x \in \mathcal{H}) \quad \nabla \underbrace{\gamma f}_{\text{Moreau envelope}} = \gamma^{-1}(\text{Id} - \text{prox}_{\gamma f}) = \underbrace{\gamma \partial f}_{\text{Yosida approximation}}.$$

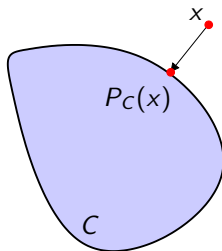
Interpretation: γf is a convex smooth lower approximation to f .

Projection

Projection :

Let $\mathcal{H}$ be a Hilbert space. Let C be a nonempty closed convex subset of $\mathcal{H}$.

$$(\forall x \in \mathcal{H}) \quad \text{prox}_{\iota_C}(x) = \underset{y \in C}{\operatorname{argmin}} \frac{1}{2} \|y - x\|^2 = P_C(x).$$



Projection

Projection :

Let $\mathcal{H}$ be a Hilbert space. Let C be a nonempty closed convex subset of $\mathcal{H}$.

$$(\forall x \in \mathcal{H}) \quad \text{prox}_{\iota_C}(x) = \underset{y \in C}{\operatorname{argmin}} \frac{1}{2} \|y - x\|^2 = P_C(x).$$

Remark :

- ▶ $p = P_C(x) \Leftrightarrow p = J_{\partial \iota_C} x \Leftrightarrow x - p \in \partial \iota_C(p) = N_C(p)$
 $\Leftrightarrow p \in C \text{ and } (\forall y \in C) \langle y - p \mid x - p \rangle \leq 0.$

Particular case: if C is a vector space: $p = P_C(x) \Leftrightarrow \begin{cases} x - p \in C^\perp \\ p \in C. \end{cases}$

- ▶ $\gamma_{\iota_C} = (2\gamma)^{-1} d_C^2$ where d_C distance to the convex set C is defined by $d_C: x \mapsto \inf_{y \in C} \|y - x\| = \|x - P_C x\|$. We have then $\nabla d_C^2 = \nabla(\frac{1}{2} \iota_C) = 2(\text{Id} - P_C).$

Exercise

For every $i \in \{1, \dots, m\}$, let $A_i: \mathcal{H} \rightarrow \mathcal{H}$ and let $\beta_i \in]0, +\infty[$.

1. Let $(\omega_i)_{1 \leq i \leq m} \in]0, +\infty[^m$ be such that $\sum_{i=1}^m \omega_i = 1$. Show that, for every $(x, y) \in \mathcal{H}^2$,

$$\omega_i \leq 1 \quad f(\sum \omega_i a_i) \leq \sum \omega_i f(a_i)$$

$$\left\| \sum_{i=1}^m \sqrt{\omega_i \beta_i} (A_i x - A_i y) \right\|^2 \leq \sum_{i=1}^m \beta_i \|A_i x - A_i y\|^2.$$

2. Assume that, for every $i \in \{1, \dots, m\}$, A_i is β_i -cocoercive. Show that $\sum_{i=1}^m A_i$ is β -cocoercive with $\beta = (\sum_{i=1}^m \beta_i^{-1})^{-1}$.

$$\langle \sum A_i(x-y), x-y \rangle \geq \beta \|x-y\|^2$$

Université Paris-Saclay – M2 Optimisation

Convex analysis and optimization

Part II.3: Forward-backward algorithm

jean-christophe@pesquet.eu, emilie.chouzenoux@inria.fr
Center for Visual Computing – OPIS Inria group
CentraleSupélec

Fixed points

In the following, $\mathcal{H}$ is a real Hilbert space.

Let $(\nu, \gamma) \in]0, +\infty[^2$. Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν -Lipschitzian gradient.

Let $T = \text{prox}_{\gamma f} \circ (\text{Id} - \gamma \nabla g)$.

1. $\text{Fix } T = \text{Argmin}(f + g)$.
2. If $\gamma \in]0, 2/\nu[$, then T is δ^{-1} -averaged with $\delta = 2 - \gamma\nu/2 \in]1, 2[$.

Proof: For every $x \in \mathcal{H}$,

$$\begin{aligned}
 x &\in \text{Fix } T \\
 &\Leftrightarrow x = \text{prox}_{\gamma f}(x - \gamma \nabla g(x)) \\
 &\Leftrightarrow (\text{Id} - \gamma \nabla g)x \in (\text{Id} + \gamma \partial f)x \\
 &\Leftrightarrow 0 \in \nabla g(x) + \partial f(x).
 \end{aligned}$$

Consequently, $\text{Fix } T = \text{zer}(\nabla g + \partial f) = \text{zer}(\partial(g + f)) = \text{Argmin}(f + g)$.

Fixed points

Let $(\nu, \gamma) \in]0, +\infty[^2$. Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν -Lipschitzian gradient.

Let $T = \text{prox}_{\gamma f} \circ (\text{Id} - \gamma \nabla g)$.

1. $\text{Fix } T = \text{Argmin}(f + g)$.

2. If $\gamma \in]0, 2/\nu[$, then T is δ^{-1} -averaged with $\delta = 2 - \gamma\nu/2 \in]1, 2[$.

Proof: $\text{prox}_{\gamma f}$ is $\alpha_1 = 1/2$ -averaged.

$\gamma \in]0, 2/\nu[\Rightarrow \text{Id} - \gamma \nabla g$ is $\alpha_2 = \gamma\nu/2$ -averaged.

It follows that T is α -averaged with

$$\begin{aligned} \alpha &= \frac{1}{1 + 1/(\frac{\alpha_1}{1-\alpha_1} + \frac{\alpha_2}{1-\alpha_2})} = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \\ &= \frac{\frac{1}{2} + \frac{\gamma\nu}{2} - 2\frac{1}{2}\frac{\gamma\nu}{2}}{1 - \frac{1}{2}\frac{\gamma\nu}{2}} \Leftrightarrow \alpha^{-1} = \delta. \end{aligned}$$

Forward-Backward with fixed step-size

Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1, 2[$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

We assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma \nabla g(x_n) \\ x_{n+1} = x_n + \lambda_n (\text{prox}_{\gamma f} y_n - x_n). \end{cases}$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges weakly to a minimizer x of $f + g$.

Exercise: Prove this result.

Forward-Backward with fixed step-size

Proof:

Let $T = \text{prox}_{\gamma f} \circ (\text{Id} - \gamma \nabla g)$. For every $n \in \mathbb{N}$,

$$x_{n+1} = x_n + \lambda_n (Tx_n - x_n).$$

T is δ^{-1} -averaged and $\text{Fix } T = \text{Argmin}(f + g)$.

The FB algorithm is thus have an instance of the generalized form of KM iteration.

Forward-Backward with fixed step-size

Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1, 2[$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

We assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma \nabla g(x_n) \\ x_{n+1} = x_n + \lambda_n (\text{prox}_{\gamma f} y_n - x_n). \end{cases}$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges weakly to a minimizer x of $f + g$.

In addition, if $T = \text{prox}_{\gamma f} \circ (\text{Id} - \gamma \nabla g)$, then

$$(\forall n \in \mathbb{N}) \quad \langle Tx_n - x \mid x_n - Tx_n \rangle \geq \gamma \langle Tx_n - x \mid \nabla g(x_n) - \nabla g(x) \rangle.$$

Forward-Backward with fixed step-size

Proof:

Let $y = x - \gamma \nabla g(x)$.

Since $\text{prox}_{\gamma f}$ is firmly nonexpansive, for every $n \in \mathbb{N}$,

$$\begin{aligned}
 & \langle \text{prox}_{\gamma f} y_n - \text{prox}_{\gamma f} y \mid y_n - y \rangle \geq \|\text{prox}_{\gamma f} y_n - \text{prox}_{\gamma f} y\|^2 \\
 \Leftrightarrow & \langle \text{prox}_{\gamma f} y_n - \text{prox}_{\gamma f} y \mid (\text{Id} - \text{prox}_{\gamma f}) y_n - (\text{Id} - \text{prox}_{\gamma f}) y \rangle \geq 0 \\
 \Leftrightarrow & \langle Tx_n - Tx \mid y_n - Tx_n - y + Tx \rangle \geq 0 \\
 \Leftrightarrow & \langle Tx_n - x \mid x_n - \gamma \nabla g(x_n) - Tx_n - x + \gamma \nabla g(x) + x \rangle \geq 0 \\
 \Leftrightarrow & \langle Tx_n - x \mid x_n - Tx_n \rangle \geq \gamma \langle Tx_n - x \mid \nabla g(x_n) - \nabla g(x) \rangle.
 \end{aligned}$$

Forward-Backward with fixed step-size

Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1, 2[$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

We assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma \nabla g(x_n) \\ x_{n+1} = x_n + \lambda_n (\text{prox}_{\gamma f} y_n - x_n). \end{cases}$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges weakly to a minimizer x of $f + g$.

In addition, $(\nabla g(x_n))_{n \in \mathbb{N}}$ converges strongly to $\nabla g(x)$.

Forward-Backward with fixed step-size

Proof:

Since T is nonexpansive, for every $n \in \mathbb{N}$,

$$\begin{aligned}
 \|x_n - x\| \|x_n - Tx_n\| &\geq \|Tx_n - Tx\| \|x_n - Tx_n\| \\
 &= \|Tx_n - x\| \|x_n - Tx_n\| \\
 &\geq \langle Tx_n - x \mid x_n - Tx_n \rangle \\
 &\geq \gamma \langle Tx_n - x \mid \nabla g(x_n) - \nabla g(x) \rangle \\
 &= \gamma \langle Tx_n - x_n \mid \nabla g(x_n) - \nabla g(x) \rangle + \gamma \langle x_n - x \mid \nabla g(x_n) - \nabla g(x) \rangle.
 \end{aligned}$$

By using the cocoercivity of ∇g ,

$$\begin{aligned}
 \|x_n - x\| \|x_n - Tx_n\| &\geq -\gamma \|Tx_n - x_n\| \|\nabla g(x_n) - \nabla g(x)\| + \gamma \nu^{-1} \|\nabla g(x_n) - \nabla g(x)\|^2 \\
 &\geq -\gamma \nu \|Tx_n - x_n\| \|x_n - x\| + \gamma \nu^{-1} \|\nabla g(x_n) - \nabla g(x)\|^2.
 \end{aligned}$$

This yields

$$\gamma \nu^{-1} \|\nabla g(x_n) - \nabla g(x)\|^2 \leq (1 + \gamma \nu) \|x_n - x\| \|x_n - Tx_n\|.$$

Forward-Backward with fixed step-size

Proof:

This yields

$$\gamma\nu^{-1}\|\nabla g(x_n) - \nabla g(x)\|^2 \leq (1 + \gamma\nu)\|x_n - x\|\|x_n - Tx_n\|.$$

Since $x_n \rightharpoonup x$, $(x_n)_{n \in \mathbb{N}}$ is bounded and so is $(\|x_n - x\|)_{n \in \mathbb{N}}$.

In addition $x_n - Tx_n \rightarrow 0$.

Therefore $\nabla g(x_n) \rightarrow \nabla g(x)$.

Some lemmas

Three point inequality

Let $(\nu, \gamma) \in]0, +\infty[^2$. Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient.

Let $h = f + g$.

Let $(x, z) \in \mathcal{H}^2$ and let $p = \text{prox}_{\gamma f}(x - \gamma \nabla g(x))$.

Then

$$h(p) \leq h(z) + \frac{1}{\gamma} \langle x - p \mid x - z \rangle - \left(\frac{1}{\gamma} - \frac{\nu}{2} \right) \|x - p\|^2.$$

Proof: According to the descent lemma,

$$g(p) \leq g(x) + \langle \nabla g(x) \mid p - x \rangle + \frac{\nu}{2} \|p - x\|^2$$

and, by the tangent inequality,

$$g(z) \geq g(x) + \langle \nabla g(x) \mid z - x \rangle.$$

We have thus

$$g(p) \leq g(z) - \langle \nabla g(x) \mid z - p \rangle + \frac{\nu}{2} \|p - x\|^2.$$

Some lemmas

Proof: We have thus

$$g(p) \leq g(z) - \langle \nabla g(x) \mid z - p \rangle + \frac{\nu}{2} \|p - x\|^2.$$

In addition

$$\begin{aligned} p &= \text{prox}_{\gamma f}(x - \gamma \nabla g(x)) \\ \Leftrightarrow x - \gamma \nabla g(x) - p &\in \gamma \partial f(p) \\ \Leftrightarrow \frac{x - p}{\gamma} - \nabla g(x) &\in \partial f(p). \end{aligned}$$

We deduce that

$$f(z) \geq f(p) + \left\langle \frac{x - p}{\gamma} - \nabla g(x) \mid z - p \right\rangle.$$

Some lemmas

Proof: We have thus

$$\begin{aligned} & \begin{cases} g(p) \leq g(z) - \langle \nabla g(x) \mid z - p \rangle + \frac{\nu}{2} \|p - x\|^2 \\ f(z) \geq f(p) + \frac{1}{\gamma} \langle x - p \mid z - p \rangle - \langle \nabla g(x) \mid z - p \rangle \end{cases} \\ \Rightarrow & h(p) \leq h(z) - \frac{1}{\gamma} \langle x - p \mid z - p \rangle + \frac{\nu}{2} \|p - x\|^2 \\ \Leftrightarrow & h(p) \leq h(z) + \frac{1}{\gamma} \langle x - p \mid p - x + x - z \rangle + \frac{\nu}{2} \|p - x\|^2 \\ \Leftrightarrow & h(p) \leq h(z) + \frac{1}{\gamma} \langle x - p \mid x - z \rangle + \left(\frac{\nu}{2} - \frac{1}{\gamma} \right) \|p - x\|^2. \end{aligned}$$

Some lemmas

Let $h \in \Gamma_0(\mathcal{H})$. Then, h is weakly lower-semicontinuous .

Proof: $h \in \Gamma_0(\mathcal{H})$ if and only if its epigraph

$$\text{epi } h = \{(x, \eta) \in \mathcal{H} \times \mathbb{R} \mid h(x) \leq \eta\}$$

is nonempty closed and convex.

Let $(x_n, \eta_n)_{n \in \mathbb{N}}$ be a sequence of epi_h such that $(x_n, \eta_n) \rightharpoonup (x, \eta)$. Since epi_h is closed and convex, $(x, \eta) \in \text{epi } h$.

This shows that $\text{epi } h$ is weakly closed, that is h is weakly lower semicontinuous.

Some lemmas

Let $h \in \Gamma_0(\mathcal{H})$. Then, h is weakly lower-semicontinuous.

Let $h \in \Gamma_0(\mathcal{H})$ be such that $\inf h > -\infty$.

Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$ which is a minimizing sequence, i.e.

$$\lim_{n \rightarrow +\infty} h(x_n) = \inf h.$$

If $x_n \rightharpoonup x$, then $x \in \text{Argmin} h$.

Proof: Since h is weakly lower-semicontinuous,

$$\inf h = \lim_{n \rightarrow +\infty} h(x_n) = \liminf_{n \rightarrow +\infty} h(x_n) \geq h(x).$$

Then $h(x) = \inf h$.

Forward-Backward with varying stepsize

Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $(\gamma_n)_{n \in \mathbb{N}}$ in $[\underline{\gamma}, \bar{\gamma}]$ where $0 < \underline{\gamma} < \bar{\gamma} < 2/\nu$ and
 let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[\underline{\lambda}, 1]$ with $0 < \underline{\lambda} \leq 1$.

We assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 \in \text{dom } f$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma_n \nabla g(x_n) \\ p_n = \text{prox}_{\gamma_n f} y_n \\ x_{n+1} = x_n + \lambda_n (p_n - x_n). \end{cases}$$

Then,

1. $(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone sequence with respect to $\text{Argmin}(f + g)$.
2. If $\hat{x} \in \text{Argmin}(f + g)$, then $\sum_{n \in \mathbb{N}} \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 < +\infty$.
3. $\sum_{n \in \mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$.

Forward-Backward with varying stepsize

Proof: Let $\hat{x} \in \text{Argmin}(f + g)$.

For every $n \in \mathbb{N}$, since $\text{prox}_{\gamma_n f}$ is firmly nonexpansive,

$$\begin{aligned}
 \|p_n - \hat{x}\|^2 &= \|\text{prox}_{\gamma_n f}(x_n - \gamma_n \nabla g(x_n)) - \text{prox}_{\gamma_n f}(\hat{x} - \gamma_n \nabla g(\hat{x}))\|^2 \\
 &\leq \|x_n - \gamma_n \nabla g(x_n) - \hat{x} + \gamma_n \nabla g(\hat{x})\|^2 \\
 &\quad - \|x_n - \gamma_n \nabla g(x_n) - p_n - \hat{x} + \gamma_n \nabla g(\hat{x}) + \hat{x}\|^2 \\
 &= \|x_n - \hat{x}\|^2 - 2\gamma_n \langle x_n - \hat{x} \mid \nabla g(x_n) - \nabla g(\hat{x}) \rangle + \gamma_n^2 \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 \\
 &\quad - \|x_n - \gamma_n \nabla g(x_n) - p_n + \gamma_n \nabla g(\hat{x})\|^2.
 \end{aligned}$$

By using the ν^{-1} -cocoercity of ∇g ,

$$\begin{aligned}
 \|p_n - \hat{x}\|^2 &\leq \|x_n - \hat{x}\|^2 - \gamma_n(2\nu^{-1} - \gamma_n) \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 \\
 &\quad - \|x_n - \gamma_n \nabla g(x_n) - p_n + \gamma_n \nabla g(\hat{x})\|^2.
 \end{aligned}$$

Forward-Backward with varying stepsize

By using the ν^{-1} -cocoercity of ∇g ,

$$\begin{aligned}\|p_n - \hat{x}\|^2 &\leq \|x_n - \hat{x}\|^2 - \gamma_n(2\nu^{-1} - \gamma_n)\|\nabla g(x_n) - \nabla g(\hat{x})\|^2 \\ &\quad - \|x_n - \gamma_n \nabla g(x_n) - p_n + \gamma_n \nabla g(\hat{x})\|^2.\end{aligned}$$

We deduce that

$$\begin{aligned}\|x_{n+1} - \hat{x}\|^2 &= \|(1 - \lambda_n)x_n + \lambda_n p_n - \hat{x}\|^2 \\ &\leq (1 - \lambda_n)\|x_n - \hat{x}\|^2 + \lambda_n\|p_n - \hat{x}\|^2 \\ &\leq \|x_n - \hat{x}\|^2 - \lambda_n \gamma_n(2\nu^{-1} - \gamma_n)\|\nabla g(x_n) - \nabla g(\hat{x})\|^2 \\ &\quad - \lambda_n\|x_n - \gamma_n \nabla g(x_n) - p_n + \gamma_n \nabla g(\hat{x})\|^2.\end{aligned}$$

Since $\gamma_n \in]0, 2/\nu[$,

$$\begin{aligned}\|x_{n+1} - \hat{x}\| &\leq \|x_n - \hat{x}\| \\ \lambda_n \gamma_n(2\nu^{-1} - \gamma_n)\|\nabla g(x_n) - \nabla g(\hat{x})\|^2 &\leq \|x_n - \hat{x}\|^2 - \|x_{n+1} - \hat{x}\|^2 \\ \lambda_n\|x_n - \gamma_n \nabla g(x_n) - p_n + \gamma_n \nabla g(\hat{x})\|^2 &\leq \|x_n - \hat{x}\|^2 - \|x_{n+1} - \hat{x}\|^2.\end{aligned}$$

Forward-Backward with varying stepsize

Since $\gamma_n \in]0, 2/\nu[$,

$$\begin{aligned}\|x_{n+1} - \hat{x}\| &\leq \|x_n - \hat{x}\| \\ \lambda_n \gamma_n (2\nu^{-1} - \gamma_n) \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 &\leq \|x_n - \hat{x}\|^2 - \|x_{n+1} - \hat{x}\|^2 \\ \lambda_n \|x_n - \gamma_n \nabla g(x_n) - p_n + \gamma_n \nabla g(\hat{x})\|^2 &\leq \|x_n - \hat{x}\|^2 - \|x_{n+1} - \hat{x}\|^2.\end{aligned}$$

This implies that $(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone sequence with respect to $\text{Argmin}(f + g)$ and

$$\begin{aligned}\underline{\lambda} \underline{\gamma} (2\nu^{-1} - \bar{\gamma}) \sum_{n \in \mathbb{N}} \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 \\ \leq \sum_{n \in \mathbb{N}} \lambda_n \gamma_n (2\nu^{-1} - \gamma_n) \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 \leq \|x_0 - \hat{x}\|^2 \\ \underline{\lambda} \sum_{n \in \mathbb{N}} \|x_n - \gamma_n \nabla g(x_n) - p_n + \gamma_n \nabla g(\hat{x})\|^2 \leq \|x_0 - \hat{x}\|^2.\end{aligned}$$

Forward-Backward with varying stepsize

This implies that $(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone sequence with respect to $\text{Argmin}(f + g)$ and

$$\begin{aligned} & \underline{\lambda} \underline{\gamma} (2\nu^{-1} - \bar{\gamma}) \sum_{n \in \mathbb{N}} \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 \\ & \leq \sum_{n \in \mathbb{N}} \lambda_n \gamma_n (2\nu^{-1} - \gamma_n) \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 \leq \|x_0 - \hat{x}\|^2 \\ & \underline{\lambda} \sum_{n \in \mathbb{N}} \|x_n - \gamma_n \nabla g(x_n) - p_n + \gamma_n \nabla g(\hat{x})\|^2 \leq \|x_0 - \hat{x}\|^2. \end{aligned}$$

Consequently, $(\nabla g(x_n) - \nabla g(\hat{x}))_{n \in \mathbb{N}} \in \ell_{\mathcal{H}}^2(\mathbb{N}) \Rightarrow$
 $(\gamma_n(\nabla g(x_n) - \nabla g(\hat{x})))_{n \in \mathbb{N}} \in \ell_{\mathcal{H}}^2(\mathbb{N})$ and
 $(x_n - \gamma_n \nabla g(x_n) - p_n + \gamma_n \nabla g(\hat{x}))_{n \in \mathbb{N}} \in \ell_{\mathcal{H}}^2(\mathbb{N}) \Rightarrow (x_n - p_n)_{n \in \mathbb{N}} \in \ell_{\mathcal{H}}^2(\mathbb{N}).$
 Since $(\forall n \in \mathbb{N}) \|x_{n+1} - x_n\| = \lambda_n \|x_n - p_n\| \leq \|x_n - p_n\|,$
 $(x_{n+1} - x_n)_{n \in \mathbb{N}} \in \ell_{\mathcal{H}}^2(\mathbb{N}).$

Forward-Backward with varying stepsize

Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $(\gamma_n)_{n \in \mathbb{N}}$ in $[\underline{\gamma}, \bar{\gamma}]$ where $0 < \underline{\gamma} < \bar{\gamma} < 2/\nu$ and let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[\underline{\lambda}, 1]$ with $0 < \underline{\lambda} \leq 1$.

We assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 \in \text{dom } f$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma_n \nabla g(x_n) \\ p_n = \text{prox}_{\gamma_n f} y_n \\ x_{n+1} = x_n + \lambda_n (p_n - x_n). \end{cases}$$

Then,

1. $(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone sequence with respect to $\text{Argmin}(f + g)$.
2. If $\hat{x} \in \text{Argmin}(f + g)$, then $\sum_{n \in \mathbb{N}} \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 < +\infty$.
3. $\sum_{n \in \mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$.
4. $(f(x_n) + g(x_n))_{n \in \mathbb{N}}$ is a decaying sequence.
5. $\sum_{n \in \mathbb{N}} ((f + g)(x_n) - \inf(f + g))^2 < +\infty$.

Forward-Backward with varying stepsize

According to the 3 point inequality, for every $z \in \mathcal{H}$ and $n \in \mathbb{N}$,

$$(f + g)(p_n) \leq (f + g)(z) + \frac{1}{\gamma_n} \langle x_n - p_n \mid x_n - z \rangle - \left(\frac{1}{\gamma_n} - \frac{\nu}{2} \right) \|x_n - p_n\|^2.$$

Setting $z = x_n$ yields

$$(f + g)(p_n) \leq (f + g)(x_n) - \left(\frac{1}{\gamma_n} - \frac{\nu}{2} \right) \|x_n - p_n\|^2.$$

Since $\gamma_n \in]0, 2/\nu[$, $(f + g)(p_n) \leq (f + g)(x_n)$. In addition, by convexity,

$$\begin{aligned} (f + g)(x_{n+1}) &= (f + g)((1 - \lambda_n)x_n + \lambda_n p_n) \\ &\leq (1 - \lambda_n)(f + g)(x_n) + \lambda_n(f + g)(p_n). \end{aligned}$$

Thus $(f + g)(x_{n+1}) \leq (f + g)(x_n)$.

Besides, $0 \leq \lambda_n((f + g)(x_n) - (f + g)(p_n)) \leq (f + g)(x_n) - (f + g)(x_{n+1})$.

Forward-Backward with varying stepsize

Besides, $0 \leq \lambda_n((f + g)(x_n) - (f + g)(p_n)) \leq (f + g)(x_n) - (f + g)(x_{n+1})$.
This yields

$$\underline{\lambda} \sum_{n \in \mathbb{N}} ((f + g)(x_n) - (f + g)(p_n)) \leq (f + g)(x_0) - \inf(f + g) < +\infty.$$

This shows that $((f + g)(x_n) - (f + g)(p_n))_{n \in \mathbb{N}} \in \ell^1(\mathbb{N}) \subset \ell^2(\mathbb{N})$.

Forward-Backward with varying stepsize

Now, set $z = \hat{x} \in \text{Argmin}(f + g)$ in the 3 point inequality.

For every $n \in \mathbb{N}$,

$$(f + g)(p_n) \leq \inf(f + g) + \frac{1}{\gamma_n} \langle x_n - p_n \mid x_n - \hat{x} \rangle - \left(\frac{1}{\gamma_n} - \frac{\nu}{2} \right) \|x_n - p_n\|^2.$$

By using Cauchy-Schwarz inequality, this leads to

$$0 \leq (f + g)(p_n) - \inf(f + g) \leq \frac{1}{\gamma_n} \|x_n - p_n\| \|x_n - \hat{x}\|.$$

Since $(x_n)_{n \in \mathbb{N}}$ is Fejér monotone with respect to $\text{Argmin}(f + g)$,

$$0 \leq (f + g)(p_n) - \inf(f + g) \leq \frac{1}{\underline{\gamma}} \|x_n - p_n\| \|x_0 - \hat{x}\|.$$

As we have proved that $(\|x_n - p_n\|)_{n \in \mathbb{N}} \in \ell^2(\mathbb{N})$,

$((f + g)(p_n) - \inf(f + g))_{n \in \mathbb{N}} \in \ell^2(\mathbb{N})$.

Knowing that $((f + g)(x_n) - (f + g)(p_n))_{n \in \mathbb{N}} \in \ell^2(\mathbb{N})$,

we conclude that $((f + g)(x_n) - \inf(f + g))_{n \in \mathbb{N}} \in \ell^2(\mathbb{N})$.

Forward-Backward with varying stepsize

Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $(\gamma_n)_{n \in \mathbb{N}}$ in $[\underline{\gamma}, \bar{\gamma}]$ where $0 < \underline{\gamma} < \bar{\gamma} < 2/\nu$ and

let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[\underline{\lambda}, 1]$ with $0 < \underline{\lambda} \leq 1$.

We assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 \in \text{dom } f$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma_n \nabla g(x_n) \\ p_n = \text{prox}_{\gamma_n f} y_n \\ x_{n+1} = x_n + \lambda_n (p_n - x_n). \end{cases}$$

Then,

1. $(x_n)_{n \in \mathbb{N}}$ is Fejér-monotone sequence with respect to $\text{Argmin}(f + g)$.
2. If $\hat{x} \in \text{Argmin}(f + g)$, then $\sum_{n \in \mathbb{N}} \|\nabla g(x_n) - \nabla g(\hat{x})\|^2 < +\infty$.
3. $\sum_{n \in \mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$.
4. $(f(x_n) + g(x_n))_{n \in \mathbb{N}}$ is a decaying sequence.
5. $\sum_{n \in \mathbb{N}} ((f + g)(x_n) - \inf(f + g))^2 < +\infty$.
6. $(x_n)_{n \in \mathbb{N}}$ converges weakly to a minimizer of $f + g$.

Forward-Backward with varying stepsize

Since $((f + g)(x_n) - \inf(f + g))_{n \in \mathbb{N}} \in \ell^2(\mathbb{N})$, $(x_n)_{n \in \mathbb{N}}$ is a minimizing sequence.

Let $(x_{n_k})_{k \in \mathbb{N}}$ be a subsequence of $(x_n)_{n \in \mathbb{N}}$ converging weakly to $x \in \mathcal{H}$. It is also a minimizing subsequence and, since $f + g \in \Gamma_0(\mathcal{H})$, $x \in \operatorname{Argmin}(f + g)$.

Thus, every weak sequential cluster point of $(x_n)_{n \in \mathbb{N}}$ belongs to $\operatorname{Argmin}(f + g)$.

It follows from the monotone convergence theorem that $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\operatorname{Argmin}(f + g)$.

Forward-Backward with varying stepsize

Let $f \in \Gamma_0(\mathcal{H})$ be strongly convex .

Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $(\gamma_n)_{n \in \mathbb{N}}$ in $[\underline{\gamma}, \overline{\gamma}]$ where $0 < \underline{\gamma} < \overline{\gamma} < 2/\nu$ and
let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[\underline{\lambda}, 1]$ with $0 < \underline{\lambda} \leq 1$.

Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma_n \nabla g(x_n) \\ p_n = \text{prox}_{\gamma_n f} y_n \\ x_{n+1} = x_n + \lambda_n (p_n - x_n). \end{cases}$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges linearly to the unique minimizer of $f + g$.

Forward-Backward with varying stepsize

Proof: If f is strongly convex, there exists $\eta \in]0, +\infty[$ and $\ell \in \Gamma_0(\mathcal{H})$ such that

$$f = \ell + \frac{\eta}{2} \|\cdot\|^2.$$

This implies that $f + g$ is strictly convex and coercive, hence has a unique minimizer $\hat{x}$.

In addition, for every $x \in \mathcal{H}$ and $\gamma \in]0, +\infty[$,

$$\text{prox}_{\gamma f} x = \text{prox}_{\frac{\gamma \ell}{1+\gamma \eta}} \left(\frac{x}{1+\gamma \eta} \right).$$

For every $n \in \mathbb{N}$,

$$\begin{aligned} & \|p_n - \hat{x}\|^2 \\ &= \|\text{prox}_{\gamma_n f}(x_n - \gamma_n \nabla g(x_n)) - \text{prox}_{\gamma_n f}(\hat{x} - \gamma_n \nabla g(\hat{x}))\|^2 \\ &\leq \frac{1}{(1 + \gamma_n \eta)^2} \|x_n - \gamma_n \nabla g(x_n) - \hat{x} + \gamma_n \nabla g(\hat{x})\|^2 \\ &\leq \frac{1}{(1 + \gamma_n \eta)^2} (\|x_n - \hat{x}\|^2 - \gamma_n (2\nu^{-1} - \gamma_n) \|\nabla g(x_n) - \nabla g(\hat{x})\|^2) \\ &\leq \frac{1}{(1 + \gamma_n \eta)^2} \|x_n - \hat{x}\|^2. \end{aligned}$$

Forward-Backward with varying stepsize

Proof: Therefore, for every $n \in \mathbb{N}$,

$$\|p_n - \hat{x}\| \leq \frac{1}{1 + \underline{\gamma}\eta} \|x_n - \hat{x}\|$$

and

$$\begin{aligned} \|x_{n+1} - \hat{x}\| &\leq (1 - \lambda_n) \|x_n - \hat{x}\| + \lambda_n \|p_n - \hat{x}\| \\ &\leq \left(1 - \lambda_n \frac{\underline{\gamma}\eta}{1 + \underline{\gamma}\eta}\right) \|x_n - \hat{x}\| \\ &\leq \chi \|x_n - \hat{x}\| \end{aligned}$$

with

$$0 \leq \chi = 1 - \lambda \frac{\underline{\gamma}\eta}{1 + \underline{\gamma}\eta} < 1.$$

We deduce that, for every $n \in \mathbb{N}$, $\|x_n - \hat{x}\| \leq \chi^n \|x_0 - \hat{x}\|$.

Forward-Backward with varying stepsize

Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $(\gamma_n)_{n \in \mathbb{N}}$ in $[\underline{\gamma}, \overline{\gamma}]$ where $0 < \underline{\gamma} < \overline{\gamma} \leq 1/\nu$.

We assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma_n \nabla g(x_n) \\ x_{n+1} = \text{prox}_{\gamma_n f} y_n \end{cases}$$

Then, there exists $M \in [0, +\infty[$ such that, for every $n \in \mathbb{N} \setminus \{0\}$,

$$(f + g)(x_n) - \inf(f + g) \leq \frac{M}{n}.$$

Forward-Backward with varying stepsize

Setting $z = \hat{x} \in \operatorname{Argmin}(f + g)$ in the 3 point inequality yields, for every $n \in \mathbb{N}$,

$$(f+g)(x_{n+1}) \leq \inf(f+g) + \frac{1}{\gamma_n} \langle x_n - x_{n+1} \mid x_n - \hat{x} \rangle - \left(\frac{1}{\gamma_n} - \frac{\nu}{2} \right) \|x_n - x_{n+1}\|^2.$$

Since $\gamma_n \in]0, 1/\nu]$, $\gamma_n^{-1} - \nu/2 \geq \gamma_n^{-1}/2$ and

$$\begin{aligned} & (f + g)(x_{n+1}) - \inf(f + g) \\ & \leq -\frac{1}{2\gamma_n} (-2 \langle x_n - x_{n+1} \mid x_n - \hat{x} \rangle + \|x_n - x_{n+1}\|^2) \\ & = -\frac{1}{2\gamma_n} (\|x_n - x_{n+1} - x_n + \hat{x}\|^2 - \|x_n - \hat{x}\|^2) \\ & \leq \frac{1}{2\underline{\gamma}} (\|x_n - \hat{x}\|^2 - \|x_{n+1} - \hat{x}\|^2). \end{aligned}$$

Forward-Backward with varying stepsize

Since $\gamma_n \in]0, 1/\nu]$, $\gamma_n^{-1} - \nu/2 \geq \gamma_n^{-1}/2$ and

$$\begin{aligned} (f + g)(x_{n+1}) - \inf(f + g) \\ \leq \frac{1}{2\underline{\gamma}} (\|x_n - \hat{x}\|^2 - \|x_{n+1} - \hat{x}\|^2). \end{aligned}$$

We deduce that, for every $N \in \mathbb{N}$,

$$\sum_{n=0}^N (f + g)(x_{n+1}) - \inf(f + g) \leq \frac{1}{2\underline{\gamma}} (\|x_0 - \hat{x}\|^2 - \|x_{N+1} - \hat{x}\|^2).$$

Since $((f + g)(x_n) - \inf(f + g))_{n \in \mathbb{N}}$ is a decaying sequence, we have thus

$$(N + 1)((f + g)(x_{N+1}) - \inf(f + g)) \leq \frac{1}{2\underline{\gamma}} \|x_0 - \hat{x}\|^2.$$

Accelerated version

Let $f \in \Gamma_0(\mathcal{H})$.

Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $\gamma \in]0, 1/\nu]$ and $\zeta \in [2, +\infty[$.

We assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 = z_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \text{prox}_{\gamma f}(z_n - \gamma \nabla g(z_n)) \\ \lambda_n = \frac{n}{n+1+\zeta} \\ z_{n+1} = x_{n+1} + \lambda_n(x_{n+1} - x_n). \end{cases}$$

Then, there exists $M \in [0, +\infty[$ such that, for every $n \in \mathbb{N} \setminus \{0\}$,

$$(f + g)(x_n) - \inf(f + g) \leq \frac{M}{n^2}.$$

In addition, if $\zeta > 2$, then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a minimizer of $f + g$.

Projected gradient algorithm

Let C be a nonempty closed convex subset of $\mathcal{H}$.

Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1, 2[$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

We assume that $\underset{x \in C}{\operatorname{Argmin}} g(x) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma \nabla g(x_n) \\ x_{n+1} = x_n + \lambda_n (P_C y_n - x_n). \end{cases}$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges weakly to a minimizer of g over C .

Gradient descent algorithm

Let $g \in \Gamma_0(\mathcal{H})$ be a differentiable function with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $(\gamma_n)_{n \in \mathbb{N}}$ in $[\underline{\gamma}, \overline{\gamma}]$ where $0 < \underline{\gamma} < \overline{\gamma} < 2/\nu$.

We assume that $\text{Argmin } g \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n \nabla g(x_n)$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges weakly to a minimizer of g .

Proximal point algorithm

Let $f \in \Gamma_0(\mathcal{H})$.

Let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $\sum_{n \in \mathbb{N}} \gamma_n = +\infty$. We assume that $\text{Argmin} f \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \text{prox}_{\gamma_n f} x_n.$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges weakly to a minimizer of f .

Reminders on the proximity operator

Let $\mathcal{H}$ be a separable Hilbert space.

Let $(b_i)_{i \in I}$ be an orthonormal basis of $\mathcal{H}$.

For every $i \in I$, let $\varphi_i \in \Gamma_0(\mathbb{R})$ such that $\varphi_i \geq 0$. For every $x \in \mathcal{H}$, if

$$f(x) = \sum_{i \in I} \varphi_i(\langle x | b_i \rangle)$$

then

$$\text{prox}_f(x) = \sum_{i \in I} \text{prox}_{\varphi_i}(\langle x | b_i \rangle) b_i.$$

Moreau decomposition formula

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$ and $\gamma \in]0, +\infty[$.

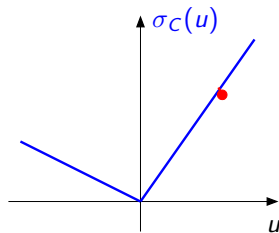
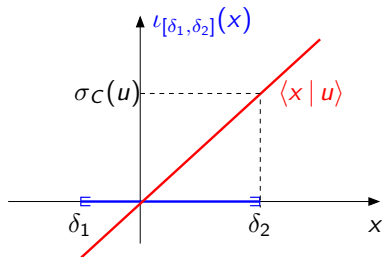
$$(\forall x \in \mathcal{H}) \quad \text{prox}_{\gamma f^*} x = x - \gamma \text{prox}_{\gamma^{-1} f}(\gamma^{-1} x).$$

Reminders on the proximity operator

Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$.

σ_C is the **support function** of C if

$$(\forall u \in \mathcal{H}) \quad \sigma_C(u) = \sup_{x \in C} \langle x | u \rangle \\ = \iota_C^*(u).$$

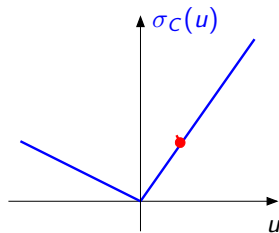
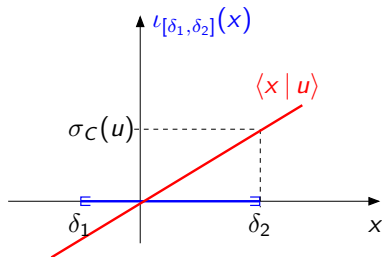


Reminders on the proximity operator

Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$.

σ_C is the **support function** of C if

$$(\forall u \in \mathcal{H}) \quad \sigma_C(u) = \sup_{x \in C} \langle x | u \rangle \\ = \iota_C^*(u).$$

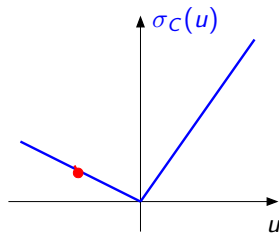
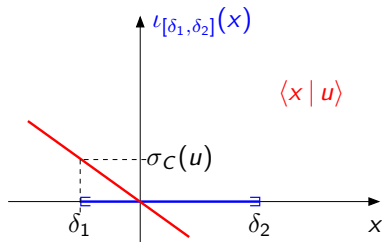


Reminders on the proximity operator

Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$.

σ_C is the **support function** of C if

$$(\forall u \in \mathcal{H}) \quad \sigma_C(u) = \sup_{x \in C} \langle x | u \rangle \\ = \iota_C^*(u).$$



Reminders on the proximity operator

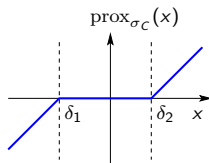
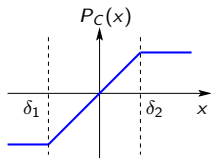
Prox of support function

Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$ be nonempty closed convex.

$$(\forall x \in \mathcal{H}) \quad \text{prox}_{\sigma_C} = \text{Id} - P_C.$$

Soft-thresholding : $\mathcal{H} = \mathbb{R}$, $\delta_1 = \inf C$ and $\delta_2 = \sup C$. For every $x \in \mathbb{R}$,

$$\sigma_C(x) = \begin{cases} \delta_1 x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ \delta_2 x & \text{if } x > 0 \end{cases} \Rightarrow \text{prox}_{\sigma_C}(x) = \text{soft}_C(x) = \begin{cases} x - \delta_1 & \text{if } x < \delta_1 \\ 0 & \text{if } x \in C \\ x - \delta_2 & \text{if } x > \delta_2. \end{cases}$$



Reminders on the proximity operator

Prox of support function

Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$ be nonempty closed convex.

$$(\forall x \in \mathcal{H}) \quad \text{prox}_{\sigma_C} = \text{Id} - P_C.$$

Soft-thresholding : $\mathcal{H} = \mathbb{R}$, $\delta_1 = \inf C$ and $\delta_2 = \sup C$. For every $x \in \mathbb{R}$,

$$\sigma_C(x) = \begin{cases} \delta_1 x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ \delta_2 x & \text{if } x > 0 \end{cases} \Rightarrow \text{prox}_{\sigma_C}(x) = \text{soft}_C(x) = \begin{cases} x - \delta_1 & \text{if } x < \delta_1 \\ 0 & \text{if } x \in C \\ x - \delta_2 & \text{if } x > \delta_2. \end{cases}$$

ℓ^1 -norm : Let $\|\cdot\|_1$ designate the ℓ^1 -norm of $\mathbb{R}^K$.

For every $\chi \in]0, +\infty[$ and $x = (\xi^{(k)})_{1 \leq k \leq K}$,

$$\|x\|_1 = \sum_{k=1}^K |\xi^{(k)}| \Rightarrow \text{prox}_{\chi \|\cdot\|_1}(x) = (\text{soft}_{[-\chi, \chi]}(\xi^{(k)}))_{1 \leq k \leq K}.$$

Iterative thresholding

Problem

Let $\mathcal{G}$ be a Hilbert space and let $L \in \mathcal{B}(\mathbb{R}^K, \mathcal{G})$.

Let $y \in \mathcal{G}$ and let $\chi \in]0, +\infty[$. We want to

$$\underset{x \in \mathbb{R}^K}{\text{minimize}} \quad \frac{1}{2} \|Lx - y\|^2 + \chi \|x\|_1.$$

Algorithm

Let $\gamma \in]0, 2/\|L\|^2[$ and $\delta = 2 - \gamma\|L\|^2/2$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma L^*(Lx_n - y) = (v_n^{(k)})_{1 \leq k \leq K} \\ p_n = (\text{soft}_{[-\gamma\chi, \gamma\chi]}(v_n^{(k)}))_{1 \leq k \leq K} \\ x_{n+1} = x_n + \lambda_n(p_n - x_n). \end{cases}$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges to a solution to the problem.

Iterative thresholding

Algorithm

Let $\gamma \in]0, 2/\|L\|^2[$ and $\delta = 2 - \gamma\|L\|^2/2$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma L^*(Lx_n - y) = (v_n^{(k)})_{1 \leq k \leq K} \\ p_n = (\text{soft}_{[-\gamma\chi, \gamma\chi]}(v_n^{(k)}))_{1 \leq k \leq K} \\ x_{n+1} = x_n + \lambda_n(p_n - x_n). \end{cases}$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges to a solution to the problem.

Proof: Set $f = \chi\|\cdot\|_1$ and $g = \frac{1}{2}\|L \cdot - y\|^2$.

$f + g \in \Gamma_0(\mathbb{R}^K)$ and is coercive, then $\text{Argmin}(f + g) \neq \emptyset$.

In addition, g is Fréchet differentiable and

$$(\forall x \in \mathbb{R}^K) \quad \nabla g(x) = L^*(Lx - y).$$

For every $(x, x') \in (\mathbb{R}^K)^2$,

$$\|\nabla g(x) - \nabla g(x')\| = \|L^*(Lx - Lx')\| \leq \|L^*L\| \|x - x'\|.$$

Thus ∇g is ν -Lipschitzian with $\nu = \|L^*L\| = \|L\|^2$.

By noticing that, for every $n \in \mathbb{N}$, $p_n = \text{prox}_{\gamma\chi\|\cdot\|_1}(y_n)$, we recognize a FB algorithm for which the conditions of convergence are met.

Exercise

Let $m \in \mathbb{N}$, $m \geq 2$. Let $(C_i)_{1 \leq i \leq m}$ be nonempty closed and convex subsets of a Hilbert space $\mathcal{H}$. We want to

$$\underset{x \in C_1}{\text{minimize}} \quad \sum_{i=2}^m d_{C_i}(x)^2$$

1. Show that, if there exists $j \in \{1, \dots, m\}$ such that C_j is bounded, then a solution exists.
2. Let

$$g: \mathcal{H} \rightarrow \mathbb{R}: x \mapsto \frac{1}{2} \sum_{i=2}^m d_{C_i}(x)^2.$$

Show that g is a Lipschitz-differentiable function.

3. Propose an algorithm to find a solution to this problem when it exists.
4. Study the case when $\cap_{i=1}^m C_i \neq \emptyset$.
5. Deduce convergence results about the POCS algorithm defined as

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = P_{C_1} P_{C_2} x_n$$

with $x_0 \in \mathcal{H}$.

Université Paris-Saclay – M2 Optimisation

Convex analysis and optimization

Part II.4: Douglas-Rachford algorithm

jean-christophe@pesquet.eu, emilie.chouzenoux@inria.fr
Center for Visual Computing – OPIS Inria group
CentraleSupélec

Optimization strategies

Motivation

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. We want to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x) + g(x).$$

Possible solutions :

- ▶ gradient descent algorithm $\Rightarrow f + g$ needs to be smooth
- ▶ proximal point algorithm $\Rightarrow f + g$ needs to be “proximable”
- ▶ Forward-Backward algorithm $\Rightarrow g$ needs to be smooth

Can we find a splitting algorithm when both f and g are nonsmooth?

Reflection of the prox

Let $\gamma \in]0, +\infty[$ and let $f \in \Gamma_0(\mathcal{H})$.

The **reflection** of the proximity operator defined as

$$\text{rprox}_{\gamma f} = 2\text{prox}_{\gamma f} - \text{Id}$$

is nonexpansive.

Fixed points

Let $\gamma \in]0, +\infty[$, let $f \in \Gamma_0(\mathcal{H})$ and let $g \in \Gamma_0(\mathcal{H})$.

We have

$$\text{zer}(\partial f + \partial g) = \text{prox}_{\gamma g}(\text{Fix } T)$$

where $T = \text{rprox}_{\gamma f} \circ \text{rprox}_{\gamma g}$.

Fixed points

Let $\gamma \in]0, +\infty[$, let $f \in \Gamma_0(\mathcal{H})$ and let $g \in \Gamma_0(\mathcal{H})$.

We have

$$\text{zer}(\partial f + \partial g) = \text{prox}_{\gamma g}(\text{Fix } T)$$

where $T = \text{rprox}_{\gamma f} \circ \text{rprox}_{\gamma g}$.

Proof: Let $x \in \mathcal{H}$.

$$\begin{aligned} 0 \in \gamma(\partial f(x) + \partial g(x)) &\Leftrightarrow (\exists y \in \mathcal{H}) \ x - y \in \gamma \partial f(x) \text{ and } y - x \in \gamma \partial g(x) \\ &\Leftrightarrow (\exists y \in \mathcal{H}) \ 2x - y \in (\text{Id} + \gamma \partial f)x \\ &\quad \text{and } x = \text{prox}_{\gamma g}(y) \\ &\Leftrightarrow (\exists y \in \mathcal{H}) \ x = \text{prox}_{\gamma f}(\text{rprox}_{\gamma g}(y)) \text{ and } x = \text{prox}_{\gamma g}(y) \\ &\Leftrightarrow (\exists y \in \mathcal{H}) \ \text{rprox}_{\gamma f}(\text{rprox}_{\gamma g}(y)) = 2x - \text{rprox}_{\gamma g}(y) = y \\ &\quad \text{and } x = \text{prox}_{\gamma g}(y) \\ &\Leftrightarrow (\exists y \in \text{Fix } T) \ x = \text{prox}_{\gamma g}(y). \end{aligned}$$

Douglas-Rachford algorithm

Let $\mathcal{H}$ be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ z_n = \text{prox}_{\gamma f}(2y_n - x_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

Douglas-Rachford algorithm

Let $\mathcal{H}$ be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.

Let $\gamma \in]0, +\infty[$ and let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 2]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$.

We assume that $\text{zer}(\partial f + \partial g) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ z_n = \text{prox}_{\gamma f} (2y_n - x_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

The following properties are satisfied:

1. $x_n \rightharpoonup \hat{x}$
2. $\hat{y} = \text{prox}_{\gamma g} \hat{x} \in \text{Argmin}(f + g)$
3. $z_n - y_n \rightarrow 0$.

Douglas-Rachford algorithm

Proof: Let $T = \text{rprox}_{\gamma f} \circ \text{rprox}_{\gamma g}$. T is nonexpansive and $\emptyset \neq \text{zer}(\partial f + \partial g) = \text{prox}_{\gamma g}(\text{Fix } T) \Rightarrow \text{Fix } T \neq \emptyset$.

Moreover, for every $n \in \mathbb{N}$,

$$\begin{aligned} x_{n+1} &= x_n + \lambda_n (\text{prox}_{\gamma f}(2\text{prox}_{\gamma g}(x_n) - x_n) - \text{prox}_{\gamma g}(x_n)) \\ &= x_n + \frac{\lambda_n}{2} (2\text{prox}_{\gamma f}(2\text{prox}_{\gamma g}(x_n) - x_n) - 2\text{prox}_{\gamma g}(x_n) + x_n - x_n) \\ &= x_n + \frac{\lambda_n}{2} (2\text{prox}_{\gamma f}(\text{rprox}_{\gamma g}(x_n)) - \text{rprox}_{\gamma g}(x_n) - x_n) \\ &= x_n + \frac{\lambda_n}{2} (Tx_n - x_n). \end{aligned}$$

$\Rightarrow$ Krasnosel'skii-Mann algorithm with relaxation factors $(\lambda_n/2)_{n \in \mathbb{N}}$.

We deduce that $Tx_n - x_n \rightarrow 0$ and $x_n \rightharpoonup \hat{x} \in \text{Fix } T$.

Douglas-Rachford algorithm

Proof: For every $n \in \mathbb{N}$,

$$z_n - y_n = \text{prox}_{\gamma f}(2\text{prox}_{\gamma g}(x_n) - x_n) - \text{prox}_{\gamma g}(x_n) = \frac{1}{2}(Tx_n - x_n) \rightarrow 0.$$

In addition, $\hat{y} = \text{prox}_{\gamma g}(\hat{x}) \in \text{zer}(\partial f + \partial g)$.

Since

$$\partial f(\hat{y}) + \partial g(\hat{y}) \subset \partial(f + g)(\hat{y}),$$

$$\hat{y} \in \text{zer}(\partial(f + g)) = \text{Argmin}(f + g).$$

Douglas-Rachford algorithm

Let $\mathcal{H}$ be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.

Let $\gamma \in]0, +\infty[$ and let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 2]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$.

We assume that $\text{zer}(\partial f + \partial g) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ z_n = \text{prox}_{\gamma f}(2y_n - x_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

The following properties are satisfied:

1. $x_n \rightharpoonup \hat{x}$
2. $\hat{y} = \text{prox}_{\gamma g} \hat{x} \in \text{Argmin}(f + g)$
3. $z_n - y_n \rightarrow 0$
4. $y_n \rightharpoonup \hat{y}, z_n \rightharpoonup \hat{y}$.

Douglas-Rachford algorithm

Proof: For simplicity, assume that $\mathcal{H}$ is finite dimensional.
 $\text{prox}_{\gamma g}$ being continuous (since nonexpansive), we have

$$y_n \rightarrow \text{prox}_{\gamma g} \hat{x} = \hat{y}.$$

Because $z_n - y_n \rightarrow 0$, we deduce that $z_n \rightarrow \text{prox}_{\gamma g} \hat{x} = \hat{y}$.

Douglas-Rachford algorithm

Let $\mathcal{H}$ be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.

Let $\gamma \in]0, +\infty[$ and let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, 2]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$.

We assume that $\text{zer}(\partial f + \partial g) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ z_n = \text{prox}_{\gamma f}(2y_n - x_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

The following properties are satisfied:

1. $x_n \rightharpoonup \hat{x}$
2. $\hat{y} = \text{prox}_{\gamma g} \hat{x} \in \text{Argmin}(f + g)$
3. $z_n - y_n \rightarrow 0$
4. $y_n \rightharpoonup \hat{y}, z_n \rightharpoonup \hat{y}$.

Remark: The limit case when $(\forall n \in \mathbb{N}) \lambda_n = 2$ is the **Peaceman-Rachford** algorithm. Its convergence is only guaranteed under some additional assumption (e.g. strong convexity of g).

Parallel form of Douglas-Rachford

Lemma

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.

Let $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$.

Then L is weakly continuous.

Proof: Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of $\mathcal{G}$ such that $x_n \rightharpoonup x \in \mathcal{G}$.

Then, for every $y \in \mathcal{H}$,

$$\begin{aligned} \langle x_n \mid L^* y \rangle &\rightarrow \langle x \mid L^* y \rangle \\ \Leftrightarrow \langle Lx_n \mid y \rangle &\rightarrow \langle Lx \mid y \rangle. \end{aligned}$$

Thus $Lx_n \rightharpoonup Lx$.

Parallel form of Douglas-Rachford

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.

Let $g \in \Gamma_0(\mathcal{H})$ and $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ be such that L^*L is an isomorphism.

Let $\gamma \in]0, +\infty[$ and let $(\lambda_n)_{n \in \mathbb{N}}$ a sequence in $[0, 2]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$.

Assume that $\text{zer}(L^* \circ \partial g \circ L) \neq \emptyset$. Let $x_0 \in \mathcal{H}$, $v_0 = (L^*L)^{-1}L^*x_0$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ c_n = (L^*L)^{-1}L^*y_n \\ x_{n+1} = x_n + \lambda_n(L(2c_n - v_n) - y_n) \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$

Then $c_n \rightharpoonup \hat{v}$ and $v_n \rightharpoonup \hat{v}$ where $\hat{v} \in \text{Argmin}(g \circ L)$.

Parallel form of Douglas-Rachford

Proof: Let $E = \text{ran } L$. E is closed.

Indeed, let $(Lw_n)_{n \in \mathbb{N}}$ be sequence of E converging to $z \in \mathcal{H}$.

By continuity,

$$\begin{aligned} L^* L w_n \rightarrow L^* z &\Rightarrow w_n \rightarrow (L^* L)^{-1} L^* z = w \\ &\Rightarrow L w_n \rightarrow L w. \end{aligned}$$

This shows that the limit of any convergent sequence of elements of E belongs to E .

Parallel form of Douglas-Rachford

Proof: Let $E = \text{ran } L$. Since E is closed, $\iota_E \in \Gamma_0(\mathcal{H})$ and

$$\begin{aligned}
 \text{zer}(L^* \circ \partial g \circ L) \neq \emptyset &\Leftrightarrow (\exists v \in \mathcal{G}) \, 0 \in L^* \partial g(Lv) \\
 &\Leftrightarrow (\exists x \in E) \, 0 \in L^* \partial g(x) \\
 &\Leftrightarrow (\exists x \in E)(\exists u \in \partial g(x)) \, 0 = L^* u \\
 &\Leftrightarrow (\exists x \in E)(\exists u \in \partial g(x)) \, -u \in \text{Ker } L^* = E^\perp \\
 &\Leftrightarrow (\exists x \in \mathcal{H})(\exists u \in \partial g(x)) \, -u \in N_E(x) = \partial \iota_E(x) \\
 &\Leftrightarrow (\exists x \in \mathcal{H}) \, 0 \in \partial \iota_E(x) + \partial g(x) \\
 &\Leftrightarrow \text{zer}(\partial \iota_E + \partial g) \neq \emptyset.
 \end{aligned}$$

We next apply Douglas-Rachford algorithm with $f = \iota_E$.

Parallel form of Douglas-Rachford

Proof: $f = \iota_E \Rightarrow \text{prox}_{\gamma f} = P_E$. The DR algorithm reads

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ z_n = P_E(2y_n - x_n) = 2P_E y_n - P_E x_n \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

For every $n \in \mathbb{N}$, $P_E y_n = Lc_n$
with

$$\begin{aligned} c_n &= \underset{c \in \mathcal{G}}{\text{argmin}} \quad \|y_n - Lc\|^2 \\ \Leftrightarrow \quad L^*(Lc_n - y_n) &= 0 \\ \Leftrightarrow \quad c_n &= (L^*L)^{-1}L^*y_n. \end{aligned}$$

Parallel form of Douglas-Rachford

Proof: $f = \iota_E \Rightarrow \text{prox}_{\gamma f} = P_E$. The DR algorithm reads

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ z_n = P_E(2y_n - x_n) = 2P_E y_n - P_E x_n \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

For every $n \in \mathbb{N}$, $P_E y_n = Lc_n$

with $c_n = (L^*L)^{-1}L^*y_n$

and $P_E x_n = Lv_n$ with $v_n = (L^*L)^{-1}L^*x_n$.

In addition

$$\begin{aligned} Lv_{n+1} &= P_E x_{n+1} = P_E x_n + \lambda_n(P_E z_n - P_E y_n) \\ &= P_E x_n + \lambda_n(z_n - P_E y_n) \\ &= P_E x_n + \lambda_n(P_E y_n - P_E x_n) \\ &= L(v_n + \lambda_n(c_n - v_n)), \end{aligned}$$

which yields $v_{n+1} = v_n + \lambda_n(c_n - v_n)$.

Parallel form of Douglas-Rachford

Proof: In summary

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ c_n = (L^* L)^{-1} L^* y_n \\ z_n = L(2c_n - v_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n) \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$

We know that

$$\begin{aligned} y_n &\rightarrow \hat{y} \in \text{Argmin}(\iota_E + g) \\ &\Leftrightarrow \hat{y} = L\hat{v} \text{ and } (\forall y \in E) \ g(y) \geq g(\hat{y}) \\ &\Leftrightarrow \hat{y} = L\hat{v} \text{ and } (\forall v \in \mathcal{G}) \ g(Lv) \geq g(L\hat{v}) \\ &\Leftrightarrow \hat{y} = L\hat{v} \text{ and } \hat{v} \in \text{Argmin}(g \circ L). \end{aligned}$$

Parallel form of Douglas-Rachford

Proof: In summary

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ c_n = (L^* L)^{-1} L^* y_n \\ z_n = L(2c_n - v_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n) \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$

We know that

$$y_n \rightharpoonup \hat{y} = L\hat{v} \text{ and } \hat{v} \in \text{Argmin}(g \circ L).$$

Because of the weak continuity of P_E ,

$$P_E y_n = L c_n \rightharpoonup P_E \hat{y} = \hat{y} = L \hat{v}$$

and, by using the weak continuity of $(L^* L)^{-1} L^*$,

$$c_n \rightharpoonup \hat{v} \in \text{Argmin}(g \circ L).$$

Parallel form of Douglas-Rachford

Proof: In summary

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ c_n = (L^* L)^{-1} L^* y_n \\ z_n = L(2c_n - v_n) \\ x_{n+1} = x_n + \lambda_n(z_n - y_n) \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$

In addition

$$z_n = L(2c_n - v_n) \rightharpoonup \hat{y} = L\hat{v}$$

and, because of the weak continuity of $(L^* L)^{-1} L^*$,

$$2c_n - v_n \rightharpoonup \hat{v} \quad \Rightarrow \quad v_n \rightharpoonup \hat{v}.$$

Parallel form of Douglas-Rachford

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.

Let $g \in \Gamma_0(\mathcal{H})$ and $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ be such that L^*L is an isomorphism.

Let $\gamma \in]0, +\infty[$ and let $(\lambda_n)_{n \in \mathbb{N}}$ a sequence in $[0, 2]$ such that

$$\sum_{n \in \mathbb{N}} \lambda_n (2 - \lambda_n) = +\infty.$$

Assume that $\text{zer}(L^* \circ \partial g \circ L) \neq \emptyset$. Let $x_0 \in \mathcal{H}$, $v_0 = (L^*L)^{-1}L^*x_0$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = \text{prox}_{\gamma g} x_n \\ c_n = (L^*L)^{-1}L^*y_n \\ x_{n+1} = x_n + \lambda_n (L(2c_n - v_n) - y_n) \\ v_{n+1} = v_n + \lambda_n (c_n - v_n). \end{cases}$$

Then $v_n \rightarrow \hat{v}$ where $\hat{v} \in \text{Argmin}(g \circ L)$.

Question: Give the particular case of the algorithm when

$$\mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$$

$$(\forall x = (x_1, \dots, x_m) \in \mathcal{H}) \quad g(x) = \sum_{i=1}^m g_i(x_i)$$

$$L: v \mapsto (L_1 v, \dots, L_m v).$$

Parallel form of Douglas-Rachford

PPXA+

Let $\mathcal{H}_1, \dots, \mathcal{H}_m$ and $\mathcal{G}$ be Hilbert spaces.

For every $i \in \{1, \dots, m\}$, let $g_i \in \Gamma_0(\mathcal{H}_i)$ and let $L_i \in \mathcal{B}(\mathcal{G}, \mathcal{H}_i)$.

Assume that $\sum_{i=1}^m L_i^* L_i$ is an isomorphism.

Let $\gamma \in]0, +\infty[$ and let $(\lambda_n)_{n \in \mathbb{N}}$ a sequence in $[0, 2]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$. Assume that $\text{zer}(\sum_{i=1}^m L_i^* \circ \partial g_i \circ L_i) \neq \emptyset$.

Let $(x_{0,i})_{1 \leq i \leq m} \in \mathcal{H}_1 \times \dots \times \mathcal{H}_m$, $v_0 = (\sum_{i=1}^m L_i^* L_i)^{-1} \sum_{i=1}^m L_i^* x_{0,i}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_{n,i} = \text{prox}_{\gamma g_i}(x_{n,i}), & i \in \{1, \dots, m\} \\ c_n = (\sum_{i=1}^m L_i^* L_i)^{-1} \sum_{i=1}^m L_i^* y_{n,i} \\ x_{n+1,i} = x_{n,i} + \lambda_n (L_i(2c_n - v_n) - y_{n,i}), & i \in \{1, \dots, m\} \\ v_{n+1} = v_n + \lambda_n (c_n - v_n). \end{cases}$$

Then $v_n \rightharpoonup \hat{v} \in \text{Argmin} \sum_{i=1}^m g_i \circ L_i$.

Exercise

Let $\mathcal{G}$ be a real Hilbert space. Let $m \in \mathbb{N}$, $m \geq 2$.

1. For every $i \in \{1, \dots, m\}$, let g_i be a function in $\Gamma_0(\mathcal{G})$ such that $\text{dom } g_1 \cap \left(\bigcap_{i=2}^m \text{int}(\text{dom } g_i)\right) \neq \emptyset$.
Propose a parallel algorithm for finding a minimizer of $\sum_{i=1}^m g_i$ when such a minimizer exists.
2. Let $p \in \{1, \dots, m-1\}$ and let $(C_i)_{1 \leq i \leq m}$ be nonempty closed convex subsets of $\mathcal{G}$. We assume that $\bigcap_{i=1}^p \text{int}(C_i) \neq \emptyset$, and for every $j \in \{p+1, \dots, m\}$, C_j is bounded.
Propose a parallel algorithm for solving the problem

$$\underset{x \in \bigcap_{i=1}^p C_i}{\text{minimize}} \quad \sum_{j=p+1}^m \sigma_{C_j}(x)$$

when this problem admits a solution.

3. Let $\mathcal{G} = \mathbb{R}^N$. Deduce an algorithm to

$$\underset{x \in \bigcap_{i=1}^p C_i}{\text{minimize}} \quad \|x\|_1.$$

Université Paris-Saclay – M2 Optimisation

Convex analysis and optimization

Part II.5: Primal-dual methods

jean-christophe@pesquet.eu, emilie.chouzenoux@inria.fr

Center for Visual Computing – OPIS Inria group
CentraleSupélec



Fenchel-Rockafellar duality

Primal problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f: \mathcal{H} \rightarrow]-\infty, +\infty]$, $g: \mathcal{G} \rightarrow]-\infty, +\infty]$. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

We want to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x) + g(Lx).$$

Dual problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f: \mathcal{H} \rightarrow]-\infty, +\infty]$, $g: \mathcal{G} \rightarrow]-\infty, +\infty]$. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

We want to

$$\underset{v \in \mathcal{G}}{\text{minimize}} \quad f^*(-L^*v) + g^*(v).$$

Fenchel-Rockafellar duality

Weak duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let f be a proper function from $\mathcal{H}$ to $] -\infty, +\infty]$, g be a proper function from $\mathcal{G}$ to $] -\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let

$$\mu = \inf_{x \in \mathcal{H}} f(x) + g(Lx) \quad \text{and} \quad \mu^* = \inf_{v \in \mathcal{G}} f^*(-L^*v) + g^*(v).$$

We have $\mu \geq -\mu^*$. If $\mu \in \mathbb{R}$, $\mu + \mu^*$ is called the **duality gap**.

Fenchel-Rockafellar duality

Weak duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let f be a proper function from $\mathcal{H}$ to $]-\infty, +\infty]$, g be a proper function from $\mathcal{G}$ to $]-\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let

$$\mu = \inf_{x \in \mathcal{H}} f(x) + g(Lx) \quad \text{and} \quad \mu^* = \inf_{v \in \mathcal{G}} f^*(-L^*v) + g^*(v).$$

We have $\mu \geq -\mu^*$. If $\mu \in \mathbb{R}$, $\mu + \mu^*$ is called the **duality gap**.

Proof: According to Fenchel-Young inequality, for every $x \in \mathcal{H}$ and $v \in \mathcal{G}$,

$$f(x) + g(Lx) + f^*(-L^*v) + g^*(v) \geq \langle x \mid -L^*v \rangle + \langle Lx \mid v \rangle = 0.$$

Fenchel-Rockafellar duality

Strong duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

If $\text{int}(\text{dom } g) \cap L(\text{dom } f) \neq \emptyset$ or $\text{dom } g \cap \text{int}(L(\text{dom } f)) \neq \emptyset$, then

$$\mu = \inf_{x \in \mathcal{H}} f(x) + g(Lx) = -\min_{v \in \mathcal{G}} f^*(-L^*v) + g^*(v) = -\mu^*.$$

Example 1: Linear programming

Let $L \in \mathbb{R}^{K \times N}$, $b \in \mathbb{R}^K$, and $c \in \mathbb{R}^N$.

The primal problem

$$\text{Primal-LP :} \quad \underset{x \in [0, +\infty[^N}{\text{minimize}} \quad \langle c \mid x \rangle \quad \text{s.t.} \quad Lx \geq b$$

is associated with the the dual problem

$$\text{Dual-LP :} \quad \underset{y \in [0, +\infty[^K}{\text{maximize}} \quad \langle b \mid y \rangle \quad \text{s.t.} \quad L^\top y \leq c.$$

In addition, if the primal problem has a solution, then strong duality holds.

Example 1: Linear programming

Let $L \in \mathbb{R}^{K \times N}$, $b \in \mathbb{R}^K$, and $c \in \mathbb{R}^N$.

The primal problem

$$\text{Primal-LP :} \quad \underset{x \in [0, +\infty[^N}{\text{minimize}} \quad \langle c \mid x \rangle \quad \text{s.t.} \quad Lx \geq b$$

is associated with the the dual problem

$$\text{Dual-LP :} \quad \underset{y \in [0, +\infty[^K}{\text{maximize}} \quad \langle b \mid y \rangle \quad \text{s.t.} \quad L^\top y \leq c.$$

In addition, if the primal problem has a solution, then strong duality holds.

Proof: Set

$$\begin{cases} (\forall x \in \mathcal{H} = \mathbb{R}^N) & f(x) = \langle c \mid x \rangle + \iota_{[0, +\infty[^N}(x), \\ (\forall z \in \mathcal{G} = \mathbb{R}^K) & g(z) = \iota_{[0, +\infty[^K}(z - b), \\ & y = -v. \end{cases}$$

Example 2: Consensus and sharing

Let $\mathcal{H}$ be a real Hilbert space.

For every $i \in \{1, \dots, m\}$, let $g_i: \mathcal{H} \rightarrow]-\infty, +\infty]$ and $h_i: \mathcal{H} \rightarrow]-\infty, +\infty]$.

The **consensus** problem is given by

$$\underset{\substack{(x_1, \dots, x_m) \in \mathcal{H}^m \\ x_1 = \dots = x_m}}{\text{minimize}} \quad \sum_{i=1}^m g_i(x_i).$$

The **sharing problem** is given by

$$\underset{\substack{(u_1, \dots, u_m) \in \mathcal{H}^m \\ u_1 + \dots + u_m = \bar{u}}}{\text{maximize}} \quad \sum_{i=1}^m h_i(u_i), \quad \bar{u} \in \mathcal{H}.$$

If, for every $i \in \{1, \dots, m\}$, $h_i = -g_i^*(\cdot - \bar{u}/m)$, then sharing is the dual problem of consensus.

Example 2: Consensus and sharing

Let $\mathcal{H}$ be a real Hilbert space.

For every $i \in \{1, \dots, m\}$, let $g_i: \mathcal{H} \rightarrow]-\infty, +\infty]$ and $h_i: \mathcal{H} \rightarrow]-\infty, +\infty]$.

The **consensus** problem is given by

$$\underset{\substack{(x_1, \dots, x_m) \in \mathcal{H}^m \\ x_1 = \dots = x_m}}{\text{minimize}} \quad \sum_{i=1}^m g_i(x_i).$$

The **sharing problem** is given by

$$\underset{\substack{(u_1, \dots, u_m) \in \mathcal{H}^m \\ u_1 + \dots + u_m = \bar{u}}}{\text{maximize}} \quad \sum_{i=1}^m h_i(u_i), \quad \bar{u} \in \mathcal{H}.$$

If, for every $i \in \{1, \dots, m\}$, $h_i = -g_i^*(\cdot - \bar{u}/m)$, then sharing is the dual problem of consensus.

Exercise: Prove this result.

Example 2: Consensus and sharing

Proof: Set $(\forall x = (x_1, \dots, x_m) \in \mathcal{H}^m)$ $g(x) = \sum_{i=1}^m g_i(x_i)$ and $\Lambda_m = \{(x_1, \dots, x_m) \in \mathcal{H}^m \mid x_1 = \dots = x_m\}$.

The consensus problem reads

$$\underset{x=(x_1, \dots, x_m) \in \mathcal{H}^m}{\text{minimize}} \quad \underbrace{\iota_{\Lambda_m}(x)}_{f(x)} + g(x).$$

The dual problem is thus ($L = \text{Id}$)

$$\underset{v \in \mathcal{H}^m}{\text{minimize}} \quad f^*(-v) + g^*(v).$$

where, for every $v = (v_i)_{1 \leq i \leq m} \in \mathcal{H}^m$,

$$f^*(v) = \sup_{x \in \Lambda_m} \langle v \mid x \rangle = \sup_{x_1 \in \mathcal{H}} \sum_{i=1}^m \langle v_i \mid x_1 \rangle = \sup_{x_1 \in \mathcal{H}} \left\langle \sum_{i=1}^m v_i \mid x_1 \right\rangle = \iota_{\{0\}} \left(\sum_{i=1}^m v_i \right).$$

Example 2: Consensus and sharing

Proof: The dual problem is thus ($L = \text{Id}$)

$$\underset{v \in \mathcal{H}^m}{\text{minimize}} \quad f^*(-v) + g^*(v).$$

where, for every $v = (v_i)_{1 \leq i \leq m} \in \mathcal{H}^m$,

$$f^*(v) = \sup_{x \in \Lambda_m} \langle v \mid x \rangle = \sup_{x_1 \in \mathcal{H}} \sum_{i=1}^m \langle v_i \mid x_1 \rangle = \sup_{x_1 \in \mathcal{H}} \left\langle \sum_{i=1}^m v_i \mid x_1 \right\rangle = \iota_{\{0\}} \left(\sum_{i=1}^m v_i \right).$$

The dual problem reads

$$\underset{\substack{(v_1, \dots, v_m) \in \mathcal{H}^m \\ v_1 + \dots + v_m = 0}}{\text{minimize}} \quad \sum_{i=1}^m g_i^*(v_i).$$

Setting $(\forall i \in \{1, \dots, m\}) u_i = v_i + \bar{u}/m$ and $h_i(u_i) = -g_i^*(u_i - \bar{u}/m)$ yields the sharing problem.

Fenchel-Rockafellar duality

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f: \mathcal{H} \rightarrow]-\infty, +\infty]$, $g: \mathcal{G} \rightarrow]-\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

If $\text{dom } g \cap L(\text{dom } f) \neq \emptyset$, then

$$(\forall x \in \mathcal{H}) \quad \partial f(x) + L^* \partial g(Lx) \subset \partial(f + g \circ L)(x).$$

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. If $\text{int}(\text{dom } g) \cap L(\text{dom } f) \neq \emptyset$ or $\text{dom } g \cap \text{int}(L(\text{dom } f)) \neq \emptyset$, then

$$\partial f + L^* \partial g L = \partial(f + g \circ L) .$$

Fenchel-Rockafellar duality

Duality theorem (1)

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

$$\text{zer}(\partial f + L^* \partial g L) \neq \emptyset \quad \Leftrightarrow \quad \text{zer}((-L) \partial f^* (-L^*) + \partial g^*) \neq \emptyset.$$

Fenchel-Rockafellar duality

Duality theorem (1)

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

$$\text{zer}(\partial f + L^* \partial g L) \neq \emptyset \quad \Leftrightarrow \quad \text{zer}((-L) \partial f^*(-L^*) + \partial g^*) \neq \emptyset.$$

Proof:

$$\begin{aligned} (\exists x \in \mathcal{H}) \quad 0 \in \partial f(x) + L^* \partial g(Lx) &\Leftrightarrow (\exists x \in \mathcal{H})(\exists v \in \mathcal{G}) \quad \begin{cases} -L^* v \in \partial f(x) \\ v \in \partial g(Lx) \end{cases} \\ &\Leftrightarrow (\exists x \in \mathcal{H})(\exists v \in \mathcal{G}) \quad \begin{cases} x \in \partial f^*(-L^* v) \\ Lx \in \partial g^*(v) \end{cases} \\ &\Leftrightarrow (\exists v \in \mathcal{G}) \quad 0 \in -L \partial f^*(-L^* v) + \partial g^*(v). \end{aligned}$$

Fenchel-Rockafellar duality

Duality theorem (2)

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

- ▶ If there exists $\hat{x} \in \mathcal{H}$ such that $0 \in \partial f(\hat{x}) + L^* \partial g(L\hat{x})$, then $\hat{x}$ is a solution to the primal problem. Moreover, there exists a solution $\hat{v}$ to the dual problem such that $-L^* \hat{v} \in \partial f(\hat{x})$ and $L\hat{x} \in \partial g^*(\hat{v})$.
- ▶ If there exists $(\hat{x}, \hat{v}) \in \mathcal{H} \times \mathcal{G}$ such that $-L^* \hat{v} \in \partial f(\hat{x})$ and $L\hat{x} \in \partial g^*(\hat{v})$ then $\hat{x}$ (resp. $\hat{v}$) is a solution to the primal (resp. dual) problem.

If $(\hat{x}, \hat{v}) \in \mathcal{H} \times \mathcal{G}$ is such that $-L^* \hat{v} \in \partial f(\hat{x})$ and $L\hat{x} \in \partial g^*(\hat{v})$, then $(\hat{x}, \hat{v})$ is called a **Kuhn-Tucker point**.

Fenchel-Rockafellar duality

Proof:

$$0 \in \partial f(\hat{x}) + L^* \partial g(L\hat{x}) \subset \partial(f + g \circ L)(\hat{x}).$$

Then, according to Fermat's rule, $\hat{x}$ is a solution to the primal problem.
In addition, there exists $\hat{v} \in \mathcal{G}$ such that

$$\begin{cases} 0 \in \partial f(\hat{x}) + L^* \hat{v} \\ \hat{v} \in \partial g(L\hat{x}) \end{cases} \Leftrightarrow \begin{cases} -L^* \hat{v} \in \partial f(\hat{x}) \\ L\hat{x} \in \partial g^*(\hat{v}). \end{cases}$$

We have also $\hat{x} \in \partial f^*(-L^* \hat{v})$, which implies that

$$0 \in -L \partial f^*(-L^* \hat{v}) + \partial g^*(\hat{v}).$$

On the other hand,

$$0 \in -L \partial f^*(-L^* \hat{v}) + \partial g^*(\hat{v}) \subset \partial(f^* \circ (-L^*) + g^*)(\hat{v})$$

$\Rightarrow \hat{v}$ solution to the dual problem.

The second assertion is shown in a similar manner.

Fenchel-Rockafellar duality

Particular case:

If $f = \varphi + \frac{1}{2} \|\cdot - z\|^2$ where $\varphi \in \Gamma_0(\mathcal{H})$ and $z \in \mathcal{H}$, then

$$\begin{aligned} -L^*\hat{v} \in \partial f(\hat{x}) &\Leftrightarrow -L^*\hat{v} \in \partial\varphi(\hat{x}) + \hat{x} - z \\ &\Leftrightarrow 0 \in \hat{x} + L^*\hat{v} - z + \partial\varphi(\hat{x}). \end{aligned}$$

Hence,

$$\hat{x} = \text{prox}_{\varphi}(-L^*\hat{v} + z).$$

Link with Lagrange duality

Minimax problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f: \mathcal{H} \rightarrow]-\infty, +\infty]$, $g: \mathcal{G} \rightarrow]-\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

The primal problem is equivalent to finding

$$\mu = \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \sup_{v \in \mathcal{G}} \mathcal{L}(x, y, v)$$

where $\mathcal{L}$ is the Lagrange function defined as

$$(\forall (x, y, v) \in \mathcal{H} \times \mathcal{G}^2) \quad \mathcal{L}(x, y, v) = f(x) + g(y) + \langle v \mid Lx - y \rangle.$$

Link with Lagrange duality

Minimax problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f: \mathcal{H} \rightarrow]-\infty, +\infty]$, $g: \mathcal{G} \rightarrow]-\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

The primal problem is equivalent to finding

$$\mu = \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \sup_{v \in \mathcal{G}} \mathcal{L}(x, y, v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$(\forall (x, y, v) \in \mathcal{H} \times \mathcal{G}^2) \quad \mathcal{L}(x, y, v) = f(x) + g(y) + \langle v \mid Lx - y \rangle.$$

Proof:

$$\begin{aligned} \mu &= \inf_{x \in \mathcal{H}} f(x) + g(Lx) = \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} f(x) + g(y) + \iota_{\{0\}}(Lx - y) \\ &= \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} f(x) + g(y) + \sup_{v \in \mathcal{G}} \langle v \mid Lx - y \rangle \\ &= \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \sup_{v \in \mathcal{G}} f(x) + g(y) + \langle v \mid Lx - y \rangle. \end{aligned}$$

Link with Lagrange duality

Maximin problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f: \mathcal{H} \rightarrow]-\infty, +\infty]$, $g: \mathcal{G} \rightarrow]-\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

The dual problem is equivalent to finding

$$-\mu^* = \sup_{v \in \mathcal{G}} \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \mathcal{L}(x, y, v)$$

where $\mathcal{L}$ is the **Lagrange function** defined as

$$(\forall (x, y, v) \in \mathcal{H} \times \mathcal{G}^2) \quad \mathcal{L}(x, y, v) = f(x) + g(y) + \langle v \mid Lx - y \rangle.$$

Proof:

$$\begin{aligned} \mu^* &= \inf_{v \in \mathcal{G}} f^*(-L^*v) + g^*(v) = \inf_{v \in \mathcal{G}} \left(\sup_{x \in \mathcal{H}} \langle x \mid -L^*v \rangle - f(x) \right) + \left(\sup_{y \in \mathcal{G}} \langle y \mid v \rangle - g(y) \right) \\ &= \inf_{v \in \mathcal{G}} - \left(\inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} f(x) + g(y) + \langle v \mid Lx - y \rangle \right) \\ &= - \sup_{v \in \mathcal{G}} \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} f(x) + g(y) + \langle v \mid Lx - y \rangle. \end{aligned}$$

Link with Lagrange duality

Maximin problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f: \mathcal{H} \rightarrow]-\infty, +\infty]$, $g: \mathcal{G} \rightarrow]-\infty, +\infty]$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

The dual problem is equivalent to finding

$$-\mu^* = \sup_{v \in \mathcal{G}} \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \mathcal{L}(x, y, v)$$

where $\mathcal{L}$ is the Lagrange function defined as

$$(\forall (x, y, v) \in \mathcal{H} \times \mathcal{G}^2) \quad \mathcal{L}(x, y, v) = f(x) + g(y) + \langle v \mid Lx - y \rangle.$$

Remark: v is called the Lagrange multiplier associated with the constraint $Lx = y$.

Link with Lagrange duality

Let $(\hat{x}, \hat{y}, \hat{v}) \in \mathcal{H} \times \mathcal{G}^2$.

$(\hat{x}, \hat{y}, \hat{v})$ is a **saddle point** of the Lagrange function $\mathcal{L}$ if

$$(\forall (x, y, v) \in \mathcal{H} \times \mathcal{G}^2) \quad \mathcal{L}(\hat{x}, \hat{y}, v) \leq \mathcal{L}(\hat{x}, \hat{y}, \hat{v}) \leq \mathcal{L}(x, y, \hat{v}).$$

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let $(\hat{x}, \hat{y}, \hat{v}) \in \mathcal{H} \times \mathcal{G}^2$.

Assume that $\text{int}(\text{dom } g) \cap L(\text{dom } f) \neq \emptyset$ or $\text{dom } g \cap \text{int}(L(\text{dom } f)) \neq \emptyset$.

$(\hat{x}, \hat{y}, \hat{v})$ is a saddle point of the Lagrange function



$(\hat{x}, \hat{v})$ is a Kuhn-Tucker point and $\hat{y} = L\hat{x}$.

Link with Lagrange duality

Proof ($\Rightarrow$): If $(\hat{x}, \hat{y}, \hat{v})$ is a saddle point of $\mathcal{L}$, then

$$\begin{aligned} & \begin{cases} 0 \in \partial_x \mathcal{L}(\hat{x}, \hat{y}, \hat{v}) = \partial f(\hat{x}) + L^* \hat{v} \\ 0 \in \partial_y \mathcal{L}(\hat{x}, \hat{y}, \hat{v}) = \partial g(\hat{y}) - \hat{v} \\ 0 = \nabla_v \mathcal{L}(\hat{x}, \hat{y}, \hat{v}) = L\hat{x} - \hat{y} \end{cases} \\ \Leftrightarrow & \begin{cases} -L^* \hat{v} \in \partial f(\hat{x}) \\ \hat{v} \in \partial g(\hat{y}) \\ \hat{y} = L\hat{x} \end{cases} \\ \Leftrightarrow & \begin{cases} -L^* \hat{v} \in \partial f(\hat{x}) \\ L\hat{x} \in \partial g^*(\hat{v}) \\ \hat{y} = L\hat{x}. \end{cases} \end{aligned}$$

Link with Lagrange duality

Proof ($\Leftarrow$): Conversely, assume that $(\hat{x}, \hat{v})$ is a Kuhn-Tucker point and $\hat{y} = L\hat{x}$. Since $\hat{x}$ (resp. $\hat{v}$) is a solution to the primal (resp. dual) problem, then

$$\begin{aligned}\mu &= \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \sup_{v \in \mathcal{G}} \mathcal{L}(x, y, v) = \sup_{v \in \mathcal{G}} \mathcal{L}(\hat{x}, \hat{y}, v) \\ -\mu^* &= \sup_{v \in \mathcal{G}} \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \mathcal{L}(x, y, v) = \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \mathcal{L}(x, y, \hat{v}).\end{aligned}$$

By strong duality, $\sup_{v \in \mathcal{G}} \mathcal{L}(\hat{x}, \hat{y}, v) = \inf_{(x,y) \in \mathcal{H} \times \mathcal{G}} \mathcal{L}(x, y, \hat{v})$, which can be rewritten as

$$(\forall (x, y, v) \in \mathcal{H} \times \mathcal{G}^2) \quad \mathcal{L}(\hat{x}, \hat{y}, v) \leq \mathcal{L}(x, y, \hat{v})$$

or equivalently

$$(\forall (x, y, v) \in \mathcal{H} \times \mathcal{G}^2) \quad \mathcal{L}(\hat{x}, \hat{y}, v) \leq \mathcal{L}(\hat{x}, \hat{y}, \hat{v}) \leq \mathcal{L}(x, y, \hat{v}).$$

Alternating-direction method of multipliers

Idea: iterations for finding a saddle point $(\hat{x}, \hat{y}, \hat{v})$:

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n \in \operatorname{Argmin} \mathcal{L}(\cdot, y_n, v_n) \\ y_{n+1} \in \operatorname{Argmin} \mathcal{L}(x_n, \cdot, v_n) \\ v_{n+1} \text{ such that } \mathcal{L}(x_n, y_{n+1}, v_{n+1}) \geq \mathcal{L}(x_n, y_{n+1}, v_n). \end{cases}$$

But the convergence is not guaranteed in general !

Alternating-direction method of multipliers

Idea: iterations for finding a saddle point $(\hat{x}, \hat{y}, \hat{v})$:

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n \in \operatorname{Argmin} \mathcal{L}(\cdot, y_n, v_n) \\ y_{n+1} \in \operatorname{Argmin} \mathcal{L}(x_n, \cdot, v_n) \\ v_{n+1} \text{ such that } \mathcal{L}(x_n, y_{n+1}, v_{n+1}) \geq \mathcal{L}(x_n, y_{n+1}, v_n). \end{cases}$$

But the convergence is not guaranteed in general !

Solution: introduce an **Augmented Lagrange function** .

Let $\gamma \in]0, +\infty[$, we define

$$(\forall (x, y, z) \in \mathcal{H} \times \mathcal{G}^2) \quad \tilde{\mathcal{L}}(x, y, z) = f(x) + g(y) + \gamma \langle z \mid Lx - y \rangle + \frac{\gamma}{2} \|Lx - y\|^2$$

The Lagrange multiplier is **$v = \gamma z$** .

Alternating-direction method of multipliers

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces.

Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let $(\hat{x}, \hat{y}, \hat{v}) \in \mathcal{H} \times \mathcal{G}^2$.

Assume that $\text{int}(\text{dom } g) \cap L(\text{dom } f) \neq \emptyset$ or $\text{dom } g \cap \text{int}(L(\text{dom } f)) \neq \emptyset$.

$(\hat{x}, \hat{y}, \hat{z})$ is a saddle point of the augmented Lagrange function



$(\hat{x}, \gamma \hat{z})$ is a Kuhn-Tucker point and $\hat{y} = L\hat{x}$.

Alternating-direction method of multipliers

Proof ($\Rightarrow$): If $(\hat{x}, \hat{y}, \hat{z})$ is a saddle point of $\tilde{\mathcal{L}}$, then

$$\begin{cases} 0 \in \partial_x \tilde{\mathcal{L}}(\hat{x}, \hat{y}, \hat{z}) = \partial f(\hat{x}) + \gamma L^* \hat{z} + \gamma L^*(L\hat{x} - \hat{y}) \\ 0 \in \partial_y \tilde{\mathcal{L}}(\hat{x}, \hat{y}, \hat{z}) = \partial g(\hat{y}) - \gamma \hat{z} + \gamma(\hat{y} - L\hat{x}) \\ 0 = \nabla_z \tilde{\mathcal{L}}(\hat{x}, \hat{y}, \hat{z}) = \gamma(L\hat{x} - \hat{y}) \end{cases}$$

$$\Leftrightarrow \begin{cases} 0 \in \partial_x \mathcal{L}(\hat{x}, \hat{y}, \gamma \hat{z}) \\ 0 \in \partial_y \mathcal{L}(\hat{x}, \hat{y}, \gamma \hat{z}) \\ 0 = \nabla_v \mathcal{L}(\hat{x}, \hat{y}, \gamma \hat{z}). \end{cases}$$

Alternating-direction method of multipliers

Proof ($\Leftarrow$): Conversely, if $(\hat{x}, \gamma\hat{z})$ is a Kuhn-Tucker point and $\hat{y} = L\hat{x}$, then it is a saddle point of $\mathcal{L}$. In addition,

$$(\forall (x, y, z) \in \mathcal{H} \times \mathcal{G}^2) \quad \tilde{\mathcal{L}}(x, y, z) = \mathcal{L}(x, y, \gamma z) + \frac{\gamma}{2} \|Lx - y\|^2.$$

It can be deduced that

$$\begin{aligned} \tilde{\mathcal{L}}(\hat{x}, \hat{y}, z) &= \mathcal{L}(\hat{x}, \hat{y}, \gamma z) \leq \mathcal{L}(\hat{x}, \hat{y}, \gamma\hat{z}) = \tilde{\mathcal{L}}(\hat{x}, \hat{y}, \hat{z}) \\ &\leq \mathcal{L}(x, y, \gamma\hat{z}) \leq \tilde{\mathcal{L}}(x, y, \hat{z}). \end{aligned}$$

Alternating-direction method of multipliers

Algorithm for finding a saddle point:

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n = \operatorname{argmin}_{x \in \mathcal{H}} \tilde{\mathcal{L}}(x, y_n, z_n) \\ y_{n+1} = \operatorname{argmin}_{y \in \mathcal{G}} \tilde{\mathcal{L}}(x_n, y, z_n) \\ z_{n+1} \text{ such that } \tilde{\mathcal{L}}(x_n, y_{n+1}, z_{n+1}) \geq \tilde{\mathcal{L}}(x_n, y_{n+1}, z_n). \end{cases}$$

By performing a gradient ascent on the Lagrange multiplier,

$$\begin{aligned} (\forall n \in \mathbb{N}) \quad & \begin{cases} x_n = \operatorname{argmin}_{x \in \mathcal{H}} f(x) + \gamma \langle z_n | Lx - y_n \rangle + \frac{\gamma}{2} \|Lx - y_n\|^2 \\ y_{n+1} = \operatorname{argmin}_{y \in \mathcal{G}} g(y) + \gamma \langle z_n | Lx_n - y \rangle + \frac{\gamma}{2} \|Lx_n - y\|^2 \\ z_{n+1} = z_n + \frac{1}{\gamma} \nabla_z \tilde{\mathcal{L}}(x_n, y_{n+1}, z_n) \end{cases} \\ \Leftrightarrow (\forall n \in \mathbb{N}) \quad & \begin{cases} x_n = \operatorname{argmin}_{x \in \mathcal{H}} \frac{1}{2} \|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x) \\ y_{n+1} = \operatorname{prox}_{\frac{\gamma}{2}}(z_n + Lx_n) \\ z_{n+1} = z_n + Lx_n - y_{n+1}. \end{cases} \end{aligned}$$

Unrelaxed Douglas-Rachford algorithm

Let $\mathcal{H}$ be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.

Let $\gamma \in]0, +\infty[$.

We assume that $\text{zer}(\partial f + \partial g) \neq \emptyset$. Let $u_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} v_n = \text{prox}_{\gamma g} u_n \\ w_n = \text{prox}_{\gamma f}(2v_n - u_n) \\ u_{n+1} = u_n + w_n - v_n. \end{cases}$$

The following properties are satisfied:

1. $u_n \rightharpoonup \hat{u}$
2. $v_n \rightharpoonup \hat{v} = \text{prox}_{\gamma g} \hat{u} \in \text{Argmin}(f + g)$.

Dual form of unrelaxed Douglas-Rachford algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be Hilbert spaces.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$.

Let $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$. Let $\gamma \in]0, +\infty[$.

We assume that $\text{zer}(\partial(f^* \circ (-L^*)) + \partial g^*) \neq \emptyset$. Let $u_0 \in \mathcal{G}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} v_n = \text{prox}_{\gamma g^*} u_n \\ w_n = \text{prox}_{\gamma f^* \circ (-L^*)}(2v_n - u_n) \\ u_{n+1} = u_n + w_n - v_n. \end{cases}$$

The following properties are satisfied:

1. $u_n \rightharpoonup \hat{u}$
2. $v_n \rightharpoonup \hat{v} = \text{prox}_{\gamma g^*} \hat{u} \in \text{Argmin}(f^* \circ (-L^*) + g^*)$.

Equivalence between DR and ADMM

Let $\mathcal{H}$ et $\mathcal{G}$ be two Hilbert spaces. Let $f \in \Gamma_0(\mathcal{H})$ et $g \in \Gamma_0(\mathcal{G})$.

Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ be such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.

The dual form of the unrelaxed Douglas-Rachford algorithm

$$(\forall n \in \mathbb{N}) \quad \begin{cases} v_n = \text{prox}_{\gamma g^*} u_n \\ w_n = \text{prox}_{\gamma f^* \circ (-L^*)} (2v_n - u_n) \\ u_{n+1} = u_n + w_n - v_n \end{cases}$$

can be reexpressed as

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n = \underset{x \in \mathcal{H}}{\text{argmin}} \quad \frac{1}{2} \|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x) \\ s_n = Lx_n \\ y_{n+1} = \text{prox}_{\frac{g}{\gamma}} (z_n + s_n) \\ z_{n+1} = z_n + s_n - y_{n+1} \end{cases}$$

by setting $y_n = \gamma^{-1}(u_n - v_n)$ and $z_n = \gamma^{-1}v_n$.

Equivalence between DR and ADMM

Proof: We have

$$w_n = \text{prox}_{\gamma f^* \circ (-L^*)}(2v_n - u_n) \Leftrightarrow 2v_n - u_n - w_n \in \gamma \partial(f^* \circ (-L^*))(w_n)$$

L^*L isomorphism $\Rightarrow \text{ran } L^* = \mathcal{H}$ and

$\text{dom } f^* \cap \text{int}(\text{ran } L^*) = \text{dom } f^* \neq \emptyset$.

D'où $\partial(f^* \circ (-L^*)) = -L \circ \partial f^* \circ (-L^*)$ and we can define x_n such that

$$\begin{cases} 2v_n - u_n - w_n = -\gamma Lx_n \Leftrightarrow w_n = 2v_n - u_n + \gamma Lx_n \\ x_n \in \partial f^*(-L^*w_n) \Leftrightarrow -L^*w_n \in \partial f(x_n) \end{cases}$$

$$\Rightarrow L^*(u_n - 2v_n - \gamma Lx_n) \in \partial f(x_n)$$

Since $y_n = (u_n - v_n)/\gamma$ et $z_n = v_n/\gamma$,

$$L^*(y_n - z_n - Lx_n) \in \frac{1}{\gamma} \partial f(x_n) \Leftrightarrow 0 \in L^*(Lx_n - y_n + z_n) + \frac{1}{\gamma} \partial f(x_n)$$

$$\Leftrightarrow x_n = \underset{x \in \mathcal{H}}{\text{argmin}} \quad \frac{1}{2} \|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x).$$

Equivalence between DR and ADMM

On the other hand,

$$\begin{aligned} v_n = \text{prox}_{\gamma g^*} u_n &\Leftrightarrow v_n = u_n - \gamma \text{prox}_{\frac{g}{\gamma}} \left(\frac{u_n}{\gamma} \right) \\ &\Leftrightarrow y_n = \frac{u_n - v_n}{\gamma} = \text{prox}_{\frac{g}{\gamma}} \left(\frac{u_n}{\gamma} \right). \end{aligned}$$

Futhermore,

$$\begin{aligned} &\begin{cases} u_{n+1} = u_n + w_n - v_n \\ 2v_n - u_n - w_n = -\gamma Lx_n \Leftrightarrow u_n + w_n = 2v_n + \gamma Lx_n \end{cases} \\ \Rightarrow \frac{u_{n+1}}{\gamma} &= \frac{v_n}{\gamma} + Lx_n = z_n + Lx_n. \end{aligned}$$

Hence

$$y_{n+1} = \text{prox}_{\frac{g}{\gamma}} \left(\frac{u_{n+1}}{\gamma} \right) = \text{prox}_{\frac{g}{\gamma}} (z_n + Lx_n).$$

Finally

$$z_{n+1} = \frac{v_{n+1}}{\gamma} = \frac{u_{n+1}}{\gamma} - y_{n+1} = z_n + Lx_n - y_{n+1}.$$

Augmented Lagrangian method

ADMM (*Alternating-direction method of multipliers*)

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$.

Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.

We assume that $\text{int}(\text{dom } g) \cap L(\text{dom } f) \neq \emptyset$ or $\text{dom } g \cap \text{int}(L(\text{dom } f)) \neq \emptyset$, and that $\text{Argmin}(f + g \circ L) \neq \emptyset$. Let $u_0 \in \mathcal{G}$, $v_0 = \text{prox}_{\gamma g^*} u_0$, $y_0 = \gamma^{-1}(u_0 - v_0)$, $z_0 = \gamma^{-1}v_0$, and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n = \underset{x \in \mathcal{H}}{\text{argmin}} \quad \frac{1}{2} \|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x) \\ s_n = Lx_n \\ y_{n+1} = \text{prox}_{\frac{g}{\gamma}}(z_n + s_n) \\ z_{n+1} = z_n + s_n - y_{n+1}. \end{cases}$$

We have:

- ▶ $x_n \rightarrow \hat{x}$ where $\hat{x} \in \text{Argmin}(f + g \circ L)$
- ▶ $\gamma z_n \rightarrow \hat{v}$ where $\hat{v} \in \text{Argmin}(f^* \circ (-L^*) + g^*)$.

Recall

Lemma

Let $\mathcal{H}$ be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$.

If $(x_n, v_n)_{n \in \mathbb{N}}$ is a sequence of $\mathcal{H}^2$ such that $v_n \in \partial f(x_n)$, $x_n \rightarrow \hat{x}$ and $v_n \rightharpoonup \hat{v}$ then $\hat{v} \in \partial f(\hat{x})$.

Augmented Lagrangian method

Proof: For simplicity, we consider the finite dimensional case.

$$\begin{aligned} \text{int}(\text{dom } g) \cap L(\text{dom } f) \neq \emptyset \text{ or } \text{dom } g \cap \text{int}(L(\text{dom } f)) \neq \emptyset \\ \Rightarrow \partial(f + g \circ L) = \partial f + L^* \partial g L. \end{aligned}$$

We deduce that

$$\text{zer}(\partial f + L^* \partial g L) = \text{zer } \partial(f + g \circ L) = \text{Argmin}(f + g \circ L) \neq \emptyset$$

$$\begin{aligned} \text{In addition, } \text{zer}(\partial f + L^* \partial g L) &\neq \emptyset \\ \Rightarrow \text{zer}((-L) \partial f^*(-L^*) + \partial g^*) &\neq \emptyset. \\ \Rightarrow \text{zer}(\partial(f^* \circ (-L^*)) + \partial g^*) &\supset \text{zer}((-L) \partial f^*(-L^*) + \partial g^*) \neq \emptyset. \end{aligned}$$

Augmented Lagrangian method

Proof: We deduce from the convergence properties of the associated dual form of the DR algorithm that

$$z_n = \gamma^{-1} v_n \rightarrow \gamma^{-1} \hat{v}, \quad y_n = \gamma^{-1}(u_n - v_n) \rightarrow \gamma^{-1}(\hat{u} - \hat{v}) = \hat{y} \text{ and}$$

$$s_n = z_{n+1} - z_n + y_{n+1} \rightarrow \hat{y}$$

where $\hat{v} \in \text{Argmin}(f^* \circ (-L^*) + g^*)$ and $\hat{v} = \text{prox}_{\gamma g^*} \hat{u}$.

In addition

$$s_n = Lx_n \Rightarrow x_n = (L^*L)^{-1}L^*s_n \rightarrow (L^*L)^{-1}L^*\hat{y} = \hat{x} \text{ and } L\hat{x} = \hat{y}.$$

$$\text{On the one hand, } \hat{v} = \text{prox}_{\gamma g^*} \hat{u} \Leftrightarrow \gamma \hat{y} = \hat{u} - \hat{v} \in \gamma \partial g^*(\hat{v})$$

$$\Leftrightarrow L\hat{x} \in \partial g^*(\hat{v}).$$

$$\text{On the other hand, } x_n \in \text{Argmin}\left(\frac{1}{2} \|L \cdot - y_n + z_n\|^2 + \frac{1}{\gamma} f\right)$$

$$\Leftrightarrow 0 \in L^*(Lx_n - y_n + z_n) + \gamma^{-1} \partial f(x_n)$$

$$\Leftrightarrow L^*(y_n - z_n - s_n) \in \gamma^{-1} \partial f(x_n).$$

$$\text{We have } x_n \rightarrow \hat{x} \text{ and } L^*(y_n - z_n - s_n) \rightarrow -\gamma^{-1} L^* \hat{v} \Rightarrow -L^* \hat{v} \in \partial f(\hat{x}).$$

In conclusion, $L\hat{x} \in \partial g^*(\hat{y})$ and $-L^*\hat{v} \in \partial f(\hat{x})$, that is

$(\hat{x}, \hat{v})$ is a pair of primal/dual solutions.

Augmented Lagrangian method

ADMM (*Alternating-direction method of multipliers*)

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$.

Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ such that L^*L is an isomorphism and let $\gamma \in]0, +\infty[$.

We assume that $\text{int}(\text{dom } g) \cap L(\text{dom } f) \neq \emptyset$ or $\text{dom } g \cap \text{int}(L(\text{dom } f)) \neq \emptyset$, and that

$\text{Argmin}(f + g \circ L) \neq \emptyset$. Let $(y_0, z_0) \in \mathcal{G}^2$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n = \underset{x \in \mathcal{H}}{\text{argmin}} \quad \frac{1}{2} \|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} f(x) \\ s_n = Lx_n \\ y_{n+1} = \text{prox}_{\frac{g}{\gamma}}(z_n + s_n) \\ z_{n+1} = z_n + s_n - y_{n+1}. \end{cases}$$

We have:

- ▶ $x_n \rightarrow \hat{x}$ where $\hat{x} \in \text{Argmin}(f + g \circ L)$
- ▶ $\gamma z_n \rightarrow \hat{v}$ where $\hat{v} \in \text{Argmin}(f^* \circ (-L^*) + g^*)$.

Primal-dual methods

Problem

Let $\mathcal{H}$ and $\mathcal{G}$ be two real Hilbert spaces, $f \in \Gamma_0(\mathcal{H})$, $h \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

It is assumed that h is differentiable and has a β -Lipschitzian gradient with $\beta \in]0, +\infty[$.

We want to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x) + g(Lx) + h(x).$$

Primal-dual methods

- ▶ ADMM algorithm: Let $\gamma \in]0, +\infty[$.

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_n = \operatorname{argmin}_{x \in \mathcal{H}} \frac{1}{2} \|Lx - y_n + z_n\|^2 + \frac{1}{\gamma} (f(x) + h(x)) \\ y_{n+1} = \operatorname{prox}_{\frac{g}{\gamma}} (z_n + Lx_n) \\ z_{n+1} = z_n + Lx_n - y_{n+1}. \end{cases}$$

- ▶ Limitations:
 - ▶ Computation of x_n at iteration $n \in \mathbb{N}$ may be complicated.
 - ▶ Convergence requires L^*L to be invertible.
 - ▶ The smoothness of h is not exploited.

Primal-dual methods

- Idea 1: the optimization problem is reformulated as finding

$$\inf_{x \in \mathcal{H}} f(x) + h(x) + \underbrace{\sup_{v \in \mathcal{G}} \langle v \mid Lx \rangle - g^*(v)}_{g(Lx)}.$$

Primal-dual methods

- Idea 1: the optimization problem is reformulated as finding

$$\inf_{x \in \mathcal{H}} \sup_{v \in \mathcal{G}} f(x) + h(x) + \langle v \mid Lx \rangle - g^*(v).$$

- Arrow-Hurwicz method: Let $(\tau_n)_{n \in \mathbb{N}}$ and $(\sigma_n)_{n \in \mathbb{N}}$ be sequences in $]0, +\infty[$.

$$(\forall n \in \mathbb{N}) \quad \begin{cases} t_n \in \partial f(x_n) \\ x_{n+1} = x_n - \tau_n(t_n + \nabla h(x_n) + L^* v_n) \\ s_n \in \partial g^*(v_n) \\ v_{n+1} = v_n - \sigma_n(s_n - Lx_{n+1}) \end{cases}$$

$\rightsquigarrow$ requires stringent conditions on the choice of the step-size (e.g. decaying to zero)

Primal-dual methods

- Idea 2: Use implicit updates

$$\begin{aligned}
 (\forall n \in \mathbb{N}) \quad & \begin{cases} t_n \in \partial f(x_{n+1}) \\ x_{n+1} = x_n - \tau_n(t_n + \nabla h(x_n) + L^* v_n) \\ s_n \in \partial g^*(v_{n+1}) \\ v_{n+1} = v_n - \sigma_n(s_n - Lx_{n+1}) \end{cases} \\
 \Leftrightarrow & \begin{cases} 0 \in x_{n+1} - x_n + \tau_n(\nabla h(x_n) + L^* v_n) + \tau_n \partial f(x_{n+1}) \\ 0 \in v_{n+1} - v_n - \sigma_n Lx_{n+1} + \sigma_n \partial g^*(v_{n+1}) \end{cases} \\
 \Leftrightarrow & \begin{cases} x_{n+1} = \text{prox}_{\tau_n f}(x_n - \tau_n(\nabla h(x_n) + L^* v_n)) \\ v_{n+1} = \text{prox}_{\sigma_n g^*}(v_n + \sigma_n Lx_{n+1}) \end{cases}
 \end{aligned}$$

$\rightsquigarrow$ still does not converge for constant values of the step-size.

Primal-dual methods

- Idea 3: Use the approximation $x_{n+1} \simeq 2x_{n+1} - x_n$

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \text{prox}_{\tau_n f}(x_n - \tau_n(\nabla h(x_n) + L^* v_n)) \\ v_{n+1} = \text{prox}_{\sigma_n g^*}(v_n + \sigma_n L(2x_{n+1} - x_n)). \end{cases}$$

Primal-dual optimization algorithm

Convergence of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$. Let $h \in \Gamma_0(\mathcal{H})$ have a β -Lipschitzian gradient.

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \text{prox}_{\tau f}(x_n - \tau(\nabla h(x_n) + L^* v_n)) \\ v_{n+1} = \text{prox}_{\sigma g^*}(v_n + \sigma L(2x_{n+1} - x_n)). \end{cases}$$

Primal-dual optimization algorithm

Convergence of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$. Let $h \in \Gamma_0(\mathcal{H})$ have a β -Lipschitzian gradient.

Let $\tau \in]0, +\infty[$ and $\sigma \in]0, +\infty[$.

We assume that $\tau^{-1} - \sigma\|L\|^2 > \beta/2$ and $\text{zer}(\partial f + \nabla h + L^* \partial g L) \neq \emptyset$.

Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \text{prox}_{\tau f}(x_n - \tau(\nabla h(x_n) + L^* v_n)) \\ v_{n+1} = \text{prox}_{\sigma g^*}(v_n + \sigma L(2x_{n+1} - x_n)). \end{cases}$$

We have:

- ▶ $x_n \rightharpoonup \hat{x} \in \text{Argmin}(f + h + g \circ L)$
- ▶ $v_n \rightharpoonup \hat{v} \in \text{Argmin}((f + h)^* \circ (-L^*) + g^*)$.

Primal-dual optimization algorithm

Convergence of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$. Let $h \in \Gamma_0(\mathcal{H})$ have a β -Lipschitzian gradient.

Let $\tau \in]0, +\infty[$ and $\sigma \in]0, +\infty[$.

We assume that $\tau^{-1} - \sigma\|L\|^2 > \beta/2$ and $\text{zer}(\partial f + \nabla h + L^* \partial g L) \neq \emptyset$.

Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \text{prox}_{\tau f}(x_n - \tau(\nabla h(x_n) + L^* v_n)) \\ v_{n+1} = \text{prox}_{\sigma g^*}(v_n + \sigma L(2x_{n+1} - x_n)). \end{cases}$$

We have:

- ▶ $x_n \rightharpoonup \hat{x} \in \text{Argmin}(f + h + g \circ L)$
- ▶ $v_n \rightharpoonup \hat{v} \in \text{Argmin}((f + h)^* \circ (-L^*) + g^*)$.

Sketch of the proof: rewrite the algorithm as a Forward-Backward iteration for solving a saddle point problem in the space $\mathcal{H} \times \mathcal{G}$.

Primal-dual optimization algorithm

Convergence of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$. Let $h \in \Gamma_0(\mathcal{H})$ have a β -Lipschitzian gradient.

Let $\tau \in]0, +\infty[$ and $\sigma \in]0, +\infty[$.

We assume that $\tau^{-1} - \sigma\|L\|^2 > \beta/2$ and $\text{zer}(\partial f + \nabla h + L^* \partial g L) \neq \emptyset$.

Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \text{prox}_{\tau f}(x_n - \tau(\nabla h(x_n) + L^* v_n)) \\ v_{n+1} = \text{prox}_{\sigma g^*}(v_n + \sigma L(2x_{n+1} - x_n)). \end{cases}$$

We have:

- ▶ $x_n \rightharpoonup \hat{x} \in \text{Argmin}(f + h + g \circ L)$
- ▶ $v_n \rightharpoonup \hat{v} \in \text{Argmin}((f + h)^* \circ (-L^*) + g^*)$.

Remark: when $g = 0$, we recover the Forward-Backward algorithm.

Relaxed form of PD algorithm

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces. Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$.

Let $h \in \Gamma_0(\mathcal{H})$ have a β -Lipschitzian gradient with $\beta \in]0, +\infty[$

Let $\tau \in]0, +\infty[$ and $\sigma \in]0, +\infty[$.

We assume that $\tau^{-1} - \sigma\|L\|^2 \geq \beta/2$. and $\text{zer}(\partial f + \nabla h + L^*\partial g L) \neq \emptyset$.

Let $\delta = 2 - \beta(\tau^{-1} - \sigma\|L\|^2)^{-1}/2 \in [1, 2]$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n \in \mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

Let $x_0 \in \mathcal{H}$, $v_0 \in \mathcal{G}$, and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} p_n = \text{prox}_{\tau f}(x_n - \tau(\nabla h(x_n) + L^*v_n)) \\ q_n = \text{prox}_{\sigma g^*}(v_n + \sigma L(2p_n - x_n)) \\ (x_{n+1}, v_{n+1}) = (x_n, v_n) + \lambda_n(p_n - x_n, q_n - v_n). \end{cases}$$

Then, $(x_n)_{n \in \mathbb{N}}$ converges weakly to a solution to a minimize of $f + h + g \circ L$.

Exercise

Let $\mathcal{H}$ and $(\mathcal{G}_i)_{1 \leq i \leq m}$ be real Hilbert spaces. Let $f \in \Gamma_0(\mathcal{H})$, let $h \in \Gamma_0(\mathcal{H})$, and, for every $i \in \{1, \dots, m\}$, let $g_i \in \Gamma_0(\mathcal{G}_i)$ and $L_i \in \mathcal{B}(\mathcal{H}, \mathcal{G}_i)$.

It is assumed that h is differentiable and has a β -Lipschitzian gradient with $\beta \in]0, +\infty[$. Propose a primal-dual algorithm to solve

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x) + \sum_{i=1}^m g_i(L_i x) + h(x).$$