Signal Processing and Networks Optimization
Part V: Fixed point algorithms

Pierre Borgnat[1], Jean-Christophe Pesquet[2], Nelly Pustelnik[1]

[1] ENS Lyon – Laboratoire de Physique – CNRS UMR 5672
pierre.borgnat@ens-lyon.fr, nelly.pustelnik@ens-lyon.fr

[2] LIGM – Univ. Paris-Est – CNRS UMR 8049
jean-christophe.pesquet@univ-paris-est.fr

Naive answer

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

If

- $\widehat{x}$ is a fixed point of T, i.e. $\widehat{x} = T\widehat{x}$
- T is a strict contraction, i.e. there exists $\rho \in [0,1[$ such that

$$\big(\forall (x,x') \in \mathbb{R}^N \times \mathbb{R}^N\big) \qquad \|Tx - Tx'\| \leq \rho \|x - x'\|$$

then $(x_n)_{n\in\mathbb{N}}$ converges to $\widehat{x}$.

Naive answer

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

If

- $\widehat{x}$ is a fixed point of T, i.e. $\widehat{x} = T\widehat{x}$
- T is a strict contraction, i.e. there exists $\rho \in [0,1[$ such that

$$\big(\forall (x,x') \in \mathbb{R}^N \times \mathbb{R}^N\big) \qquad \|Tx - Tx'\| \leq \rho \|x - x'\|$$

then $(x_n)_{n\in\mathbb{N}}$ converges to $\widehat{x}$.

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

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

Consequently, $\|x_n - \widehat{x}\| \leq \rho^n \|x_0 - \widehat{x}\|$. Hence, we have proved that $(x_n)_{n\in\mathbb{N}}$ converges linearly to $\widehat{x}$.

Objective of this part

- 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.
  $\rightsquigarrow$ How to relate T to the objective function f ?

Fixed point algorithm

Fixed point algorithm: convergence

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

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

$$\lim_{n\to+\infty} \|x_n - \widehat{x}\| = 0.$$

  It is denoted by $x_n \to \widehat{x}$.

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

$$(\forall y \in \mathcal{H}) \qquad \lim_{n\to+\infty} \langle y \mid x_n - \widehat{x}\rangle = 0.$$

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

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

Fixed point algorithm: convergence

Let $\mathcal{H}$ be a Hilbert space.
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
- $(x_n)_{n\in\mathbb{N}}$ posseses at most one sequential cluster point in the weak topology.

- $\widehat{x}$ 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 $\widehat{x}$.

Fixed point algorithm: convergence

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

- $(x_n)_{n\in\mathbb{N}}$ posseses at most one sequential cluster point in the weak topology.

Illustration:

x_0	x_1	x_2	x_3	x_4	x_5	...
1	-1	1	-1	1	-1	...

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

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

Fixed point algorithm: convergence

Lemma 1

Let $\mathcal{H}$ be a Hilbert space and $D \subset \mathcal{H}$ nonempty.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in D.
$(x_n)_{n\in\mathbb{N}}$ 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.

Fixed point algorithm: convergence

<u>Proof</u>:

If $(\|x_n - x\|)_{n\in\mathbb{N}}$ 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 $x_{n_k} \rightharpoonup \widehat{x}$ and $x_{n_\ell} \rightharpoonup \widehat{x}'$ where $(\widehat{x}, \widehat{x}') \in D^2$. For every $n \in \mathbb{N}$,

$$2\left\langle x_n \mid \widehat{x}' - \widehat{x}\right\rangle = \|x_n - \widehat{x}\|^2 - \|x_n - \widehat{x}'\|^2 - \|\widehat{x}\|^2 + \|\widehat{x}'\|^2.$$

Because $(\|x_n - \widehat{x}\|)_{n\in\mathbb{N}}$ and $(\|x_n - \widehat{x}'\|)_{n\in\mathbb{N}}$ converge, there exists $\alpha \in \mathbb{R}$ such that $\langle x_n \mid \widehat{x}' - \widehat{x}\rangle \to \alpha$ and thus
$\langle x_{n_k} \mid \widehat{x}' - \widehat{x}\rangle \to \langle \widehat{x} \mid \widehat{x}' - \widehat{x}\rangle = \alpha$. Similarly, $\langle \widehat{x}' \mid \widehat{x}' - \widehat{x}\rangle = \alpha$.
Consequently, $\|\widehat{x}' - \widehat{x}\|^2 = 0 \Rightarrow \widehat{x} = \widehat{x}'$.

Fixed point algorithm: Fejér-monotone sequence

Let $\mathcal{H}$ be a Hilbert space and D be a nonempty subset of $\mathcal{H}$.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ is **Fejér-monotone** with respect to D if

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

Let $\mathcal{H}$ be a Hilbert space and D be a nonempty subset of $\mathcal{H}$.
Let $(x_n)_{n\in\mathbb{N}}$ be Fejér-monotone with respect to D then

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

Fixed point algorithm: Fejér-monotone sequence

Fejér-monotone convergence

Let $\mathcal{H}$ be a Hilbert space and let D be a nonempty subset of $\mathcal{H}$.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$.
$(x_n)_{n\in\mathbb{N}}$ 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.

Fixed point algorithm: Fejér-monotone sequence

Lemma 2

Let $\mathcal{H}$ be a Hilbert space and 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 $\widehat{x}$ then $\widehat{x} \in C$.

Fixed point algorithm: Fejér-monotone sequence

Lemma 2

Let $\mathcal{H}$ be a Hilbert space and 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 $\widehat{x}$ then $\widehat{x} \in C$.

Proof:

We have $\widehat{x} - P_C\widehat{x} \in N_C\big(P_C\widehat{x}\big)$.

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

$$\langle x_n - P_C\widehat{x} \mid \widehat{x} - P_C\widehat{x}\rangle \leq 0.$$

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

Nonexpansive operator: definition

Let $\mathcal{H}$ be a Hilbert space. Let $T : \mathcal{H} \to 2^{\mathcal{H}}$.
The set of fixed points of B is

$$\operatorname{Fix} T = \{x \in \mathcal{H} \mid x \in Tx\}$$

Let $\mathcal{H}$ be a Hilbert space and let $C \subset \mathcal{H}$ be a nonempty closed convex set.
Let $T : C \to \mathcal{H}$.
T is a nonexpansive operator if $\big(\forall (x, y) \in C^2\big) \quad \|Tx - Ty\| \leq \|x - y\|$.

Nonexpansive operator: fixed point algorithm

Demiclosedness principle

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$.
Let $T\colon C \to \mathcal{H}$ be a nonexpansive operator.
If $(x_n)_{n\in\mathbb{N}}$ is a sequence in C that converges weakly to $\widehat{x}$ and if $Tx_n - x_n \to 0$ then $\widehat{x} \in \operatorname{Fix} T$.

Nonexpansive operator: fixed point algorithm

Demiclosedness principle

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$. Let $T\colon C \to \mathcal{H}$ be a nonexpansive operator.
If $(x_n)_{n\in\mathbb{N}}$ is a sequence in C that converges weakly to $\widehat{x}$ and if $Tx_n - x_n \to 0$ then $\widehat{x} \in \operatorname{Fix} T$.

Proof:

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

$$\|x_n - T\widehat{x}\|^2 = \|x_n - \widehat{x}\|^2 + \|\widehat{x} - T\widehat{x}\|^2 + 2\,\langle x_n - \widehat{x} \mid \widehat{x} - T\widehat{x}\rangle$$

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

$$\Rightarrow \|\widehat{x} - T\widehat{x}\|^2 = \|x_n - Tx_n\|^2 + \|Tx_n - T\widehat{x}\|^2 - \|x_n - \widehat{x}\|^2 + 2\,\langle x_n - Tx_n \mid Tx_n - T\widehat{x}\rangle - 2\,\langle x_n - \widehat{x} \mid \widehat{x} - T\widehat{x}\rangle$$

Nonexpansive operator: fixed point algorithm

Demiclosedness principle

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$. Let $T\colon C \to \mathcal{H}$ be a nonexpansive operator.
If $(x_n)_{n\in\mathbb{N}}$ is a sequence in C that converges weakly to $\widehat{x}$ and if $Tx_n - x_n \to 0$ then $\widehat{x} \in \operatorname{Fix} T$.

Proof:

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

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

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

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

Nonexpansive operator: fixed point algorithm

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty subset of $\mathcal{H}$.
Let $T\colon C \to C$ be a nonexpansive operator

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

Nonexpansive operator: fixed point algorithm

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty subset of $\mathcal{H}$.
Let $T\colon C \to C$ be a nonexpansive operator

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

If $x_n - Tx_n \to 0$,

Nonexpansive operator: fixed point algorithm

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty subset of $\mathcal{H}$.
Let $T\colon C \to C$ be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$.
Let $x_0 \in C$,

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

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

Proof :
For every $n \in \mathbb{N}$ and $y \in \operatorname{Fix} T$, $\|x_{n+1} - y\| \leq \|Tx_n - Ty\| \leq \|x_n - y\|$.
$(x_n)_{n\in\mathbb{N}}$ is Fejér-monotone with respect to $\operatorname{Fix} T$.
Let $(x_{n_k})_{k\in\mathbb{N}}$ be a subsequence of $(x_n)_{n\in\mathbb{N}}$ such that $x_{n_k} \rightharpoonup \widehat{x}$ where $\widehat{x} \in \mathcal{H}$.
By assumption $x_{n_k} - Tx_{n_k} \to 0$ and thus, according to the demiclosedness principle, $\widehat{x} \in \operatorname{Fix} T$.
This shows the weak convergence of $(x_n)_{n\in\mathbb{N}}$.

Fixed point algorithm: Fejér-monotone sequence

Krasnosel'skii-Mann algorithm

Let $\mathcal{H}$ be a Hilbert space and C be a nonempty closed convex subset of $\mathcal{H}$.
Let $T\colon C \to C$ be a nonexpansive operator such that $\operatorname{Fix} T \neq \varnothing$.
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 $\operatorname{Fix} T$.
- $(Tx_n - x_n)_{n\in\mathbb{N}}$ converges strongly to 0.
- $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.

Fixed point algorithm: Fejér-monotone sequence

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

Fejr-monotony with respect to $\operatorname{Fix} T$: $(\forall x \in \operatorname{Fix} T)(\forall n \in \mathbb{N})$

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

Fixed point algorithm: Fejér-monotone sequence

Proof :

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

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

$$\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n)\|Tx_n - x_n\|^2 \leq \|x_0 - x\|^2$$

$$\Rightarrow (\forall n \in \mathbb{N}) \qquad \inf_{k\geq n} \|Tx_k - x_k\|^2 \sum_{k=n}^{+\infty} \lambda_k(1-\lambda_k) \to 0.$$

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

Fixed point algorithm: Fejér-monotone sequence

Proof :
The assumptions on $(\lambda_n)_{n\in\mathbb{N}}$ lead to $\liminf_{n\to+\infty} \|Tx_n - x_n\| = 0$.
Moreover, as T is a nonexpansive operator

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

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

$$Tx_n - x_n \to 0.$$

Fixed point algorithm: Fejér-monotone sequence

Proof :
Let $(x_{n_k})_{k\in\mathbb{N}}$ be a subsequence of $(x_n)_{n\in\mathbb{N}}$ such that $x_{n_k} \rightharpoonup \widehat{x}$. Considering the demiclosedness principle, $Tx_{n_k} - x_{n_k} \to 0$, leads to $\widehat{x} \in \operatorname{Fix} T$. The weak convergence of $(x_n)_{n\in\mathbb{N}}$ to $\widehat{x}$ results from the Fejr-monotony of $(x_n)_{n\in\mathbb{N}}$ with respect to $\operatorname{Fix} T$.

α-averaged operator: definition

Let $\mathcal{H}$ be Hilbert space and let $C \subset \mathcal{H}$ nonempty closed convex set.
Let $A : C \to \mathcal{H}$ and let $\alpha \in]0,1[$.
A is a **α-averaged** operator if there exists a nonexpansive operator $R\colon C \to \mathcal{H}$ such that

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

Let $\mathcal{H}$ be Hilbert space and let $C \subset \mathcal{H}$ nonempty closed convex set.
Let $A : C \to \mathcal{H}$and let $\alpha \in]0,1[$.
A is a **α-averaged** operator if

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

α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$, $\nu \in]0, 2[$.
If f is differentiable with a ν-lipschitzian gradient then, $\mathrm{Id} - \nabla f$ is a $\nu/2$-averaged operator.

Remark : $\mathrm{Id} - \nabla f$ denotes the gradient descent operator.

α-averaged operator: example

Proof : 1) Descent lemma

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

$$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt$$

$$\Leftrightarrow \; f(y) - f(x) - \langle y - x \mid \nabla f(x)\rangle = \int_0^1 \langle y - x \mid \nabla f(x + t(y - x)) - \nabla f(x)\rangle \, dt.$$

But, according to the Cauchy-Schwarz inequality,

$$\langle y - x \mid \nabla f(x + t(y - x)) - \nabla f(x)\rangle$$
$$\leq \|y - x\| \|\nabla f(x + t(y - x)) - \nabla f(x)\| \leq t\nu \|y - x\|^2.$$

This leads to

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

α-averaged operator: example

Proof : 2) $\mathrm{Id} - \nabla f$ is α-averaged

From the descent lemma, for every $(x, y, z) \in \mathcal{H}^3$,

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

Moreover, according to the Fenchel-Young inequality,

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

Thus,

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

α-averaged operator: example

Proof : 2) $\mathrm{Id} - \nabla f$ is α-averaged

Thus,

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

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

This yields

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

Consequently,

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

α-averaged operator: example

Proof : 2) $\mathrm{Id} - \nabla f$ is α-averaged

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

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

It results that

$$\boxed{\|(\mathrm{Id} - \nabla f)x - (\mathrm{Id} - \nabla f)y\|^2 + \frac{1-\nu/2}{\nu/2}\|\nabla f(x) - \nabla f(y)\|^2 \leq \|x - y\|^2.}$$

α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$, $\nu \in \,]0, +\infty[$ and $\gamma \in]0, 2/\nu[$.
If f is differentiable with a ν-lipschitzian gradient then $\mathrm{Id} - \gamma \nabla f$ is a $\gamma\nu/2$-averaged operator.

Remark : $\mathrm{Id} - \gamma \nabla f$ denotes the gradient descent operator.

α-averaged operator: example

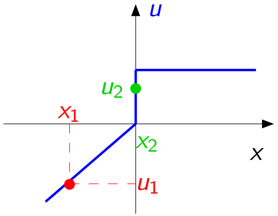
Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$.
prox_f is a $1/2$-averaged operator.

α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$.
prox_f is a $1/2$-averaged operator.

Proof:

- We recall that : $\partial f(x) = \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\}$

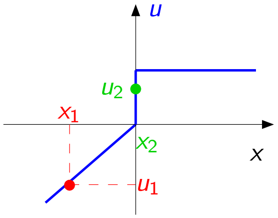


α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$.
prox_f is a $1/2$-averaged operator.

Proof:

- We recall that : $\partial f(x) = \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\}$
- Let $u_1 \in \partial f(x_1)$ and $u_2 \in \partial f(x_2)$.
  By definition:
  $$\langle x_2 - x_1 | u_1 \rangle + f(x_1) \leq f(x_2)$$
  $$\langle x_1 - x_2 | u_2 \rangle + f(x_2) \leq f(x_1)$$
  it results that $\boxed{\langle x_1 - x_2 \,|\, u_1 - u_2 \rangle \geq 0}$.

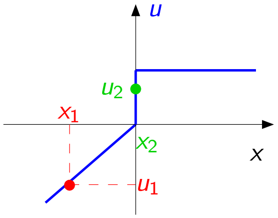


α-averaged operator: example

Let $\mathcal{H}$ be a Hilbert space, $f \in \Gamma_0(\mathcal{H})$.
prox_f is a $1/2$-averaged operator.

<u>Proof</u>:

- We recall that : $\partial f(x) = \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle y - x | u \rangle + f(x) \leq f(y)\}$
- Let $u_1 \in \partial f(x_1)$ and $u_2 \in \partial f(x_2)$.
  By definition:
  $$\langle x_2 - x_1 | u_1 \rangle + f(x_1) \leq f(x_2)$$
  $$\langle x_1 - x_2 | u_2 \rangle + f(x_2) \leq f(x_1)$$
  it results that $\boxed{\langle x_1 - x_2 \,|\, u_1 - u_2 \rangle \geq 0}$.



- *Remark* : ∂f is a monotone operator.

α-averaged operator: example

Proof :

Let $u_1 \in \partial f(x_1)$ and $u_2 \in \partial f(x_2)$

$$\langle x_1 - x_2 \mid u_1 - u_2 \rangle \geq 0 \Leftrightarrow \quad \langle x_1 - x_2 \mid x_1 - x_2 + u_1 - u_2 \rangle \geq \|x_1 - x_2\|^2$$

We consider $u_1' \in (\mathrm{Id} + \partial f)x_1$ et $u_2' \in (\mathrm{Id} + \partial f)x_2$, it results that

$$\langle x_1 - x_2 \mid u_1' - u_2' \rangle \geq \|x_1 - x_2\|^2$$

Then, from the definition of the proximity operator,

$$\langle \mathrm{prox}_f u_1' - \mathrm{prox}_f u_2' \mid u_1' - u_2' \rangle \geq \|\mathrm{prox}_f u_1' - \mathrm{prox}_f u_2'\|^2$$

We can deduce that prox_f is a $1/2$-averaged operator, i.e,

$$\| u_1' - u_2' \|^2 \geq \|\mathrm{prox}_f u_1' - \mathrm{prox}_f u_2'\|^2 + \|(\mathrm{Id} - \mathrm{prox}_f)u_1' - (\mathrm{Id} - \mathrm{prox}_f)u_2'\|^2$$

Fixed point algorithm: α-averaged operator

Let $\mathcal{H}$ be a Hilbert space and let $\alpha \in]0,1[$.
Let $T\colon \mathcal{H} \to \mathcal{H}$ be an α-averaged operator with $\alpha \in]0,1[$ such that $\mathrm{Fix}\, T \neq \varnothing$.
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 $\mathrm{Fix}\, T$.
- $(Tx_n - x_n)_{n\in\mathbb{N}}$ converge strongly to 0.
- $(x_n)_{n\in\mathbb{N}}$ converge weakly to a point in $\mathrm{Fix}\, T$.

Fixed point algorithm: α-averaged operator

Proof :
Since T is α-average, there exists a non expansive operator R such that $T = (1-\alpha)\mathrm{Id} + \alpha R$.
Let $(\forall n \in \mathbb{N})$ $\mu_n = \alpha\lambda_n \in [0,1]$.
The iterations can be written as

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

Moreover, $\mathrm{Fix} R = \mathrm{Fix}\, T$.
$+$ Krasnosel'skii-Mann algorithm.

Optimization algorithm: *Forward-Backward*

Let $\mathcal{H}$ be a Hilbert space.
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 $\gamma \in]0, 2/\nu[$ and $\delta = \min\{1, 1/(\nu\gamma)\} + 1/2$.
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 $\mathrm{Argmin}(f + g) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

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

Optimization algorithm: *Forward-Backward*

Proof: Let $T = \mathrm{prox}_{\gamma f} \circ (\mathrm{Id} - \gamma\nabla g)$. For every $x \in \mathcal{H}$,

$$x \in \mathrm{Fix}\, T \Leftrightarrow (\mathrm{Id} - \gamma\nabla g)x \in (\mathrm{Id} + \gamma\partial f)x \Leftrightarrow 0 \in \nabla g(x) + \partial f(x).$$

Consequently, $\mathrm{Fix}\, T = \mathrm{zer}\,(\nabla g + \partial f) \neq \varnothing$. Moreover, for every $n \in \mathbb{N}$,

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

$\mathrm{prox}_{\gamma f}$ is $1/2$-average and $\mathrm{Id} - \gamma\nabla g$ is $\gamma\nu/2$-averaged.
It follows that T is α-averaged with

$$\alpha = \frac{2}{1 + \frac{1}{\max\{\frac{1}{2}, \frac{\gamma\nu}{2}\}}} \quad \Leftrightarrow \quad \alpha^{-1} = \delta.$$

Optimization algorithm: projected gradient

Let $\mathcal{H}$ be a Hilbert space.
Let C 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 $\delta = \min\{1, 1/(\nu\gamma)\} + 1/2$.
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 $\operatorname{Argmin}_{x\in C} g(x) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

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

Optimization algorithm: gradient descent

Let $\mathcal{H}$ be a Hilbert space.
Let $g \in \Gamma_0(\mathcal{H})$ be a differentiable function with a ν-lipschitzian gradient where $\nu \in \,]0, +\infty[$.
Le $\gamma \in]0, 2/\nu[$.
We assume that $\operatorname{Argmin} g \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

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

Optimization algorithm: Douglas-Rachford

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

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

Optimization algorithm: Douglas-Rachford

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 $\operatorname{zer}(\partial f + \partial g) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

The following properties are satisfied:

- $x_n \rightharpoonup \widehat{x}$
- $z_n - y_n \to 0$, $y_n \rightharpoonup \widehat{y}$, $z_n \rightharpoonup \widehat{y}$ where $\widehat{y} = \operatorname{prox}_{\gamma g}\widehat{x} \in \operatorname{Argmin}(f+g)$.

Optimization algorithm: Douglas-Rachford

Proof: follows from Krasnosel'skii-Mann algorithm (skipped)

Optimization algorithm: Douglas-Rachford

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $g \in \Gamma_0(\mathcal{H})$ and $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ such that $\mathrm{ran}\, L$ is closed and L^*L is a isomorphism.
Let $\gamma \in \,]0, +\infty[$ and let $(\lambda_n)_{n\in\mathbb{N}}$ a sequence in $[0, 2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$.
We assume that $\mathrm{zer}\,(L^* \circ \partial g \circ L) \neq \varnothing$. Let $x_0 \in \mathcal{H}$, $v_0 = (L^*L)^{-1}L^*x_0$ et

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

We have then:
$v_n \rightharpoonup \widehat{v}$ where $\widehat{v} \in \mathrm{Argmin}(g \circ L)$.

Optimization algorithm: Douglas-Rachford

Sketch of proof:

$$\underset{v\in\mathcal{G}}{\text{minimize}}\ g(Lv) \ \Leftrightarrow\ \underset{x\in\mathcal{H}}{\text{minimize}}\ \iota_E(x)+g(x)$$

where $E=\operatorname{ran} L$.
We apply Douglas-Rachford algorithm with
$f=\iota_E \Rightarrow \operatorname{prox}_{\gamma f}=P_E$ by setting

$$(\forall n\in\mathbb{N})\qquad P_E y_n = Lc_n \ \text{ and }\ P_E x_n = Lv_n$$

where $c_n=\underset{c\in\mathcal{H}}{\operatorname{argmin}}\ \|y_n-Lc\|^2=(L^*L)^{-1}L^*y_n$.

Optimization algorithm: Douglas-Rachford

Particular case of Douglas-Rachford algorithm:
$\mathcal{H} = \mathcal{H}_1 \times \cdots \mathcal{H}_m$ where $\mathcal{H}_1, \ldots, \mathcal{H}_m$ Hilbert spaces
$\big(\forall x = (x_1, \ldots, x_m) \in \mathcal{H}\big)\ g(x) = \sum_{i=1}^m g_i(x_i)$
where $(\forall i \in \{1, \ldots, m\}\ g_i \in \Gamma_0(\mathcal{H}_i)$
$L\colon v \mapsto (L_1 v, \ldots, L_m v)$ where $(\forall i \in \{1, \ldots, m\})\ L_i \in \mathcal{B}(\mathcal{G}, \mathcal{H}_i)$.

PPXA+ algorithm

Let $(x_{0,i})_{1 \le i \le m} \in \mathcal{H}$, $v_0 = \big(\sum_{i=1}^m L_i^* L_i\big)^{-1} \sum_{i=1}^m L_i^* x_{0,i}$ and

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

We have then $v_n \rightharpoonup \widehat{v} \in \operatorname{Argmin} \sum_{i=1}^m g_i \circ L_i$.

Optimization algorithm: Douglas-Rachford

Particular case of Douglas-Rachford algorithm:

$\mathcal{H} = \mathcal{H}_1 \times \cdots \mathcal{H}_m$ where $\mathcal{H}_1 = \ldots = \mathcal{H}_m$ Hilbert spaces

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

where $(\forall i \in \{1, \ldots, m\}\ g_i \in \Gamma_0(\mathcal{H}_i)$

$L\colon v \mapsto (L_1 v, \ldots, L_m v)$ where $L_1 = \ldots = L_m = \mathrm{Id}.$

PPXA algorithm

Let $(x_{0,i})_{1\le i\le m} \in \mathcal{H}$, $v_0 = \frac{1}{m}\sum_{i=1}^m x_{0,i}$ and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_{n,i} = \mathrm{prox}_{\gamma g_i} x_{n,i}, \quad i \in \{1, \ldots, m\} \\ c_n = \frac{1}{m}\sum_{i=1}^m y_{n,i} \\ x_{n+1,i} = x_{n,i} + \lambda_n\big(2c_n - v_n - y_{n,i}\big), \quad i \in \{1, \ldots, m\} \\ v_{n+1} = v_n + \lambda_n(c_n - v_n). \end{cases}$$

We have then $v_n \rightharpoonup \widehat{v} \in \mathrm{Argmin} \sum_{i=1}^m g_i$.