Optimization
– Algorithms

Nelly Pustelnik

CNRS, Laboratoire de Physique de l'ENS de Lyon, France

Fixed point algorithm

Fixed point algorithm: zeros and fixed points

$2^{\mathcal{H}}$ is the power set of $\mathcal{H}$, i.e. the family of all subsets of $\mathcal{H}$.

Let $\mathcal{H}$ be a Hilbert space. Let $\Phi\colon \mathcal{H} \to 2^{\mathcal{H}}$.

The set of **fixed points** of Φ is : $\mathrm{Fix}\Phi = \{x \in \mathcal{H} \,|\, x \in \Phi x\}$.

The set of **zeros** of Φ is : $\mathrm{zer}\Phi = \{x \in \mathcal{H} \,|\, 0 \in \Phi x\}$.

Fixed point algorithm: convergence

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

Let $(x_k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and $\widehat{x} \in \mathcal{H}$.

- $(x_k)_{k\in\mathbb{N}}$ **converges strongly** to $\widehat{x}$ if

$$\lim_{k\to+\infty} \|x_k - \widehat{x}\| = 0.$$

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

- $(x_k)_{k\in\mathbb{N}}$ **converges weakly** to $\widehat{x}$ if

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

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

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

Banach-Picard theorem

An operator $\Phi\colon \mathcal{H} \to \mathcal{H}$ is **$\omega-$Lipschitz continuous** for some $\omega \in [0, +\infty[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|\Phi x - \Phi y\| \leq \omega \|x - y\|.$$

Φ is **nonexpansive** if it is $1-$Lipschitz continuous.

Banach-Picard theorem Let $\omega \in [0, 1[$, let $\Phi\colon \mathcal{H} \to \mathcal{H}$ be a $\omega-$Lipschitz continuous operator, and let $x_0 \in \mathcal{H}$. Set

$$(\forall k \in \mathbb{N}) \quad x_{k+1} = \Phi x_k.$$

Then, $\mathrm{Fix}\Phi = \{\widehat{x}\}$ for some $\widehat{x} \in \mathcal{H}$ and we have

$$(\forall k \in \mathbb{N}) \quad \|x_k - \widehat{x}\| \leq \omega^k \|x_0 - \widehat{x}\|.$$

Moreover, $(x_k)_{k \in \mathbb{N}}$ converges strongly to $\widehat{x}$ with linear convergence rate ω.

Averaged nonexpansive operator

An operator $\Phi\colon \mathcal{H} \to \mathcal{H}$ is $\mu-$**averaged nonexpansive** for some $\mu \in]0,1]$ if, for every $x \in \mathcal{H}$ and $y \in \mathcal{H}$,

$$\|\Phi x - \Phi y\|^2 \leq \|x - y\|^2 - \left(\frac{1-\mu}{\mu}\right) \|(\mathrm{Id} - \Phi)x - (\mathrm{Id} - \Phi)y\|^2,$$

Φ is **firmly nonexpansive** if it is $1/2-$averaged.

Φ is **nonexpansive** if and only if Φ is $1-$averaged.

Theorem Let $\mu \in]0,1[$, let $\Phi\colon \mathcal{H} \to \mathcal{H}$ be a $\mu-$averaged nonexpansive operator such that $\mathrm{Fix}\Phi \neq \varnothing$, and let $x_0 \in \mathcal{H}$. Set

$$(\forall k \in \mathbb{N}) \quad x_{k+1} = \Phi x_k.$$

Then $(x_k)_{k\in\mathbb{N}}$ converges weakly to a point in $\mathrm{Fix}\Phi$.

Nonlinear operators

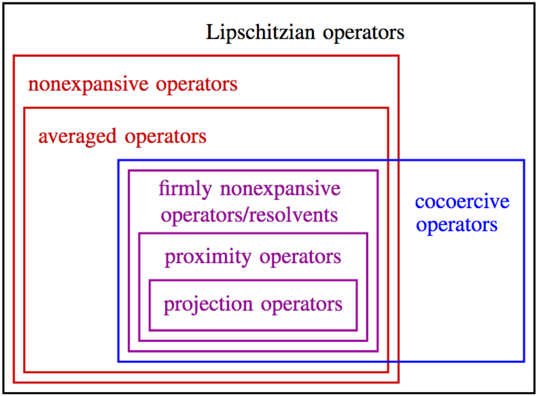


Fig. 3: Classes of nonlinear operators.

$\rightarrow$ extracted from [Combettes-Pesquet, 2020]

Averaged operator: example

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

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

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,

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

This leads to

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

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,

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

Averaged operator: example

Proof : 2) $\mathrm{Id} - \nabla f$ is μ-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,

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

Averaged operator: example

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

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

Averaged operator: example

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

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

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 $\boxed{\langle x_1 - x_2 \,|\, u_1 - u_2 \rangle \geq 0}$.

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 $\boxed{\langle x_1 - x_2 \,|\, u_1 - u_2 \rangle \geq 0}$.
- *Remark* : ∂f is a monotone operator.

Averaged operator: example

Proof :

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

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

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

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

Then, from the definition of the proximity operator,

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

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

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

Composition of averaged operator

Theorem Let D be a nonempty subset of $\mathcal{H}$. Let $\alpha_1 \in]0,1[$ and $\alpha_2 \in]0,1[$. Let $\Phi_1 : D \to D$ be α_1-averaged and $\Phi_2 : D \to D$ be α_2-averaged. Then $\Phi = \Phi_1\Phi_2$ is α-averaged with

$$\alpha = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \in]0,1[.$$

Proof: Extracted from Theorem 26.14 [Bauschke-Combettes, 2017]

Gradient method

Let $f \in \Gamma_0(\mathcal{H})$ and f is ν-smooth. We set, for some $\tau > 0$,

$$\Phi := \mathrm{Id} - \tau \nabla f$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = x_k - \tau \nabla f(x_k)$.

Gradient method

Let $f \in \Gamma_0(\mathcal{H})$ and f is ν-smooth. We set, for some $\tau > 0$,

$$\Phi := \mathrm{Id} - \tau \nabla f$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = x_k - \tau \nabla f(x_k)$.
- For every $\tau > 0$, $\operatorname{zer} \nabla f = (\nabla f)^{-1}(0) = \mathrm{Fix}\Phi$.

Gradient method

Let $f \in \Gamma_0(\mathcal{H})$ and f is ν-smooth. We set, for some $\tau > 0$,

$$\Phi := \mathrm{Id} - \tau \nabla f$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = x_k - \tau \nabla f(x_k)$.
- For every $\tau > 0$, $\mathrm{zer}\, \nabla f = (\nabla f)^{-1}(0) = \mathrm{Fix}\Phi$.
- $\Phi = \mathrm{Id} - \tau \nabla f$ is a $\tau\nu/2$-averaged operator.

  $\rightarrow$ cf. s.7 or Proposition 4.39 in [Bauschke-Combettes, 2017]

Gradient method

Let $f \in \Gamma_0(\mathcal{H})$ and f is ν-smooth. We set, for some $\tau > 0$,

$$\Phi := \mathrm{Id} - \tau \nabla f$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = x_k - \tau \nabla f(x_k)$.
- For every $\tau > 0$, $\mathrm{zer}\, \nabla f = (\nabla f)^{-1}(0) = \mathrm{Fix}\Phi$.
- $\Phi = \mathrm{Id} - \tau \nabla f$ is a $\tau\nu/2$-averaged operator.

  $\rightarrow$ cf. s.7 or Proposition 4.39 in [Bauschke-Combettes, 2017]
- For every $\tau \in \left]0, 2\nu^{-1}\right[$, the gradient method **converges** to a point in $\mathrm{zer}\, \nabla f$.

  $\rightarrow$ cf. s.5

Proximal Point Algorithm (PPA)

Let $f \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau f} = (\mathrm{Id} + \tau \partial f)^{-1}.$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = \mathrm{prox}_{\gamma f}(x_k)$.

Proximal Point Algorithm (PPA)

Let $f \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau f} = (\mathrm{Id} + \tau \partial f)^{-1}.$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = \mathrm{prox}_{\gamma f}(x_k).$
- For every $\tau > 0$, $\mathrm{Fix}\Phi = \mathrm{zer}\,\partial f$.

  Proof:

$$\begin{aligned} x = \mathrm{prox}_{\tau f} x &\Leftrightarrow x \in (\mathrm{I} + \tau \partial f) x \\ &\Leftrightarrow x \in x + \tau \partial f(x) \\ &\Leftrightarrow 0 \in \partial f \end{aligned}$$

Proximal Point Algorithm (PPA)

Let $f \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau f} = (\mathrm{Id} + \tau \partial f)^{-1}.$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = \mathrm{prox}_{\gamma f}(x_k)$.
- For every $\tau > 0$, $\mathrm{Fix}\Phi = \mathrm{zer}\,\partial f$.
- For every $\tau > 0$ and any $f \in \Gamma_0(\mathcal{H})$, $\mathrm{prox}_{\tau f}$ is 1/2-averaged.

  $\rightarrow$ cf. s.10 or Proposition 23.8 in [Bauschke-Combettes, 2017]

Proximal Point Algorithm (PPA)

Let $f \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau f} = (\mathrm{Id} + \tau \partial f)^{-1}.$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = \mathrm{prox}_{\gamma f}(x_k).$
- For every $\tau > 0$, $\mathrm{Fix}\Phi = \mathrm{zer}\, \partial f$.
- For every $\tau > 0$ and any $f \in \Gamma_0(\mathcal{H})$, $\mathrm{prox}_{\tau f}$ is 1/2-averaged.

  → cf. s.10 or Proposition 23.8 in [Bauschke-Combettes, 2017]
- The **PPA method converges** to a point in $\mathrm{zer}\, \partial f$.

  → cf. p.6

Forward-backward splitting

Let $f \in \Gamma_0(\mathcal{H})$ and ν-smooth and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f) = (\mathrm{Id} + \tau \partial g)^{-1}(\mathrm{Id} - \tau \nabla f)$$

- **Iterations**:$(\forall k \in \mathbb{N})$ $\quad x_{k+1} = \mathrm{prox}_{\tau g}(x_k - \tau \nabla f(x_k))$.

Forward-backward splitting

Let $f \in \Gamma_0(\mathcal{H})$ and ν-smooth and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f) = (\mathrm{Id} + \tau \partial g)^{-1}(\mathrm{Id} - \tau \nabla f)$$

- **Iterations**:$(\forall k \in \mathbb{N})$ $\quad x_{k+1} = \mathrm{prox}_{\tau g}(x_k - \tau \nabla f(x_k))$.
- Roots in projected gradient method [Levitin 1966] when $g = \iota_C$ for some closed convex set C

Forward-backward splitting

Let $f \in \Gamma_0(\mathcal{H})$ and ν-smooth and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f) = (\mathrm{Id} + \tau \partial g)^{-1}(\mathrm{Id} - \tau \nabla f)$$

- **Iterations**:$(\forall k \in \mathbb{N})\quad x_{k+1} = \mathrm{prox}_{\tau g}(x_k - \tau \nabla f(x_k))$.
- For every $\tau > 0$, $\mathrm{zer}\,(\nabla f + \partial g) = \mathrm{Fix}\Phi$.
  Proof:
  $$x \in \mathrm{Fix}\Phi \Leftrightarrow (\mathrm{Id} - \gamma \nabla f)x \in (\mathrm{Id} + \gamma \partial g)x$$
  $$\Leftrightarrow 0 \in \nabla f(x) + \partial g(x).$$

Forward-backward splitting

Let $f \in \Gamma_0(\mathcal{H})$ and ν-smooth and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f) = (\mathrm{Id} + \tau \partial g)^{-1}(\mathrm{Id} - \tau \nabla f)$$

- **Iterations**:$(\forall k \in \mathbb{N})$ $\quad x_{k+1} = \mathrm{prox}_{\tau g}(x_k - \tau \nabla f(x_k))$.
- For every $\tau > 0$, $\mathrm{zer}\,(\nabla f + \partial g) = \mathrm{Fix}\Phi$.
- $f \in \Gamma_0(\mathcal{H})$, ∇f is $\nu^{-1}-$cocoercive and $\tau \in \left]0, 2\nu^{-1}\right[$, $\mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f)$ is α-averaged nonexpansive where

$$\alpha = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2}.$$

where $\alpha_2 = \tau\nu/2$ and $\alpha_1 = 1/2$ leading to

$$\alpha = \frac{1}{2 - \tau\nu/2} \in]0, 1[.$$

Leading to

$$\gamma < 2/\nu.$$

Forward-backward splitting

Let $f \in \Gamma_0(\mathcal{H})$ and ν-smooth and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f) = (\mathrm{Id} + \tau \partial g)^{-1}(\mathrm{Id} - \tau \nabla f)$$

- **Iterations**:$(\forall k \in \mathbb{N}) \quad x_{k+1} = \mathrm{prox}_{\tau g}(x_k - \tau \nabla f(x_k))$.
- For every $\tau > 0$, $\mathrm{zer}\,(\nabla f + \partial g) = \mathrm{Fix}\Phi$.
- $f \in \Gamma_0(\mathcal{H})$, ∇f is $\nu^{-1}-$cocoercive and $\tau \in \left]0, 2\nu^{-1}\right[$, $\mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f)$ is α-averaged nonexpansive where

  $$\alpha = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2}.$$

  where $\alpha_2 = \tau\nu/2$ and $\alpha_1 = 1/2$ leading to

  $$\alpha = \frac{1}{2 - \tau\nu/2} \in]0, 1[.$$

  Leading to

  $$\gamma < 2/\nu.$$

- For every $\tau \in \left]0, 2\nu^{-1}\right[$, the **FBS method converges** to a point in $\mathrm{zer}\,(\nabla f + \partial g)$.

Peaceman-Rachford splitting

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := (2\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id})$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = 2\mathrm{prox}_{\tau g}(2\mathrm{prox}_{\tau f} x_k - x_k) - 2\mathrm{prox}_{\tau f} x_k + x_k.$

Peaceman-Rachford splitting

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := (2\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id})$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = 2\mathrm{prox}_{\tau g}(2\mathrm{prox}_{\tau f} x_k - x_k) - 2\mathrm{prox}_{\tau f} x_k + x_k$.
- For every $\tau > 0$, $\mathrm{zer}\,(\nabla f + \partial g) = \mathrm{prox}_{\tau f}(\mathrm{Fix}\Phi)$.

  $\rightarrow$ cf. Proposition 26.1(iii)(b)[Bauschke-Combettes, 2017]

Peaceman-Rachford splitting

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := (2\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id})$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = 2\mathrm{prox}_{\tau g}(2\mathrm{prox}_{\tau f} x_k - x_k) - 2\mathrm{prox}_{\tau f} x_k + x_k$.
- For every $\tau > 0$, $\mathrm{zer}\,(\nabla f + \partial g) = \mathrm{prox}_{\tau f}(\mathrm{Fix}\Phi)$.

  $\rightarrow$ cf. Proposition 26.1(iii)(b)[Bauschke-Combettes, 2017]
- Weak **convergence not guaranteed in this general setting**.
- Weak **convergence guaranteed if f is ν-smooth.**

  $\rightarrow$ cf. Corollary 1& Remark 2(2) [Lions-Mercier, 1979]

Douglas-Rachford splitting

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g} \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id}) + \mathrm{Id} - \mathrm{prox}_{\tau f}$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = \mathrm{prox}_{\tau g}(2\mathrm{prox}_{\tau f} x_k - x_k) + x_k - \mathrm{prox}_{\tau f} x_k$.

Douglas-Rachford splitting

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g} \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id}) + \mathrm{Id} - \mathrm{prox}_{\tau f}$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = \mathrm{prox}_{\tau g}(2\mathrm{prox}_{\tau f} x_k - x_k) + x_k - \mathrm{prox}_{\tau f} x_k$.
- For every $\tau > 0$, $\mathrm{zer}\,(\nabla f + \partial g) = \mathrm{prox}_{\tau f}(\mathrm{Fix}\Phi)$.

  $\rightarrow$ cf. Proposition 26.1(iii)(b)[Bauschke-Combettes, 2017]

Douglas-Rachford splitting

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau > 0$,

$$\Phi := \mathrm{prox}_{\tau g} \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id}) + \mathrm{Id} - \mathrm{prox}_{\tau f}$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x_{k+1} = \mathrm{prox}_{\tau g}(2\mathrm{prox}_{\tau f} x_k - x_k) + x_k - \mathrm{prox}_{\tau f} x_k$.
- For every $\tau > 0$, $\mathrm{zer}\,(\nabla f + \partial g) = \mathrm{prox}_{\tau f}(\mathrm{Fix}\Phi)$.

  $\rightarrow$ cf. Proposition 26.1(iii)(b)[Bauschke-Combettes, 2017]
- **Weak convergence** insured as

  $\rightarrow$ cf. Corollary 1& Remark 2(2) [Lions-Mercier, 1979]

Additional properties of smooth functions

Let $f \in \Gamma_0(\mathcal{H})$ and ν-smooth. Then

$$(\forall x, y \in \mathcal{H}) \qquad \langle \nabla f(x) - \nabla f(y) | x - y \rangle \leq \nu \|x - y\|^2$$

Proof: From Cauchy-Schwarz inequality

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

Additional properties of smooth functions

Let $f \in \Gamma_0(\mathcal{H})$ and ν-smooth. Then

$$(\forall x, y \in \mathcal{H}) \qquad \langle \nabla f(x) - \nabla f(y) | x - y \rangle \leq \nu \|x - y\|^2$$

Proof: From Cauchy-Schwarz inequality

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

Let $f \in \Gamma_0(\mathcal{H})$ and ν-smooth. Then

$$(\forall x, y \in \mathcal{H}) \qquad \|\nabla f(x) - \nabla f(y)\|^2 \leq \nu \langle \nabla f(x) - \nabla f(y) | x - y \rangle$$

Proof: cf. s. 8.

Strong convexity

Let $f \in \Gamma_0(\mathcal{H})$. f is **ρ-strongly** convex with $\rho > 0$ if

$$f - \frac{\rho}{2}\|\cdot\|_2^2$$

is convex.

Properties:

- If f is ρ-strongly convex then

$$(\forall x, y \in \mathcal{H}) \qquad \langle \nabla f(x) - \nabla f(y) | x - y \rangle \geq \rho \|x - y\|^2$$

- If f is twice differentiable, then f is ρ-strongly convex if and only if all the eigenvalues of the Hessian matrix of f are at most equal to ρ.

Gradient method: linear convergence

Let $f \in \Gamma_0(\mathcal{H})$ and f is ν-smooth and ρ strongly convex.
We set, for some $\tau \in]0, 2/\nu[$,

$$\Phi := \mathrm{Id} - \tau \nabla f$$

Then,

$$(\forall k \in \mathbb{N}) \quad \|x_k - \widehat{x}\| \leq \omega^k \|x_0 - \widehat{x}\|.$$

with

$$\omega = \max\{|1 - \tau\rho|, |1 - \tau\nu|\}$$

Moreover, $(x_k)_{k\in\mathbb{N}}$ converges strongly to $\widehat{x}$ with linear convergence rate ω.

Gradient method: linear convergence

Proof:

- Since f is ν smooth and ρ strongly convex with $\nu > \rho$:

$$(\forall x, y \in \mathcal{H}) \qquad \langle \nabla f(x) - \nabla f(y) | x - y \rangle \leq \nu \|x - y\|^2$$

by substracting $-\rho\|x-y\|^2$:

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

$\rightarrow \widetilde{f} = f - \frac{\rho}{2}\|\cdot\|^2$ is $(\nu - \rho)$-smooth.

- Gradient descent on f:

$$\begin{aligned} G_{\tau \nabla f} &= \mathrm{Id} - \tau \nabla f \\ &= \mathrm{Id} - \tau \nabla \widetilde{f} - \tau \rho \mathrm{Id} \\ &= (1 - \tau\rho)\mathrm{Id} - \tau \nabla \widetilde{f} \end{aligned}$$

Gradient method: linear convergence

Proof:

- Let $\tau \in]0, \frac{2}{\nu}[$ and $\begin{cases} p = G_{\tau \nabla f} x \\ q = G_{\tau \nabla f} y \end{cases}$

$$
\begin{aligned}
\|p - q\|^2 &= \|(1 - \tau\rho)(x - y) - \tau \nabla \widetilde{f}(x) - \widetilde{f}(y)\|^2 \\
&= (1 - \tau\rho)^2 \|(x - y)\|^2 + \tau^2 \|\nabla \widetilde{f}(x) - \nabla \widetilde{f}(y)\|^2 \\
&\qquad - 2(1 - \tau\rho)\tau \langle (x - y) | \nabla \widetilde{f}(x) - \nabla \widetilde{f}(y) \rangle \\
&\leq (1 - \tau\rho)^2 \|(x - y)\|^2 \\
&\qquad + \big(\tau^2 \nu - 2(1 - \tau\rho)\tau\big) \langle (x - y) | \nabla \widetilde{f}(x) - \nabla \widetilde{f}(y) \rangle \\
&\leq (1 - \tau\rho)^2 \|(x - y)\|^2 \\
&\qquad + \max\{0, \tau\big(\tau(\nu + \rho) - 2\big)\}(\nu - \rho)\|x - y\|^2 \\
&\leq \max\{(1 - \tau\rho)^2, (1 - \tau\nu)^2\} \|x - y\|^2
\end{aligned}
$$

Forward-Backward method: linear convergence

Let $f \in \Gamma_0(\mathcal{H})$ and f is ν-smooth and ρ strongly convex and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau \in]0, 2/\nu[$,

$$\Phi := \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f)$$

Then,

$$(\forall k \in \mathbb{N}) \quad \|x_k - \widehat{x}\| \leq \omega^k \|x_0 - \widehat{x}\|.$$

with

$$\omega = \max\{|1 - \tau\rho|, |1 - \tau\nu|\}$$

Moreover, $(x_k)_{k\in\mathbb{N}}$ converges strongly to $\widehat{x}$ with linear convergence rate ω.

Remark: Same rate as gradient descent.

Forward-Backward method: linear convergence

Let $f \in \Gamma_0(\mathcal{H})$ and f is ν-smooth and ρ strongly convex and $g \in \Gamma_0(\mathcal{H})$. We set, for some $\tau \in]0, 2/\nu[$,

$$\Phi := \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f)$$

Then,

$$(\forall k \in \mathbb{N}) \quad \|x_k - \widehat{x}\| \leq \omega^k \|x_0 - \widehat{x}\|.$$

with

$$\omega = \max\{|1 - \tau\rho|, |1 - \tau\nu|\}$$

Moreover, $(x_k)_{k\in\mathbb{N}}$ converges strongly to $\widehat{x}$ with linear convergence rate ω.

Proof: The proximity is a non expansive operator, that is for

$$\begin{cases} \tilde{p} = \mathrm{prox}_{\tau g}(G_{\tau\nabla f} x) \\ \tilde{q} = \mathrm{prox}_{\tau g}(G_{\tau\nabla f} y) \end{cases} \quad \text{leading to } \|\tilde{p} - \tilde{q}\| \leq \|p - q\|$$

Comparisons

Let $f \in C^{1,1}_{1/\alpha}(\mathcal{H})$ and $g \in C^{1,1}_{1/\beta}(\mathcal{H})$, for some $\alpha > 0$ and $\beta > 0$.
The problem is to

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

under the assumption that solutions exist.

Example: **Smooth TV denoising**

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \; \frac{1}{2}\|x - z\|_2^2 + \chi h_\mu(Lx),$$

- $L \in \mathbb{R}^{N-1 \times N}$ denotes the first order discrete difference operator

  $$(\forall n \in \{1, \ldots, N-1\}) \quad (Lx)_n = \tfrac{1}{2}(x_n - x_{n-1})$$

- h_μ: Huber loss, the smooth approximation of the ℓ_1−norm parametrized by $\mu > 0$.

  $h_\mu \in C^{1,1}_{1/\mu}(\mathbb{R}^{N-1})$.

  Closed form expression of prox_{h_μ}.

Comparisons

Proposition (see [**Briceno-Arias, Pustelnik, 2021**] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

1. **Gradient descent** Suppose that $\tau \in]0, 2\beta\alpha/(\beta+\alpha)[$. Then, $\mathrm{Id} - \tau(\nabla g + \nabla f)$ is $r_G(\tau)-$Lipschitz continuous, where

$$r_G(\tau) := \max\left\{|1-\tau\rho|, |1-\tau(\beta^{-1}+\alpha^{-1})|\right\} \in]0,1[\,. \tag{1}$$

In particular, the minimum in (1) is achieved at

$$\tau^* = \frac{2}{\rho + \alpha^{-1} + \beta^{-1}}$$

and

$$r_G(\tau^*) = \frac{\alpha^{-1} + \beta^{-1} - \rho}{\alpha^{-1} + \beta^{-1} + \rho}.$$

Comparisons

Proposition (see [**Briceno-Arias, Pustelnik, 2021**] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

1. **FBS** Suppose that $\tau \in]0, 2\alpha[$. Then $\mathrm{prox}_{\tau g}(\mathrm{Id} - \tau\nabla f)$ is $r_{T_1}(\tau)-$Lipschitz continuous, where

$$r_{T_1}(\tau) := \max\left\{|1-\tau\rho|, |1-\tau\alpha^{-1}|\right\} \in]0,1[\,. \tag{1}$$

In particular, the minimum in (1) is achieved at

$$\tau^* = \frac{2}{\rho + \alpha^{-1}} \quad \text{and} \quad r_{T_1}(\tau^*) = \frac{\alpha^{-1} - \rho}{\alpha^{-1} + \rho}.$$

Comparisons

Proposition (see [**Briceno-Arias, Pustelnik, 2021**] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

1. **FBS** Suppose that $\tau \in]0, 2\beta]$. Then $\mathrm{prox}_{\tau f}(\mathrm{Id} - \tau\nabla g)$ is $r_{T_2}(\tau)-$Lipschitz continuous, where $\mathrm{r}_{T_2}(\tau) := \dfrac{1}{1+\tau\rho} \in]0,1[$ .In particular, the minimum in (1) is achieved at

$$\tau^* = 2\beta \quad \text{and} \quad r_{T_2}(\tau^*) = \frac{1}{1+2\beta\rho}.$$

Comparisons

Proposition (see [**Briceno-Arias, Pustelnik, 2021**] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

1. **PRS** $(2\mathrm{prox}_{\tau g} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau f} - \mathrm{Id})$ and $(2\mathrm{prox}_{\tau f} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau g} - \mathrm{Id})$ are $r_R(\tau)-$Lipschitz continuous, where

$$r_R(\tau) = \max\left\{\frac{1-\tau\rho}{1+\tau\rho}, \frac{\tau\alpha^{-1}-1}{\tau\alpha^{-1}+1}\right\} \in \left]0,1\right[. \tag{1}$$

In particular, the minimum in (1) is achieved at

$$\tau^* = \sqrt{\frac{\alpha}{\rho}} \quad \text{and} \quad r_R(\tau^*) = \frac{1-\sqrt{\alpha\rho}}{1+\sqrt{\alpha\rho}}.$$

Comparisons

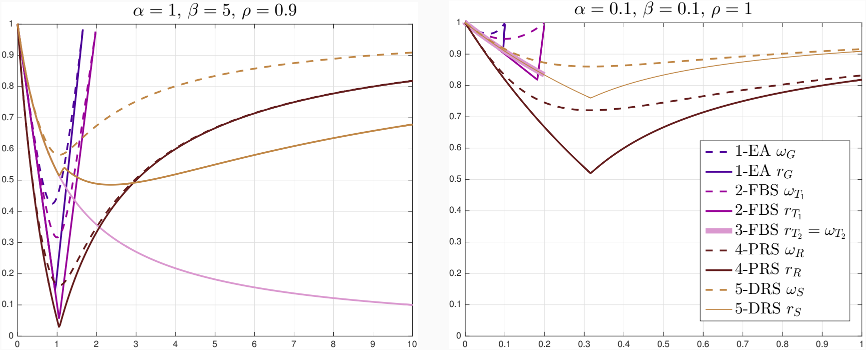
Proposition (see [**Briceno-Arias, Pustelnik, 2021**] for detailed references)
In the context of Problem p.16, suppose that f **is $\rho-$strongly convex**, for some $\rho \in \left]0, \alpha^{-1}\right[$, and let $\tau > 0$. Then, the following holds:

1. **DRS** $S_{\tau\nabla g, \tau\nabla f}$ and $S_{\tau\nabla f, \tau\nabla g}$ are $r_S(\tau)-$Lipschitz continuous, where

$$r_S(\tau) = \min\left\{\frac{1 + r_R(\tau)}{2}, \frac{\beta + \tau^2\rho}{\beta + \tau\beta\rho + \tau^2\rho}\right\} \in]0, 1[\tag{1}$$

and r_R is defined in p.16. In particular, the optimal step-size and the minimum in (1) are

$$(\tau^*, r_S(\tau^*)) = \begin{cases} \left(\sqrt{\frac{\alpha}{\rho}}, \frac{1}{1+\sqrt{\alpha\rho}}\right), & \text{if } \beta \leq 4\alpha; \\ \left(\sqrt{\frac{\beta}{\rho}}, \frac{2}{2+\sqrt{\beta\rho}}\right), & \text{otherwise.} \end{cases}$$



Comparison of the convergence rates of EA, FBS, PRS, DRS for two choices of α, β, and ρ. Note that optimization rates are better than cocoercive rates in general.

Example: Smooth TV denoising

- **First formulation**: $\underset{x\in\mathbb{R}^N}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|x-z\|_2^2}_{f(x)} + \underbrace{\chi h(Lx)}_{g(x)}$

  $\rightarrow f$ is $\rho = 1$ strongly convex, $\alpha = 1$, and $\beta = \frac{\mu}{\chi\|L\|^2}$.

1- **EA:** Use $G_{\tau(\nabla g + \nabla f)}$

2- **FBS:** Use $T_{\tau\nabla f, \tau\nabla g}$

- **Second formulation**: $\min_{x\in\mathcal{H}} \underbrace{\frac{1}{2}\|x-z\|_2^2 + \chi h_{\mathbb{I}_1}(L_{\mathbb{I}_1}x)}_{\tilde{f}(x)} + \underbrace{\chi h_{\mathbb{I}_2}(L_{\mathbb{I}_2}x)}_{\tilde{g}(x)}$

  $\rightarrow \widetilde{f}$ is $\rho = 1$ strongly convex, $\alpha = \frac{\mu}{\mu + \chi\|L_{\mathbb{I}_2}\|^2}$, and $\beta = \frac{\mu}{\chi\|L_{\mathbb{I}_1}\|^2}$.

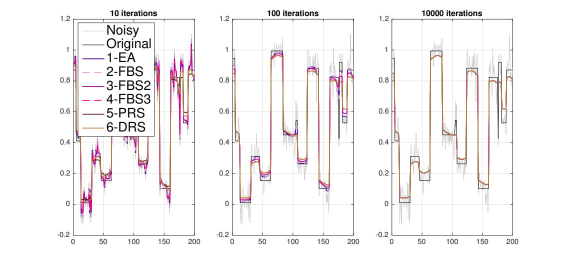
3- **FBS 2:** Use $T_{\tau\nabla\widetilde{g}, \tau\nabla\widetilde{f}}$

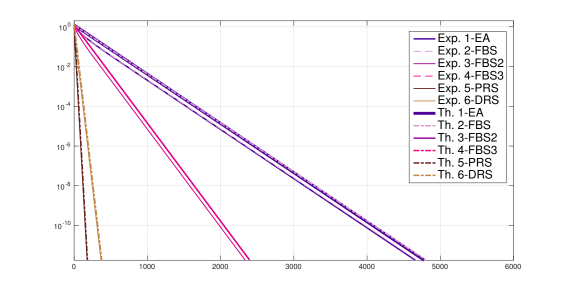
4- **FBS 3:** Use $T_{\tau\nabla\widetilde{f}, \tau\nabla\widetilde{g}}$

5- **PRS:** Use $R_{\tau\nabla\widetilde{f}, \tau\nabla\widetilde{g}}$

6- **DRS:** Use $S_{\tau\nabla\widetilde{f}, \tau\nabla\widetilde{g}}$

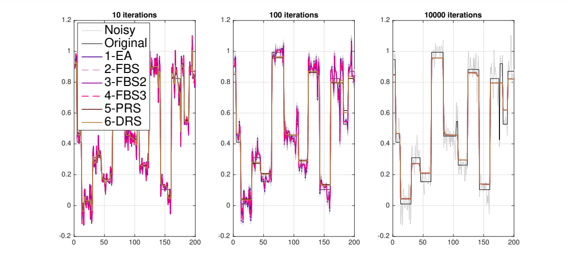
Numerical and theoretical comparisons

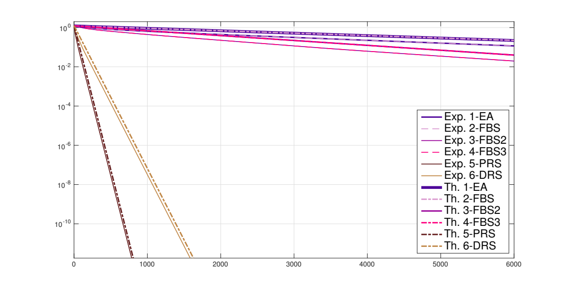




Piecewise constant denoising estimates after 10, 100, and 10000 iterations with $\chi = 0.7$ and $\mu = 0.002$ when considering gradient descent, FBS, PRS or DRS. Associated theoretical and numerical convergence rates.

Numerical and theoretical comparisons





Piecewise constant denoising estimates after 10, 100, and 10000 iterations with $\chi = 0.7$ and $\mu = 0.0001$ when considering gradient descent, FBS, PRS or DRS. Associated theoretical and numerical convergence rates.

Optimization algorithms

Forward-Backward	$f+g$	f ζ-smooth prox_g	[Combettes,Wajs,2005]
ISTA	$f+g$	f ζ-smooth $g=\lambda\|\cdot\|_1$	[Daubechies et al, 2003]
Projected gradient	$f+g$	f ζ-smooth g$=\iota_C$	
Gradient descent	$f+g$	f ζ-smooth $f_2=0$	
Douglas-Rachford	$f+g$	prox_f prox_{f_2}	[Combettes,Pesquet, 2007]
PPXA	$\sum_i f_i$	prox_{f_i}	[Combettes,Pesquet, 2008]

Forward-Backward method: Worst case complexity

Let $f \in \Gamma_0(\mathcal{H})$ and f is ζ-smooth and $g \in \Gamma_0(\mathcal{H})$.
We set, for some $\tau \in]0, 1/\beta]$ where $\beta \geq \zeta$,

$$\Phi := \mathrm{prox}_{\tau g}(\mathrm{Id} - \tau \nabla f)$$

Then,

$$(\forall k \geq 1) \qquad f(x_k) - f(\widehat{x}) \leq \frac{\beta}{2k} \|x_0 - \widehat{x}\|^2.$$

Make algorithms faster

Acceleration: Nesterov acceleration

Beck-Teboulle proximal gradient algorithm (FISTA)

[Beck,Teboulle,2009]:

- **Goal**: $\min_{x\in\mathcal{H}} f(x)+g(x)$ with f is ν-Lipschitz differentiable.
- **Iterations**: Let $\gamma \in]0,1/\nu[$, $x_0 = z_0 \in \mathcal{H}$, $t_0 = 1$, and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = \mathrm{prox}_{\gamma g}(z_n - \gamma \nabla f(z_n)) \\ t_{n+1} = \frac{1+\sqrt{4t_n^2+1}}{2} \\ \lambda_n = \frac{t_n - 1}{t_{n+1}} \\ z_{n+1} = x_{n+1} + \lambda_n (x_{n+1} - x_n). \end{cases}$$

- **Guarantees**:
  - Convergence of $(f(x_n)+g(x_n))_{n\in\mathbb{N}}$ with rate $O(1/n^2)$.
  - Convergence of $(x_n)_{n\in\mathbb{N}}$ not secured theoretically.

Acceleration: Nesterov acceleration

Chambolle-Dossal proximal gradient algorithm

[Chambolle,Dossal,2015]:

- **Goal**: $\min_{x\in\mathcal{H}} f(x)+g(x)$ with f is ν-Lipschitz differentiable.

- **Iterations**: Let $\gamma \in]0,1/\nu[$, $x_0 = z_0 \in \mathcal{H}$, $a > 2$, and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} x_{n+1} = \mathrm{prox}_{\beta^{-1}g}(z_n - \beta^{-1}\nabla f(z_n)) \\ \lambda_n = \frac{n-1}{n+a} \\ z_{n+1} = x_{n+1} + \lambda_n(x_{n+1} - x_n). \end{cases}$$

- **Guarantees**:

  - Convergence of $(f(x_n)+g(x_n))_{n\in\mathbb{N}}$ with rate $O(1/n^2)$.
  - Convergence of $(x_n)_{n\in\mathbb{N}}$.