Lecture Series
on
"Convex analysis with applications in inverse problems"

Lecture 2: Proximal methods in convex optimization

Radu Ioan Boţ

Chemnitz University of Technology
Department of Mathematics
09107 Chemnitz
www.tu-chemnitz.de/∼rabot

Institute for Numerical and Applied Mathematics
University of Goettingen
June 13, 2012

TECHNISCHE UNIVERSITÄT CHEMNITZ

Deutsche Forschungsgemeinschaft DFG

Lecture Series
on
"Convex analysis with applications in inverse problems"

- Lecture 1: Convex analysis: basics, conjugation and duality (Monday, June 11, 2012)
- Lecture 2: Proximal methods in convex optimization (Wednesday, June 13, 2012)
- Lecture 3: Convex regularization techniques for linear inverse problems (Thursday, June 14, 2012)

Contents

Moreau envelope and proximal mapping

Strongly convex functions

Let $(H, \langle \cdot, \cdot \rangle)$ be a real Hilbert space and $\| \cdot \| := \sqrt{\langle \cdot, \cdot \rangle}$.

Strictly convex versus strongly convex function

A function $f : H \to \overline{\mathbb{R}}$ is said to be

- strictly convex, if
$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y) \ \forall x, y \in \operatorname{dom} f, x \neq y \ \forall \lambda \in (0,1);$$
- strongly convex (with modulus $\beta > 0$), if
$$f(\lambda x + (1-\lambda)y) + \lambda(1-\lambda)\frac{\beta}{2}\|x-y\|^2 \leq \lambda f(x) + (1-\lambda)f(y) \ \forall x, y \in \operatorname{dom} f \ \forall \lambda \in (0,1).$$

Strongly convex function: characterization

A function $f : H \to \overline{\mathbb{R}}$ is strongly convex with modulus $\beta > 0$ if and only if $f - \frac{\beta}{2}\| \cdot \|^2$ is convex.

Coercive versus supercoercive function

A function $f : H \to \overline{\mathbb{R}}$ is said to be

- coercive, if $\displaystyle\lim_{\|x\| \to +\infty} f(x) = +\infty$;
- supercoercive, if $\displaystyle\lim_{\|x\| \to +\infty} \frac{f(x)}{\|x\|} = +\infty$.

- Obviously, every supercoercive function is coercive.

Coercivity versus boundedness

A function $f : H \to \overline{\mathbb{R}}$ is coercive if and only if for every $\lambda \in \mathbb{R}$ its upper level set $\{x \in H : f(x) \leq \lambda\}$ is bounded.

Strong convexity implies supercoercivity

Every proper, strongly convex and lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ is supercoercive.

- Indeed, there exists $\beta > 0$ such that

$$f(x) = \frac{\beta}{2}\|x\|^2 + \left(f(x) - \frac{\beta}{2}\|x\|^2\right) \ \forall x \in H.$$

Since $f - \frac{\beta}{2}\|\cdot\|^2$ is proper, convex and lower semicontinuous, there exists $(x^*, \alpha) \in H \times \mathbb{R}$ such that

$$f(x) \geq \frac{\beta}{2}\|x\|^2 + \langle x^*, x\rangle + \alpha \geq \frac{\beta}{2}\|x\|^2 - \|x\|\|x^*\| + \alpha \ \forall x \in H.$$

Minimization of strongly convex functions

Every proper, convex, lower semicontinuous and coercive function $f : H \to \overline{\mathbb{R}}$ has a minimizer in H. Thus, every proper, strongly convex, lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ has exactly one minimizer in H.

- Indeed, there exists $\lambda \in \mathbb{R}$ such that $\{x \in H : f(x) \leq \lambda\} \neq \emptyset$ and

$$\inf_{x \in H} f(x) = \inf_{\{y \in H : f(y) \leq \lambda\}} f(x).$$

Since $\{y \in H : f(y) \leq \lambda\}$ is bounded and closed, it is weakly compact, thus f, being weakly lower semicontinuous, has at least one minimizer in $\{y \in H : f(y) \leq \lambda\}$, which is actually a minimizer of f in H. When f is strongly convex, the uniqueness of the minimizer follows from the fact that f is strictly convex, too.

The strong convexity of the conjugate

For a proper, convex and lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ the following properties are equivalent:

- $\operatorname{dom} f = H$, f is Fréchet differentiable on H and ∇f is β-Lipschitz continuous (Lipschitz continuous with Lipschitz constant $\beta > 0$);
- $\operatorname{dom} f = H$, f is Fréchet differentiable on H and the descent formula holds

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\beta}{2}\|x - y\|^2 \ \forall x, y \in H;$$

- f^* is strongly convex with modulus $\frac{1}{\beta}$.

Moreau envelope

Moreau envelope of parameter $\gamma > 0$

For a proper, convex and lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ and $\gamma > 0$ the Moreau envelope of f of parameter γ is the convex function

$$^{\gamma}f(x) = \left(f \Box \frac{1}{2\gamma}\|\cdot\|^2\right)(x) = \inf_{y \in H}\left\{f(y) + \frac{1}{2\gamma}\|x-y\|^2\right\} \ \forall x \in H.$$

For all $x \in H$ the function $y \mapsto f(y) + \frac{1}{2\gamma}\|x-y\|^2$ is proper, strongly convex and lower semicontinuous, thus the infimum is attained and $^{\gamma}f(x) \in \mathbb{R}$. This means that

$$^{\gamma}f : H \to \mathbb{R}.$$

Proximal point

For a proper, convex and lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ and $x \in H$, the unique minimum of

$$y \mapsto f(y) + \frac{1}{2}\|x-y\|^2$$

is called proximal point of f at x and it is denoted by $\operatorname{prox}_f(x)$. The mapping

$$\operatorname{prox}_f : H \to H$$

is well-defined and is said to be the proximal mapping of f.

Since for $\gamma > 0$ one has ${}^{\gamma}f = \frac{1}{\gamma}{}^{1}(\gamma f)$, it holds

$$^{\gamma}f(x) = f(\mathrm{prox}_{\gamma f}(x)) + \frac{1}{2\gamma}\|x - \mathrm{prox}_{\gamma f}(x)\|^2 \ \forall x \in H.$$

Example

When $S \subseteq H$ is a nonempty, convex and closed set and $\gamma > 0$, one has

$$^{\gamma}\delta_S(x) = \inf_{y \in S}\left\{\frac{1}{2\gamma}\|x - y\|^2\right\} = \frac{1}{2\gamma}d_S^2(x) \ \forall x \in H.$$

Consequently,

$$\mathrm{prox}_{\delta_S}(x) = P_S(x) \ \forall x \in H,$$

where $P_S : H \to S$ denotes the metric projection on S.

Characterization of the proximal mapping

For a proper, convex and lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ and $x, p \in H$ one has

$$p = \mathrm{prox}_f(x) \Leftrightarrow x - p \in \partial f(p).$$

- Indeed, $p = \mathrm{prox}_f(x) \Leftrightarrow 0 \in \partial\left(f + \frac{1}{2}\|x - \cdot\|^2\right)(p) = \partial f(p) + \nabla\left(\frac{1}{2}\|x - \cdot\|^2\right)(p)$
$\Leftrightarrow 0 \in \partial f(p) + p - x \Leftrightarrow x - p \in \partial f(p)$.

Firmly nonexpansive operators versus nonexpansive operators

For a nonempty set $D \subseteq H$, an operator $T : D \to H$ is said to be

- firmly nonexpansive, if

$$\|T(x) - T(y)\|^2 + \|(\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y)\|^2 \leq \|x - y\|^2 \ \forall x, y \in D;$$

- nonexpansive, if it is 1-Lipschitz continuous, i.e.,

$$\|T(x) - T(y)\| \leq \|x - y\| \ \forall x, y \in D.$$

Here, $\mathrm{Id} : H \to H, \mathrm{Id}(x) = x \ \forall x \in H$, denotes the identity operator on H.

- Obviously, every firmly nonexpansive operator is nonexpansive.

Firmly nonexpansive operator: equivalent characterizations

For a nonempty set $D \subseteq H$ let be $T : D \to H$. The following statements are equivalent:

- T is firmly nonexpansive;
- $\mathrm{Id} - T$ is firmly nonexpansive;
- $2T - \mathrm{Id}$ is firmly nonexpansive;
- $\|T(x) - T(x)\|^2 \leq \langle T(x) - T(y), x - y \rangle \ \forall x, y \in D.$

Example

For a proper, convex and lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ the operators

$$\operatorname{prox}_f : H \to H \text{ and } \operatorname{Id} - \operatorname{prox}_f : H \to H$$

are firmly nonexpansive, thus nonexpansive.

Fixed points of the proximal mapping

For a proper, convex and lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ one has

$$\operatorname{Fix} \operatorname{prox}_f = \operatorname{argmin} f,$$

where, for an operator $T : D \to H$, by $\operatorname{Fix} T := \{x \in D : T(x) = x\}$ we denote the set of fixed points of T.

- Indeed, $x \in \operatorname{argmin} f \Leftrightarrow 0 \in \partial f(x) \Leftrightarrow x - x \in \partial f(x) \Leftrightarrow \operatorname{prox}_f(x) = x$.

Differentiability of the Moreau envelope

Let $f : H \to \overline{\mathbb{R}}$ be a proper, convex and lower semicontinuous function and $\gamma > 0$. Then ${}^{\gamma}f : H \to \mathbb{R}$ is Fréchet differentiable on H and it holds

$$\nabla({}^{\gamma}f)(x) = \frac{1}{\gamma}(x - \operatorname{prox}_{f}(x)) = \frac{1}{\gamma}(\mathrm{Id} - \operatorname{prox}_{f})(x) \ \forall x \in H.$$

Consequently,

$$\|\nabla({}^{\gamma}f)(x) - \nabla({}^{\gamma}f)(y)\| \leq \frac{1}{\gamma}\|x - y\| \ \forall x, y \in H.$$

Notice also that

$$x \in \operatorname{argmin}({}^{\gamma}f) \Leftrightarrow \nabla({}^{\gamma}f)(x) = 0 \Leftrightarrow x = \operatorname{prox}_{f}(x) \Leftrightarrow x \in \operatorname{Fix}\operatorname{prox}_{f} \Leftrightarrow x \in \operatorname{argmin} f,$$

in which case

$${}^{\gamma}f(x) = f(x).$$

Example

When $S \subseteq H$ is a nonempty, convex and closed set and $\gamma > 0$, then d_S^2 is Fréchet differentiable on H and it holds

$$\nabla(d_S^2)(x) = 2\gamma\nabla({}^{\gamma}\delta_S)(x) = 2(\mathrm{Id} - P_S)(x) \ \forall x \in H.$$

Further properties of the proximal mapping

Moreau's decomposition

For a proper, convex and lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ and $\gamma > 0$ the following statements hold:

- ${}^{\gamma}f + {}^{\frac{1}{\gamma}}(f^*) \circ \frac{1}{\gamma}\,\mathrm{Id} = \frac{1}{2\gamma}\|\cdot\|^2$;
- $\operatorname{prox}_{\gamma f} + \gamma \operatorname{prox}_{f^*/\gamma} \circ \frac{1}{\gamma}\,\mathrm{Id} = \mathrm{Id}$;
- $\operatorname{prox}_{f^*/\gamma}(x/\gamma) \in \partial f(\operatorname{prox}_{\gamma f}(x))\ \forall x \in H$.

The case $\gamma = 1$

For a proper, convex and lower semicontinuous function $f : H \to \overline{\mathbb{R}}$ it holds:

$$\operatorname{prox}_f + \operatorname{prox}_{f^*} = \mathrm{Id}\,.$$

Example (the proximal mapping of the norm)

For $f = \|\cdot\|$ we have $f^* = \delta_{\overline{B}(0,1)}$. For all $x \in H$ it holds

$$\operatorname{prox}_f(x) = x - \operatorname{prox}_{f^*}(x) = x - P_{\overline{B}(0,1)}(x) = \begin{cases} \left(1 - \frac{1}{\|x\|}\right)x, & \text{if } \|x\| > 1,\\ 0, & \text{if } \|x\| \le 1. \end{cases}$$

Example (the proximal mapping of the support function)

Let $S \subseteq H$ be a nonempty, convex and closed set. Then

$$\operatorname{prox}_{\sigma_S}(x) = x - \operatorname{prox}_{\delta_S}(x) = x - P_S(x) \ \forall x \in H.$$

Further formulae for the proximal mapping

Let $f : H \to \overline{\mathbb{R}}$ be a proper, convex and lower semicontinuous function. Then

- for $\bar{x} \in H$ and $g(x) = f(x - \bar{x})$ it holds $\operatorname{prox}_g(x) = \bar{x} + \operatorname{prox}_f(x - \bar{x})$ for all $x \in H$;
- for $\rho \neq 0$ and $g(x) = f\left(\frac{1}{\rho}x\right)$ it holds $\operatorname{prox}_g(x) = \rho \operatorname{prox}_{\frac{1}{\rho^2}f}\left(\frac{1}{\rho}x\right)$ for all $x \in H$;
- for $g(x) = f(-x)$ it holds $\operatorname{prox}_g(x) = -\operatorname{prox}_f(-x)$ for all $x \in H$.

The proximal mapping of the Moreau envelope

Let $f : H \to \overline{\mathbb{R}}$ be a proper, convex and lower semicontinuous function. Then

$$\operatorname{prox}_{({}^1 f)}(x) = \frac{1}{2}(x + \operatorname{prox}_{2f}(x)) \ \forall x \in H.$$

Example

Let $S \subseteq H$ be a nonempty, convex and closed set. Then

$$\operatorname{prox}_{\frac{1}{2}d_S^2}(x) = \frac{1}{2}(x + \operatorname{prox}_{2\delta_S}(x)) = \frac{1}{2}(x + P_S(x)) \ \forall x \in H.$$

Regularization algorithms

For $f : H \to \overline{\mathbb{R}}$ a proper, convex and lower semicontinuous function we discuss several regularization algorithms for solving the optimization problem

$$\inf_{x \in H} f(x).$$

- Let be $v(P) := \inf_{x \in H} f(x) \in \mathbb{R} \cup \{-\infty\}$ and $\operatorname{argmin} f := \{x \in H : f(x) = v(P)\}$ the (possibly empty) set of optimal solutions.

The proximal point algorithm

Proximal point algorithm

Initialization: Choose $x_0 \in \operatorname{dom} f$ and set $k := 0$

For $k \geq 0$: Choose $\gamma_k > 0$ and set

$$x_{k+1} := \operatorname{prox}_{\gamma_k f}(x_k) = \operatorname*{argmin}_{x \in H} \left\{ f(x) + \frac{1}{2\gamma_k} \|x - x_k\|^2 \right\}$$

- The proximal point algorithm is well-defined!

Notations

For $k \geq 0$ let be

$$s_{k+1} := \frac{x_k - x_{k+1}}{\gamma_{k+1}} \text{ and } \sigma_k := \sum_{j=0}^{k} \gamma_j.$$

Facts on the proximal point algorithm

- For every $k \geq 0$ it holds $s_{k+1} \in \partial f(x_{k+1})$, since

$$0 \in \partial \left(f + \frac{1}{2\gamma_k} \|\cdot - x_k\|^2 \right)(x_{k+1}) = \partial f(x_{k+1}) + \frac{1}{\gamma_k}(x_{k+1} - x_k) = \partial f(x_{k+1}) - s_{k+1};$$

- For every $k \geq 1$ it holds $\|s_{k+1}\| \leq \|s_k\|$;
- Fundamental estimate: For every $k \geq 1$ and every $x \in H$ it holds

$$f(x_k) - f(x) \leq \frac{\|x - x_0\|^2}{2\sigma_{k-1}} - \frac{\|x - x_k\|^2}{2\sigma_{k-1}} - \frac{\sigma_{k-1}}{2}\|s_k\|^2.$$

Convergence of the proximal point algorithm

For every $k \geq 1$ and every $x \in H$ it holds

$$f(x_k) - f(x) \leq \frac{\|x - x_0\|^2}{2\sigma_{k-1}}.$$

- If $\lim_{k \to +\infty} \sigma_k = +\infty$, then $f(x_k) \to v(P) (k \to +\infty)$;
- If, additionally, $\operatorname{argmin} f \neq \emptyset$, then $(x_k)_{k \geq 0}$ converges weakly to a minimizer of f and

$$f(x_k) - v(P) \leq \frac{d^2_{\operatorname{argmin} f}(x_0)}{2\sigma_{k-1}} \ \forall k \geq 1.$$

- The assumption $\lim_{k\to+\infty} \sigma_k = +\infty$ is, for instance, fulfilled when $(\gamma_k)_{k\geq 0}$ is a constant sequence.

Strong convergence of the proximal point algorithm

If f is strongly convex and $\lim_{k\to+\infty} \sigma_k = +\infty$, then $(x_k)_{k\geq 0}$ converges strongly to the unique minimizer of f.

Improved convergence rate

If $\lim_{k\to+\infty} \sigma_k = +\infty$, $\operatorname{argmin} f \neq \emptyset$ and $(x_k)_{k\geq 0}$ converges strongly to a minimizer of f, then

$$\lim_{k\to+\infty} \sigma_{k-1}(f(x_k) - v(P)) = 0.$$

Weak convergence versus strong convergence (Güler, 1991)

There exists a proper, convex and lower semicontinuous function $f : \ell^2 \to \overline{\mathbb{R}}$ such that given any bounded positive sequence $(\gamma_k)_{k\geq 0}$, there exists a point $x_0 \in \operatorname{dom} f$ for which the proximal point algorithm with

$$x_{k+1} := \operatorname{prox}_{\gamma_k f}(x_k) \ \forall k \geq 0$$

converges weakly, but not strongly to a minimizer of f.

A proximal-like algorithm with Bregman functions

Bregman distance

We call Bregman distance

$$D_\psi : \mathrm{cl}(Z) \times Z \to \mathbb{R},\ D_\psi(x,y) = \psi(x) - \psi(y) - \langle \nabla\psi(y), x - y\rangle$$

where $Z \subseteq \mathbb{R}^n$ is an open set and $\psi : \mathrm{cl}(Z) \to \mathbb{R}$ is a so-called Bregman function with zone Z, namely it has the following properties:

- ψ is continuously differentiable on Z;
- ψ is strictly convex and continuous on $\mathrm{cl}(Z)$;
- the partial upper level sets $\{x \in \mathrm{cl}(Z) : D_\psi(y,x) \le \lambda\}$ and $\{y \in Z : D_\psi(x,y) \le \lambda\}$ are bounded for every $\lambda \in \mathbb{R}$, $y \in Z$ and $x \in \mathrm{cl}(Z)$;
- if $(y_k)_{k\ge 0} \subseteq Z$ converges to y, then $D_\psi(y, y^k) \to 0 (k \to +\infty)$;
- if $(y_k)_{k\ge 0}$ is a sequence converging to $y \in \mathrm{cl}(Z)$ and $(x_k)_{k\ge 0}$ is a bounded sequence such that $D_\psi(x^k, y^k) \to 0 (k \to +\infty)$, then $x^k \to y (k \to +\infty)$.

Some properties of the Bregman distance

- $D_\psi(\cdot,\cdot)$ is not a distance (it might not be symmetric and might not satisfy the triangle inequality);
- since ψ is strictly convex, it holds $D_\psi(x,y) \ge 0$ for all $(x,y) \in Z \times \mathrm{cl}(Z)$ and $D_\psi(x,y) = 0$ if and only if $x = y$.

Example (half square Euclidean distance)

For $Z = \mathbb{R}^n$ and $\psi : \mathbb{R}^n \to \mathbb{R}$ and $\psi(x) = \frac{1}{2}\|x\|^2$ one obtains

$$D_\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},\ D_\psi(x,y) = \frac{1}{2}\|x-y\|^2.$$

Example (Kullback-Leibler relative entropy distance)

For $Z = \operatorname{int}(\mathbb{R}^n_+)$ and $\psi : \mathbb{R}^n_+ \to \mathbb{R}$ and $\psi(x) = \sum_{i=1}^n x_i \log x_i - x_i$ (with the convention $0 \log 0 = 0$) one obtains

$$D_\psi : \mathbb{R}^n_+ \times \operatorname{int}(\mathbb{R}^n_+) \to \mathbb{R},\ D_\psi(x,y) = \sum_{i=1}^n x_i \log \frac{x_i}{y_i} + y_i - x_i.$$

Proximal-like algorithm with Bregman functions for $H = \mathbb{R}^n$

Initialization: Choose $x_0 \in \operatorname{dom} f$ and set $k := 0$

For $k \geq 0$: Choose $\gamma_k > 0$ and set

$$x_{k+1} := \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{\gamma_k} D_\psi(x, x_k) \right\}$$

- The proximal point algorithm is well-defined!

Facts on the proximal-like algorithm with Bregman functions

If $\operatorname{dom} f \subseteq Z$, the following statements are true:

- For every $k \geq 0$ it holds $f(x_{k+1}) \leq f(x_k)$;
- For every $k \geq 1$ and every $x \in \operatorname{argmin} f$ it holds $D_\psi(x, x_{k+1}) \leq D_\psi(x, x_k)$;
- Fundamental estimate: For every $k \geq 1$ and every $x \in \operatorname{cl}(Z)$ it holds

$$f(x_k) - f(x) \leq \frac{D_\psi(x, x_0)}{\sigma_{k-1}} - \frac{D_\psi(x, x_k)}{\sigma_{k-1}} - \sum_{j=1}^{k-1} \frac{\sigma_j}{\gamma_j} D_\psi(x_j, x_{j-1}).$$

Convergence of the proximal-like algorithm with Bregman functions

For every $k \geq 1$ and every $x \in \operatorname{cl}(Z)$ it holds

$$f(x_k) - f(x) \leq \frac{D_\psi(x, x_0)}{\sigma_{k-1}}.$$

- If $\lim_{k \to +\infty} \sigma_k = +\infty$, then $f(x_k) \to v(P)(k \to +\infty)$;
- If, additionally, $\operatorname{argmin} f \neq \emptyset$, then $(x_k)_{k \geq 0}$ converges to a minimizer of f and

$$f(x_k) - v(P) \leq \frac{D_\psi(x, x_0)}{2\sigma_{k-1}} \ \forall k \geq 1 \ \forall x \in \operatorname{argmin} f.$$

Tikhonov regularization algorithm

Tikhonov regularization algorithm

```
Initialization:   Choose x_0 ∈ dom f and set k := 0
      For k ≥ 0 :  Choose ε_k > 0 and set
```

$$x_{k+1} := \operatorname*{argmin}_{x\in H}\left\{f(x)+\frac{\varepsilon_k}{2}\|x\|^2\right\}$$

- The Tikhonov regularization algorithm is well-defined!

Facts on the Tikhonov regularization algorithm

or every $k \geq 0$ let be $s_{k+1} := -\varepsilon_k x_{k+1}$. It holds $s_{k+1} \in \partial f(x_{k+1})$, since

$$0 \in \partial\left(f+\frac{\varepsilon_k}{2}\|\cdot\|^2\right)(x_{k+1}) = \partial f(x_{k+1}) + \varepsilon_k x_{k+1} = \partial f(x_{k+1}) - s_{k+1}.$$

Convergence of the Tikhonov regularization algorithm

Let $(\varepsilon_k)_{k\geq 0}$ be such that $\lim_{k\to+\infty}\varepsilon_k = 0$.

- Then $(x_k)_{k\geq 0}$ converges strongly if and only if $\operatorname{argmin} f$ is nonempty;
- In this case, $(x_k)_{k\geq 0}$ converges strongly to $P_{\operatorname{argmin} f}(0)$, which is nothing else than the unique optimal solution of the problem

$$\inf_{x\in\operatorname{argmin} f}\|x\|.$$

Example (Moore-Penrose inverse)

For K a real Hilbert space, $A : H \to K$ a linear continuous operator with $\operatorname{ran} A := A(H)$ closed and $y \in K$, the equation $Az = y$ has at least one least-squares solution, i.e., an optimal solution of the problem

$$\min_{x \in H} \frac{1}{2}\|Ax - y\|^2.$$

The element $x \in H$ is a least-squares solution to $Az = y$ if and only if $A^*Ax = A^*y$. The Moore-Penrose inverse of A is the linear continuous operator $A^\dagger : K \to H$ defined as

$$A^\dagger(y) = P_{\{x \in H : A^*Ax = A^*y\}}(0).$$

- If A^*A is invertible, then $A^\dagger = (A^*A)^{-1}A$. If A is invertible, then $A^\dagger = A^{-1}$.
- Let be $\varepsilon_k > 0$ for all $k \geq 0$ with $\lim_{k \to +\infty} \varepsilon_k = 0$. Then

$$x_k = (A^*A + \varepsilon_k \operatorname{Id})^{-1}A^*y = \operatorname*{argmin}_{x \in H} \left\{ \tfrac{1}{2}\|Ax - y\|^2 + \tfrac{\varepsilon_k}{2}\|x\|^2 \right\} \ \forall k \geq 0.$$

Consequently,

$$\lim_{k \to +\infty} (A^*A + \varepsilon_k \operatorname{Id})^{-1}A^*y = A^\dagger(y).$$

Generalized regularization function

Let $r : H \to \overline{\mathbb{R}}$ be a proper, strictly convex, coercive and lower semicontinuous function with $\operatorname{argmin} f \cap \operatorname{dom} r \neq \emptyset$. Then the optimization problem

$$\min_{x \in \operatorname{argmin} f} r(x)$$

has an unique optimal solution.

Generalized Tikhonov-type regularization algorithm

Initialization: Choose $x_0 \in \operatorname{dom} f$ and set $k := 0$

For $k \geq 0$: Choose $\varepsilon_k > 0$ and set

$$x_{k+1} := \operatorname*{argmin}_{x \in H} \{f(x) + \varepsilon_k r(x)\}$$

- The generalized Tikhonov-type regularization algorithm is well-defined!

Convergence of the generalized Tikhonov-type regularization algorithm

Let $(\varepsilon_k)_{k \geq 0}$ be such that $\lim_{k \to +\infty} \varepsilon_k = 0$.

- Then $(x_k)_{k \geq 0}$ converges weakly to $\operatorname{argmin}_{\operatorname{argmin} f} r$. Moreover, $\lim_{k \to +\infty} r(x_k) = r(\operatorname{argmin}_{\operatorname{argmin} f} r)$;
- If r is strongly convex, then $(x_k)_{k \geq 0}$ converges strongly to $\operatorname{argmin}_{\operatorname{argmin} f} r$.

Splitting methods

The splitting methods are motivated by the need to solve optimization problems of the form

$$\inf_{x\in H}\{f(x)+g(x)\}.$$

One should notice that usable formulae for

$$\operatorname{prox}_{f+g},$$

namely, formulae involving prox_f and prox_g are in general not available!

Forward-Backward splitting

Let $f : H \to \overline{\mathbb{R}}$ be a proper, convex and lower semicontinuous function and $g : H \to \mathbb{R}$ a convex and Fréchet differentiable function with β-Lipschitz continuous gradient. We consider the optimization problem

$$\inf_{x\in H}\{f(x)+g(x)\},$$

for which we assume that $\operatorname{argmin}(f+g)\neq\emptyset$.

A characterization of the optimal solution as starting point

For $\gamma > 0$ one has

$$x \in \operatorname{argmin}(f+g) \Leftrightarrow 0 \in \partial(f+g)(x) \Leftrightarrow 0 \in \partial f(x) + \partial g(x) \Leftrightarrow -\nabla g(x) \in \partial f(x)$$

$$\Leftrightarrow (x-\gamma\nabla g(x)) - x \in \partial(\gamma f)(x) \Leftrightarrow x = \operatorname{prox}_{\gamma f}(x-\gamma\nabla g(x)).$$

Forward-Backward algorithm

`Initialization:`	`Choose` $x_0 \in H$ `and set` $k := 0$
`For` $k \geq 0$:	`Choose` $\gamma_k > 0$ `and set` $x_{k+1} := \text{prox}_{\gamma_k f}(x_k - \gamma_k \nabla g(x_k))$

The case $f = 0$

The Forward-Backward algorithm reduces to the gradient method:

`Initialization:`	`Choose` $x_0 \in H$ `and set` $k := 0$
`For` $k \geq 0$:	`Choose` $\gamma_k > 0$ `and set` $x_{k+1} := x_k - \gamma_k \nabla g(x_k)$

The case $g = 0$

The Forward-Backward algorithm reduces to the proximal point algorithm:

`Initialization:`	`Choose` $x_0 \in H$ `and set` $k := 0$
`For` $k \geq 0$:	`Choose` $\gamma_k > 0$ `and set` $x_{k+1} := \text{prox}_{\gamma_k f}(x_k)$

Convergence of the Forward-Backward algorithm

Let $(\gamma_k)_{k \geq 0}$ be such that

$$0 < \inf_{k \geq 0} \gamma_k \leq \sup_{k \geq 0} \gamma_k < \frac{2}{\beta}.$$

- Then $(x_k)_{k \geq 0}$ converges weakly to an element in $\text{argmin}(f + g)$ and $(\nabla g(x_k))_{k \geq 0}$ converges to $\nabla g(x)$ for every $x \in \text{argmin}(f + g)$;
- If f or g is strongly convex, then $(x_k)_{k \geq 0}$ converges strongly to the unique element in $\text{argmin}(f + g)$.

Linear convergence of the Forward-Backward algorithm

Assuming, additionally, that $f : H \to \overline{\mathbb{R}}$ is strongly convex with modulus $\alpha > 0$ and that we are in one of the following two situations:

- $(\gamma_k)_{k\geq 0}$ is such that

$$0 < \inf_{k\geq 0} \gamma_k \text{ and } \xi := \sup_{k\geq 0} \left(\frac{\sqrt{1+\gamma_k^2\beta^2}}{1+\alpha\gamma_k} \right) < 1;$$

- $(\gamma_k)_{k\geq 0}$ is such that $\gamma_k := \gamma \in \left(0, \frac{2}{\beta}\right)$ for all $k \geq 0$ and $\xi := \frac{1}{1+\alpha\gamma}$;

then $(x_k)_{k\geq 0}$ converges (strongly) linear with constant $\xi \in (0,1)$ to the unique element in $\bar{x} \in \operatorname{argmin}(f+g)$, namely

$$\|x_{k+1} - \bar{x}\| \leq \xi \|x_k - \bar{x}\| \ \forall k \geq 0.$$

Forward-Backward algorithm: a variant incorporating relaxation parameters

Initialization:	Choose $x_0 \in H$ and set $k := 0$
For $k \geq 0$:	Choose $\gamma_k > 0$ and set $y_k := x_k - \gamma_k \nabla g(x_k)$
	Choose $\lambda_k > 0$ and set $x_{k+1} := x_k + \lambda_k(\operatorname{prox}_{\gamma_k f}(y_k) - x_k)$

- If $\lambda_k = 1$ for all $k \geq 0$, then one rediscovers the classical version of the Forward-Backward algorithm;

Let be $\varepsilon \in \left(0, \min\{1, \frac{1}{\beta}\}\right)$ fixed and $(\lambda_k)_{k\geq 0}$ and $(\gamma_k)_{k\geq 0}$ such that

$$\gamma_k \in \left[\varepsilon, \frac{2}{\beta} - \varepsilon\right] \text{ and } \lambda_k \in [\varepsilon, 1] \ \forall k \geq 0.$$

- Then $(x_k)_{k\geq 0}$ converges weakly to an element in $\operatorname{argmin}(f+g)$ and $(\nabla g(x_k))_{k\geq 0}$ converges to $\nabla g(x)$ for every $x \in \operatorname{argmin}(f+g)$;
- If f or g is strongly convex, then $(x_k)_{k\geq 0}$ converges strongly to the unique element in $\operatorname{argmin}(f+g)$.

Particular instance (of the classical version): the projected gradient algorithm

Assuming that $f := \delta_S$, where $S \subseteq H$ is a convex closed set such that $\operatorname{argmin}_S g \neq \emptyset$, the problem to be solved becomes

$$\inf_{x \in S} g(x).$$

Since $\operatorname{prox}_{\gamma f} = \operatorname{prox}_f = P_S$ for all $\gamma > 0$, the Forward-Backward algorithm gives in this case rise to the so-called projected gradient algorithm:

Initialization:	Choose $x_0 \in H$ and set $k := 0$
For $k \geq 0$:	Choose $\gamma_k > 0$ and set $x_{k+1} := P_S(x_k - \gamma_k \nabla g(x_k))$.

Particular instance (of the classical version): the Backward-Backward algorithm

For $f, g : H \to \overline{\mathbb{R}}$ be a proper, convex and lower semicontinuous functions we consider the optimization problem

$$\inf_{x \in H}\{f(x) + {}^1 g(x)\},$$

for which we assume that $\operatorname{argmin}(f + {}^1 g) \neq \emptyset$.

- Recall that ${}^1 g : H \to \mathbb{R}$ is convex and Fréchet differentiable in H, $\nabla({}^1 g)(x) = x - \operatorname{prox}_g(x)$ for all $x \in H$ and $\nabla({}^1 g)$ is 1-Lipschitz continuous.

Taking $\gamma_k = 1$ for all $k \geq 0$, the Forward-Backward algorithm gives rise to the so-called Backward-Backward algorithm:

```
Initialization:   Choose x_0 ∈ H and set k := 0
      For k ≥ 0 :   Set x_{k+1} := prox_f(prox_g(x_k))
```

- $(x_k)_{k \geq 0}$ converges weakly to an element in $\operatorname{argmin}(f + {}^1 g)$;
- If f is strongly convex, then $(x_k)_{k \geq 0}$ converges strongly to the unique element in $\operatorname{argmin}(f + {}^1 g)$.

Particular instance (of the Backward-Backward algorithm): the alternating projections algorithm

Taking $f := \delta_S$ and $g := \delta_T$, where $S, T \subseteq H$ are convex closed sets, such that one of them is bounded, the problem

$$\inf_{x \in H} \{f(x) +^1 g(x)\},$$

becomes

$$\inf_{x \in S} \tfrac{1}{2} d_T^2,$$

which amounts to finding an element in S at closest distance from T.

- If S is bounded, then $\operatorname{argmin}_S \left(\frac{1}{2} d_T^2\right) = \operatorname{argmin}_S d_T \neq \emptyset$, since $\frac{1}{2} d_T^2$ is continuous;
- If T is bounded, then $\operatorname{argmin}_S \left(\frac{1}{2} d_T^2\right) = \operatorname{argmin}_S d_T \neq \emptyset$, since $\frac{1}{2} d_T^2$ is coercive.

Since $\operatorname{prox}_f = P_S$ and $\operatorname{prox}_g = P_T$, the Backward-Backward algorithm yield the alternating projections algorithm:

```
Initialization:   Choose x_0 ∈ H and set k := 0
      For k ≥ 0 :   Set x_{k+1} := P_S(P_T(x_k))
```

- $(x_k)_{k \geq 0}$ converges weakly to an element in $\operatorname{argmin}_S d_T \neq \emptyset$.

The case when g is composed with a linear continuous operator

Let $f : H \to \overline{\mathbb{R}}$ be a proper, convex and lower semicontinuous function, $g : K \to \mathbb{R}$ a convex and Fréchet differentiable function with β-Lipschitz continuous gradient and $A : H \to K$ a nonzero linear and continuous operator. We consider the optimization problem

$$\inf_{x \in H} \{f(x) + g(Ax)\},$$

for which we assume that $\operatorname{argmin}(f + g \circ A) \neq \emptyset$.

- Notice that $g \circ A : H \to \mathbb{R}$ is convex and Fréchet differentiable in H, $\nabla(g \circ A)(x) = A^*(\nabla g(Ax))$ for all $x \in H$ and $\nabla(g \circ A)$ is $\beta\|A\|^2$-Lipschitz continuous.

The Forward-Backward algorithm reads:

Initialization: Choose $x_0 \in H$ and set $k := 0$
For $k \geq 0$: Choose $\gamma_k > 0$ and set $x_{k+1} := \operatorname{prox}_{\gamma_k f}(x_k - \gamma_k A^* \nabla g(Ax_k))$

Let $(\gamma_k)_{k \geq 0}$ be such that

$$0 < \inf_{k \geq 0} \gamma_k \leq \sup_{k \geq 0} \gamma_k < \frac{2}{\beta\|A\|^2}.$$

- Then $(x_k)_{k \geq 0}$ converges weakly to an element in $\operatorname{argmin}(f + g \circ A)$ and $(\nabla g(Ax_k))_{k \geq 0}$ converges to $\nabla g(Ax)$ for every $x \in \operatorname{argmin}(f + g \circ A)$;
- If f is strongly convex, then $(x_k)_{k \geq 0}$ converges strongly to the unique element in $\operatorname{argmin}(f + g \circ A)$.

Particular instance: the projected Landweber algorithm

Taking $f = \delta_S$, where $S \subseteq H$ is a nonempty, convex and closed set, and $g : K \to \mathbb{R}$, $g(z) = \frac{1}{2}\|z - y\|^2$, where $y \in K$, which is a convex and Fréchet differentiable function with 1-Lipschitz continuous gradient, the problem

$$\inf_{x \in H} \{f(x) + g(Ax)\}$$

yields the constrained least-squares problem

$$\inf_{x \in S} \frac{1}{2}\|Ax - y\|^2.$$

We assume that $\operatorname{argmin}_S \left(\frac{1}{2}\|A(\cdot) - y\|^2\right) \neq \emptyset$.

The above Forward-Backward algorithm gives rise to the projected Landweber algorithm:

```
Initialization:    Choose x_0 ∈ H and set k := 0
      For k ≥ 0 :   Choose γ_k > 0 and set x_{k+1} := P_S(x_k + γ_k A^*(y − Ax_k))
```

Let $(\gamma_k)_{k \geq 0}$ be such that

$$0 < \inf_{k \geq 0} \gamma_k \leq \sup_{k \geq 0} \gamma_k < \frac{2}{\|A\|^2}.$$

- Then $(x_k)_{k \geq 0}$ converges weakly to an element in $\operatorname{argmin}_S \left(\frac{1}{2}\|A(\cdot) - y\|^2\right)$.

Douglas-Rachford algorithm

Let $f, g : H \to \overline{\mathbb{R}}$ be a proper, convex and lower semicontinuous functions with $\operatorname{dom} f \cap \operatorname{dom} g \neq \emptyset$ and the optimization problem

$$\inf_{x \in H} \{f(x) + g(x)\}.$$

We assume that $\operatorname{argmin}(f+g) \neq \emptyset$ and that one of the following qualification conditions:

- (AB): $0 \in \operatorname{sqri}(\operatorname{dom} f - \operatorname{dom} g)$;
- (B): $\operatorname{epi} f^* + \operatorname{epi} g^*$ is closed;

is fulfilled.

- Recall that, in this circumstances, $\partial(f+g)(x) = \partial f(x) + \partial g(x)$ for all $x \in H$.

A characterization of the optimal solution as starting point

For $\gamma > 0$ one has

$$x \in \operatorname{argmin}(f+g) \Leftrightarrow 0 \in \partial(f+g)(x) \Leftrightarrow 0 \in \partial f(x) + \partial g(x) \Leftrightarrow \exists y \in H : x - y \in \gamma \partial f(x)$$

$$\text{and } y - x \in \gamma \partial g(x) \Leftrightarrow \exists y \in H : (2x - y) - x \in \partial(\gamma f)(x) \text{ and } y - x \in \partial(\gamma g)(x)$$

$$\Leftrightarrow \exists y \in H : x = \operatorname{prox}_{\gamma f}(2x - y) \text{ and } x = \operatorname{prox}_{\gamma g}(y)$$

$$\Leftrightarrow \exists y \in H : x = \operatorname{prox}_{\gamma g}(y) \text{ and } 0 = \operatorname{prox}_{\gamma f}(2x - y) - x$$

$$\Leftrightarrow \exists y \in H : x = \operatorname{prox}_{\gamma g}(y) \text{ and } y = y + \operatorname{prox}_{\gamma f}(2x - y) - x.$$

Douglas-Rachford algorithm

`Initialization:` `Choose` $\gamma > 0$, $y_0 \in H$ `and set` $k := 0$

`For` $k \geq 0$: `Set` $x_k := \text{prox}_{\gamma g}(y_k)$

$$y_{k+1} := y_k + \left(\text{prox}_{\gamma f}(2x_k - y_k) - x_k\right)$$

The case $f = 0$

The Douglas-Rachford algorithm reduces to the proximal point algorithm:

`Initialization:` `Choose` $\gamma > 0$, $y_0 \in H$ `and set` $k := 0$

`For` $k \geq 0$: `Set` $x_k := \text{prox}_{\gamma g}(y_k)$

$$y_{k+1} := x_k$$

The case $g = 0$

The Douglas-Rachford algorithm reduces to the proximal point algorithm:

`Initialization:` `Choose` $\gamma > 0$, $y_0 \in H$ `and set` $k := 0$

`For` $k \geq 0$: `Set` $x_k := y_k$

$$y_{k+1} := \text{prox}_{\gamma g}(x_k)$$

Convergence of the Douglas-Rachford algorithm

There exists $x \in H$ with $\text{prox}_{\gamma g}(x) \in \text{argmin}(f + g)$ such that $(y_k)_{k \geq 0}$ converges weakly to x and:

- $(x_k)_{k \geq 0}$ converges weakly to $\text{prox}_{\gamma g}(x)$;
- whenever f or g is strongly convex, $(x_k)_{k \geq 0}$ converges strongly to $\text{prox}_{\gamma g}(x)$, which is the unique element in $\text{argmin}(f + g)$.

Douglas-Rachford algorithm: a variant incorporating relaxation parameters

```
Initialization:   Choose γ > 0, y_0 ∈ H and set k := 0
     For k ≥ 0 :  Set x_k := prox_{γg}(y_k)
                  Choose λ_k > 0 and set y_{k+1} := y_k + λ_k(prox_{γf}(2x_k − y_k) − x_k)
```

- If $\lambda_k = 1$ for all $k \geq 0$, then one rediscovers the classical version of the Douglas-Rachford algorithm;

Let be $\varepsilon \in (0,1)$ fixed and $(\lambda_k)_{k\geq 0}$ such that

$$\lambda_k \in [\varepsilon, 2-\varepsilon] \ \forall k \geq 0.$$

There exists $x \in H$ with $\operatorname{prox}_{\gamma g}(x) \in \operatorname{argmin}(f+g)$ such that $(y_k)_{k\geq 0}$ converges weakly to x and:

- $(x_k)_{k\geq 0}$ converges weakly to $\operatorname{prox}_{\gamma g}(x)$;
- whenever f or g is strongly convex, $(x_k)_{k\geq 0}$ converges strongly to $\operatorname{prox}_{\gamma g}(x)$, which is the unique element in $\operatorname{argmin}(f+g)$.

Particular instance (of the classical version): determining an element in a set S at closest distance from a set T

Taking $f := \delta_S$ and $g := \frac{1}{2}d_T^2$, where $S, T \subseteq H$ are convex closed sets, such that one of them is bounded, the problem

$$\inf_{x \in H} \{f(x) + g(x)\},$$

becomes

$$\inf_{x \in S} \frac{1}{2}d_T^2,$$

which amounts to finding an element in S at closest distance from T.

- If S is bounded, then $\operatorname{argmin}_S \left(\frac{1}{2}d_T^2\right) = \operatorname{argmin}_S d_T \neq \emptyset$, since $\frac{1}{2}d_T^2$ is continuous;
- If T is bounded, then $\operatorname{argmin}_S \left(\frac{1}{2}d_T^2\right) = \operatorname{argmin}_S d_T \neq \emptyset$, since $\frac{1}{2}d_T^2$ is coercive.
- Since $\operatorname{dom} g = H$, the qualification condition (AB) is fulfilled;

The Douglas-Rachford algorithm with $\gamma = 1$ yields, since $\operatorname{prox}_f = P_S$ and $\operatorname{prox}_g(x) = \frac{1}{2}(x + P_T(x))$:

```
Initialization:   Choose y_0 ∈ H and set k := 0
    For k ≥ 0 :   Set x_k := 1/2 (x_k + P_T(x_k))
                  y_{k+1} := P_S(P_T(y_k)) + y_k − x_k
```

There exists $x \in H$ with $\frac{1}{2}(x + P_T(x)) \in \operatorname{argmin}_S d_T$ such that $(y_k)_{k \geq 0}$ converges weakly to x and $(x_k)_{k \geq 0}$ converges weakly to $\frac{1}{2}(x + P_T(x))$.

The case when g is composed with a linear continuous operator

Let $f, g : H \to \overline{\mathbb{R}}$ be proper, convex and lower semicontinuous functions, $A : H \to K$ a nonzero linear and continuous operator with $A(\operatorname{dom} f) \cap \operatorname{dom} g \neq \emptyset$ and the optimization problem

$$\inf_{x \in H} \{f(x) + g(Ax)\}.$$

We assume that $\operatorname{argmin}(f + g \circ A) \neq \emptyset$, $AA^* = \alpha \operatorname{Id}$, for $\alpha > 0$, and that one of the following qualification conditions:

- (AB): $0 \in \operatorname{sqri}(A(\operatorname{dom} f) - \operatorname{dom} g)$;
- (B): $A^* \times \operatorname{id}_{\mathbb{R}}(\operatorname{epi} f^*) + \operatorname{epi} g^*$ is closed;

is fulfilled.

- Recall that, in this circumstances, $\partial(f + g \circ A)(x) = \partial f(x) + \partial A^*(g(Ax))$ for all $x \in H$.

For $\gamma > 0$ one has

$$x \in \operatorname{argmin}(f + g \circ A) \Leftrightarrow \exists y \in H : x = \operatorname{prox}_{\gamma g \circ A}(y) \text{ and } x = \operatorname{prox}_{\gamma f}(2x - y)$$

$$\Leftrightarrow \exists y \in H : x = y + \frac{1}{\alpha} A^*(\operatorname{prox}_{\alpha \gamma g}(Ay) - Ay) \text{ and } x = \operatorname{prox}_{\gamma f}(2x - y).$$

The case when g is composed with a linear continuous operator (continuation)

The Douglas-Rachford algorithm reads:

`Initialization:` `Choose` $\gamma > 0$, $y_0 \in H$ `and set` $k := 0$

`For` $k \geq 0$: `Set` $x_k := y_k + \frac{1}{\alpha} A^*(\text{prox}_{\alpha\gamma g}(Ay_k) - Ay_k)$

$$y_{k+1} := y_k + \Big(\text{prox}_{\gamma f}(2x_k - y_k) - x_k\Big)$$

There exists $x \in H$ with $\text{prox}_{\gamma g \circ A}(x) \in \text{argmin}(f + g)$ such that $(y_k)_{k\geq 0}$ converges weakly to x and:

- $(x_k)_{k\geq 0}$ converges weakly to $\text{prox}_{\gamma g \circ A}(x)$;
- whenever f is strongly convex, $(x_k)_{k\geq 0}$ converges strongly to $\text{prox}_{\gamma g \circ A}(x)$, which is the unique element in $\text{argmin}(f + g)$.

Particular instance

By taking $g = \delta_{\{y\}}$, for $y \in A(\operatorname{dom} f)$, the optimization problem

$$\inf_{x \in H} \{f(x) + g(Ax)\}$$

becomes

$$\inf_{\substack{x \in H \\ Ax = y}} f(x).$$

We assume that $\operatorname{argmin}_{A^{-1}(y)} f \neq \emptyset$, where $A^{-1}(y) = \{x \in H : Ax = y\}$, $AA^* = \alpha \operatorname{Id}$, for $\alpha > 0$, and that one of the following qualification conditions:

- (AB): $y \in \operatorname{sqri}(A(\operatorname{dom} f))$;
- (B): $A^* \times \operatorname{id}_{\mathbb{R}}(\operatorname{epi} f^*) + \{(y^*, r) \in K^* \times \mathbb{R} : \langle y^*, y \rangle \leq r\}$ is closed;

is fulfilled.

Since $\operatorname{prox}_{\alpha\gamma g}(z) = y$ for all $z \in K$, the Douglas-Rachford algorithm reads:

```
Initialization:   Choose γ > 0, y_0 ∈ H and set k := 0
       For k ≥ 0 :   Set x_k := y_k + (1/α) A*(y − A y_k)
                     y_{k+1} := y_k + (prox_{γf}(2x_k − y_k) − x_k)
```

There exists $x \in H$ with $P_{A^{-1}(y)}(x) \in \operatorname{argmin}_{A^{-1}(y)} f$ such that $(y_k)_{k \geq 0}$ converges weakly to x and:

- $(x_k)_{k \geq 0}$ converges weakly to $P_{A^{-1}(y)}(x)$;
- whenever f is strongly convex, $(x_k)_{k \geq 0}$ converges strongly to $P_{A^{-1}(y)}(x)$, which is the unique element in $\operatorname{argmin}_{A^{-1}(y)} f$.

References

- H.H. Bauschke, P.-L. Combettes (2011): *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer-Verlag, New York
- D.P. Bertsekas (1999): *Nonlinear Programming*, Athena Scientific, Nashua
- G. Chen, M. Teboulle (1993): *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM Journal on Optimization 3(3), 538-543
- J. Douglas, H.H. Rachford (1956): *On the numerical solution of the heat conduction problem in 2 and 3 space variables*, Transactions of the American Mathematical Society 82, 421-439
- O. Güler (1991): *On the convergence of the proximal point algorithm for convex minimization*, SIAM Journal on Control and Optimization 29(2), 403-419
- P.L. Lions, B. Mercier (1979): *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis 16(6), 964-979
- R.T. Rockafellar (1976): *Monotone operators and the proximal point algorihm*, SIAM Journal on Control and Optimization 29(2), 877-898