Numerical Optimization Methods in Imaging

Part III: Forward-backward and Douglas-Rachford algorithms

jean-christophe@pesquet.eu

PhD Summer School MMLIA – Bologna

Fixed points

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

Let $(\nu,\gamma) \in \,]0,+\infty[^2$. Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν-Lipschitzian gradient.
Let $T = \mathrm{prox}_{\gamma f} \circ (\mathrm{Id} - \gamma \nabla g)$.

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

Fixed points

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

Let $(\nu, \gamma) \in \,]0, +\infty[^2$. Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν-Lipschitzian gradient.
Let $T = \mathrm{prox}_{\gamma f} \circ (\mathrm{Id} - \gamma \nabla g)$.

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

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

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

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

Fixed points

Let $(\nu,\gamma) \in \ ]0,+\infty[^2$. Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν-Lipschitzian gradient.
Let $T = \operatorname{prox}_{\gamma f} \circ (\mathrm{Id} - \gamma \nabla g)$.

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

Proof: $\operatorname{prox}_{\gamma f}$ is $\alpha_1 = 1/2$-averaged.
$\gamma \in]0,2/\nu[\Rightarrow \mathrm{Id} - \gamma\nabla g$ is $\alpha_2 = \gamma\nu/2$-averaged.
It follows that T is α-averaged with

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

Forward-Backward with fixed step-size

Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν-Lipschitzian gradient where $\nu \in \,]0, +\infty[$.
Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1, 2[$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
We assume that $\operatorname{Argmin}(f + g) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

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

Forward-Backward with fixed step-size

Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν-Lipschitzian gradient where $\nu \in\ ]0,+\infty[$.
Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1,2[$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,\delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
We assume that $\operatorname{Argmin}(f+g) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

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

<u>Exercice</u>: Prove this result.

Forward-Backward with fixed step-size

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

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

T is δ^{-1}-averaged and $\mathrm{Fix}\, T = \mathrm{Argmin}(f+g)$.
The FB algorithm is thus have an instance of the generalized form of KM iteration.

Forward-Backward with fixed step-size

Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν-Lipschitzian gradient where $\nu \in \,]0,+\infty[$.
Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1,2[$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,\delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
We assume that $\operatorname{Argmin}(f+g) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

Then, $(x_n)_{n\in\mathbb{N}}$ converges weakly to a minimizer x of $f+g$.
In addition, if $T = \operatorname{prox}_{\gamma f} \circ (\operatorname{Id} - \gamma \nabla g)$, then

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

Forward-Backward with fixed step-size

Proof:

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

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

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

Forward-Backward with fixed step-size

Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be Fréchet differentiable with a ν-Lipschitzian gradient where $\nu \in \,]0,+\infty[$.
Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1,2[$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,\delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
We assume that $\operatorname{Argmin}(f+g) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

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

Forward-Backward with fixed step-size

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

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

By using the cocoercivity of ∇g,

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

This yields

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

Forward-Backward with fixed step-size

Proof:
This yields

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

Since $x_n \rightharpoonup x$, $(x_n)_{n\in\mathbb{N}}$ is bounded and so is $(\|x_n - x\|)_{n\in\mathbb{N}}$.
In addition $x_n - Tx_n \to 0$.
Therefore $\nabla g(x_n) \to \nabla g(x)$.

A useful lemma

Three point inequality

Let $(\nu, \gamma) \in \,]0, +\infty[^2$. Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν-Lipschitzian gradient.
Let $h = f + g$.
Let $(x, z) \in \mathcal{H}^2$ and let $p = \mathrm{prox}_{\gamma f}(x - \gamma \nabla g(x))$.
Then

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

A useful lemma

Three point inequality

Let $(\nu,\gamma) \in \,]0,+\infty[^2$. Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν-Lipschitzian gradient.
Let $h = f + g$.
Let $(x,z) \in \mathcal{H}^2$ and let $p = \mathrm{prox}_{\gamma f}(x - \gamma\nabla g(x))$.
Then

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

Proof: According to the descent lemma,

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

and, by the tangent inequality,

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

We have thus

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

A useful lemma

Proof: We have thus

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

In addition

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

We deduce that

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

A useful lemma

<u>Proof</u>: We have thus

$$\begin{cases} g(p) \leq g(z) - \langle \nabla g(x) \mid z - p \rangle + \frac{\nu}{2}\|p - x\|^2 \\ f(z) \geq f(p) + \frac{1}{\gamma} \langle x - p \mid z - p \rangle - \langle \nabla g(x) \mid z - p \rangle \end{cases}$$

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

Forward-Backward with varying stepsize

Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in \,]0, +\infty[$.
Let $(\gamma_n)_{n\in\mathbb{N}}$ in $[\underline{\gamma}, \overline{\gamma}]$ where $0 < \underline{\gamma} < \overline{\gamma} < 2/\nu$ and
let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\underline{\lambda}, 1]$ with $0 < \underline{\lambda} \leq 1$.
We assume that $\operatorname{Argmin}(f+g) \neq \varnothing$. Let $x_0 \in \operatorname{dom} f$ and

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

Then,

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

Forward-Backward with varying stepsize

Let $f \in \Gamma_0(\mathcal{H})$ be strongly convex.
Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in \left]0, +\infty\right[$.
Let $(\gamma_n)_{n\in\mathbb{N}}$ in $[\underline{\gamma}, \overline{\gamma}]$ where $0 < \underline{\gamma} < \overline{\gamma} < 2/\nu$ and
let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\underline{\lambda}, 1]$ with $0 < \underline{\lambda} \leq 1$.
Let $x_0 \in \mathcal{H}$ and

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

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

Forward-Backward with varying stepsize

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

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

This implies that $f+g$ is strictly convex and coercive, hence has a unique minimizer $\widehat{x}$.
In addition, for every $x \in \mathcal{H}$ and $\gamma \in \,]0,+\infty[$,

$$\mathrm{prox}_{\gamma f} x = \mathrm{prox}_{\frac{\gamma \ell}{1+\gamma\eta}}\Big(\frac{x}{1+\gamma\eta}\Big).$$

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

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

Forward-Backward with varying stepsize

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

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

and

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

with

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

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

Forward-Backward with varying stepsize

Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in \,]0, +\infty[$.
Let $(\gamma_n)_{n\in\mathbb{N}}$ in $[\underline{\gamma}, \overline{\gamma}]$ where $0 < \underline{\gamma} < \overline{\gamma} \leq 1/\nu$.
We assume that $\mathrm{Argmin}(f+g) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

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

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

Forward-Backward with varying stepsize

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

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

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

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

Forward-Backward with varying stepsize

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

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

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

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

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

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

Accelerated version

Let $f \in \Gamma_0(\mathcal{H})$.
Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in \,]0,+\infty[$.
Let $\gamma \in]0, 1/\nu]$ and $\zeta \in [2,+\infty[$.
We assume that $\operatorname{Argmin}(f+g) \neq \varnothing$. Let $x_0 = z_0 \in \mathcal{H}$ and

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

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

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

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

Projected gradient algorithm

Let C be a nonempty closed convex subset of $\mathcal{H}$.
Let $g \in \Gamma_0(\mathcal{H})$ be differentiable with a ν-Lipschitzian gradient where $\nu \in \,]0, +\infty[$.
Let $\gamma \in]0, 2/\nu[$ and $\delta = 2 - \gamma\nu/2 \in]1, 2[$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
We assume that $\underset{x\in C}{\mathrm{Argmin}}\; g(x) \neq \varnothing$. 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.

Gradient descent algorithm

Let $g \in \Gamma_0(\mathcal{H})$ be a differentiable function with a ν-Lipschitzian gradient where $\nu \in \,]0, +\infty[$.
Let $(\gamma_n)_{n\in\mathbb{N}}$ in $[\underline{\gamma}, \overline{\gamma}]$ where $0 < \underline{\gamma} < \overline{\gamma} < 2/\nu$.
We assume that $\operatorname{Argmin} g \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

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

Proximal point algorithm

Let $f \in \Gamma_0(\mathcal{H})$.
Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$. We assume that $\mathrm{Argmin} f \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

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

Iterative thresholding

Problem

Let $\mathcal{G}$ be a Hilbert space and let $L \in \mathcal{B}(\mathbb{R}^K, \mathcal{G})$.
Let $y \in \mathcal{G}$ and let $\chi \in \,]0, +\infty[$. We want to

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

Algorithm

Let $\gamma \in]0, 2/\|L\|^2[$ and $\delta = 2 - \gamma\|L\|^2/2$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
Let $x_0 \in \mathcal{H}$ and

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

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

Iterative thresholding

Algorithm

Let $\gamma \in]0, 2/\|L\|^2[$ and $\delta = 2 - \gamma\|L\|^2/2$.
Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.
Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = x_n - \gamma L^*(Lx_n - y) = (\upsilon_n^{(k)})_{1\le k\le K} \\ p_n = \left(\mathrm{soft}_{[-\gamma\chi,\gamma\chi]}(\upsilon_n^{(k)})\right)_{1\le k\le K} \\ x_{n+1} = x_n + \lambda_n(p_n - x_n). \end{cases}$$

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

Proof: Set $f = \chi\|\cdot\|_1$ and $g = \frac{1}{2}\|L\cdot - y\|^2$.
$f + g \in \Gamma_0(\mathbb{R}^K)$ and is coercive, then $\mathrm{Argmin}(f+g) \neq \varnothing$.
In addition, g is Fréchet differentiable and

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

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

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

Thus ∇g is ν-Lipschitzian with $\nu = \|L^*L\| = \|L\|^2$.
By noticing that, for every $n \in \mathbb{N}$, $p_n = \mathrm{prox}_{\gamma\chi\|\cdot\|_1}(y_n)$, we recognize a FB algorithm for which the conditions of convergence are met.

Graph of a set-valued operator

Let $\mathcal{H}$ be a real Hilbert space.
Let $A : \mathcal{H} \to 2^{\mathcal{H}}$.
The graph of A is

$$\mathrm{gra}A = \big\{\big(x, u\big) \in \mathcal{H}^2 \mid u \in Ax\big\}.$$

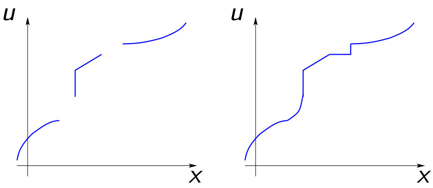
Let $\mathcal{H}$ be a Hilbert space.
Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$.
A^{-1} is the operator from $\mathcal{H}$ to $2^{\mathcal{H}}$ the graph of which is

$$\mathrm{gra}(A^{-1}) = \big\{(u, x) \;\big|\; (x, u) \in \mathrm{gra}A\big\}.$$

Monotone operator: definition

Let $\mathcal{H}$ be a real Hilbert space.
Let $A : \mathcal{H} \to 2^{\mathcal{H}}$.
A is monotone if

$$\big(\forall (x_1, u_1) \in \operatorname{gra} A\big)\big(\forall (x_2, u_2) \in \operatorname{gra} A\big) \qquad \langle u_1 - u_2 \mid x_1 - x_2 \rangle \geq 0.$$



Monotone operator: definition

Let $\mathcal{H}$ be a real Hilbert space.
Let $A: \mathcal{H} \to 2^{\mathcal{H}}$.
A is monotone if

$$\big(\forall (x_1, u_1) \in \operatorname{gra} A\big)\big(\forall (x_2, u_2) \in \operatorname{gra} A\big) \qquad \langle u_1 - u_2 \mid x_1 - x_2 \rangle \geq 0.$$

Let $\mathcal{H}$ be a Hilbert space.
Let $A: \mathcal{H} \to 2^{\mathcal{H}}$.
A is maximally monotone if A is monotone and if there exists no monotone operator $B: \mathcal{H} \to 2^{\mathcal{H}}$ (different from A) such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$.

Resolvent

Let $\mathcal{H}$ be a Hilbert space.
Let $A : \mathcal{H} \to 2^{\mathcal{H}}$.
The revolvent of A is

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

Resolvent

Let $\mathcal{H}$ be a Hilbert space.
Let $A : \mathcal{H} \to 2^{\mathcal{H}}$.
The revolvent of A is

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

Minty theorem

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

$$A \text{ is maximally monotone} \quad \Rightarrow \quad J_A \colon \mathcal{H} \to \mathcal{H}$$

Resolvent

Let $\mathcal{H}$ be a Hilbert space.
Let $A : \mathcal{H} \to 2^{\mathcal{H}}$.
The revolvent of A is

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

Minty theorem

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

$$A \text{ is maximally monotone} \quad \Rightarrow \quad J_A \colon \mathcal{H} \to \mathcal{H}$$

- Example : if $f \in \Gamma_0(\mathcal{H})$, then ∂f is maximally monotone and $J_{\partial f} = \mathrm{prox}_f$.

Forward-Backward algorithm

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

Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator .

Let $B\colon \mathcal{H} \to \mathcal{H}$ be a ν^{-1}-cocoercive operator with $\nu \in]0, +\infty[$.

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

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

We assume that $\operatorname{zer}(A + B) \neq \varnothing$. Let $x_0 \in \mathcal{H}$ and

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

$(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(A + B)$.

Example : $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{H})$ with ν-Lipschitzian gradient
$\Rightarrow$ Proximal gradient algorithm

Recap of optimization strategies

Motivation

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

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

Possible solutions :

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

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

Reflection of the prox

Let $\gamma \in \,]0, +\infty[$ and let $f \in \Gamma_0(\mathcal{H})$.
The reflection of the proximity operator defined as

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

is nonexpansive.

Fixed points

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

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

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

Fixed points

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

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

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

<u>Proof:</u> Let $x \in \mathcal{H}$.

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

Douglas-Rachford algorithm

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

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

Douglas-Rachford algorithm

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

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

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

We assume that $\operatorname{zer}(\partial f + \partial g) \neq \varnothing$. 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:

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

Douglas-Rachford algorithm

Proof: Let $T = \mathrm{rprox}_{\gamma f} \circ \mathrm{rprox}_{\gamma g}$. T is nonexpansive and $\varnothing \neq \mathrm{zer}\,(\partial f + \partial g) = \mathrm{prox}_{\gamma g}(\mathrm{Fix}\, T) \Rightarrow \mathrm{Fix}\, T \neq \varnothing$.
Moreover, for every $n \in \mathbb{N}$,

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

$\Rightarrow$ Krasnosel'skii-Mann algorithm with relaxation factors $(\lambda_n/2)_{n\in\mathbb{N}}$.
We deduce that $Tx_n - x_n \to 0$ and $x_n \rightharpoonup \widehat{x} \in \mathrm{Fix}\, T$.

Douglas-Rachford algorithm

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

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

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

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

$\widehat{y} \in \mathrm{zer}\,\big(\partial(f+g)\big) = \mathrm{Argmin}(f+g)$.

Douglas-Rachford algorithm

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

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

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

We assume that $\operatorname{zer}(\partial f + \partial g) \neq \varnothing$. 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:

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

Douglas-Rachford algorithm

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

$$y_n \to \mathrm{prox}_{\gamma g}\widehat{x} = \widehat{y}.$$

Because $z_n - y_n \to 0$, we deduce that $z_n \to \mathrm{prox}_{\gamma g}\widehat{x} = \widehat{y}$.

Douglas-Rachford algorithm

Let $\mathcal{H}$ be a Hilbert space.
Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.
Let $\gamma \in \,]0, +\infty[$ and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0,2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$.
We assume that $\operatorname{zer}(\partial f + \partial g) \neq \varnothing$. 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:

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

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

Example

Problem

Let $u \in \mathbb{R}^K$, let $\Phi_u \in \Gamma_0(\mathbb{R}^K)$, and let $\chi \in \,]0,+\infty[$. We want to

$$\underset{x \in \mathbb{R}^K}{\text{minimize}} \;\; \Phi_u(x) + \chi \|x\|_1.$$

Algorithm

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 $\inf \Phi_u > -\infty$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = \operatorname{prox}_{\gamma \Phi_u}(x_n) \\ z_n = \big(\operatorname{soft}_{[-\gamma\chi,\gamma\chi]}(2y_n^{(k)} - x_n^{(k)})\big)_{1\leq k\leq K} \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

Then, $(y_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge to the same solution to the problem.

Example

Algorithm

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 $\inf \Phi_u > -\infty$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} y_n = \mathrm{prox}_{\gamma\Phi_u}(x_n) \\ z_n = \left(\mathrm{soft}_{[-\gamma\chi,\gamma\chi]}(2y_n^{(k)} - x_n^{(k)})\right)_{1\le k\le K} \\ x_{n+1} = x_n + \lambda_n(z_n - y_n). \end{cases}$$

Then, $(y_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ converge to the same solution to the problem.

Proof: Set $f = \Phi_u$ and $g = \|\cdot\|_1$.
$f+g \in \Gamma_0(\mathbb{R}^K)$ and is coercive, then $\mathrm{zer}\,(\partial(f+g)) = \mathrm{Argmin}(f+g) \neq \varnothing$.
In addition since $\|\cdot\|_1$ is finite-valued, $\mathrm{zer}\,(\partial f + \partial g) = \mathrm{zer}\,(\partial(f+g)) \neq \varnothing$.
We recognize a DR algorithm for which the conditions of convergence are met.

Parallel form of Douglas-Rachford

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

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

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

Parallel form of Douglas-Rachford

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

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

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

Remark: $(L^*L)^{-1}L^*$ is the pseudo-inverse of L

Parallel form of Douglas-Rachford

Proof: Let $E = \operatorname{ran} L$. E is closed.
Indeed, let $(Lw_n)_{n\in\mathbb{N}}$ be sequence of E converging to $z \in \mathcal{H}$.
By continuity,

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

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

Parallel form of Douglas-Rachford

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

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

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

Parallel form of Douglas-Rachford

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

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

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

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

Parallel form of Douglas-Rachford

<u>Proof</u>: $f = \iota_E \Rightarrow \mathrm{prox}_{\gamma f} = P_E$. The DR algorithm reads

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

For every $n \in \mathbb{N}$, $P_E y_n = Lc_n$
with $c_n = (L^*L)^{-1}L^* y_n$
and $P_E x_n = Lv_n$ with $v_n = (L^*L)^{-1}L^* x_n$.
In addition

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

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

Parallel form of Douglas-Rachford

Proof: In summary

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

We know that

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

Parallel form of Douglas-Rachford

Proof: In summary

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

We know that

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

Because of the weak continuity of P_E,

$$P_E y_n = Lc_n \rightharpoonup P_E \widehat{y} = \widehat{y} = L\widehat{v}$$

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

$$c_n \rightharpoonup \widehat{v} \in \operatorname{Argmin}(g \circ L).$$

Parallel form of Douglas-Rachford

Proof: In summary

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

In addition

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

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

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

Parallel form of Douglas-Rachford

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

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

Then $v_n \rightharpoonup \widehat{v}$ where $\widehat{v} \in \operatorname{Argmin}(g \circ L)$.

Question: Give the particular case of the algorithm when
$\mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$
$(\forall x = (x_1,\ldots,x_m) \in \mathcal{H})\ g(x) = \sum_{i=1}^m g_i(x_i)$
$L\colon v \mapsto (L_1 v,\ldots,L_m v)$.

Parallel form of Douglas-Rachford

PPXA+

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

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

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

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

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

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

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