Mathematics of Computational Imaging Systems

Proximal algorithms

Nelly Pustelnik

CNRS, ENSL, Laboratoire de physique, F-69342 Lyon, France

Proximal algorithms

➡ **Minimisation problem** :

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}}{\mathrm{Argmin}}\ f_1(\mathrm{x}) + f_2(\mathrm{x})$$

with f_1 and f_2 either diff. with Lipschitz gradient or proximable.

➡ **Design of a recursive sequence of the form:**

$$(\forall k \in \mathbb{N}) \qquad \mathrm{x}^{[k+1]} = \mathbf{T}\mathrm{x}^{[k]},$$

Gradient descent	$\mathbf{T} = \mathrm{Id} - \tau(\nabla f_1 + \nabla f_2)$
Proximal point	$\mathbf{T} = \mathrm{prox}_{\tau(f_1+f_2)}$
Forward-Backward	$\mathbf{T} = \mathrm{prox}_{\tau f_2}(\mathrm{Id} - \tau\nabla f_1)$
Peaceman-Rachford	$\mathbf{T} = (2\mathrm{prox}_{\tau f_2} - \mathrm{Id}) \circ (2\mathrm{prox}_{\tau f_1} - \mathrm{Id})$
Douglas-Rachford	$\mathbf{T} = \mathrm{prox}_{\tau f_2}(2\mathrm{prox}_{\tau f_1} - \mathrm{Id}) + \mathrm{Id} - \mathrm{prox}_{\tau f_1}$

Fixed point algorithm: zeros and fixed points

Let $\mathcal{H}$ be a Hilbert space. Let $\mathbf{\Phi}\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $\mathbf{T}\colon \mathcal{H} \to 2^{\mathcal{H}}$.
The set of **fixed points** of $\mathbf{T}$ is : $\mathrm{Fix}\mathbf{T} = \{\mathrm{x} \in \mathcal{H} \,|\, \mathrm{x} \in \mathbf{T}\mathrm{x}\}$.
The set of **zeros** of $\mathbf{\Phi}$ is : $\mathrm{zer}\mathbf{\Phi} = \{\mathrm{x} \in \mathcal{H} \,|\, 0 \in \mathbf{\Phi}\mathrm{x}\}$.

Minimisation problem and Fermat rule

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}\in\mathbb{R}^N}{\mathrm{Argmin}}\; f(\mathrm{x}) \quad \Leftrightarrow \quad \nabla f(\widehat{\mathrm{x}}) = 0 \quad \Leftrightarrow \quad \widehat{\mathrm{x}} \in \mathrm{zer}\nabla f$$

Fixed point algorithm: zeros and fixed points

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

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

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

Minimisation problem and Fermat rule

$$\widehat{\mathrm{x}} \in \underset{\mathrm{x}\in\mathbb{R}^N}{\mathrm{Argmin}}\; f(\mathrm{x}) \quad \Leftrightarrow \quad \nabla f(\widehat{\mathrm{x}}) = 0 \quad \Leftrightarrow \quad \widehat{\mathrm{x}} \in \mathrm{zer}\nabla f$$

Fix point and zeros for gradient descent:

$$\widehat{\mathrm{x}} \in \mathrm{Fix}(\mathrm{Id} - \nabla f) \quad \Leftrightarrow \quad \widehat{\mathrm{x}} = (\mathrm{Id} - \nabla f)\widehat{\mathrm{x}} \quad \Leftrightarrow \quad \widehat{\mathrm{x}} \in \mathrm{zer}\nabla f$$

Remark: $\mathbf{T} = \mathrm{Id} - \nabla f$ and $\mathbf{\Phi} = \nabla f$

Fixed point algorithm: convergence

Let $\mathcal{H}$ be a Hilbert space, $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and $\widehat{\mathrm{x}} \in \mathcal{H}$.

- $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ **converges strongly** to $\widehat{\mathrm{x}}$ if

$$\lim_{k\to} \|\mathrm{x}^{[k]} - \widehat{\mathrm{x}}\| = 0.$$

It is denoted by $\mathrm{x}^{[k]} \to \widehat{\mathrm{x}}$.

- $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ **converges weakly** to $\widehat{\mathrm{x}}$ if

$$(\forall \mathsf{u} \in \mathcal{H}) \qquad \lim_{n\to\inf} \langle \mathsf{u}, \mathrm{x}^{[k]} - \widehat{\mathrm{x}}\rangle = 0.$$

It is denoted by $\mathrm{x}^{[k]} \rightharpoonup \widehat{\mathrm{x}}$.

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

Banach-Picard theorem

$\mathbf{T}\colon \mathcal{H} \to \mathcal{H}$ is **$\omega-$Lipschitz continuous** for some $\omega > 0$ if

$$(\forall \mathrm{x} \in \mathcal{H})(\forall \mathrm{u} \in \mathcal{H}) \qquad \|\mathbf{T}\mathrm{x} - \mathbf{T}\mathrm{u}\| \leq \omega \|\mathrm{x} - \mathrm{u}\|.$$

$\mathbf{T}$ is **nonexpansive** if it is $1-$Lipschitz continuous.

Banach-Picard theorem:

Let $\omega \in [0, 1[$, $\mathbf{T}\colon \mathcal{H} \to \mathcal{H}$ be a $\omega-$Lipschitz continuous operator, and $\mathrm{x}^{[0]} \in \mathcal{H}$. Set

$$(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathbf{T}\mathrm{x}^{[k]}.$$

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

$$(\forall k \in \mathbb{N}) \quad \|\mathrm{x}^{[k]} - \widehat{\mathrm{x}}\| \leq \omega^k \|\mathrm{x}_0 - \widehat{\mathrm{x}}\|.$$

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

Averaged nonexpansive operator

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

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

$\mathbf{T}$ is **firmly nonexpansive** if it is $1/2-$averaged.

$\mathbf{T}$ is **nonexpansive** if and only if $\mathbf{T}$ is $1-$averaged.

Theorem:

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

$$(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathbf{T}\mathrm{x}^{[k]}.$$

Then $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ converges weakly to a point in $\mathrm{Fix}\mathbf{T}$.

Nonlinear operators

Properties of f	**T**	ω-Lipschitz	μ-averaged
f convex ∇f β-Lipschitz	$\mathrm{Id} - \tau \nabla f$ $\tau \in (0, 2\beta^{-1})$	$\omega = 1$	$\mu = \frac{\tau\beta}{2}$
f ρ-strongly convex ∇f β-Lipschitz	$\mathrm{Id} - \tau \nabla f$ $\tau \in (0, 2\beta^{-1})$	$\omega = \max\{(1-\tau\rho), (\tau\beta - 1)\}$	$\mu = \frac{1+\omega}{2}$
$f \in \Gamma_0$	$\mathrm{prox}_{\tau f}$ $\tau > 0$	$\omega = 1$	$\mu = \frac{1}{2}$
f ρ-strongly convex	$\mathrm{prox}_{\tau f}$ $\tau > 0$	$\omega = (1+\tau\rho)^{-1}$	$\mu = \frac{1+\omega}{2}$

[Taylor, J. M. Hendrickx, and F. Glineur] [Bauschke,Combettes,2017]

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.

Let $f \in \Gamma_0(\mathbb{R}^N)$ differentiable with a β-Lipschitz gradient. We set, for some $\tau > 0$,

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

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x^{[k+1]} = x^{[k]} - \tau \nabla f(x^{[k]})$.

Let $f \in \Gamma_0(\mathbb{R}^N)$ differentiable with a β-Lipschitz gradient. We set, for some $\tau > 0$,

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

- **Iterations**: $(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{x}^{[k]} - \tau \nabla f(\mathrm{x}^{[k]})$.
- For every $\tau > 0$, $\mathrm{zer} \nabla f = \mathrm{Fix}\mathbf{T}$.

Let $f \in \Gamma_0(\mathbb{R}^N)$ differentiable with a β-Lipschitz gradient. We set, for some $\tau > 0$,

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

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x^{[k+1]} = x^{[k]} - \tau\nabla f(x^{[k]})$.
- For every $\tau > 0$, $\mathrm{zer}\nabla f = \mathrm{Fix}\mathbf{T}$.
- $\mathbf{T} = \mathrm{Id} - \tau\nabla f$ is a $\tau\beta/2$-averaged operator.

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

Gradient descent $\quad \widehat{\mathrm{x}} \in \operatorname{Argmin}_{\mathrm{x}\in\mathbb{R}^N} f(\mathrm{x})$

Let $f \in \Gamma_0(\mathbb{R}^N)$ differentiable with a β-Lipschitz gradient. We set, for some $\tau > 0$,

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

- **Iterations**: $(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{x}^{[k]} - \tau\nabla f(\mathrm{x}^{[k]})$.
- For every $\tau > 0$, $\mathrm{zer}\nabla f = \mathrm{Fix}\mathbf{T}$.
- $\mathbf{T} = \mathrm{Id} - \tau\nabla f$ is a $\tau\beta/2$-averaged operator.

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

Let $f \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f}$$

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

Let $f \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \text{prox}_{\tau f}$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x^{[k+1]} = \text{prox}_{\gamma f}(x^{[k]})$.
- For every $\tau > 0$, $\text{Fix}\mathbf{T} = \text{zer}\partial f$.
  Proof:

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

Let $f \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \text{prox}_{\tau f}$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad x^{[k+1]} = \text{prox}_{\gamma f}(x^{[k]})$.
- For every $\tau > 0$, $\text{Fix}\mathbf{T} = \text{zer}\partial f$.
- For every $\tau > 0$ and any $f \in \Gamma_0(\mathcal{H})$, $\text{prox}_{\tau f}$ is $1/2$-averaged.

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

Proximal Point Algorithm (PPA) $\widehat{\mathrm{x}} \in \mathrm{Argmin}_{\mathrm{x}\in\mathbb{R}^N} f(\mathrm{x})$

Let $f \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f}$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{prox}_{\gamma f}(\mathrm{x}^{[k]})$.
- For every $\tau > 0$, $\mathrm{Fix}\mathbf{T} = \mathrm{zer}\partial f$.
- 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]
- The **PPA method converges** to a point in $\mathrm{zer}\partial f$.

FB algorithm $\widehat{\mathrm{x}} \in \operatorname{Argmin}_{\mathrm{x}\in\mathbb{R}^N} \left\{ f(\mathrm{x}) = f_1(\mathrm{x}) + f_2(\mathrm{x}) \right\}$

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathsf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau f_2}(\mathrm{x}^{[k]} - \tau \nabla f_1(\mathrm{x}^{[k]}))$.

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau f_2}(\mathrm{x}^{[k]} - \tau \nabla f_1(\mathrm{x}^{[k]}))$.
- Roots in projected gradient method [Levitin 1966] when $g = \iota_C$ for some closed convex set C.

Objective Let $f_1\colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau f_2}(\mathrm{x}^{[k]} - \tau \nabla f_1(\mathrm{x}^{[k]}))$.
- For every $\tau > 0$, $\mathrm{zer}(\nabla f_1 + \partial \mathrm{g}) = \mathrm{Fix}\mathbf{T}$.

  Proof:

$$\begin{aligned} \mathrm{x} \in \mathrm{Fix}\mathbf{T} &\Leftrightarrow (\mathrm{Id} - \tau\nabla f_1)\mathrm{x} \in (\mathrm{Id} + \tau\partial f_2)\mathrm{x} \\ &\Leftrightarrow 0 \in \nabla f_1(\mathrm{x}) + \partial f_2(\mathrm{x}). \end{aligned}$$

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau f_2}(\mathrm{x}^{[k]} - \tau \nabla f_1(\mathrm{x}^{[k]}))$.
- For every $\tau > 0$, $\mathrm{zer}(\nabla f_1 + \partial \mathrm{g}) = \mathrm{Fix}\mathbf{T}$.
- $\mathrm{prox}_{\tau f_2}(\mathrm{Id} - \tau \nabla f_1)$ is μ-averaged nonexpansive where $\mu = \frac{\mu_1 + \mu_2 - 2\mu_1\mu_2}{1 - \mu_1\mu_2}$ with $\mu_2 = \tau\beta/2$ and $\mu_1 = 1/2$ leading to $\mu = \frac{1}{2 - \tau\beta/2} \in]0, 1[$ and $\tau < 2/\beta$.

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

- **Iterations**: $(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau f_2}(\mathrm{x}^{[k]} - \tau \nabla f_1(\mathrm{x}^{[k]}))$.
- For every $\tau > 0$, $\mathrm{zer}(\nabla f_1 + \partial \mathrm{g}) = \mathrm{Fix}\mathbf{T}$.
- $\mathrm{prox}_{\tau f_2}(\mathrm{Id} - \tau \nabla f_1)$ is μ-averaged nonexpansive where $\mu = \frac{\mu_1+\mu_2-2\mu_1\mu_2}{1-\mu_1\mu_2}$ with $\mu_2 = \tau\beta/2$ and $\mu_1 = 1/2$ leading to $\mu = \frac{1}{2-\tau\beta/2} \in]0,1[$ and $\tau < 2/\beta$.
- For every $\tau \in]0, 2/\beta[$, the **FBS converges** to a point in $\mathrm{zer}(\nabla f_1 + \partial f_2)$.

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

FB algorithm $\widehat{\mathrm{x}} \in \operatorname{Argmin}_{\mathrm{x}\in\mathbb{R}^N} \Big\{ f(\mathrm{x}) = f_1(\mathrm{x}) + f_2(\mathrm{x}) \Big\}$

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \operatorname{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

Theorem [Combettes & Wajs, 2005]: Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be a sequence generated by the FB algorithm. Let $0 < \tau < 2\beta^{-1}$. Then

- $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ converges to a minimiser $\widehat{\mathrm{x}}$ of f
- $\big(f(\mathrm{x}^{[k]})\big)_{k\in\mathbb{N}}$ is a non-increasing sequence converging to $f(\widehat{\mathrm{x}})$.

Objective Let $f_1 : \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

Theorem [Briceno-Arias & Pustelnik, 2005]:

If additionally, f_1 is ρ-strongly convex. Suppose that $\tau \in \left]0, 2\beta^{-1}\right[$. Then $\mathbf{T}$ is $\omega(\tau)-$Lipschitz continuous with

$$\omega(\tau) := \max\Big\{|1-\tau\rho|, |1-\tau\beta|\Big\} \in \left]0,1\right[.$$

Objective Let $f_1 : \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

Theorem [Briceno-Arias & Pustelnik, 2005]:

If additionally, f_1 is ρ-strongly convex. Suppose that $\tau \in \left]0, 2\beta^{-1}\right[$. Then **T** is $\omega(\tau)-$Lipschitz continuous with

$$\omega(\tau) := \max\left\{|1-\tau\rho|, |1-\tau\beta|\right\} \in \left]0,1\right[.$$

FB algorithm $\widehat{\mathrm{x}} \in \mathrm{Argmin}_{\mathrm{x}\in\mathbb{R}^N}\Big\{f(\mathrm{x}) = f_1(\mathrm{x}) + f_2(\mathrm{x})\Big\}$

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

Theorem [Briceno-Arias & Pustelnik, 2005]:

If additionally, f_1 is ρ-strongly convex. Suppose that $\tau \in \left]0, 2\beta^{-1}\right[$. Then $\mathbf{T}$ is $\omega(\tau)-$Lipschitz continuous with

$$\omega(\tau) := \max\Big\{|1-\tau\rho|, |1-\tau\beta|\Big\} \in \left]0,1\right[.$$

In particular, the minimum is achieved at

$$\tau^* = \frac{2}{\rho+\beta} \quad \text{and} \quad \omega(\tau^*) = \frac{\beta-\rho}{\beta+\rho}.$$

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathsf{T} := \operatorname{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

Theorem [Briceno-Arias & Pustelnik, 2023][Briceno-Arias , 2025]:

If additionally, f_2 is ρ-strongly convex. Suppose that $\tau \in \left]0, 2\beta^{-1}\right]$. Then T is $\omega(\tau)-$Lipschitz continuous with

$$\omega(\tau) := \frac{1}{1+\tau\rho} \in \left]0,1\right[.$$

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

Theorem [Beck & Teboulle, 2009]:

Suppose that $\tau \in \left]0, \beta^{-1}\right]$ and let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ the sequence generated by $\mathrm{x}^{[k+1]} = \mathbf{T}\mathrm{x}^{[k]}$. Then,

$$f(\mathrm{x}^{[k]}) - f(\widehat{\mathrm{x}}) \leq \frac{\beta}{2k} \|\mathrm{x}^{[0]} - \widehat{\mathrm{x}}\|^2$$

FB algorithm $\widehat{x} \in \text{Argmin}_{x \in \mathbb{R}^N} \left\{ f(x) = f_1(x) + f_2(x) \right\}$

Objective Let $f_1 \colon \mathbb{R}^N \to \mathbb{R}$ a convex, proper and β-Lipschitz differentiable function and $f_2 \in \Gamma_0(\mathbb{R}^N)$. We set, for some $\tau > 0$,

$$\mathbf{T} := \text{prox}_{\tau f_2} \circ (\text{Id} - \tau \nabla f_1)$$

- Convergence may be slow in practice...
  - ☛ Use Nesterov acceleration (*inertia*)
  - ☛ Use second order information (*preconditioning*)
  - ☛ Use multilevel strategy
- What if $\text{prox}_{\gamma_k \mathbf{g}}$ does not have a closed form?
  - ☛ Use sub-iterations (e.g. dual FB algorithm)
  - ☛ Use more advanced methods (e.g. primal-dual algorithms)
  - ☛ Use more advanced methods (e.g. primal-dual algorithms)

What is inertia?

Goal: Inertia aims to use information from the **previous iterate(s)** $(\mathrm{x}^{[k']})_{k' \leq k}$ to build the next iterate $\mathrm{x}^{[k+1]}$.

Why? Use memory to go faster!

For FB we have:

$$(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathsf{T}_k(\mathrm{x}^{[k]}) \text{ where } \mathsf{T}_k = \mathrm{prox}_{\tau f_2} \circ (\mathrm{Id} - \tau \nabla f_1)$$

Introducing inertia would lead to

$$(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \widetilde{\mathsf{T}}_k(\mathrm{x}_1, \ldots, \mathrm{x}^{[k]})$$

QUESTION: How to choose $\widetilde{\mathsf{T}}_k$?

REMARK: In general $\widetilde{\mathsf{T}}_k$ only depends on $(\mathrm{x}^{[k]}, \mathrm{x}_{k-1})$ to avoid memory issues

Particular case: Inertia for GD algorithm

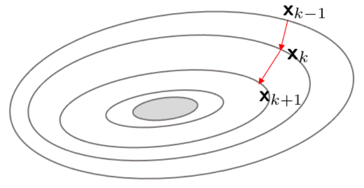
Let $f_2 \equiv 0$. In this case $\mathrm{prox}_{f_2} = \mathrm{Id}$.

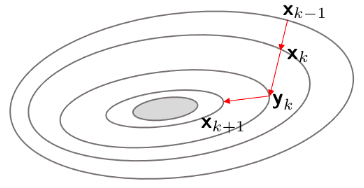
The *path* taken by the iterates $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ is determined by the opposite of the gradient direction:

$$(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{x}^{[k]} - \tau_k \nabla f_1(\mathrm{x}^{[k]})$$

Acceleration: *Nesterov-type accelerated GD algorithm* [Nesterov, 83]

$$(\forall k \in \mathbb{N}) \qquad \begin{aligned} &\mathrm{x}^{[k+1]} = \mathrm{y}^{[k]} - \tau \nabla f_1(\mathrm{y}^{[k]}) \quad \text{with } \tau \in]0, 1/\beta] \\ &\mathrm{y}_{k+1} = \mathrm{x}^{[k+1]} + \alpha_k(\mathrm{x}^{[k+1]} - \mathrm{x}_k) \end{aligned}$$





Particular case: Inertia for GD algorithm

Let $f_2 \equiv 0$. In this case $\mathrm{prox}_{f_2} = \mathrm{Id}$.

The *path* taken by the iterates $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ is determined by the opposite of the gradient direction:

$$(\forall k \in \mathbb{N}) \quad \mathrm{x}^{[k+1]} = \mathrm{x}^{[k]} - \tau_k \nabla f_1(\mathrm{x}^{[k]})$$

Acceleration: *Nesterov-type accelerated GD algorithm* [Nesterov, 83]

$$(\forall k \in \mathbb{N}) \qquad \begin{aligned} &\mathrm{x}^{[k+1]} = \mathrm{y}^{[k]} - \tau \nabla f_1(\mathrm{y}^{[k]}) \quad \text{with } \tau \in]0, 1/\beta] \\ &\mathrm{y}_{k+1} = \mathrm{x}^{[k+1]} + \alpha_k(\mathrm{x}^{[k+1]} - \mathrm{x}_k) \end{aligned}$$

- Each iteration takes nearly the same computational cost as GD
- **not** a *descent* method (i.e. we may not have $f_1(\mathrm{x}^{[k+1]}) \leq f_1(\mathrm{x}^{[k]})$)

Inertial FB

Inertial FB

For $k = 0, 1, \dots$

$$
\begin{array}{|l}
\text{Let } \gamma_k \in]0, 1/\beta] \\
\mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau_k f_2}\left(\mathrm{y}^{[k]} - \tau_k \nabla f_1(\mathrm{y}^{[k]})\right) \\
\mathrm{y}^{[k+1]} = \mathrm{x}^{[k+1]} + \alpha_k(\mathrm{x}^{[k+1]} - \mathrm{x}^{[k]})
\end{array}
$$

- [Beck and Teboulle, 2009]
  Adopt the inertia (momentum) strategy proposed by Nesterov

$$\alpha_k = \frac{\theta_k - 1}{\theta_{k+1}} \quad \text{with} \quad \theta_{k+1} = \frac{1 + \sqrt{1 + 4\theta_k^2}}{2}$$

- [Chambolle and Dossal, 2015]
  Different rule

$$\alpha_k = \frac{\theta_k - 1}{\theta_{k+1}} \quad \text{with} \quad \theta_{k+1} = \left(\frac{k + a}{a}\right)^d$$

with $d \in]0, 1]$ and $a > \max\{1, (2d)^{1/d}\}$.

Convergence of Inertial FB

Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be generated by FB iterations with $\tau \in]0,\beta^{-1}[$.
$(f(\mathrm{x}^{[k]}))_{k\in\mathbb{N}}$ converges to $f(\widehat{\mathrm{x}})$ at the rate $O(1/k)$:

$$f(\mathrm{x}^{[k]}) - f(\widehat{\mathrm{x}}) \leq \frac{\beta}{2k}\|\mathrm{x}^{[0]} - \widehat{\mathrm{x}}\|^2$$

Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be generated by Inertial FB.
$(f(\mathrm{x}^{[k]}))_{k\in\mathbb{N}}$ converges to $f(\widehat{\mathrm{x}})$ at the rate $O(1/k^2)$:

$$f(\mathrm{x}^{[k]}) - f(\widehat{\mathrm{x}}) \leq \frac{2\beta}{(k+1)^2}\|\mathrm{x}^{[0]} - \widehat{\mathrm{x}}\|^2$$

Convergence of Inertial FB

Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be generated by FB iterations with $\tau \in]0, \beta^{-1}[$.
$(f(\mathrm{x}^{[k]}))_{k\in\mathbb{N}}$ converges to $f(\widehat{\mathrm{x}})$ at the rate $O(1/k)$:

$$f(\mathrm{x}^{[k]}) - f(\widehat{\mathrm{x}}) \leq \frac{\beta}{2k}\|\mathrm{x}^{[0]} - \widehat{\mathrm{x}}\|^2$$

Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be generated by Inertial FB.
$(f(\mathrm{x}^{[k]}))_{k\in\mathbb{N}}$ converges to $f(\widehat{\mathrm{x}})$ at the rate $O(1/k^2)$:

$$f(\mathrm{x}^{[k]}) - f(\widehat{\mathrm{x}}) \leq \frac{2\beta}{(k+1)^2}\|\mathrm{x}^{[0]} - \widehat{\mathrm{x}}\|^2$$

Proof: A complete proof is provided in [Beck and Teboulle, 2009].

Convergence of Inertial FB

Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be generated by FB iterations with $\tau \in]0, \beta^{-1}[$.
$(f(\mathrm{x}^{[k]}))_{k\in\mathbb{N}}$ converges to $f(\widehat{\mathrm{x}})$ at the rate $O(1/k)$:

$$f(\mathrm{x}^{[k]}) - f(\widehat{\mathrm{x}}) \leq \frac{\beta}{2k}\|\mathrm{x}^{[0]} - \widehat{\mathrm{x}}\|^2$$

Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be generated by Inertial FB.
$(f(\mathrm{x}^{[k]}))_{k\in\mathbb{N}}$ converges to $f(\widehat{\mathrm{x}})$ at the rate $O(1/k^2)$:

$$f(\mathrm{x}^{[k]}) - f(\widehat{\mathrm{x}}) \leq \frac{2\beta}{(k+1)^2}\|\mathrm{x}^{[0]} - \widehat{\mathrm{x}}\|^2$$

- Improved iteration complexity:
  - FB: $\mathsf{f}(\mathrm{x}^{[k]}) - \mathsf{f}(\widehat{\mathrm{x}}) \approx O(1/k)$
  - FISTA: $\mathsf{f}(\mathrm{x}^{[k]}) - \mathsf{f}(\widehat{\mathrm{x}}) \approx O(1/k^2)$
- (Almost) Same computational complexity per iteration as FB
- Issue: Convergence guaranties of the sequence $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$?

Convergence of Inertial FB

Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be generated by FB iterations with $\tau \in]0,\beta^{-1}[$.
$(f(\mathrm{x}^{[k]}))_{k\in\mathbb{N}}$ converges to $f(\widehat{\mathrm{x}})$ at the rate $O(1/k)$:

$$f(\mathrm{x}^{[k]}) - f(\widehat{\mathrm{x}}) \leq \frac{\beta}{2k}\|\mathrm{x}^{[0]} - \widehat{\mathrm{x}}\|^2$$

Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be generated by Inertial FB.
$(f(\mathrm{x}^{[k]}))_{k\in\mathbb{N}}$ converges to $f(\widehat{\mathrm{x}})$ at the rate $O(1/k^2)$:

$$f(\mathrm{x}^{[k]}) - f(\widehat{\mathrm{x}}) \leq \frac{2\beta}{(k+1)^2}\|\mathrm{x}^{[0]} - \widehat{\mathrm{x}}\|^2$$

Let $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ be generated by Inertial FB with Chambolle-Dossal rule $\alpha_k = \frac{\theta_k - 1}{\theta_{k+1}}$ with $\theta_{k+1} = \left(\frac{k+a}{a}\right)^d$ with $d \in]0,1]$ and $a > \max\{1, (2d)^{1/d}\}$
Then the sequence $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ converges to a minimiser of f.

Duality

Minimization problem

Find

$$\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f_1(x) + f_2(x) + g(Lx)$$

- $f_1 : \mathbb{R}^N \to \mathbb{R}$ is convex and β-Lipschitz differentiable
- $f_2 \in \Gamma_0(\mathbb{R}^N)$
- $g \in \Gamma_0(\mathbb{R}^M)$ and $L \in \mathbb{R}^{M \times N}$

Use FB algorithm ?

For $k = 0, 1, \ldots$

$$x^{[k+1]} = \mathrm{prox}_{\tau(f_2 + g \circ L)}\big(x^{[k]} - \tau \nabla f_1(x^{[k]})\big)$$

How to compute $\mathrm{prox}_{\tau(f_2 + g \circ \mathbf{L})}$?

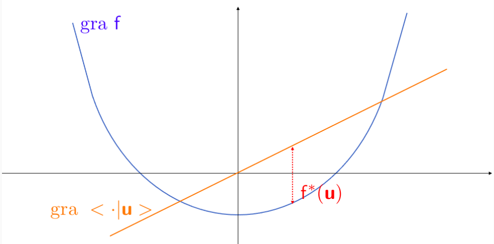
➤ Use primal-dual methods

Conjugate function

The conjugate of a function $f\colon \mathbb{R}^N \to]-\infty, +\infty]$ is the function f^* defined as

$$f^*\colon \begin{array}{ccc} \mathbb{R}^N & \to & [-\infty, +\infty] \\ \mathrm{u} & \mapsto & \sup_{\mathrm{x}\in\mathbb{R}^N} \langle \mathrm{x} \mid \mathrm{u} \rangle - f(\mathrm{x}) \end{array}$$

Graphical illustration: $f^*(\mathrm{u})$ is the supremum of the signed vertical distance between the graph of f and that of the continuous linear functional $\cdot\mathrm{u}$

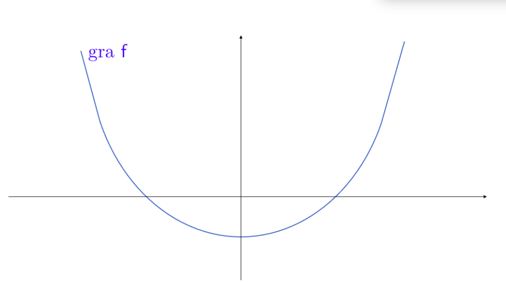


Conjugate function

The conjugate of a function $f\colon \mathbb{R}^N \to]-\infty, +\infty]$ is the function f^* defined as

$$f^*\colon \begin{array}{ccc} \mathbb{R}^N & \to & [-\infty, +\infty] \\ \mathrm{u} & \mapsto & \sup_{\mathrm{x}\in\mathbb{R}^N} \langle \mathrm{x} \mid \mathrm{u}\rangle - f(\mathrm{x}) \end{array}$$

Graphical illustration: Second interpretation - How to build $f^*(\mathrm{u})$?

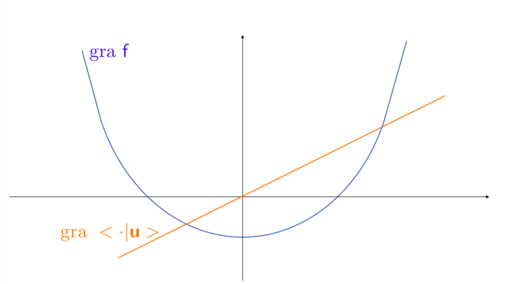


Conjugate function

The conjugate of a function $f : \mathbb{R}^N \to]-\infty, +\infty]$ is the function f^* defined as

$$\begin{array}{rccl} f^* : & \mathbb{R}^N & \to & [-\infty, +\infty] \\ & \mathrm{u} & \mapsto & \sup_{\mathrm{x} \in \mathbb{R}^N} \langle \mathrm{x} \mid \mathrm{u} \rangle - f(\mathrm{x}) \end{array}$$

Graphical illustration: Second interpretation - How to build $f^*(\mathrm{u})$?

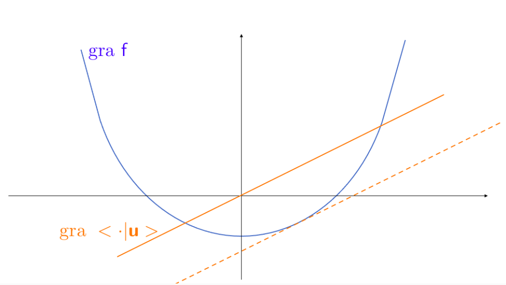


Conjugate function

The conjugate of a function $f\colon \mathbb{R}^N \to]-\infty, +\infty]$ is the function f^* defined as

$$f^*\colon \quad \begin{array}{rcl} \mathbb{R}^N & \to & [-\infty, +\infty] \\ \mathrm{u} & \mapsto & \sup_{\mathrm{x}\in\mathbb{R}^N} \langle \mathrm{x} \mid \mathrm{u}\rangle - f(\mathrm{x}) \end{array}$$

Graphical illustration: Second interpretation - How to build $f^*(\mathrm{u})$?

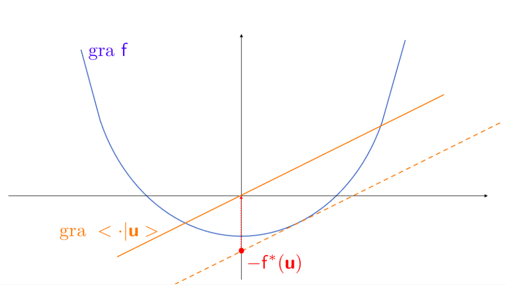


Conjugate function

The conjugate of a function $f\colon \mathbb{R}^N \to]-\infty, +\infty]$ is the function f^* defined as

$$f^*\colon \begin{array}{ccc} \mathbb{R}^N & \to & [-\infty, +\infty] \\ \mathrm{u} & \mapsto & \sup_{\mathrm{x}\in\mathbb{R}^N} \langle \mathrm{x} \mid \mathrm{u} \rangle - f(\mathrm{x}) \end{array}$$

Graphical illustration: Second interpretation - How to build $f^*(\mathrm{u})$?



Conjugate function

The conjugate of a function $f\colon \mathbb{R}^N \to]-\infty,+\infty]$ is the function f^* defined as

$$
\begin{array}{rccl}
f^*\colon & \mathbb{R}^N & \to & [-\infty,+\infty] \\
 & \mathrm{u} & \mapsto & \sup_{\mathrm{x}\in\mathbb{R}^N} \langle \mathrm{x} \mid \mathrm{u}\rangle - f(\mathrm{x})
\end{array}
$$

Examples :

- $f = \frac{1}{2}\|\cdot\|^2 \Rightarrow f^* = \frac{1}{2}\|\cdot\|^2$

  Proof : For every $(\mathrm{x},\mathrm{u}) \in \mathcal{H}^2$,
  $\langle \mathrm{x} \mid \mathrm{u}\rangle - \frac{1}{2}\|\mathrm{x}\|^2 = \frac{1}{2}\|\mathrm{u}\|^2 - \frac{1}{2}\|\mathrm{u}-\mathrm{x}\|^2$ is maximum at $\mathrm{x}=\mathrm{u}$.
  Consequently, $f^*(\mathrm{u}) = \frac{1}{2}\|\mathrm{u}\|^2$.

Conjugate function

The conjugate of a function $f\colon \mathbb{R}^N \to]-\infty,+\infty]$ is the function f^* defined as

$$\begin{array}{rccl} f^*\colon & \mathbb{R}^N & \to & [-\infty,+\infty] \\ & \mathrm{u} & \mapsto & \sup_{\mathrm{x}\in\mathbb{R}^N} \langle \mathrm{x} \mid \mathrm{u}\rangle - f(\mathrm{x}) \end{array}$$

Moreau-Fenchel theorem

Let $\mathcal{H}$ be a Hilbert space and $f\colon \mathcal{H} \to]-\infty,+\infty]$ be a proper function.

$$f \text{ is l.s.c. and convex} \Leftrightarrow f^{**} = f.$$

Conjugate: properties

Fenchel-Young inequality: If f is proper, then

1. $\left(\forall(\mathrm{x},\mathrm{u})\in\mathcal{H}^2\right) \qquad f(\mathrm{x})+f^*(\mathrm{u})\geq\langle\mathrm{x}\mid\mathrm{u}\rangle$

2. $\left(\forall(\mathrm{x},\mathrm{u})\in\mathcal{H}^2\right) \qquad \mathrm{u}\in\partial f(\mathrm{x})\Leftrightarrow \quad f(\mathrm{x})+f^*(\mathrm{u})=\langle\mathrm{x}\mid\mathrm{u}\rangle.$

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

$$\left(\forall(\mathrm{x},\mathrm{u})\in\mathcal{H}^2\right) \qquad \mathrm{u}\in\partial f(\mathrm{x}) \;\Leftrightarrow\; \mathrm{x}\in\partial f^*(\mathrm{u}).$$

Conjugate: Moreau decomposition

Moreau decomposition formula

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

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

Proof:

$$\begin{aligned}
\mathrm{p} = \mathrm{prox}_{\gamma f^*}\mathrm{x} &\Leftrightarrow \mathrm{x} - \mathrm{p} \in \gamma \partial f^*(\mathrm{p}) \\
&\Leftrightarrow \mathrm{p} \in \partial f\left(\frac{\mathrm{x}-\mathrm{p}}{\gamma}\right) \\
&\Leftrightarrow \frac{\mathrm{x}}{\gamma} - \frac{\mathrm{x}-\mathrm{p}}{\gamma} \in \frac{1}{\gamma}\partial f\left(\frac{\mathrm{x}-\mathrm{p}}{\gamma}\right) \\
&\Leftrightarrow \frac{\mathrm{x}-\mathrm{p}}{\gamma} = \mathrm{prox}_{\gamma^{-1} f}(\gamma^{-1}\mathrm{x}) \\
&\Leftrightarrow \mathrm{p} = \mathrm{x} - \gamma \mathrm{prox}_{\gamma^{-1} f}(\gamma^{-1}\mathrm{x}).
\end{aligned}$$

Conjugate: Moreau decomposition

Moreau decomposition formula

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

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

<u>Example</u>: If $\mathcal{H} = \mathbb{R}^N$, $f = \frac{1}{q}\|\cdot\|_q^q$ with $q \in]1, +\infty[$, then $f^* = \frac{1}{q^*}\|\cdot\|_{q^*}^{q^*}$ with $1/q + 1/q^* = 1$, and

$$(\forall \mathrm{x} \in \mathbb{R}^N) \qquad \mathrm{prox}_{\frac{\gamma}{q^*}\|\cdot\|_{q^*}^{q^*}}\mathrm{x} = \mathrm{x} - \gamma \mathrm{prox}_{\frac{1}{\gamma q}\|\cdot\|_q^q}(\gamma^{-1}\mathrm{x}).$$

Example: Support function

Let $\mathcal{C}$ be a subset of $\mathbb{R}^N$. The support function of $\mathcal{C}$, denoted by $\sigma_{\mathcal{C}}$, is

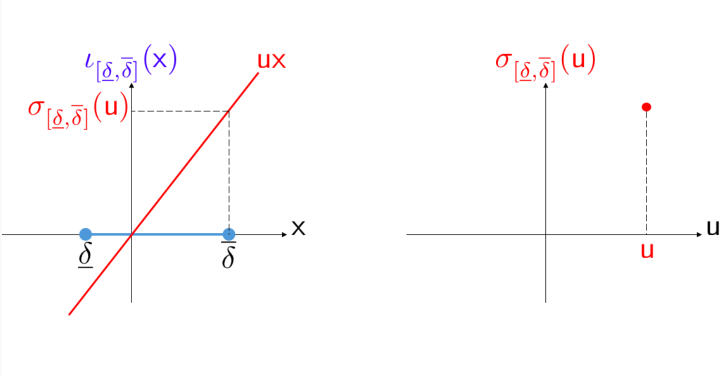
$$(\forall \mathrm{u} \in \mathbb{R}^N) \quad \sigma_{\mathcal{C}}(\mathrm{u}) = \sup_{\mathrm{x} \in \mathcal{C}} \langle \mathrm{x}, \mathrm{u} \rangle$$

Example: Support function

Let $\mathcal{C}$ be a subset of $\mathbb{R}^N$. The support function of $\mathcal{C}$, denoted by $\sigma_{\mathcal{C}}$, is

$$(\forall \mathrm{u} \in \mathbb{R}^N) \quad \sigma_{\mathcal{C}}(\mathrm{u}) = \sup_{\mathrm{x}\in\mathcal{C}} \langle \mathrm{x}, \mathrm{u} \rangle$$

Graphical illustration: $\mathcal{C} = [\underline{\delta}, \overline{\delta}]$

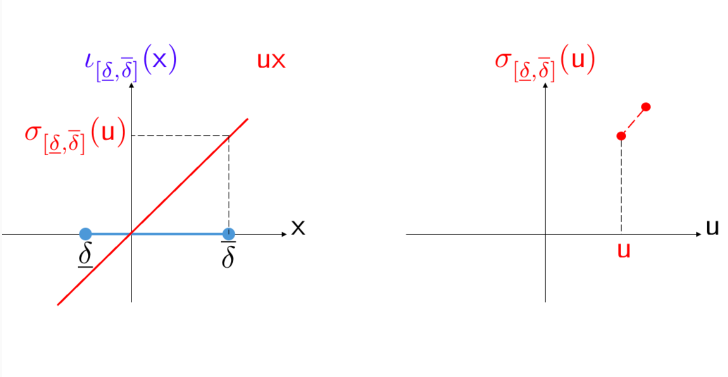


Example: Support function

Let $\mathcal{C}$ be a subset of $\mathbb{R}^N$. The support function of $\mathcal{C}$, denoted by $\sigma_{\mathcal{C}}$, is

$$(\forall \mathrm{u} \in \mathbb{R}^N) \quad \sigma_{\mathcal{C}}(\mathrm{u}) = \sup_{\mathrm{x} \in \mathcal{C}} \langle \mathrm{x}, \mathrm{u} \rangle$$

Graphical illustration: $\mathcal{C} = [\underline{\delta}, \overline{\delta}]$

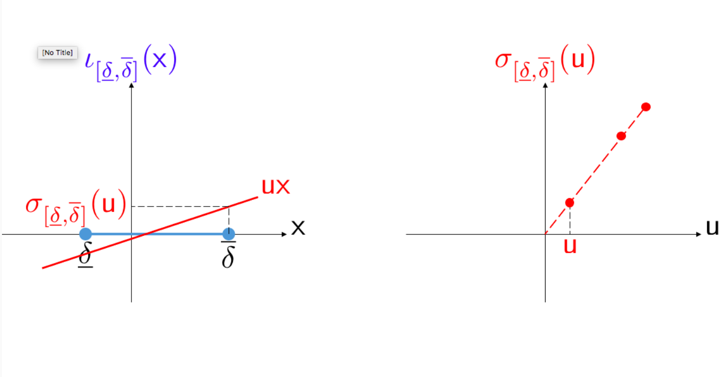


Example: Support function

Let $\mathcal{C}$ be a subset of $\mathbb{R}^N$. The support function of $\mathcal{C}$, denoted by $\sigma_\mathcal{C}$, is

$$(\forall \mathrm{u} \in \mathbb{R}^N) \quad \sigma_\mathcal{C}(\mathrm{u}) = \sup_{\mathrm{x}\in\mathcal{C}} \langle \mathrm{x}, \mathrm{u} \rangle$$

Graphical illustration: $\mathcal{C} = [\underline{\delta}, \overline{\delta}]$

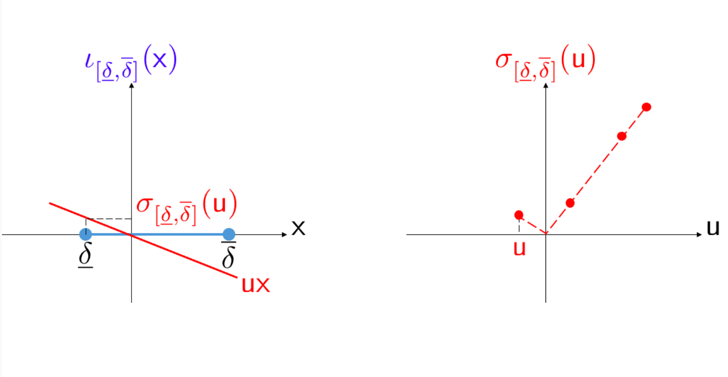


Example: Support function

Let $\mathcal{C}$ be a subset of $\mathbb{R}^N$. The support function of $\mathcal{C}$, denoted by $\sigma_{\mathcal{C}}$, is

$$(\forall \mathrm{u} \in \mathbb{R}^N) \quad \sigma_{\mathcal{C}}(\mathrm{u}) = \sup_{\mathrm{x} \in \mathcal{C}} \langle \mathrm{x}, \mathrm{u} \rangle$$

Graphical illustration: $\mathcal{C} = [\underline{\delta}, \overline{\delta}]$

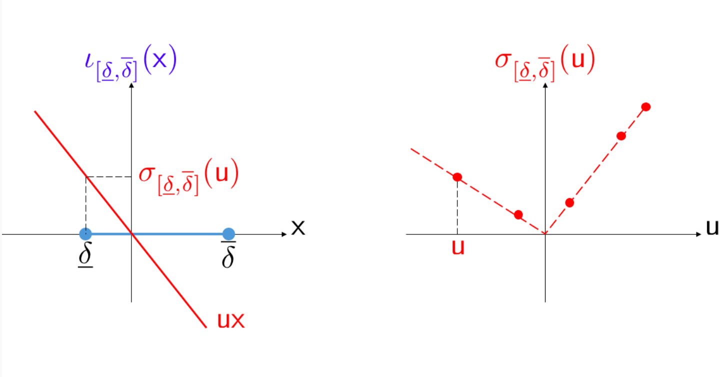


Example: Support function

Let $\mathcal{C}$ be a subset of $\mathbb{R}^N$. The support function of $\mathcal{C}$, denoted by $\sigma_{\mathcal{C}}$, is

$$(\forall \mathrm{u} \in \mathbb{R}^N) \quad \sigma_{\mathcal{C}}(\mathrm{u}) = \sup_{\mathrm{x} \in \mathcal{C}} \langle \mathrm{x}, \mathrm{u} \rangle$$

Graphical illustration: $\mathcal{C} = [\underline{\delta}, \overline{\delta}]$



Example: Support function

Let $\mathcal{C}$ be a subset of $\mathbb{R}^N$. The support function of $\mathcal{C}$, denoted by $\sigma_{\mathcal{C}}$, is

$$(\forall \mathrm{u} \in \mathbb{R}^N) \quad \sigma_{\mathcal{C}}(\mathrm{u}) = \sup_{\mathrm{x} \in \mathcal{C}} \langle \mathrm{x}, \mathrm{u} \rangle$$

Remarks:

- We have $\sigma_{\mathcal{C}} = \iota_{\mathcal{C}}^*$
- If $\mathcal{C}$ is a closed, convex, non-empty subset of $\mathbb{R}^N$, then $\sigma_{\mathcal{C}}^* = \iota_{\mathcal{C}}^{**} = \iota_{\mathcal{C}}$
- Let $-\infty \leq \underline{\delta} < \overline{\delta} \leq +\infty$, and $\mathcal{C} = [\underline{\delta}, \overline{\delta}]$. Then

$$(\forall \mathsf{x} \in \mathbb{R}) \quad \sigma_{\mathcal{C}}(\mathsf{x}) = \begin{cases} \underline{\delta}\mathsf{x} & \text{if } \mathsf{x} < 0 \\ 0 & \text{if } \mathsf{x} = 0 \\ \overline{\delta}\mathsf{x} & \text{if } \mathsf{x} > 0 \end{cases}$$

As a consequence we have $(\forall \delta > 0)(\forall \mathsf{x} \in \mathbb{R})$ $f(\mathsf{x}) = \delta|\mathsf{x}| = \sigma_{[-\delta,+\delta]}(\mathsf{x})$

and $f^* = \iota_{[-\delta,+\delta]} = \iota_{\mathcal{B}_\infty(0,\delta)}$

More generally, for $f = \delta\|\cdot\|_1$, we have $f^* = \iota_{\mathcal{B}_\infty(\mathbf{0},\delta)}$

Dual methods

Fenchel-Rockafellar duality

Primal problem

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

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

We want to

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

Dual problem

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

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

We want to

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

Weak duality

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

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

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

Fenchel-Rockafellar duality

Weak duality

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

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

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

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

Proof: According to Fenchel-Young inequality,

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

Fenchel-Rockafellar duality

Strong duality

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

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

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

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

Fenchel-Rockafellar duality

Duality theorem (1)

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

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

$$\operatorname{zer}(\partial f + \mathrm{L}\partial g \mathrm{L}^*) \neq \emptyset \quad \Leftrightarrow \quad \operatorname{zer}\big((-L)\partial f^*(-\mathrm{L}^*) + \partial g^*\big) \neq \emptyset.$$

Fenchel-Rockafellar duality

Duality theorem (1)

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

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

$$\operatorname{zer}(\partial f + \mathrm{L}\partial g \mathrm{L}^*) \neq \emptyset \quad \Leftrightarrow \quad \operatorname{zer}\big((-L)\partial f^*(-\mathrm{L}^*) + \partial g^*\big) \neq \emptyset.$$

Proof:

$$(\exists x \in \mathcal{H}) \;\; 0 \in \partial f(\mathrm{x}) + \mathrm{L}^*\partial g(\mathrm{Lx})$$

$$\Leftrightarrow \quad (\exists \mathrm{x} \in \mathcal{H})(\exists u \in \mathcal{G}) \qquad \begin{cases} -\mathrm{L}^*\mathrm{u} \in \partial f(\mathrm{x}) \\ \mathrm{u} \in \partial g(\mathrm{Lx}) \end{cases}$$

$$\Leftrightarrow \quad (\exists \mathrm{x} \in \mathcal{H})(\exists \mathrm{u} \in \mathcal{G}) \qquad \begin{cases} \mathrm{x} \in \partial f^*(-\mathrm{L}^*\mathrm{u}) \\ \mathrm{Lx} \in \partial g^*(\mathrm{u}) \end{cases}$$

$$\Leftrightarrow \quad (\exists \mathrm{u} \in \mathcal{G}) \qquad 0 \in -\mathrm{L}\partial f^*(-\mathrm{L}^*\mathrm{u}) + \partial g^*(\mathrm{u}).$$

Fenchel-Rockafellar duality

Duality theorem (2)

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

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

- If there exists $\widehat{\mathrm{x}} \in \mathcal{H}$ such that $0 \in \partial f(\widehat{\mathrm{x}}) + \mathrm{L}^* \partial g(\mathrm{L}\widehat{\mathrm{x}})$, then $\widehat{\mathrm{x}}$ is a solution to the primal problem. Moreover, there exists a solution $\widehat{\mathrm{u}}$ to the dual problem such that $-\mathrm{L}^*\widehat{\mathrm{u}} \in \partial f(\widehat{\mathrm{x}})$ and $\mathrm{L}\widehat{\mathrm{x}} \in \partial g^*(\widehat{\mathrm{u}})$.
- If there exists $(\widehat{\mathrm{x}}, \widehat{\mathrm{u}}) \in \mathcal{H} \times \mathcal{G}$ such that $-\mathrm{L}^*\widehat{\mathrm{u}} \in \partial f(\widehat{\mathrm{x}})$ and $\mathrm{L}\widehat{\mathrm{x}} \in \partial g^*(\widehat{\mathrm{u}})$ then $\widehat{\mathrm{x}}$ (resp. $\widehat{\mathrm{u}}$) is a solution to the primal (resp. dual) problem.

If $(\widehat{\mathrm{x}}, \widehat{\mathrm{u}}) \in \mathcal{H} \times \mathcal{G}$ is such that $-\mathrm{L}^*\widehat{\mathrm{u}} \in \partial f(\widehat{\mathrm{x}})$ and $\mathrm{L}\widehat{\mathrm{x}} \in \partial g^*(\widehat{\mathrm{u}})$, then $(\widehat{\mathrm{x}}, \widehat{\mathrm{u}})$ is called a **Kuhn-Tucker point**.

Fenchel-Rockafellar duality

Proof:

$$0 \in \partial f(\widehat{\mathrm{x}}) + \mathrm{L}^* \partial g(\mathrm{L}\widehat{\mathrm{x}}) \subset \partial(f + g \circ \mathrm{L})(\widehat{\mathrm{x}}).$$

Then, according to Fermat rule, $\widehat{\mathrm{x}}$ is a solution to the primal problem.
In addition, there exists $\widehat{\mathrm{u}} \in \mathcal{G}$ such that

$$\begin{cases} 0 \in \partial f(\widehat{\mathrm{x}}) + \mathrm{L}^* \widehat{\mathrm{u}} \\ \widehat{\mathrm{u}} \in \partial g(\mathrm{L}\widehat{\mathrm{x}}) \end{cases} \quad \Leftrightarrow \quad \begin{cases} -\mathrm{L}^* \widehat{\mathrm{u}} \in \partial f(\widehat{\mathrm{x}}) \\ \mathrm{L}\widehat{\mathrm{x}} \in \partial g^*(\widehat{\mathrm{u}}). \end{cases}$$

We have also $\widehat{\mathrm{x}} \in \partial f^*(-\mathrm{L}^* \widehat{\mathrm{u}})$, which implies that

$$0 \in -\mathrm{L} \partial f^*(-\mathrm{L}^* \widehat{\mathrm{u}}) + \partial g^*(\widehat{\mathrm{u}}).$$

On the other hand,

$$0 \in -\mathrm{L} \partial f^*(-\mathrm{L}^* \widehat{\mathrm{u}}) + \partial g^*(\widehat{\mathrm{u}}) \subset \partial\big(f^* \circ (-\mathrm{L}^*) + g^*\big)(\widehat{\mathrm{u}})$$

$\Rightarrow \widehat{\mathrm{u}}$ solution to the dual problem.
The second assertion is shown in a similar manner.

Fenchel-Rockafellar duality

Particular case :

If $f = \varphi + \frac{1}{2}\| \cdot - z\|^2$ where $\varphi \in \Gamma_0(\mathcal{H})$ and $\mathrm{z} \in \mathcal{H}$, then

$$\begin{aligned} -\mathrm{L}^*\widehat{\mathrm{u}} \in \partial f(\widehat{\mathrm{x}}) &\Leftrightarrow -\mathrm{L}^*\widehat{\mathrm{u}} \in \partial\varphi(\widehat{\mathrm{x}}) + \widehat{\mathrm{x}} - \mathrm{z} \\ &\Leftrightarrow 0 \in \widehat{\mathrm{x}} + \mathrm{L}^*\widehat{\mathrm{u}} - \mathrm{z} + \partial\varphi(\widehat{\mathrm{x}}). \end{aligned}$$

Hence,

$$\widehat{\mathrm{x}} = \mathrm{prox}_{\varphi}(-\mathrm{L}^*\widehat{\mathrm{u}} + \mathrm{z}).$$

Dual FB algorithm

Let $\mathrm{z} \in \mathbb{R}^N$, g $\in \Gamma_0(\mathbb{R}^M)$ and $\mathrm{L} \in \mathbb{R}^{M \times N}$.

Primal problem:

$$\widehat{\mathrm{x}} = \arg\min_{\mathrm{x} \in \mathbb{R}^N} \frac{1}{2}\|\mathrm{x} - \mathrm{z}\|^2 + g(\mathrm{Lx})$$

Dual problem:

$$\widehat{\mathrm{u}} \in \operatorname*{Argmin}_{\mathrm{u} \in \mathbb{R}^M} \frac{1}{2}\|\mathrm{z} - \mathrm{L}^*\mathrm{u}\| + g^*(\mathrm{u})$$

Choose $\mathrm{u}^{[0]} \in \mathbb{R}^M$ and $\tau \in]0, 2/\|\mathrm{L}\|^2[$.

For $k = 0, 1, \ldots$

$$\mathrm{x}^{[k]} = \mathrm{z} - \mathrm{L}^*\mathrm{u}^{[k]}$$

$$\mathrm{u}_{k+1} = \mathrm{prox}_{\tau g^*}\left(\mathrm{u}^{[k]} + \tau \mathrm{L}\mathrm{x}^{[k]}\right)$$

[Combettes, Dung, Vũ, 2011]

- The sequence $(\mathrm{u}^{[k]})_{k \in \mathbb{N}}$ converges to a solution to the dual problem $\widehat{\mathrm{u}}$.
- The sequence $(\mathrm{x}^{[k]})_{k \in \mathbb{N}}$ converges to a solution to the primal problem

$$\widehat{\mathrm{x}} = \mathrm{z} - \mathrm{L}^*\widehat{\mathrm{u}}.$$

Dual FB algorithm

Let $\mathrm{z} \in \mathbb{R}^N$, $f \in \Gamma_0(\mathbb{R}^N)$, g $\in \Gamma_0(\mathbb{R}^M)$ and $\mathrm{L} \in \mathbb{R}^{M \times N}$.

Primal problem:

$$\widehat{\mathrm{x}} = \arg \min_{\mathrm{x} \in \mathbb{R}^N} f(\mathrm{x}) + \frac{1}{2}\|\mathrm{x} - \mathrm{z}\|^2 + g(\mathrm{L}\mathrm{x})$$

Dual problem:

$$\widehat{\mathrm{u}} \in \mathrm{Argmin}_{\mathrm{u} \in \mathbb{R}^M} \widetilde{f^*}(\mathrm{z} - \mathrm{L}^*\mathrm{u}) + g^*(\mathrm{u})$$

Dual FB algorithm

Let $z \in \mathbb{R}^N$, $f \in \Gamma_0(\mathbb{R}^N)$, g $\in \Gamma_0(\mathbb{R}^M)$ and $L \in \mathbb{R}^{M\times N}$.

Primal problem:

$$\widehat{x} = \arg\min_{x\in\mathbb{R}^N} f(x) + \frac{1}{2}\|x - z\|^2 + g(Lx)$$

Dual problem:

$$\widehat{u} \in \text{Argmin}_{u\in\mathbb{R}^M} \widetilde{f^*}(z - L^* u) + g^*(u)$$

REMARK: [(Lem. 2.5) Combettes *et al*, 2010]

Let $\varphi = f + \frac{1}{2}\|\cdot - z\|^2$. Then $\varphi^* = \widetilde{f^*}(\cdot + z) - \frac{1}{2}\|y\|^2$

Where $\widetilde{f^*}$ is the Moreau enveloppe of f^*: $\widetilde{f^*}(v) = \min_y f^*(y) + \frac{1}{2}\|y - v\|^2$

DUAL PROBLEM: Find $\widehat{u} \in \text{Argmin}_{u\in\mathbb{R}^M} \widetilde{f^*}(z - L^* u) + g^*(u)$

- $\widetilde{f^*}$ is differentiable and $\nabla\widetilde{f^*} = \text{prox}_f = -\text{prox}_{f^*}$ [Moreau, 1965]
- Use FB on the dual problem!

Dual FB algorithm

Let $z \in \mathbb{R}^N$, $f \in \Gamma_0(\mathbb{R}^N)$, g $\in \Gamma_0(\mathbb{R}^M)$ and $L \in \mathbb{R}^{M \times N}$.

Primal problem:

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} f(x) + \frac{1}{2}\|x - z\|^2 + g(Lx)$$

Dual problem:

$$\widehat{u} \in \mathrm{Argmin}_{u \in \mathbb{R}^M} \widetilde{f^*}(z - L^* u) + g^*(u)$$

Choose $u_0 \in \mathbb{R}^M$ and $\tau \in]0, 2/\|L\|^2[$.

For $k = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x^{[k]} = \mathrm{prox}_f\left(z - L^* u^{[k]}\right) \\ u^{[k+1]} = \mathrm{prox}_{\tau g^*}\left(u^{[k]} + \tau L x^{[k]}\right) \end{array} \right.$$

[Combettes, Dung, Vũ, 2011]

The sequence $(u^{[k]})_{k \in \mathbb{N}}$ converges to a solution to the dual problem $\widehat{u}$.

The sequence $(x^{[k]})_{k \in \mathbb{N}}$ converges to a solution to the primal problem

$$\widehat{x} = \mathrm{prox}_f(z - L^* \widehat{u}).$$

Primal-dual methods

Augmented Lagrangian method

ADMM algorithm (*Alternating-direction method of multipliers*)
⇒ **Lagrangian interpretation**

$$\underset{\mathrm{x}\in\mathcal{H}}{\text{minimize}}\; f(\mathrm{x}) + g(\mathrm{Lx}) \quad \Leftrightarrow \quad \underset{\substack{\mathrm{x}\in\mathcal{H},\mathrm{u}\in\mathcal{G}\\ \mathrm{Lx}=\mathrm{u}}}{\text{minimize}}\; f(\mathrm{x}) + g(\mathrm{u})$$

- Lagrange function : $\mathcal{L}(\mathrm{x}, \mathrm{u}, \mathrm{v}) = f(\mathrm{x}) + g(\mathrm{u}) + \langle \mathrm{v} \mid \mathrm{Lx} - \mathrm{u} \rangle$

$\Rightarrow$ $\mathrm{v} \in \mathcal{G}$ denotes the Lagrange multiplier.

Augmented Lagrangian method

ADMM algorithm (*Alternating-direction method of multipliers*)
$\Rightarrow$ **Lagrangian interpretation**

$$\underset{\mathrm{x}\in\mathcal{H}}{\text{minimize}}\; f(\mathrm{x}) + g(\mathrm{Lx}) \quad\Leftrightarrow\quad \underset{\substack{\mathrm{x}\in\mathcal{H},\mathrm{u}\in\mathcal{G}\\ \mathrm{Lx}=\mathrm{u}}}{\text{minimize}}\; f(\mathrm{x}) + g(\mathrm{u})$$

- Lagrange function: $\mathcal{L}(\mathrm{x},\mathrm{u},\mathrm{v}) = f(\mathrm{x}) + g(\mathrm{u}) + \langle \mathrm{v} \mid \mathrm{Lx} - \mathrm{u}\rangle$

$\Rightarrow$ $\mathrm{v} \in \mathcal{G}$ denotes the Lagrange multiplier.

- Idea: iterations for finding a saddle point $(\widehat{\mathrm{x}}, \widehat{\mathrm{u}}, \widehat{\mathrm{v}})$:

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} \mathrm{x}^{[k]} \in \operatorname{Argmin} \mathcal{L}(\cdot, \mathrm{u}^{[k]}, \mathrm{v}^{[k]}) \\ \mathrm{u}^{[k+1]} \in \operatorname{Argmin} \mathcal{L}(\mathrm{x}^{[k]}, \cdot, \mathrm{v}^{[k]}) \\ \mathrm{v}^{[k+1]} \text{ such that } \mathcal{L}(\mathrm{x}^{[k]}, \mathrm{u}^{[k+1]}, \mathrm{v}^{[k+1]}) \geq \mathcal{L}(\mathrm{x}^{[k]}, \mathrm{y}^{[k+1]}, \mathrm{v}^{[k]}). \end{cases}$$

But **the convergence is not guaranteed in general !**

Augmented Lagrangian method

ADMM algorithm (*Alternating-direction method of multipliers*)
$\Rightarrow$ **Lagrangian interpretation**

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ f(x)+g(Lx) \quad \Leftrightarrow \quad \underset{\substack{x\in\mathcal{H},u\in\mathcal{G}\\ Lx=u}}{\text{minimize}}\ f(x)+g(u)$$

- Lagrange function: $\mathcal{L}(x,u,v) = f(x) + g(u) + \langle v \mid Lx - u \rangle$

$\Rightarrow$ $v \in \mathcal{G}$ denotes the Lagrange multiplier.

- Solution: introduce an **Augmented Lagrange function**:

$$\widetilde{\mathcal{L}}(x,u,w) = f(x) + g(u) + \gamma \langle w \mid Lx - u \rangle + \frac{\gamma}{2}\|Lx - u\|^2$$

$\Rightarrow$ The Lagrange multiplier is $v = \gamma w$ with $\gamma > 0$.

Alternating-direction method of multipliers

Algorithm for finding a saddle point:

$$(\forall k \in \mathbb{N}) \qquad \begin{cases} \mathrm{x}^{[k]} \in \underset{\mathrm{x}\in\mathcal{H}}{\mathrm{Argmin}}\ \widetilde{\mathcal{L}}(\mathrm{x}, \mathrm{y}^{[k]}, \mathrm{w}^{[k]}) \\ \mathrm{y}^{[k+1]} \in \underset{\mathrm{y}\in\mathcal{G}}{\mathrm{Argmin}}\ \widetilde{\mathcal{L}}(\mathrm{x}^{[k]}, \mathrm{y}, \mathrm{w}^{[k]}) \\ \mathrm{w}^{[k+1]} \text{ such that } \widetilde{\mathcal{L}}(\mathrm{x}^{[k]}, \mathrm{y}^{[k+1]}, \mathrm{w}^{[k+1]}) \geq \widetilde{\mathcal{L}}(\mathrm{x}^{[k]}, \mathrm{y}^{[k+1]}, \mathrm{w}^{[k]}). \end{cases}$$

By performing a gradient ascent on the Lagrange multiplier,

$$(\forall k \in \mathbb{N}) \qquad \begin{cases} \mathrm{x}^{[k]} \in \underset{\mathrm{x}\in\mathcal{H}}{\mathrm{Argmin}}\ f(\mathrm{x}) + \gamma \left\langle \mathrm{w}^{[k]} \mid \mathrm{Lx} - \mathrm{y}^{[k]} \right\rangle + \frac{\gamma}{2}\|\mathrm{Lx} - \mathrm{y}^{[k]}\|^2 \\ \mathrm{y}^{[k+1]} \in \underset{\mathrm{y}\in\mathcal{G}}{\mathrm{Argmin}}\ g(\mathrm{y}) + \gamma \left\langle \mathrm{w}^{[k]} \mid \mathrm{Lx}^{[k]} - \mathrm{y} \right\rangle + \frac{\gamma}{2}\|\mathrm{Lx}^{[k]} - \mathrm{y}\|^2 \\ \mathrm{w}^{[k+1]} = \mathrm{w}^{[k]} + \frac{1}{\gamma}\nabla_{\mathrm{w}}\widetilde{\mathcal{L}}(\mathrm{x}^{[k]}, \mathrm{y}^{[k+1]}, \mathrm{w}^{[k]}) \end{cases}$$

$$\Leftrightarrow \quad (\forall k \in \mathbb{N}) \qquad \begin{cases} \mathrm{x}^{[k]} \in \underset{\mathrm{x}\in\mathcal{H}}{\mathrm{Argmin}}\ \frac{1}{2}\left\|\mathrm{Lx} - \mathrm{y}^{[k]} + \mathrm{w}^{[k]}\right\|^2 + \frac{1}{\gamma}f(\mathrm{x}) \\ \mathrm{y}^{[k+1]} = \mathrm{prox}_{\frac{g}{\gamma}}\left(\mathrm{w}^{[k]} + \mathrm{Lx}^{[k]}\right) \\ \mathrm{w}^{[k+1]} = \mathrm{w}^{[k]} + \mathrm{Lx}^{[k]} - \mathrm{y}^{[k+1]}. \end{cases}$$

Augmented Lagrange method

ADMM algorithm (*Alternating-direction method of multipliers*)

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

Let $\mathrm{L} \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ such that $\mathrm{L}^*\mathrm{L}$ is an isomorphism and let $\gamma > 0$.

$$(\forall k \in \mathbb{N}) \qquad \begin{cases} \mathrm{x}^{[k]} \in \underset{\mathrm{x} \in \mathcal{H}}{\mathrm{Argmin}}\ \frac{1}{2} \left\| \mathrm{Lx} - \mathrm{y}^{[k]} + \mathrm{w}^{[k]} \right\|^2 + \frac{1}{\gamma} f(\mathrm{x}) \\ \mathrm{s}^{[k]} = \mathrm{Lx}^{[k]} \\ \mathrm{y}^{[k+1]} = \mathrm{prox}_{\frac{g}{\gamma}} \left(\mathrm{w}^{[k]} + \mathrm{s}^{[k]} \right) \\ \mathrm{w}^{[k+1]} = \mathrm{w}^{[k]} + \mathrm{s}^{[k]} - \mathrm{y}^{[k+1]}. \end{cases}$$

ADMM algorithm (*Alternating-direction method of multipliers*)

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

Let $\mathrm{L} \in \mathcal{B}(\mathcal{H},\mathcal{G})$ such that $\mathrm{L}^*\mathrm{L}$ is an isomorphism and let $\gamma > 0$.

We assume that $(\operatorname{dom} g) \cap \mathrm{L}(\operatorname{dom} f) \neq \emptyset$ or $\operatorname{dom} g \cap \big(\mathrm{L}(\operatorname{dom} f)\big) \neq \emptyset$ and that $\operatorname{Argmin}(f + g \circ \mathrm{L}) \neq \emptyset$. Let

$$(\forall k \in \mathbb{N}) \qquad \begin{cases} \mathrm{x}^{[k]} \in \underset{\mathrm{x}\in\mathcal{H}}{\operatorname{Argmin}} \ \frac{1}{2}\left\|\mathrm{Lx} - \mathrm{y}^{[k]} + \mathrm{w}^{[k]}\right\|^2 + \frac{1}{\gamma} f(\mathrm{x}) \\ \mathrm{s}^{[k]} = \mathrm{Lx}^{[k]} \\ \mathrm{y}^{[k+1]} = \operatorname{prox}_{\frac{g}{\gamma}}\left(\mathrm{w}^{[k]} + \mathrm{s}^{[k]}\right) \\ \mathrm{w}^{[k+1]} = \mathrm{w}^{[k]} + \mathrm{s}^{[k]} - \mathrm{y}^{[k+1]}. \end{cases}$$

We have:

- $\mathrm{x}^{[k]} \rightharpoonup \widehat{\mathrm{x}}$ where $\widehat{\mathrm{x}} \in \operatorname{Argmin}(f + g \circ \mathrm{L})$
- $\gamma \mathrm{w}^{[k]} \rightharpoonup \widehat{\mathrm{v}}$ where $\widehat{\mathrm{v}} \in \operatorname{Argmin}(f^* \circ (-\mathrm{L}^*) + g^*)$.

Augmented Lagrange method

ADMM algorithm (*Alternating-direction method of multipliers*)

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

Let $\mathrm{L} \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ such that $\mathrm{L}^*\mathrm{L}$ is an isomorphism and let $\gamma > 0$.

We assume that $(\mathrm{dom}\, g) \cap \mathrm{L}(\mathrm{dom}\, f) \neq \emptyset$ or $\mathrm{dom}\, g \cap \big(\mathrm{L}(\mathrm{dom}\, f)\big) \neq \emptyset$ and that $\mathrm{Argmin}(f + g \circ \mathrm{L}) \neq \emptyset$. Let

$$(\forall k \in \mathbb{N}) \qquad \begin{cases} \mathrm{x}^{[k]} \in \underset{\mathrm{x} \in \mathcal{H}}{\mathrm{Argmin}}\ \frac{1}{2} \left\| \mathrm{Lx} - \mathrm{y}^{[k]} + \mathrm{w}^{[k]} \right\|^2 + \frac{1}{\gamma} f(\mathrm{x}) \\ \mathrm{s}^{[k]} = \mathrm{Lx}^{[k]} \\ \mathrm{y}^{[k+1]} = \mathrm{prox}_{\frac{g}{\gamma}} \left(\mathrm{w}^{[k]} + \mathrm{s}^{[k]} \right) \\ \mathrm{w}^{[k+1]} = \mathrm{w}^{[k]} + \mathrm{s}^{[k]} - \mathrm{y}^{[k+1]}. \end{cases}$$

We have:

- $\mathrm{x}^{[k]} \rightharpoonup \widehat{\mathrm{x}}$ where $\widehat{\mathrm{x}} \in \mathrm{Argmin}(f + g \circ \mathrm{L})$
- $\gamma \mathrm{w}^{[k]} \rightharpoonup \widehat{\mathrm{v}}$ where $\widehat{\mathrm{v}} \in \mathrm{Argmin}(f^* \circ (-\mathrm{L}^*) + g^*)$.

Problem formulation

Let $f_1 \in \Gamma_0(\mathbb{R}^N)$, $f_2 \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^M)$ and $\mathrm{L} \in \mathbb{R}^{M \times N}$.

Primal problem: $$\widehat{\mathrm{x}} \in \underset{\mathrm{x} \in \mathbb{R}^N}{\mathrm{Argmin}}\ f_1(\mathrm{x}) + f_2(\mathrm{x}) + g(\mathrm{L}\mathrm{x})$$

Dual problem: $$\widehat{\mathrm{u}} \in \underset{\mathrm{u} \in \mathbb{R}^M}{\mathrm{Argmin}} (f_1 + f_2)^*(\mathrm{L}^*\mathrm{u}) + g^*(\mathrm{u})$$

Problem formulation

Let $f_1 \in \Gamma_0(\mathbb{R}^N)$, $f_2 \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^M)$ and $\mathrm{L} \in \mathbb{R}^{M \times N}$.

Primal problem: $$\widehat{\mathrm{x}} \in \underset{\mathrm{x} \in \mathbb{R}^N}{\mathrm{Argmin}}\, f_1(\mathrm{x}) + f_2(\mathrm{x}) + g(\mathrm{Lx})$$

Dual problem: $$\widehat{\mathrm{u}} \in \underset{\mathrm{u} \in \mathbb{R}^M}{\mathrm{Argmin}} (f_1 + f_2)^*(\mathrm{L}^*\mathrm{u}) + g^*(\mathrm{u})$$

Lagrangian formulation Another formulation of the Primal-Dual problem is to combine them into the search of a **saddle point of the Lagrangian:**

$$(\widehat{\mathrm{x}}, \widehat{\mathrm{u}}) \in \underset{\mathrm{x} \in \mathbb{R}^N}{\mathrm{Argmin}}\, \max_{\mathrm{u} \in \mathbb{R}^M} f_1(\mathrm{x}) + f_2(\mathrm{x}) - g^*(\mathrm{u}) + \langle \mathrm{Lx}, \mathrm{u} \rangle$$

Problem formulation

Let $f_1 \in \Gamma_0(\mathbb{R}^N)$, $f_2 \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^M)$ and $\mathrm{L} \in \mathbb{R}^{M \times N}$.

Primal problem: $$\widehat{\mathrm{x}} \in \underset{\mathrm{x} \in \mathbb{R}^N}{\mathrm{Argmin}}\ f_1(\mathrm{x}) + f_2(\mathrm{x}) + g(\mathrm{Lx})$$

Dual problem: $$\widehat{\mathrm{u}} \in \underset{\mathrm{u} \in \mathbb{R}^M}{\mathrm{Argmin}} (f_1 + f_2)^*(\mathrm{L}^*\mathrm{u}) + g^*(\mathrm{u})$$

Lagrangian formulation Another formulation of the Primal-Dual problem is to combine them into the search of a **saddle point of the Lagrangian:**

$$(\widehat{\mathrm{x}}, \widehat{\mathrm{u}}) \in \underset{\mathrm{x} \in \mathbb{R}^N}{\mathrm{Argmin}}\ \underset{\mathrm{u} \in \mathbb{R}^M}{\max}\ f_1(\mathrm{x}) + f_2(\mathrm{x}) - g^*(\mathrm{u}) + \langle \mathrm{Lx}, \mathrm{u} \rangle$$

REMARK: Recall that, for $\psi \in \Gamma_0(\mathbb{R}^N)$, $\widehat{\mathrm{x}} \in \mathrm{Argmin}\ \psi \Leftrightarrow \mathbf{0} \in \partial\psi(\widehat{\mathrm{x}})$

Do we have similar conditions for the primal-dual problem?

$\rightsquigarrow$ Look at the Lagrangian saddle point problem and derive optimal conditions for $\widehat{\mathrm{x}}$, and for $\widehat{\mathrm{u}}$ *alternatively*

$\rightsquigarrow$ These are called KKT conditions

Problem formulation

Let $f_1 \in \Gamma_0(\mathbb{R}^N)$, $f_2 \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^M)$ and $\mathrm{L} \in \mathbb{R}^{M \times N}$.

Primal problem: $$\widehat{\mathrm{x}} \in \underset{\mathrm{x} \in \mathbb{R}^N}{\mathrm{Argmin}}\ f_1(\mathrm{x}) + f_2(\mathrm{x}) + g(\mathrm{Lx})$$

Dual problem: $$\widehat{\mathrm{u}} \in \underset{\mathrm{u} \in \mathbb{R}^M}{\mathrm{Argmin}} (f_1 + f_2)^*(\mathrm{L}^*\mathrm{u}) + g^*(\mathrm{u})$$

Lagrangian formulation Another formulation of the Primal-Dual problem is to combine them into the search of a **saddle point of the Lagrangian:**

$$(\widehat{\mathrm{x}}, \widehat{\mathrm{u}}) \in \underset{\mathrm{x} \in \mathbb{R}^N}{\mathrm{Argmin}}\ \underset{\mathrm{u} \in \mathbb{R}^M}{\max}\ f_1(\mathrm{x}) + f_2(\mathrm{x}) - g^*(\mathrm{u}) + \langle \mathrm{Lx}, \mathrm{u} \rangle$$

Karush-Kuhn-Tucker conditions

Assume that $\mathrm{dom} g \cap \mathrm{L}(\mathrm{dom} f) \neq \emptyset$ and f_2 differentiable.
$(\widehat{\mathrm{x}}, \widehat{\mathrm{u}}) \in \mathbb{R}^N \times \mathbb{R}^M$ is a solution to the Primal-Dual problem if and only if

$$\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \in \begin{pmatrix} \partial f_1(\widehat{\mathrm{x}}) + \mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}}) \\ -\mathrm{L}\widehat{\mathrm{x}} + \partial g^*(\widehat{\mathrm{u}}) \end{pmatrix}$$

From KKT to fixed-point equations...

KKT:

$$\begin{cases} \mathbf{0} \in \partial f_1(\widehat{\mathrm{x}}) + \mathrm{L}^* \widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}}) \\ \mathbf{0} \in -\mathrm{L}\widehat{\mathrm{x}} + \partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

From KKT to fixed-point equations...

KKT:

$$\begin{cases} \mathbf{0} \in \partial f_1(\widehat{\mathrm{x}}) + \mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}}) \\ \mathbf{0} \in -\mathrm{L}\widehat{\mathrm{x}} + \partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Multiply by $\tau > 0$ the first equation and $\sigma > 0$ the second equation:

$$\begin{cases} -\tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big) \in \tau\partial f_1(\widehat{\mathrm{x}}) \\ \sigma\mathrm{L}\widehat{\mathrm{x}} \in \sigma\partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

From KKT to fixed-point equations...

KKT:

$$\begin{cases} \mathbf{0} \in \partial f_1(\widehat{\mathrm{x}}) + \mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}}) \\ \mathbf{0} \in -\mathrm{L}\widehat{\mathrm{x}} + \partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Multiply by $\tau > 0$ the first equation and $\sigma > 0$ the second equation:

$$\begin{cases} -\tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big) \in \tau\partial f_1(\widehat{\mathrm{x}}) \\ \sigma\mathrm{L}\widehat{\mathrm{x}} \in \sigma\partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Since $\widehat{\mathrm{x}} - \widehat{\mathrm{x}} = \mathbf{0}$, and $\widehat{\mathrm{u}} - \widehat{\mathrm{u}} = \mathbf{0}$, the last equations are equivalent to

$$\begin{cases} \widehat{\mathrm{x}} - \tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big) - \widehat{\mathrm{x}} \in \tau\partial f_1(\ \widehat{\mathrm{x}}\ ) \\ \\ \widehat{\mathrm{u}} + \sigma\mathrm{L}\widehat{\mathrm{x}} - \widehat{\mathrm{u}} \in \sigma\partial g^*(\ \widehat{\mathrm{u}}\ ) \end{cases}$$

From KKT to fixed-point equations...

KKT:

$$\begin{cases} \mathbf{0} \in \partial f_1(\widehat{\mathrm{x}}) + \mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}}) \\ \mathbf{0} \in -\mathrm{L}\widehat{\mathrm{x}} + \partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Multiply by $\tau > 0$ the first equation and $\sigma > 0$ the second equation:

$$\begin{cases} -\tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big) \in \tau\partial f_1(\widehat{\mathrm{x}}) \\ \sigma\mathrm{L}\widehat{\mathrm{x}} \in \sigma\partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Since $\widehat{\mathrm{x}} - \widehat{\mathrm{x}} = \mathbf{0}$, and $\widehat{\mathrm{u}} - \widehat{\mathrm{u}} = \mathbf{0}$, the last equations are equivalent to

$$\begin{cases} \widehat{\mathrm{x}} - \tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big) - \ \widehat{\mathrm{x}} \ \in \tau\partial f_1(\ \widehat{\mathrm{x}}\ ) \\ \\ \widehat{\mathrm{u}} + \sigma\mathrm{L}(2\widehat{\mathrm{x}} - \widehat{\mathrm{x}}) - \ \widehat{\mathrm{u}} \ \in \sigma\partial g^*(\ \widehat{\mathrm{u}}\ ) \end{cases}$$

From KKT to fixed-point equations...

KKT:

$$\begin{cases} \mathbf{0} \in \partial f_1(\widehat{\mathrm{x}}) + \mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}}) \\ \mathbf{0} \in -\mathrm{L}\widehat{\mathrm{x}} + \partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Multiply by $\tau > 0$ the first equation and $\sigma > 0$ the second equation:

$$\begin{cases} -\tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big) \in \tau\partial f_1(\widehat{\mathrm{x}}) \\ \sigma\mathrm{L}\widehat{\mathrm{x}} \in \sigma\partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Since $\widehat{\mathrm{x}} - \widehat{\mathrm{x}} = \mathbf{0}$, and $\widehat{\mathrm{u}} - \widehat{\mathrm{u}} = \mathbf{0}$, the last equations are equivalent to

$$\begin{cases} \underbrace{\widehat{\mathrm{x}} - \tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big)}_{\overline{\mathrm{x}}} - \underbrace{\widehat{\mathrm{x}}}_{\overline{\mathbf{p}}} \in \tau\partial f(\underbrace{\widehat{\mathrm{x}}}_{\overline{\mathbf{p}}}) \quad \rightsquigarrow \mathrm{prox}_{\tau f_1} \\ \underbrace{\widehat{\mathrm{u}} + \sigma\mathrm{L}(2\widehat{\mathrm{x}} - \widehat{\mathrm{x}})}_{\overline{\mathrm{x}}} - \underbrace{\widehat{\mathrm{u}}}_{\overline{\mathbf{p}}} \in \sigma\partial g^*(\underbrace{\widehat{\mathrm{u}}}_{\overline{\mathbf{p}}}) \quad \rightsquigarrow \mathrm{prox}_{\sigma g^*} \end{cases}$$

Prox characterisation: $\overline{\mathrm{x}} - \overline{\mathbf{p}} \in \gamma\partial\psi(\overline{\mathbf{p}}) \Leftrightarrow \overline{\mathbf{p}} = \mathrm{prox}_{\gamma\psi}(\overline{\mathrm{x}})$

From KKT to fixed-point equations...

KKT:

$$\begin{cases} \mathbf{0} \in \partial f_1(\widehat{\mathrm{x}}) + \mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}}) \\ \mathbf{0} \in -\mathrm{L}\widehat{\mathrm{x}} + \partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Multiply by $\tau > 0$ the first equation and $\sigma > 0$ the second equation:

$$\begin{cases} -\tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big) \in \tau\partial f_1(\widehat{\mathrm{x}}) \\ \sigma\mathrm{L}\widehat{\mathrm{x}} \in \sigma\partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Since $\widehat{\mathrm{x}} - \widehat{\mathrm{x}} = \mathbf{0}$, and $\widehat{\mathrm{u}} - \widehat{\mathrm{u}} = \mathbf{0}$, the last equations are equivalent to

$$\begin{cases} \underbrace{\widehat{\mathrm{x}} - \tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big)}_{\overline{\mathrm{x}}} - \underbrace{\widehat{\mathrm{x}}}_{\overline{\mathbf{p}}} \in \tau\partial f(\underbrace{\widehat{\mathrm{x}}}_{\overline{\mathbf{p}}}) \quad \rightsquigarrow \mathrm{prox}_{\tau f_1} \\ \underbrace{\widehat{\mathrm{u}} + \sigma\mathrm{L}(2\widehat{\mathrm{x}} - \widehat{\mathrm{x}})}_{\overline{\mathrm{x}}} - \underbrace{\widehat{\mathrm{u}}}_{\overline{\mathbf{p}}} \in \sigma\partial g^*(\underbrace{\widehat{\mathrm{u}}}_{\overline{\mathbf{p}}}) \quad \rightsquigarrow \mathrm{prox}_{\sigma g^*} \end{cases}$$

$$\Leftrightarrow \begin{cases} \widehat{\mathrm{x}} = \mathrm{prox}_{\tau f_1}\Big(\widehat{\mathrm{x}} - \tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big)\Big) \\ \widehat{\mathrm{u}} = \mathrm{prox}_{\sigma g^*}\Big(\widehat{\mathrm{u}} + \sigma\mathrm{L}(2\widehat{\mathrm{x}} - \widehat{\mathrm{x}})\Big) \end{cases}$$

From KKT to fixed-point equations...

KKT:

$$\begin{cases} \mathbf{0} \in \partial f_1(\widehat{\mathrm{x}}) + \mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}}) \\ \mathbf{0} \in -\mathrm{L}\widehat{\mathrm{x}} + \partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Multiply by $\tau > 0$ the first equation and $\sigma > 0$ the second equation:

$$\begin{cases} -\tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big) \in \tau\partial f_1(\widehat{\mathrm{x}}) \\ \sigma\mathrm{L}\widehat{\mathrm{x}} \in \sigma\partial g^*(\widehat{\mathrm{u}}) \end{cases}$$

Since $\widehat{\mathrm{x}} - \widehat{\mathrm{x}} = \mathbf{0}$, and $\widehat{\mathrm{u}} - \widehat{\mathrm{u}} = \mathbf{0}$, the last equations are equivalent to

$$\begin{cases} \underbrace{\widehat{\mathrm{x}} - \tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big)}_{\overline{\mathrm{x}}} - \underbrace{\widehat{\mathrm{x}}}_{\overline{\mathbf{p}}} \in \tau\partial f(\underbrace{\widehat{\mathrm{x}}}_{\overline{\mathbf{p}}}) & \rightsquigarrow \mathrm{prox}_{\tau f_1} \\ \underbrace{\widehat{\mathrm{u}} + \sigma\mathrm{L}(2\widehat{\mathrm{x}} - \widehat{\mathrm{x}})}_{\overline{\mathrm{x}}} - \underbrace{\widehat{\mathrm{u}}}_{\overline{\mathbf{p}}} \in \sigma\partial g^*(\underbrace{\widehat{\mathrm{u}}}_{\overline{\mathbf{p}}}) & \rightsquigarrow \mathrm{prox}_{\sigma g^*} \end{cases}$$

$$\Leftrightarrow \begin{cases} \widehat{\mathrm{x}} = \mathrm{prox}_{\tau f_1}\Big(\widehat{\mathrm{x}} - \tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big)\Big) \\ \widehat{\mathrm{u}} = \mathrm{prox}_{\sigma g^*}\Big(\widehat{\mathrm{u}} + \sigma\mathrm{L}(2\widehat{\mathrm{x}} - \widehat{\mathrm{x}})\Big) \end{cases} \quad \rightsquigarrow \textit{Fixed-point equations}$$

Fixed-point algorithm

From the fixed-point equations:

$$\begin{cases} \widehat{\mathrm{x}} = \mathrm{prox}_{\tau f_1}\Big(\widehat{\mathrm{x}} - \tau\big(\mathrm{L}^*\widehat{\mathrm{u}} + \nabla f_2(\widehat{\mathrm{x}})\big)\Big) \\ \widehat{\mathrm{u}} = \mathrm{prox}_{\sigma g^*}\Big(\widehat{\mathrm{u}} + \sigma \mathrm{L}(2\widehat{\mathrm{x}} - \widehat{\mathrm{x}})\Big) \end{cases}$$

We can derive a fixed-point algorithm:

For $k = 0, 1, \ldots$

$$\begin{array}{l} \mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau f_1}\Big(\mathrm{x}^{[k]} - \tau\big(\mathrm{L}^*\mathrm{u}^{[k]} + \nabla f_2(\mathrm{x}^{[k]})\big)\Big) \\ \mathrm{u}^{[k+1]} = \mathrm{prox}_{\sigma g^*}\Big(\mathrm{u}^{[k]} + \sigma \mathrm{L}(2\mathrm{x}^{[k+1]} - \mathrm{x}^{[k]})\Big) \end{array}$$

REMARKS:

- This algorithm is known as the Condat-Vũ algorithm

Step-size and convergence of Condat-Vu algorithm

Let $f_1 \in \Gamma_0(\mathbb{R}^N)$, $f_2 \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^M)$ and $\mathrm{L} \in \mathbb{R}^{M\times N}$.

Primal problem: $$\widehat{\mathrm{x}} \in \underset{\mathrm{x}\in\mathbb{R}^N}{\mathrm{Argmin}}\ f_1(\mathrm{x}) + f_2(\mathrm{x}) + g(\mathrm{Lx})$$

Dual problem: $$\widehat{\mathrm{u}} \in \underset{\mathrm{u}\in\mathbb{R}^M}{\mathrm{Argmin}} (f_1 + f_2)^*(\mathrm{L}^*\mathrm{u}) + g^*(\mathrm{u})$$

Choose $\tau > 0$ and $\sigma > 0$ such that $\frac{1}{\tau} - \sigma\|\mathrm{L}\|^2 > \frac{\beta}{2}$ with f_2 β-Lipschitz gradient.

For $k = 0, 1, \dots$

$$\left\lfloor \begin{array}{l} \mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau f_1}\left(\mathrm{x}^{[k]} - \tau\left(\nabla f_2(\mathrm{x}^{[k]}) + \mathrm{L}^*\mathrm{u}^{[k]}\right)\right) \\ \mathrm{u}^{[k+1]} = \mathrm{prox}_{\sigma g^*}\left(\mathrm{u}^{[k]} + \sigma\mathrm{L}(2\mathrm{x}^{[k+1]} - \mathrm{x}^{[k]})\right) \end{array}\right.$$

The sequence $(\mathrm{x}^{[k]})_{k\in\mathbb{N}}$ converges to a solution to the primal problem.

The sequence $(\mathrm{u}^{[k]})_{k\in\mathbb{N}}$ converges to a solution to the dual problem.

[Vũ, 2013][Condat, 2013]

Particular cases

CONDAT-VŨ ALGORITHM: [Vũ, 2013][Condat, 2013]

PROBLEM: Find $\widehat{x} \in \operatorname{Argmin}_{x \in \mathbb{R}^N} f_1(x) + f_2(x) + g(Lx)$

Choose $\tau > 0$ and $\sigma > 0$ such that $\frac{1}{\tau} - \sigma \|L\|^2 > \frac{\beta}{2}$ with f_2 β-Lipschitz.

For $k = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} x^{[k+1]} = \operatorname{prox}_{\tau f_1}\Big(x^{[k]} - \tau\big(\nabla f_2(x^{[k]}) + L^* u^{[k]}\big)\Big) \\ u^{[k+1]} = \operatorname{prox}_{\sigma g^*}\Big(u^{[k]} + \sigma L(2x^{[k+1]} - x^{[k]})\Big) \end{array} \right.$$

Particular cases

Condat-Vũ algorithm: [Vũ, 2013][Condat, 2013]

Problem: Find $\widehat{\mathrm{x}} \in \operatorname{Argmin}_{\mathrm{x}\in\mathbb{R}^N} f_1(\mathrm{x}) + f_2(\mathrm{x}) + g(\mathrm{Lx})$

Choose $\tau > 0$ and $\sigma > 0$ such that $\frac{1}{\tau} - \sigma\|\mathrm{L}\|^2 > \frac{\beta}{2}$ with f_2 β-Lipschitz.

For $k = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} \mathrm{x}^{[k+1]} = \operatorname{prox}_{\tau f_1}\left(\mathrm{x}^{[k]} - \tau\left(\nabla f_2(\mathrm{x}^{[k]}) + \mathrm{L}^*\mathrm{u}^{[k]}\right)\right) \\ \mathrm{u}^{[k+1]} = \operatorname{prox}_{\sigma g^*}\left(\mathrm{u}^{[k]} + \sigma\mathrm{L}(2\mathrm{x}^{[k+1]} - \mathrm{x}^{[k]})\right) \end{array}\right.$$

Chambolle-Pock algorithm: $f_2 \equiv 0$ [Chambolle & Pock, 2011]

Problem: Find $\widehat{\mathrm{x}} \in \operatorname{Argmin}_{\mathrm{x}\in\mathbb{R}^N} f_1(\mathrm{x}) + g(\mathrm{Lx})$

Choose $\tau > 0$ and $\sigma > 0$ such that $\sigma\tau\|\mathrm{L}\|^2 < 1$.

For $k = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} \mathrm{x}^{[k+1]} = \operatorname{prox}_{\tau f_1}(\mathrm{x}^{[k]} - \tau\mathrm{L}^*\mathrm{u}^{[k]}) \\ \mathrm{u}^{[k+1]} = \operatorname{prox}_{\sigma g^*}(\mathrm{u}^{[k]} + \sigma\mathrm{L}(2\mathrm{x}^{[k+1]} - \mathrm{x}^{[k]})) \end{array}\right.$$

Particular cases

Condat-Vũ algorithm: [Vũ, 2013][Condat, 2013]

Problem: Find $\widehat{\mathrm{x}} \in \mathrm{Argmin}_{\mathrm{x}\in\mathbb{R}^N} f_1(\mathrm{x}) + f_2(\mathrm{x}) + g(\mathrm{Lx})$

Choose $\tau > 0$ and $\sigma > 0$ such that $\frac{1}{\tau} - \sigma\|\mathrm{L}\|^2 > \frac{\beta}{2}$ with f_2 β-Lipschitz.

For $k = 0, 1, \ldots$

$$\mathrm{x}^{[k+1]} = \mathrm{prox}_{\tau f_1}\Big(\mathrm{x}^{[k]} - \tau\big(\nabla f_2(\mathrm{x}^{[k]}) + \mathrm{L}^*\mathrm{u}^{[k]}\big)\Big)$$

$$\mathrm{u}^{[k+1]} = \mathrm{prox}_{\sigma g^*}\Big(\mathrm{u}^{[k]} + \sigma\mathrm{L}(2\mathrm{x}^{[k+1]} - \mathrm{x}^{[k]})\Big)$$

Douglas-Rachford algorithm: $f_2 \equiv 0$, $\mathrm{L} = \mathrm{Id}$ and $\tau = 1/\sigma$

Problem: Find $\widehat{\mathrm{x}} \in \mathrm{Argmin}_{\mathrm{x}\in\mathbb{R}^N} f_1(\mathrm{x}) + g(\mathrm{x})$

Choose $\sigma > 0$.

For $k = 0, 1, \ldots$

$$\mathrm{x}^{[k+1]} = \mathrm{prox}_{\sigma^{-1} f_1}(\mathbf{s}_k)$$

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \mathrm{x}^{[k+1]} - \mathrm{prox}_{\sigma^{-1} g}(2\mathrm{x}^{[k+1]} - \mathbf{s}_k)$$

Chambolle-Pock algorithm and strong convexity

CHAMBOLLE-POCK ALGORITHM: [Chambolle & Pock, 2011]

PROBLEM: Find $\widehat{\mathrm{x}} \in \operatorname{Argmin}_{\mathrm{x}\in\mathbb{R}^N} f_1(\mathrm{x}) + g(\mathrm{Lx})$

Choose $\tau > 0$ and $\sigma > 0$ such that $\sigma\tau\|\mathrm{L}\|^2 < 1$.

For $k = 0, 1, \ldots$

$$\begin{aligned}
\mathrm{x}^{[k+1]} &= \operatorname{prox}_{\tau f_1}\left(\mathrm{x}^{[k]} - \tau \mathrm{L}^* \mathrm{u}^{[k]}\right)\\
\mathrm{u}^{[k+1]} &= \operatorname{prox}_{\sigma g^*}\left(\mathrm{u}^{[k]} + \sigma \mathrm{L}(2\mathrm{x}^{[k+1]} - \mathrm{x}^{[k]})\right)
\end{aligned}$$

ACCELERATED VERSION: f_1 ρ-strongly convex [Chambolle & Pock, 2011]

Choose $\tau_0 > 0$ and $\sigma_0 > 0$ such that $\sigma_0\tau_0\|\mathrm{L}\|^2 < 1$.

For $k = 0, 1, \ldots$

$$\begin{aligned}
\mathrm{x}^{[k+1]} &= \operatorname{prox}_{\tau_k f_1}\left(\mathrm{x}^{[k]} - \tau_k \mathrm{L}^* \mathrm{u}^{[k]}\right)\\
\alpha_k &= (1 + 2\rho\tau_k)^{-1/2}\\
\tau_{k+1} &= \alpha_k \tau_k\\
\sigma_k &= \sigma_k \alpha_k^{-1/2}\\
\mathrm{y}^{[k+1]} &= \mathrm{x}^{[k+1]} + \alpha_k(\mathrm{x}^{[k+1]} - \mathrm{x}^{[k]})\\
\mathrm{u}^{[k+1]} &= \operatorname{prox}_{\sigma_{k+1} g^*}\left(\mathrm{u}^{[k]} + \sigma \mathrm{L}\mathrm{y}^{[k+1]}\right)
\end{aligned}$$

Optimization algorithms

Forward-Backward	$f_1 + f_2$	f_1 grad. Lipschitz prox_{f_2}	[Combettes,Wajs,2005]
ISTA	$f_1 + f_2$	f_1 grad. Lipschitz $f_2 = \lambda\|\cdot\|_1$	[Daubechies et al, 2003]
Douglas-Rachford	$f_1 + f_2$	prox_{f_1} prox_{f_2}	[Combettes,Pesquet, 2007]
PPXA	$\sum_i f_i$	prox_{f_i}	[Combettes,Pesquet, 2008]
PPXA+	$\sum_i g_i \circ \mathrm{L}_i$	prox_{g_i} $(\sum_{i=1}^m \mathrm{L}_i^* \mathrm{L}_i)^{-1}$	[Pesquet, Pustelnik, 2012]
ADMM	$f + g \circ \mathrm{L}$	prox_f $(\mathrm{L}^*\mathrm{L})^{-1}$	[Eckstein, Yao, 2015]
Chambolle-Pock	$f + g \circ \mathrm{L}$	prox_f prox_g	[Chambolle, Pock, 2011]
Condat-Vũ	$f_1 + f_2 + g \circ \mathrm{L}$	prox_f prox_g f_2 grad. Lipschitz	[Condat, 2013][Vũ, 2013]