

# Un petit tutoriel sur les méthodes primales-duales proximales pour l'optimisation convexe

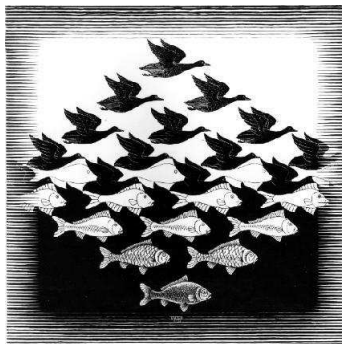
Audrey Repetti, Émilie Chouzenoux et Jean-Christophe Pesquet

Laboratoire d'Informatique Gaspard Monge - CNRS  
Univ. Paris-Est Marne-la-Vallée, France

GRETSI 2015 – Jeudi 10 Septembre



# Introduction



*“In the realm of convexity, almost every mathematical object can be paired with another, said to be dual to it.”*

*(R. T. Rockafellar & R. J.-B. Wets, Variational Analysis)*

## Un vieux problème

**OBJECTIF:** Trouver une solution au problème d'optimisation

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{minimiser}} \quad f(\mathbf{x})$$

où

- $\mathcal{H}$  est un espace hilbertien réel,
- $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  est convexe.

## Des questions d'actualité

- ▶ Que faire si  $f$  n'est pas lisse ?
- ▶ Tous les coûts  $f$  d'intérêt sont-ils fortement convexes ?
- ▶ Peut-on optimiser efficacement un million de variables ?
- ▶ Comment transformer un problème d'optimisation complexe en une suite de problèmes simples ?
- ▶ Comment résoudre un problème inverse sans inverser de matrice ?
- ▶ Les méthodes de lagrangiens augmentés sont-elles les seules solutions viables ?
- ▶ Si  $(\mathbf{x}_k)_{k \in \mathbb{N}}$  est une suite générée par un algorithme d'optimisation, quelle foi peut-on avoir en les résultats de convergence sur  $(f(\mathbf{x}_k))_{k \in \mathbb{N}}$  ?

# Outils fondamentaux de l'analyse convexe

## Notations et définitions

Soit  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$ .

- ▶ Le **domaine** de la fonction  $f$  est donné par

$$\text{dom } f = \{\mathbf{x} \in \mathcal{H} \mid f(\mathbf{x}) < +\infty\}$$

Si  $\text{dom } f \neq \emptyset$ , la fonction  $f$  est dite **propre**.

- ▶ La fonction  $f$  est dite **convexe** si

$$(\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{H}^2)(\forall \lambda \in [0, 1]) \quad f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}).$$

- ▶ La fonction  $f$  est dite **semi-continue inférieurement** (sci) sur  $\mathcal{H}$  si, pour tout  $\mathbf{x} \in \mathcal{H}$ , pour toute suite  $(\mathbf{x}_k)_{k \in \mathbb{N}}$  de  $\mathcal{H}$ ,

$$\mathbf{x}_k \rightarrow \mathbf{x} \quad \Rightarrow \quad \liminf f(\mathbf{x}_k) \geq f(\mathbf{x}).$$

## Notations et définitions

Soit  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$ .

- ▶ Le **domaine** de la fonction  $f$  est donné par

$$\text{dom } f = \{\mathbf{x} \in \mathcal{H} \mid f(\mathbf{x}) < +\infty\}$$

Si  $\text{dom } f \neq \emptyset$ , la fonction  $f$  est dite **propre**.

- ▶ La fonction  $f$  est dite **convexe** si

$$(\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{H}^2)(\forall \lambda \in [0, 1]) \quad f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}).$$

- ▶ La fonction  $f$  est dite **semi-continue inférieurement** (sci) sur  $\mathcal{H}$  si, pour tout  $\mathbf{x} \in \mathcal{H}$ , pour toute suite  $(\mathbf{x}_k)_{k \in \mathbb{N}}$  de  $\mathcal{H}$ ,

$$\mathbf{x}_k \longrightarrow \mathbf{x} \quad \Rightarrow \quad \liminf f(\mathbf{x}_k) \geq f(\mathbf{x}).$$

★ *Par la suite, toutes les fonctions considérées seront supposées convexes, semi-continue inférieurement et propres.*

## Notations et définitions

L'ensemble des opérateurs linéaires auto-adjoints bornés fortement positifs de  $\mathcal{H}$  dans  $\mathcal{H}$  sera noté  $\mathcal{S}^+(\mathcal{H})$ .

Soit  $\mathbf{U} \in \mathcal{S}^+(\mathcal{H})$ . La norme pondérée par  $\mathbf{U}$  est définie comme

$$\|\cdot\|_{\mathbf{U}} = \sqrt{\langle \cdot | \mathbf{U} \cdot \rangle},$$

avec la convention  $\|\cdot\| = \|\cdot\|_{\text{Id}}$ .



## Notations et définitions

L'ensemble des opérateurs linéaires auto-adjoints bornés fortement positifs de  $\mathcal{H}$  dans  $\mathcal{H}$  sera noté  $\mathcal{S}^+(\mathcal{H})$ .

Soit  $\mathbf{U} \in \mathcal{S}^+(\mathcal{H})$ . La norme pondérée par  $\mathbf{U}$  est définie comme

$$\|\cdot\|_{\mathbf{U}} = \sqrt{\langle \cdot | \mathbf{U} \cdot \rangle},$$

avec la convention  $\|\cdot\| = \|\cdot\|_{\text{Id}}$ .

Soit  $f: \mathcal{H} \rightarrow ]-\infty, +\infty[$ . La fonction  $f$  est dite de gradient  $\beta$ -Lipschitz si elle est différentiable sur  $\mathcal{H}$  et si son gradient vérifie

$$(\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{H}^2) \quad \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq \beta \|\mathbf{x} - \mathbf{y}\|,$$

avec  $\beta \in ]0, +\infty[$ .

## Notations et définitions

L'inf-convolution de  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  avec  $g : \mathcal{H} \rightarrow ]-\infty, +\infty]$  est

$$f \square g : \mathcal{H} \rightarrow [-\infty, +\infty]$$

$$\mathbf{x} \mapsto \inf_{\mathbf{y} \in \mathcal{H}} f(\mathbf{y}) + g(\mathbf{x} - \mathbf{y}).$$

La convolution de  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  avec  $g : \mathcal{H} \rightarrow ]-\infty, +\infty]$  est

$$f * g : \mathcal{H} \rightarrow [-\infty, +\infty]$$

$$\mathbf{x} \mapsto \int_{\mathcal{H}} f(\mathbf{y})g(\mathbf{x} - \mathbf{y})d\mathbf{y}.$$

## Notations et définitions

L'inf-convolution de  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  avec  $g : \mathcal{H} \rightarrow ]-\infty, +\infty]$  est

$$f \square g : \mathcal{H} \rightarrow ]-\infty, +\infty]$$

$$\mathbf{x} \mapsto \inf_{\mathbf{y} \in \mathcal{H}} f(\mathbf{y}) + g(\mathbf{x} - \mathbf{y}).$$

La convolution de  $f : \mathcal{H} \rightarrow ]-\infty, +\infty]$  avec  $g : \mathcal{H} \rightarrow ]-\infty, +\infty]$  est

$$f * g : \mathcal{H} \rightarrow ]-\infty, +\infty]$$

$$\mathbf{x} \mapsto \int_{\mathcal{H}} f(\mathbf{y})g(\mathbf{x} - \mathbf{y})d\mathbf{y}.$$

★ CAS PARTICULIER:  $f \square \iota_{\{0\}} = f,$

où, pour  $C \subset \mathcal{H},$

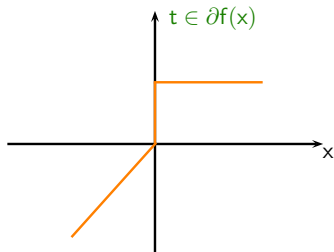
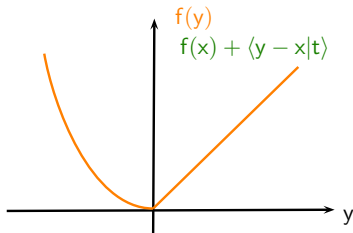
$$(\forall \mathbf{x} \in \mathcal{H}) \quad \iota_C(\mathbf{x}) = \begin{cases} 0 & \text{si } \mathbf{x} \in C, \\ +\infty & \text{sinon.} \end{cases}$$

## Sous-différentiel

Le **sous-différentiel** de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  en  $\mathbf{x}$  est l'ensemble

$$\partial f(\mathbf{x}) = \{\mathbf{t} \in \mathcal{H} \mid (\forall \mathbf{y} \in \mathcal{H}) f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{t} \mid \mathbf{y} - \mathbf{x} \rangle\}$$

Un élément  $\mathbf{t}$  de  $\partial f(\mathbf{x})$  est appelé **sous-gradient** de  $f$  en  $\mathbf{x}$ .

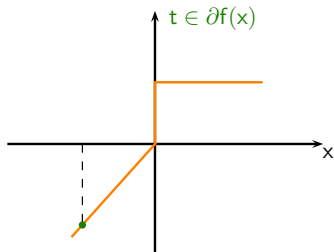
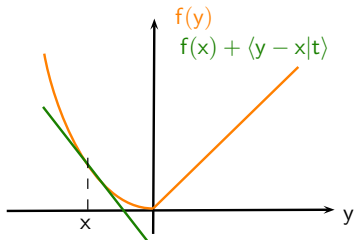


## Sous-différentiel

Le sous-différentiel de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  en  $x$  est l'ensemble

$$\partial f(x) = \{t \in \mathcal{H} \mid (\forall y \in \mathcal{H}) f(y) \geq f(x) + \langle t \mid y - x \rangle\}$$

Un élément  $t$  de  $\partial f(x)$  est appelé sous-gradient de  $f$  en  $x$ .

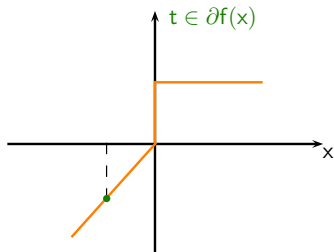
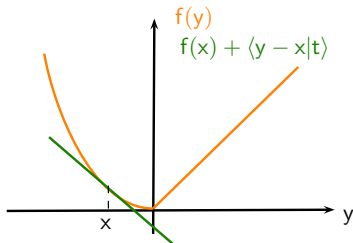


## Sous-différentiel

Le sous-différentiel de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  en  $x$  est l'ensemble

$$\partial f(x) = \{t \in \mathcal{H} \mid (\forall y \in \mathcal{H}) f(y) \geq f(x) + \langle t \mid y - x \rangle\}$$

Un élément  $t$  de  $\partial f(x)$  est appelé sous-gradient de  $f$  en  $x$ .

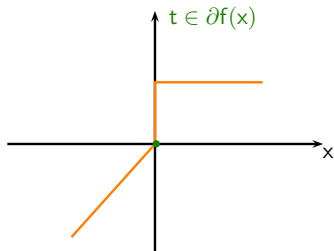
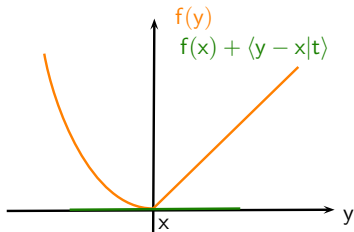


## Sous-différentiel

Le **sous-différentiel** de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  en  $\mathbf{x}$  est l'ensemble

$$\partial f(\mathbf{x}) = \{\mathbf{t} \in \mathcal{H} \mid (\forall \mathbf{y} \in \mathcal{H}) f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{t} \mid \mathbf{y} - \mathbf{x} \rangle\}$$

Un élément  $\mathbf{t}$  de  $\partial f(\mathbf{x})$  est appelé **sous-gradient** de  $f$  en  $\mathbf{x}$ .

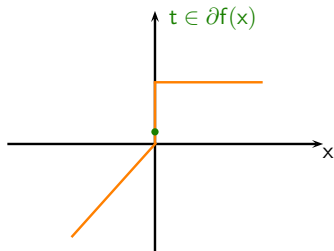
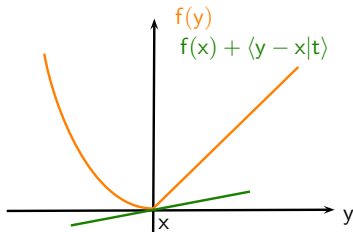


## Sous-différentiel

Le sous-différentiel de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  en  $\mathbf{x}$  est l'ensemble

$$\partial f(\mathbf{x}) = \{\mathbf{t} \in \mathcal{H} \mid (\forall \mathbf{y} \in \mathcal{H}) f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{t} \mid \mathbf{y} - \mathbf{x} \rangle\}$$

Un élément  $\mathbf{t}$  de  $\partial f(\mathbf{x})$  est appelé sous-gradient de  $f$  en  $\mathbf{x}$ .



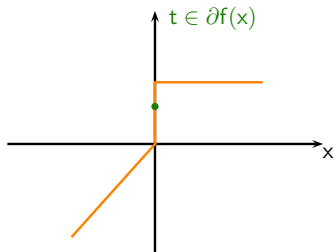
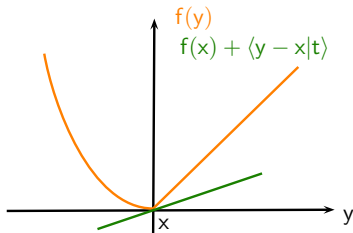


## Sous-différentiel

Le sous-différentiel de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  en  $\mathbf{x}$  est l'ensemble

$$\partial f(\mathbf{x}) = \{\mathbf{t} \in \mathcal{H} \mid (\forall \mathbf{y} \in \mathcal{H}) f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{t} \mid \mathbf{y} - \mathbf{x} \rangle\}$$

Un élément  $\mathbf{t}$  de  $\partial f(\mathbf{x})$  est appelé sous-gradient de  $f$  en  $\mathbf{x}$ .

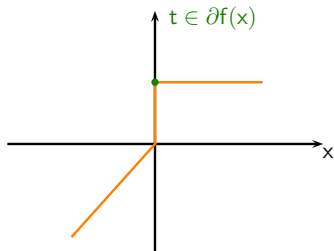
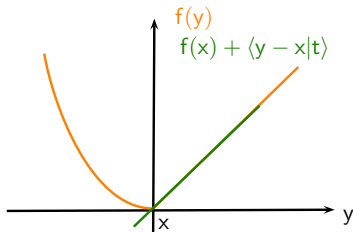


## Sous-différentiel

Le **sous-différentiel** de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  en  $\mathbf{x}$  est l'ensemble

$$\partial f(\mathbf{x}) = \{\mathbf{t} \in \mathcal{H} \mid (\forall \mathbf{y} \in \mathcal{H}) f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{t} \mid \mathbf{y} - \mathbf{x} \rangle\}$$

Un élément  $\mathbf{t}$  de  $\partial f(\mathbf{x})$  est appelé **sous-gradient** de  $f$  en  $\mathbf{x}$ .

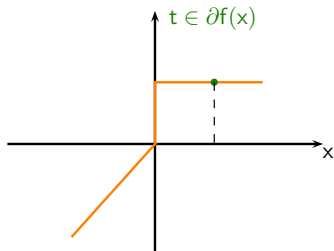
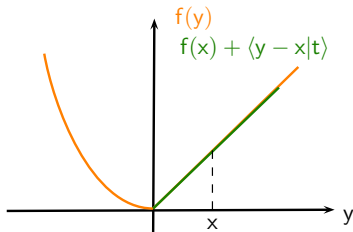


## Sous-différentiel

Le **sous-différentiel** de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  en  $\mathbf{x}$  est l'ensemble

$$\partial f(\mathbf{x}) = \{\mathbf{t} \in \mathcal{H} \mid (\forall \mathbf{y} \in \mathcal{H}) f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{t} \mid \mathbf{y} - \mathbf{x} \rangle\}$$

Un élément  $\mathbf{t}$  de  $\partial f(\mathbf{x})$  est appelé **sous-gradient** de  $f$  en  $\mathbf{x}$ .



## Sous-différentiel

### EXEMPLES:

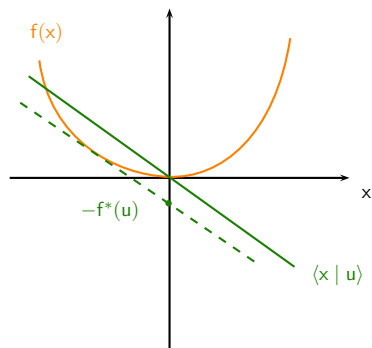
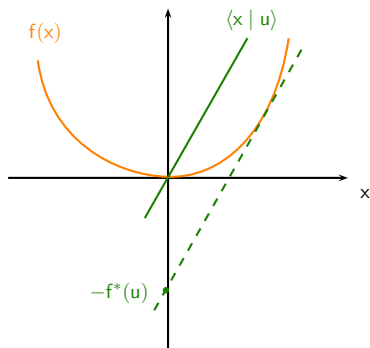
- ▶ Si  $f$  est différentiable en  $\mathbf{x} \in \mathcal{H}$  alors  $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$ .
- ▶ Si  $f = |\cdot|$  alors

$$(\forall x \in \mathbb{R}) \quad \partial f(x) = \begin{cases} \{\text{sign}(x)\} & \text{si } x \neq 0 \\ [-1, 1] & \text{si } x = 0 \end{cases}$$

## Fonction conjuguée

La **conjuguée** de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  est  $f^*: \mathcal{H} \rightarrow [-\infty, +\infty]$  telle que

$$(\forall \mathbf{u} \in \mathcal{H}) \quad f^*(\mathbf{u}) = \sup_{\mathbf{x} \in \mathcal{H}} (\langle \mathbf{x} | \mathbf{u} \rangle - f(\mathbf{x})).$$



## Fonction conjuguée

La **conjuguée** de  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$  est  $f^*: \mathcal{H} \rightarrow [-\infty, +\infty]$  telle que

$$(\forall \mathbf{u} \in \mathcal{H}) \quad f^*(\mathbf{u}) = \sup_{\mathbf{x} \in \mathcal{H}} (\langle \mathbf{x} | \mathbf{u} \rangle - f(\mathbf{x})).$$

Origine:

Alexis Claude Clairaut (1713-1765),

Adrien-Marie Legendre (1752-1833),

Werner Fenchel (1905-1988)

# Propriétés de la conjugaison (et parallèle avec Fourier)

	Conjuguées		Transformées de Fourier	
	$f(\mathbf{x})$	$f^*(\mathbf{u})$	$f(\mathbf{x})$	$f(\boldsymbol{\nu})$
fonction invariante	$\frac{1}{2}\ \mathbf{x}\ ^2$	$\frac{1}{2}\ \mathbf{u}\ ^2$	$\exp(-\pi\ \mathbf{x}\ ^2)$	$\exp(-\pi\ \boldsymbol{\nu}\ ^2)$
translation $\mathbf{z} \in \mathcal{H}$	$f(\mathbf{x} - \mathbf{z})$	$f^*(\mathbf{u}) + \langle \mathbf{u}   \mathbf{z} \rangle$	$f(\mathbf{x} - \mathbf{z})$	$\exp(-j2\pi\langle \boldsymbol{\nu}   \mathbf{z} \rangle)\widehat{f}(\boldsymbol{\nu})$
translation duale $\mathbf{z} \in \mathcal{H}$	$f(\mathbf{x}) + \langle \mathbf{x}   \mathbf{z} \rangle$	$f^*(\mathbf{u} - \mathbf{z})$	$f(\mathbf{x}) + \langle \mathbf{x}   \mathbf{z} \rangle$	$\widehat{f}(\boldsymbol{\nu} - \mathbf{z})$
multiplication scalaire $\alpha \in ]0, +\infty[$	$\alpha f(\mathbf{x})$	$\alpha f^*\left(\frac{\mathbf{u}}{\alpha}\right)$	$\alpha f(\mathbf{x})$	$\widehat{\alpha f}(\boldsymbol{\nu})$
mise à l'échelle $\alpha \in \mathbb{R}^*$	$f\left(\frac{\mathbf{x}}{\alpha}\right)$	$f^*(\alpha\mathbf{u})$	$f\left(\frac{\mathbf{x}}{\alpha}\right)$	$ \alpha \widehat{f}(\alpha\boldsymbol{\nu})$
isomorphisme $\mathbf{L} \in \mathcal{B}(\mathcal{G}, \mathcal{H})$	$f(\mathbf{L}\mathbf{x})$	$f^*(\mathbf{L}^{-\top}\mathbf{u})$	$f(\mathbf{L}\mathbf{x})$	$\frac{1}{ \det(\mathbf{L}) }\widehat{f}(\mathbf{L}^{-\top}\boldsymbol{\nu})$
réflexion	$f(-\mathbf{x})$	$f^*(-\mathbf{u})$	$f(-\mathbf{x})$	$\widehat{f}(-\boldsymbol{\nu})$
séparabilité	$\sum_{n=1}^N \varphi_n(\mathbf{x}^{(n)})$ $\mathbf{x} = (\mathbf{x}^{(n)})_{1 \leq n \leq N}$	$\sum_{n=1}^N \varphi_n^*(\mathbf{u}^{(n)})$ $\mathbf{u} = (\mathbf{u}^{(n)})_{1 \leq n \leq N}$	$\prod_{n=1}^N \varphi_n(\mathbf{x}^{(n)})$ $\mathbf{x} = (\mathbf{x}^{(n)})_{1 \leq n \leq N}$	$\prod_{n=1}^N \widehat{\varphi}_n(\boldsymbol{\nu}^{(n)})$ $\boldsymbol{\nu} = (\boldsymbol{\nu}^{(n)})_{1 \leq n \leq N}$
isotropie	$\psi(\ \mathbf{x}\ )$	$\psi^*(\ \mathbf{u}\ )$	$\psi(\ \mathbf{x}\ )$	$\psi(\ \boldsymbol{\nu}\ )$
inf-convolution	$(f \square g)(\mathbf{x})$	$f^*(\mathbf{u}) + g^*(\mathbf{u})$	$(f * g)(\mathbf{x})$	$\widehat{f}(\boldsymbol{\nu})\widehat{g}(\boldsymbol{\nu})$
somme	$f(\mathbf{x}) + g(\mathbf{x})$ $(f, g) \in (\Gamma_0(\mathcal{H}))^2$ $\text{dom } f \cap \text{dom } g \neq \emptyset$	$(f^* \square g^*)(\mathbf{u})$	$f(\mathbf{x}) + g(\mathbf{x})$	$(\widehat{f} * \widehat{g})(\boldsymbol{\nu})$
élément neutre de la convolution	$\iota_{\{0\}}(\mathbf{x})$	0	$\delta(\mathbf{x})$	1
élément neutre de l'addition	0	$\iota_{\{0\}}(\mathbf{u})$	1	$\delta(\boldsymbol{\nu})$

# Opérateur proximal

## ALGORITHME DE SOUS-GRADIENT

Soit  $\gamma \in ]0, +\infty[$ .

Explicite:  $(\forall k \in \mathbb{N}) \quad \mathbf{x}_k - \mathbf{x}_{k+1} \in \gamma \partial f(\mathbf{x}_k)$



# Opérateur proximal

## ALGORITHME DE SOUS-GRADIENT

Soit  $\gamma \in ]0, +\infty[$ .

$$\text{Implicite: } (\forall k \in \mathbb{N}) \quad \mathbf{x}_k - \mathbf{x}_{k+1} \in \gamma \partial f(\mathbf{x}_{k+1})$$

# Opérateur proximal

## ALGORITHME DE SOUS-GRADIENT

Soit  $\mathbf{U} \in \mathcal{S}^+(\mathcal{H})$ .

Implicite préconditionné:  $(\forall k \in \mathbb{N}) \quad \mathbf{x}_k - \mathbf{x}_{k+1} \in \mathbf{U}^{-1} \partial f(\mathbf{x}_{k+1})$

# Opérateur proximal

## ALGORITHME DE SOUS-GRADIENT

Soit  $\mathbf{U} \in \mathcal{S}^+(\mathcal{H})$ .

Implicite préconditionné:  $(\forall k \in \mathbb{N}) \quad \mathbf{x}_k - \mathbf{x}_{k+1} \in \mathbf{U}^{-1} \partial f(\mathbf{x}_{k+1})$

## CARACTÉRISATION DE L'OPÉRATEUR PROXIMAL

$(\forall \mathbf{x} \in \mathcal{H}) \quad \hat{\mathbf{y}} = \text{prox}_{\mathbf{U}, f}(\mathbf{x}) \Leftrightarrow \mathbf{x} - \hat{\mathbf{y}} \in \mathbf{U}^{-1} \partial f(\hat{\mathbf{y}})$ .

## Opérateur proximal

### CARACTÉRISATION DE L'OPÉRATEUR PROXIMAL

$$(\forall \mathbf{x} \in \mathcal{H}) \quad \hat{\mathbf{y}} = \text{prox}_{\mathbf{U},f}(\mathbf{x}) \Leftrightarrow \mathbf{x} - \hat{\mathbf{y}} \in \mathbf{U}^{-1} \partial f(\hat{\mathbf{y}}).$$

L'opérateur proximal  $\text{prox}_{\mathbf{U},f}(\mathbf{x})$  de  $f$  en  $\mathbf{x} \in \mathcal{H}$  relatif à la métrique induite par  $\mathbf{U}$  est l'unique vecteur  $\hat{\mathbf{y}} \in \mathcal{H}$  tel que

$$f(\hat{\mathbf{y}}) + \frac{1}{2} \|\hat{\mathbf{y}} - \mathbf{x}\|_{\mathbf{U}}^2 = \inf_{\mathbf{y} \in \mathcal{H}} f(\mathbf{y}) + \frac{1}{2} \langle \mathbf{y} - \mathbf{x} \mid \mathbf{U}(\mathbf{y} - \mathbf{x}) \rangle.$$

# Propriétés de l'opérateur proximal

	$f(\mathbf{x})$	$\text{prox}_f(\mathbf{x}) = \text{prox}_{\text{Id}, f}(\mathbf{x})$
translation $\mathbf{z} \in \mathcal{H}$	$f(\mathbf{x} - \mathbf{z})$	$\mathbf{z} + \text{prox}_f(\mathbf{x} - \mathbf{z})$
perturbation quadratique $\mathbf{z} \in \mathcal{H}, \alpha > 0, \gamma \in \mathbb{R}$	$f(\mathbf{x}) + \alpha \ \mathbf{x}\ ^2/2 + \langle \mathbf{x}   \mathbf{z} \rangle + \gamma$	$\text{prox}_{\frac{f}{\alpha+1}}\left(\frac{\mathbf{x}-\mathbf{z}}{\alpha+1}\right)$
mise à l'échelle $\rho \in \mathbb{R}^*$	$f(\rho\mathbf{x})$	$\frac{1}{\rho} \text{prox}_{\rho^2 f}(\rho\mathbf{x})$
fonction quadratique $\mathbf{L} \in \mathcal{B}(\mathcal{G}, \mathcal{H}), \gamma > 0, \mathbf{z} \in \mathcal{G}$	$\gamma \ \mathbf{L}\mathbf{x} - \mathbf{z}\ ^2/2$	$(\text{Id} + \gamma \mathbf{L}\mathbf{L}^*)^{-1}(\mathbf{x} - \gamma \mathbf{L}^* \mathbf{z})$
isomorphisme $\mathbf{L} \in \mathcal{B}(\mathcal{G}, \mathcal{H}), \mathbf{L}\mathbf{L}^* = \mu \text{Id}, \mu > 0$	$f(\mathbf{L}\mathbf{x})$	$\mathbf{x} - \mu^{-1} \mathbf{L}^*(\mathbf{x} - \text{prox}_{\mu f}(\mathbf{L}\mathbf{x}))$
réflexion	$f(-\mathbf{x})$	$-\text{prox}_f(-\mathbf{x})$
séparabilité	$\sum_{n=1}^N \varphi_n(\mathbf{x}^{(n)})$ $\mathbf{x} = (\mathbf{x}^{(n)})_{1 \leq n \leq N}$	$(\text{prox}_{\varphi_n}(\mathbf{x}^{(n)}))_{1 \leq n \leq N}$
fonction indicatrice	$\iota_C(\mathbf{x})$	$P_C(\mathbf{x})$
fonction support	$\iota_C^*(\mathbf{x}) = \sigma_C(\mathbf{x})$	$\mathbf{x} - P_C(\mathbf{x})$

# Algorithmes proximaux primaux-duaux

# Problème primal-dual

## PROBLÈME PRIMAL

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{minimiser}} \quad f(\mathbf{x}) + g(\mathbf{L}\mathbf{x}) + h(\mathbf{x}).$$

- ▶  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$ ,  $g: \mathcal{G} \rightarrow ]-\infty, +\infty]$ ,
- ▶  $h: \mathcal{H} \rightarrow \mathbb{R}$   $\beta$ -Lipschitz différentiable avec  $\beta \in ]0, +\infty[$ ,
- ▶  $\mathcal{H}$  et  $\mathcal{G}$  espaces de Hilbert réels,
- ▶  $\mathbf{L}: \mathcal{H} \rightarrow \mathcal{G}$  linéaire et borné.

# Problème primal-dual

## PROBLÈME PRIMAL

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{minimiser}} f(\mathbf{x}) + g(\mathbf{L}\mathbf{x}) + h(\mathbf{x}).$$

- ▶  $f: \mathcal{H} \rightarrow ]-\infty, +\infty]$ ,  $g: \mathcal{G} \rightarrow ]-\infty, +\infty]$ ,
- ▶  $h: \mathcal{H} \rightarrow \mathbb{R}$   $\beta$ -Lipschitz différentiable avec  $\beta \in ]0, +\infty[$ ,
- ▶  $\mathcal{H}$  et  $\mathcal{G}$  espaces de Hilbert réels,
- ▶  $\mathbf{L}: \mathcal{H} \rightarrow \mathcal{G}$  linéaire et borné.

## PROBLÈME DUAL

$$\underset{\mathbf{v} \in \mathcal{G}}{\text{minimiser}} (f^* \square h^*)(-\mathbf{L}^*\mathbf{v}) + g^*(\mathbf{v}).$$



## Problème primal-dual

### FORMULATION LAGRANGIENNE

Le problème primal-dual est équivalent à la recherche de point selle du Lagrangien :

$$(\forall (\mathbf{x}, \mathbf{y}, \mathbf{v}) \in \mathcal{H} \times \mathcal{H} \times \mathcal{G}) \quad \mathcal{L}_{\mathbf{v}}(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + h(\mathbf{x}) + g(\mathbf{y}) + \langle \mathbf{v} \mid \mathbf{L}\mathbf{x} - \mathbf{y} \rangle.$$

# Problème primal-dual

## FORMULATION LAGRANGIENNE

Le problème primal-dual est équivalent à la recherche de point selle du Lagrangien :

$$(\forall (\mathbf{x}, \mathbf{y}, \mathbf{v}) \in \mathcal{H} \times \mathcal{H} \times \mathcal{G}) \quad \mathcal{L}_{\mathbf{v}}(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + h(\mathbf{x}) + g(\mathbf{y}) + \langle \mathbf{v} \mid \mathbf{L}\mathbf{x} - \mathbf{y} \rangle.$$

## CONDITIONS KKT

Si un couple  $(\hat{\mathbf{x}}, \hat{\mathbf{v}})$  vérifie les conditions de Karush-Kuhn-Tucker :

$$-\mathbf{L}^*\hat{\mathbf{v}} - \nabla h(\hat{\mathbf{x}}) \in \partial f(\hat{\mathbf{x}}) \quad \text{et} \quad \mathbf{L}\hat{\mathbf{x}} \in \partial g^*(\hat{\mathbf{v}}),$$

alors  $\left\{ \begin{array}{l} \hat{\mathbf{x}} \text{ est solution du problème primal} \\ \hat{\mathbf{v}} \text{ est solution du problème dual.} \end{array} \right.$

## Algorithme proximal primal-dual

pour  $k = 0, 1, \dots$

$$\left[ \begin{array}{l} \mathbf{y}_k = \text{prox}_{\tau f} \left( \mathbf{x}_k - \tau (\nabla h(\mathbf{x}_k) + \mathbf{c}_k + \mathbf{L}^* \mathbf{v}_k) \right) + \mathbf{a}_k \\ \mathbf{u}_k = \text{prox}_{\sigma g^*} \left( \mathbf{v}_k + \sigma \mathbf{L} (2\mathbf{y}_k - \mathbf{x}_k) \right) + \mathbf{b}_k \\ (\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) = (\mathbf{x}_k, \mathbf{v}_k) + \lambda_k ((\mathbf{y}_k, \mathbf{u}_k) - (\mathbf{x}_k, \mathbf{v}_k)) \end{array} \right.$$

## Algorithme proximal primal-dual

pour  $k = 0, 1, \dots$

$$\begin{cases} \mathbf{y}_k = \text{prox}_{\tau f}(\mathbf{x}_k - \tau(\nabla h(\mathbf{x}_k) + \mathbf{c}_k + \mathbf{L}^* \mathbf{v}_k)) + \mathbf{a}_k \\ \mathbf{u}_k = \text{prox}_{\sigma g^*}(\mathbf{v}_k + \sigma \mathbf{L}(2\mathbf{y}_k - \mathbf{x}_k)) + \mathbf{b}_k \\ (\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) = (\mathbf{x}_k, \mathbf{v}_k) + \lambda_k((\mathbf{y}_k, \mathbf{u}_k) - (\mathbf{x}_k, \mathbf{v}_k)) \end{cases}$$

- ▶  $(\tau, \sigma) \in ]0, +\infty[^2$  tels que  $\tau^{-1} - \sigma \|\mathbf{L}\|^2 > \beta/2$ ,

## Algorithme proximal primal-dual

pour  $k = 0, 1, \dots$

$$\left[ \begin{array}{l} \mathbf{y}_k = \text{prox}_{\tau f} \left( \mathbf{x}_k - \tau (\nabla h(\mathbf{x}_k) + \mathbf{c}_k + \mathbf{L}^* \mathbf{v}_k) \right) + \mathbf{a}_k \\ \mathbf{u}_k = \text{prox}_{\sigma g^*} \left( \mathbf{v}_k + \sigma \mathbf{L} (2\mathbf{y}_k - \mathbf{x}_k) \right) + \mathbf{b}_k \\ (\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) = (\mathbf{x}_k, \mathbf{v}_k) + \lambda_k ((\mathbf{y}_k, \mathbf{u}_k) - (\mathbf{x}_k, \mathbf{v}_k)) \end{array} \right.$$

- ▶  $(\tau, \sigma) \in ]0, +\infty[^2$  tels que  $\tau^{-1} - \sigma \|\mathbf{L}\|^2 > \beta/2$ ,
- ▶ pour tout  $k \in \mathbb{N}$ ,  $\lambda_k \in ]0, \bar{\lambda}[$ , où  $\bar{\lambda} = 2 - \beta(\tau^{-1} - \sigma \|\mathbf{L}\|^2)^{-1}/2 \in [1, 2[$ ,

## Algorithme proximal primal-dual

pour  $k = 0, 1, \dots$

$$\begin{cases} \mathbf{y}_k = \text{prox}_{\tau f} \left( \mathbf{x}_k - \tau (\nabla h(\mathbf{x}_k) + \mathbf{c}_k + \mathbf{L}^* \mathbf{v}_k) \right) + \mathbf{a}_k \\ \mathbf{u}_k = \text{prox}_{\sigma g^*} \left( \mathbf{v}_k + \sigma \mathbf{L} (2\mathbf{y}_k - \mathbf{x}_k) \right) + \mathbf{b}_k \\ (\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) = (\mathbf{x}_k, \mathbf{v}_k) + \lambda_k ((\mathbf{y}_k, \mathbf{u}_k) - (\mathbf{x}_k, \mathbf{v}_k)) \end{cases}$$

- ▶  $(\tau, \sigma) \in ]0, +\infty[^2$  tels que  $\tau^{-1} - \sigma \|\mathbf{L}\|^2 > \beta/2$ ,
- ▶ pour tout  $k \in \mathbb{N}$ ,  $\lambda_k \in ]0, \bar{\lambda}[$ , où  $\bar{\lambda} = 2 - \beta(\tau^{-1} - \sigma \|\mathbf{L}\|^2)^{-1}/2 \in [1, 2[$ ,
- ▶  $\sum_{k \in \mathbb{N}} \|\mathbf{a}_k\| < +\infty$ ,  $\sum_{k \in \mathbb{N}} \|\mathbf{b}_k\| < +\infty$ , et  $\sum_{k \in \mathbb{N}} \|\mathbf{c}_k\| < +\infty$ ,

## Algorithme proximal primal-dual

pour  $k = 0, 1, \dots$

$$\begin{cases} \mathbf{y}_k = \text{prox}_{\tau f} \left( \mathbf{x}_k - \tau (\nabla h(\mathbf{x}_k) + \mathbf{c}_k + \mathbf{L}^* \mathbf{v}_k) \right) + \mathbf{a}_k \\ \mathbf{u}_k = \text{prox}_{\sigma g^*} \left( \mathbf{v}_k + \sigma \mathbf{L} (2\mathbf{y}_k - \mathbf{x}_k) \right) + \mathbf{b}_k \\ (\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) = (\mathbf{x}_k, \mathbf{v}_k) + \lambda_k ((\mathbf{y}_k, \mathbf{u}_k) - (\mathbf{x}_k, \mathbf{v}_k)) \end{cases}$$

- ▶  $(\tau, \sigma) \in ]0, +\infty[^2$  tels que  $\tau^{-1} - \sigma \|\mathbf{L}\|^2 > \beta/2$ ,
- ▶ pour tout  $k \in \mathbb{N}$ ,  $\lambda_k \in ]0, \bar{\lambda}[$ , où  $\bar{\lambda} = 2 - \beta(\tau^{-1} - \sigma \|\mathbf{L}\|^2)^{-1}/2 \in [1, 2[$ ,
- ▶  $\sum_{k \in \mathbb{N}} \|\mathbf{a}_k\| < +\infty$ ,  $\sum_{k \in \mathbb{N}} \|\mathbf{b}_k\| < +\infty$ , et  $\sum_{k \in \mathbb{N}} \|\mathbf{c}_k\| < +\infty$ ,
- ▶ il existe  $\bar{\mathbf{x}} \in \mathcal{H}$  tel que  $0 \in \partial f(\bar{\mathbf{x}}) + \nabla h(\bar{\mathbf{x}}) + \mathbf{L}^* \partial g(\mathbf{L}\bar{\mathbf{x}})$ ,

Alors:

- ★  $\mathbf{x}_k \rightarrow \hat{\mathbf{x}}$  où  $\hat{\mathbf{x}}$  est une solution du problème primal,
- ★  $\mathbf{v}_k \rightarrow \hat{\mathbf{v}}$  où  $\hat{\mathbf{v}}$  est une solution du problème dual.

## Algorithme proximal primal-dual préconditionné

pour  $k = 0, 1, \dots$

$$\begin{cases} \mathbf{y}_k = \text{prox}_{\mathbf{W}^{-1}, f}(\mathbf{x}_k - \mathbf{W}(\nabla h(\mathbf{x}_k) + \mathbf{c}_k + \mathbf{L}^* \mathbf{v}_k)) + \mathbf{a}_k \\ \mathbf{u}_k = \text{prox}_{\mathbf{U}^{-1}, g^*}(\mathbf{v}_k + \mathbf{U}\mathbf{L}(2\mathbf{y}_k - \mathbf{x}_k)) + \mathbf{b}_k \\ (\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) = (\mathbf{x}_k, \mathbf{v}_k) + \lambda_k((\mathbf{y}_k, \mathbf{u}_k) - (\mathbf{x}_k, \mathbf{v}_k)) \end{cases}$$

- ▶  $\mathbf{W} \in \mathcal{S}^+(\mathcal{H})$  et  $\mathbf{U} \in \mathcal{S}^+(\mathcal{G})$  tels que  $1 - \|\mathbf{U}^{1/2} \mathbf{L} \mathbf{W}^{1/2}\| > \frac{\beta}{2} \|\mathbf{W}\|$ ,
- ▶ pour tout  $k \in \mathbb{N}$ ,  $\lambda_k \in ]0, 1]$ , et  $\inf_{k \in \mathbb{N}} \lambda_k > 0$ ,
- ▶  $\sum_{k \in \mathbb{N}} \|\mathbf{a}_k\| < +\infty$ ,  $\sum_{k \in \mathbb{N}} \|\mathbf{b}_k\| < +\infty$ , et  $\sum_{k \in \mathbb{N}} \|\mathbf{c}_k\| < +\infty$ ,
- ▶ il existe  $\bar{\mathbf{x}} \in \mathcal{H}$  tel que  $0 \in \partial f(\bar{\mathbf{x}}) + \nabla h(\bar{\mathbf{x}}) + \mathbf{L}^* \partial g(\mathbf{L}\bar{\mathbf{x}})$ .

Alors:

- ★  $\mathbf{x}_k \rightarrow \hat{\mathbf{x}}$  où  $\hat{\mathbf{x}}$  est une solution du problème primal,
- ★  $\mathbf{v}_k \rightarrow \hat{\mathbf{v}}$  où  $\hat{\mathbf{v}}$  est une solution du problème dual.



## Variantes de l'algorithme proximal primal-dual

### FORME SYMMÉTRIQUE

pour  $k = 0, 1, \dots$

$$\begin{cases} \mathbf{u}_k = \text{prox}_{\mathbf{U}^{-1}, g^*}(\mathbf{v}_k + \mathbf{U}\mathbf{L}\mathbf{x}_k) + \mathbf{b}_k \\ \mathbf{y}_k = \text{prox}_{\mathbf{W}^{-1}, f}(\mathbf{x}_k - \mathbf{W}(\nabla h(\mathbf{x}_k) + \mathbf{c}_k + \mathbf{L}^*(2\mathbf{u}_k - \mathbf{v}_k))) + \mathbf{a}_k \\ (\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) = (\mathbf{x}_k, \mathbf{v}_k) + \lambda_k((\mathbf{y}_k, \mathbf{u}_k) - (\mathbf{x}_k, \mathbf{v}_k)) \end{cases}$$

## Variantes de l'algorithme proximal primal-dual

### FORME SYMMÉTRIQUE

pour  $k = 0, 1, \dots$

$$\left[ \begin{array}{l} \mathbf{u}_k = \text{prox}_{\mathbf{U}^{-1}, g^*}(\mathbf{v}_k + \mathbf{UL}\mathbf{x}_k) + \mathbf{b}_k \\ \mathbf{y}_k = \text{prox}_{\mathbf{W}^{-1}, f}(\mathbf{x}_k - \mathbf{W}(\nabla h(\mathbf{x}_k) + \mathbf{c}_k + \mathbf{L}^*(2\mathbf{u}_k - \mathbf{v}_k))) + \mathbf{a}_k \\ (\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) = (\mathbf{x}_k, \mathbf{v}_k) + \lambda_k((\mathbf{y}_k, \mathbf{u}_k) - (\mathbf{x}_k, \mathbf{v}_k)) \end{array} \right.$$

### FORME SIMPLIFIÉE ( $h \equiv 0$ )

pour  $k = 0, 1, \dots$

$$\left[ \begin{array}{l} \mathbf{y}_k = \text{prox}_{\mathbf{W}^{-1}, f}(\mathbf{x}_k - \mathbf{WL}^*\mathbf{v}_k) + \mathbf{a}_k \\ \mathbf{u}_k = \text{prox}_{\mathbf{U}^{-1}, g^*}(\mathbf{v}_k + \mathbf{UL}(2\mathbf{y}_k - \mathbf{x}_k)) + \mathbf{b}_k \\ (\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) = (\mathbf{x}_k, \mathbf{v}_k) + \lambda_k((\mathbf{y}_k, \mathbf{u}_k) - (\mathbf{x}_k, \mathbf{v}_k)) \end{array} \right.$$

# Algorithme proximal primal-dual parallèle

PROBLÈME PRIMAL

$$\underset{x \in \mathcal{H}}{\text{minimiser}} \quad f(x) + g(Lx) + h(x).$$

# Algorithme proximal primal-dual parallèle

## PROBLÈME PRIMAL

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{minimiser}} \quad f(\mathbf{x}) + \sum_{n=1}^q g_n(\mathbf{L}_n \mathbf{x}) + h(\mathbf{x}).$$

avec, pour tout  $n \in \{1, \dots, q\}$ ,  $g_n: \mathcal{G}_n \rightarrow ]-\infty, +\infty]$ ,  $(\mathcal{G}_n)_{1 \leq n \leq q}$   
espaces de Hilbert réels

tels que  $\mathcal{G} = \mathcal{G}_1 \oplus \dots \oplus \mathcal{G}_q$ , et  $\mathbf{L}_n: \mathcal{H} \rightarrow \mathcal{G}_n$  opérateurs linéaires bornés.

# Algorithme proximal primal-dual parallèle

## PROBLÈME PRIMAL

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{minimiser}} \quad f(\mathbf{x}) + \sum_{n=1}^q g_n(\mathbf{L}_n \mathbf{x}) + h(\mathbf{x}).$$

avec, pour tout  $n \in \{1, \dots, q\}$ ,  $g_n: \mathcal{G}_n \rightarrow ]-\infty, +\infty]$ ,  $(\mathcal{G}_n)_{1 \leq n \leq q}$   
espaces de Hilbert réels

tels que  $\mathcal{G} = \mathcal{G}_1 \oplus \dots \oplus \mathcal{G}_q$ , et  $\mathbf{L}_n: \mathcal{H} \rightarrow \mathcal{G}_n$  opérateurs linéaires bornés.

## PROBLÈME DUAL

$$\underset{(v^{(1)}, \dots, v^{(q)}) \in \mathcal{G}}{\text{minimiser}} \quad (f^* \square h^*)\left(-\sum_{n=1}^q \mathbf{L}_n^* v^{(n)}\right) + \sum_{n=1}^q g_n^*(v^{(n)}).$$

# Algorithme proximal primal-dual parallèle

## PROBLÈME PRIMAL

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{minimiser}} \quad f(\mathbf{x}) + \sum_{n=1}^q g_n(\mathbf{L}_n \mathbf{x}) + h(\mathbf{x}).$$

avec, pour tout  $n \in \{1, \dots, q\}$ ,  $g_n: \mathcal{G}_n \rightarrow ]-\infty, +\infty]$ ,  $(\mathcal{G}_n)_{1 \leq n \leq q}$   
espaces de Hilbert réels

tels que  $\mathcal{G} = \mathcal{G}_1 \oplus \dots \oplus \mathcal{G}_q$ , et  $\mathbf{L}_n: \mathcal{H} \rightarrow \mathcal{G}_n$  opérateurs linéaires bornés.

## PROBLÈME DUAL

$$\underset{(v^{(1)}, \dots, v^{(q)}) \in \mathcal{G}}{\text{minimiser}} \quad (f^* \square h^*)\left(-\sum_{n=1}^q \mathbf{L}_n^* v^{(n)}\right) + \sum_{n=1}^q g_n^*(v^{(n)}).$$

La **séparabilité** de  $g^*$  implique que  $\text{prox}_{g^*}$  peut être obtenu en calculant  
en **parallèle** les opérateurs proximaux des fonctions  $(g_n^*)_{1 \leq n \leq q}$

⇒ SCHÉMAS PRIMAUX-DUAUX PARALLÈLES.

## Algorithme proximal primal-dual parallèle

pour  $k = 0, 1, \dots$

$$\mathbf{y}_k = \text{prox}_{\mathbf{W}^{-1}, f} \left( \mathbf{x}_k - \mathbf{W}(\nabla h(\mathbf{x}_k) + \mathbf{c}_k + \sum_{n=1}^q \mathbf{L}_n^* \mathbf{v}_k^{(n)}) \right) + \mathbf{a}_k$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k (\mathbf{y}_k - \mathbf{x}_k)$$

pour  $n = 1, \dots, q$

$$\begin{cases} \mathbf{u}_k^{(n)} = \text{prox}_{\mathbf{U}_n^{-1}, g_n^*} (\mathbf{v}_k^{(n)} + \mathbf{U}_n \mathbf{L}_n (2\mathbf{y}_k - \mathbf{x}_k)) + \mathbf{b}_k^{(n)} \\ \mathbf{v}_{k+1}^{(n)} = \mathbf{v}_k^{(n)} + \lambda_k (\mathbf{u}_k^{(n)} - \mathbf{v}_k^{(n)}) \end{cases}$$

## Remarques bibliographiques

- ★ Travail précurseur : Méthode de Arrow-Hurwicz-Uzawa pour la recherche de point selle [Arrow, Hurwicz and Uzawa - 1958] [Nedić and Ozdaglar - 2009]
- ★ Méthodes basées sur un schéma explicite-implicite :
  - type I : [Vu - 2013][Condat - 2013] (extensions de [Esser *et al.* - 2010][Chambolle and Pock - 2011])
  - type II: [Combettes *et al.* - 2014] (extensions de [Loris and Verhoeven - 2011][Chen *et al.* - 2014])
- ★ Méthodes basées sur un schéma explicite-implicite-explicite [Combettes and Pesquet - 2012] [Boţ and Hendrich, 2014]
- ★ Méthodes basées sur des projections [Alotaibi *et al.* - 2013]
- ★ ...



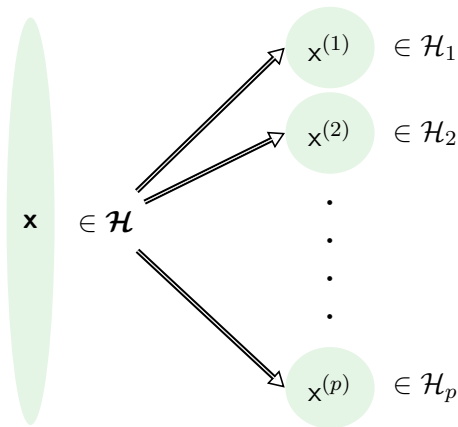
# Schémas primaux-duaux alternés par blocs

## Schémas primaux-duaux alternés par blocs

► **Idée:** éclatement des variables.

## Schémas primaux-duaux alternés par blocs

► **Idée: éclatement des variables.**



$$\mathcal{H} = \times_{j=1}^p \mathcal{H}_j$$

$\mathcal{H}_1, \dots, \mathcal{H}_p$  sont des espace hilbertiens réels séparables

## Schémas primaux-duaux alternés par blocs

► Hypothèse simplificatrice:  $f$  et  $h$  fonctions séparables par blocs.

$$f(x) = f(x^{(1)}, x^{(2)}, \dots, x^{(p)}) = \sum_{j=1}^p f_j(x^{(j)})$$

## Schémas primaux-duaux alternés par blocs

► Hypothèse simplificatrice:  $f$  et  $h$  fonctions **séparables par blocs**.

$$h(x) = h \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(p)} \end{pmatrix} = \sum_{j=1}^p h_j(x^{(j)})$$

$(\forall j \in \{1, \dots, p\})$   $h_j$   $\beta_j$ -Lipschitz différentiable avec  $\beta_j \in ]0, +\infty[$ .

## Schémas primaux-duaux alternés par blocs

★ A chaque itération  $k \in \mathbb{N}$ , mise à jour d'un sous-ensemble d'éléments ( $\sim$  approche Gauss-Seidel).

### AVANTAGES:

- ✓ Coût calculatoire par itération réduit,
- ✓ Occupation mémoire limitée,
- ✓ Plus grande flexibilité.

# Problème primal-dual

## PROBLÈME PRIMAL

Soit  $F$  l'ensemble de solutions du problème

$$\underset{x^{(1)} \in \mathcal{H}_1, \dots, x^{(p)} \in \mathcal{H}_p}{\text{minimiser}} \quad \sum_{j=1}^p \left( f_j(x^{(j)}) + h_j(x^{(j)}) \right) + \sum_{n=1}^q g_n \left( \sum_{j=1}^p L_{n,j} x^{(j)} \right)$$

$(\forall j \in \{1, \dots, p\})(\forall n \in \{1, \dots, q\})$

- ▶  $\mathcal{H}_j$  et  $\mathcal{G}_n$  espaces de Hilbert réels séparables,
- ▶  $f_j: \mathcal{H}_j \rightarrow ]-\infty, +\infty]$
- ▶  $h_j: \mathcal{H}_j \rightarrow \mathbb{R}$   $\beta_j$ -Lipschitz différentiable, avec  $\beta_j \in ]0, +\infty[$
- ▶  $g_n: \mathcal{G}_n \rightarrow ]-\infty, +\infty]$
- ▶  $L_{n,j}: \mathcal{H}_j \rightarrow \mathcal{G}_n$  linéaire et borné,
- ▶  $\mathbb{L}_n = \{j \in \{1, \dots, p\} \mid L_{n,j} \neq 0\} \neq \emptyset$ , et  $\mathbb{L}_j^* = \{n \in \{1, \dots, q\} \mid L_{n,j} \neq 0\} \neq \emptyset$ .





## Algorithme primal-dual alterné par blocs

Pour  $k = 0, 1, \dots$

**pour**  $j = 1, \dots, p$

$$y_k^{(j)} = \varepsilon_k^{(j)} \left( \text{prox}_{W_j^{-1}, f_j} \left( x_k^{(j)} - W_j (\nabla h_j(x_k^{(j)})) + c_k^{(j)} + \sum_{n=1}^q L_{n,j}^* v_k^{(n)} \right) + a_k^{(j)} \right)$$

$$x_{k+1}^{(j)} = x_k^{(j)} + \lambda_k \varepsilon_k^{(j)} (y_k^{(j)} - x_k^{(j)})$$

**pour**  $n = 1, \dots, q$

$$u_k^{(n)} = \varepsilon_k^{(p+n)} \left( \text{prox}_{U_n^{-1}, g_n^*} \left( v_k^{(n)} + U_n \sum_{j=1}^p L_{n,j} (2y_k^{(j)} - x_k^{(j)}) \right) + b_k^{(n)} \right)$$

$$v_{k+1}^{(n)} = v_k^{(n)} + \lambda_k \varepsilon_k^{(p+n)} (u_k^{(n)} - v_k^{(n)}),$$

$(\varepsilon_k)_{k \in \mathbb{N}}$  est un vecteur de  $p + q$  variables booléennes choisies aléatoirement à chaque itération  $k \in \mathbb{N}$  afin d'indiquer les blocs de variables **primales** et **duales** à mettre à jour.

## Algorithme primal-dual alterné par blocs

Supposons que

- ▶ Les variables  $(\varepsilon_k)_{k \in \mathbb{N}}$  sont **indépendantes, identiquement distribuées**, telles que  $(\forall j \in \{1, \dots, p\}) \mathbb{P}[\varepsilon_0^{(j)} = 1] > 0$ , et pour tout  $k \in \mathbb{N}$ ,  
 $(\forall n \in \{1, \dots, q\}) \quad \varepsilon_k^{(p+n)} = \max_{1 \leq j \leq p} \{\varepsilon_k^{(j)} \mid n \in \mathbb{L}_j^*\},$

## Algorithme primal-dual alterné par blocs

Supposons que

- ▶ Les variables  $(\varepsilon_k)_{k \in \mathbb{N}}$  sont **indépendantes, identiquement distribuées**, telles que  $(\forall j \in \{1, \dots, p\}) P[\varepsilon_0^{(j)} = 1] > 0$ , et pour tout  $k \in \mathbb{N}$ ,  
 $(\forall n \in \{1, \dots, q\}) \quad \varepsilon_k^{(p+n)} = \max_{1 \leq j \leq p} \{\varepsilon_k^{(j)} \mid n \in \mathbb{L}_j^*\},$
- ▶  $(\forall j \in \{1, \dots, p\}) W_j \in \mathcal{S}^+(\mathcal{H}_j)$ ,  $(\forall n \in \{1, \dots, q\}) U_n \in \mathcal{S}^+(\mathcal{G}_n)$ , avec  
 $1 - \left( \sum_{j=1}^p \sum_{n=1}^q \|U_n^{1/2} L_{n,j} W_j^{1/2}\|^2 \right)^{1/2} > \frac{1}{2} \max\{(\|W_j\| \beta_j)_{1 \leq j \leq p}\},$
- ▶  $(\forall k \in \mathbb{N}) \lambda_k \in ]0, 1]$  tel que  $\inf_{k \in \mathbb{N}} \lambda_k > 0$ ,
- ▶  $\sum_{k \in \mathbb{N}} \sqrt{\mathbb{E}\|a_k\|^2} < +\infty$ ,  $\sum_{k \in \mathbb{N}} \sqrt{\mathbb{E}\|b_k\|^2} < +\infty$  et  $\sum_{k \in \mathbb{N}} \sqrt{\mathbb{E}\|c_k\|^2} < +\infty$  p.s.

- ▶  $(x_k)_{k \in \mathbb{N}}$  converge faiblement p.s. vers une variable aléatoire à valeurs dans  $\mathbf{F}$ .
- ▶  $(v_k)_{k \in \mathbb{N}}$  converge faiblement p.s. vers une variable aléatoire à valeurs dans  $\mathbf{F}^*$ .

**Preuve:** Basée sur les propriétés des suites quasi-Fejér stochastiques  
 [Combettes & Pesquet – 2015].

# Illustration de la stratégie d'activation des blocs

Sélection des blocs ( $\forall k \in \mathbb{N}$ )

$x_k^{(1)}$  activé quand  $\varepsilon_k^{(1)} = 1$

$x_k^{(2)}$  activé quand  $\varepsilon_k^{(2)} = 1$

$x_k^{(3)}$  activé quand  $\varepsilon_k^{(3)} = 1$

$x_k^{(4)}$  activé quand  $\varepsilon_k^{(4)} = 1$

$x_k^{(5)}$  activé quand  $\varepsilon_k^{(5)} = 1$

$x_k^{(6)}$  activé quand  $\varepsilon_k^{(6)} = 1$

Comment choisir ( $\forall k \in \mathbb{N}$ ) la variable  $\varepsilon_k = (\varepsilon_k^{(1)}, \dots, \varepsilon_k^{(6)})$ ?

# Illustration de la stratégie d'activation des blocs

Sélection des blocs ( $\forall k \in \mathbb{N}$ )

$x_k^{(1)}$  activé quand  $\varepsilon_k^{(1)} = 1$

$x_k^{(2)}$  activé quand  $\varepsilon_k^{(2)} = 1$

$x_k^{(3)}$  activé quand  $\varepsilon_k^{(3)} = 1$

$x_k^{(4)}$  activé quand  $\varepsilon_k^{(4)} = 1$

$x_k^{(5)}$  activé quand  $\varepsilon_k^{(5)} = 1$

$x_k^{(6)}$  activé quand  $\varepsilon_k^{(6)} = 1$

Comment choisir ( $\forall k \in \mathbb{N}$ ) la variable  $\varepsilon_k = (\varepsilon_k^{(1)}, \dots, \varepsilon_k^{(6)})$ ?

$$P[\varepsilon_k = (1, 1, 0, 0, 0, 0)] = 0.1$$

# Illustration de la stratégie d'activation des blocs

Sélection des blocs ( $\forall n \in \mathbb{N}$ )

$x_k^{(1)}$  activé quand  $\varepsilon_k^{(1)} = 1$

$x_k^{(2)}$  activé quand  $\varepsilon_k^{(2)} = 1$

$x_k^{(3)}$  activé quand  $\varepsilon_k^{(3)} = 1$

$x_k^{(4)}$  activé quand  $\varepsilon_k^{(4)} = 1$

$x_k^{(5)}$  activé quand  $\varepsilon_k^{(5)} = 1$

$x_k^{(6)}$  activé quand  $\varepsilon_k^{(6)} = 1$

Comment choisir ( $\forall k \in \mathbb{N}$ ) la variable  $\varepsilon_k = (\varepsilon_k^{(1)}, \dots, \varepsilon_k^{(6)})$ ?

$$P[\varepsilon_k = (1, 1, 0, 0, 0, 0)] = 0.1$$

$$P[\varepsilon_k = (1, 0, 1, 0, 0, 0)] = 0.2$$

# Illustration de la stratégie d'activation des blocs

## Sélection des blocs ( $\forall k \in \mathbb{N}$ )

$x_k^{(1)}$  activé quand  $\varepsilon_k^{(1)} = 1$

$x_k^{(2)}$  activé quand  $\varepsilon_k^{(2)} = 1$

$x_k^{(3)}$  activé quand  $\varepsilon_k^{(3)} = 1$

$x_k^{(4)}$  activé quand  $\varepsilon_k^{(4)} = 1$

$x_k^{(5)}$  activé quand  $\varepsilon_k^{(5)} = 1$

$x_k^{(6)}$  activé quand  $\varepsilon_k^{(6)} = 1$

Comment choisir ( $\forall k \in \mathbb{N}$ ) la variable  $\varepsilon_k = (\varepsilon_k^{(1)}, \dots, \varepsilon_k^{(6)})$ ?

$$P[\varepsilon_k = (1, 1, 0, 0, 0, 0)] = 0.1$$

$$P[\varepsilon_k = (1, 0, 1, 0, 0, 0)] = 0.2$$

$$P[\varepsilon_k = (1, 0, 0, 1, 1, 0)] = 0.2$$

# Illustration de la stratégie d'activation des blocs

Sélection des blocs ( $\forall k \in \mathbb{N}$ )

$x_k^{(1)}$  activé quand  $\varepsilon_k^{(1)} = 1$

$x_k^{(2)}$  activé quand  $\varepsilon_k^{(2)} = 1$

$x_k^{(3)}$  activé quand  $\varepsilon_k^{(3)} = 1$

$x_k^{(4)}$  activé quand  $\varepsilon_k^{(4)} = 1$

$x_k^{(5)}$  activé quand  $\varepsilon_k^{(5)} = 1$

$x_k^{(6)}$  activé quand  $\varepsilon_k^{(6)} = 1$

Comment choisir ( $\forall k \in \mathbb{N}$ ) la variable  $\varepsilon_k = (\varepsilon_k^{(1)}, \dots, \varepsilon_k^{(6)})$ ?

$$P[\varepsilon_k = (1, 1, 0, 0, 0, 0)] = 0.1$$

$$P[\varepsilon_k = (1, 0, 1, 0, 0, 0)] = 0.2$$

$$P[\varepsilon_k = (1, 0, 0, 1, 1, 0)] = 0.2$$

$$P[\varepsilon_k = (0, 1, 1, 1, 1, 1)] = 0.5$$



## Vers les algorithmes distribués ...

### PROBLÈME D'OPTIMISATION

Trouver un élément de l'ensemble  $\mathbf{F}$  des solutions de

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{minimiser}} \sum_{i=1}^m f_i(\mathbf{x}) + h_i(\mathbf{x}) + g_i(\mathbf{M}_i \mathbf{x})$$

avec  $\mathcal{H}, \mathcal{G}_1, \dots, \mathcal{G}_m$  espaces hilbertiens séparables et, pour tout  $i \in \{1, \dots, m\}$ :

- ▶  $h_i$   $\beta_i$ -Lipschitz différentiable avec  $\beta_i \in ]0, +\infty[$ ,
- ▶  $\mathbf{M}_i: \mathcal{H} \rightarrow \mathcal{G}_i$  opérateurs linéaires non nuls, bornés.

## Vers les algorithmes distribués ...

### PROBLÈME D'OPTIMISATION

Trouver un élément de l'ensemble  $\mathbf{F}$  des solutions de

$$\underset{\mathbf{x} \in \mathcal{H}}{\text{minimiser}} \quad \sum_{i=1}^m f_i(\mathbf{x}) + h_i(\mathbf{x}) + g_i(\mathbf{M}_i \mathbf{x})$$



### PROBLÈME D'OPTIMISATION

Trouver  $(x_1, \dots, x_m) \in \mathcal{H}^m$  tel que

$$\underset{(x_i)_{1 \leq i \leq m} \in \Lambda_m}{\text{minimiser}} \quad \sum_{i=1}^m f_i(x_i) + h_i(x_i) + g_i(\mathbf{M}_i x_i)$$

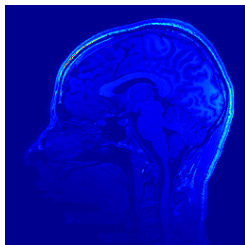
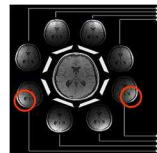
avec  $\Lambda_m = \{(x_i)_{1 \leq i \leq m} \in \mathcal{H}^m \mid x_1 = \dots = x_m\}$

# Applications

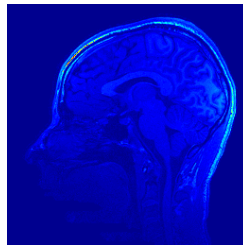
# Imagerie par Résonance Magnétique Parallèle

*(Collaboration avec le CEA Neurospin)*

- ★ Acquisition parallèle et échantillonnage compressé
  - ★ Signaux à valeurs complexes
- ⇒ Méthode primale-duale pour minimiser le critère pénalisé



Originale

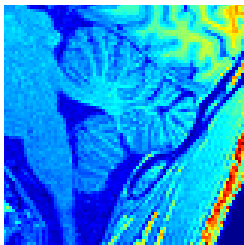
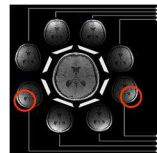


Reconstruite

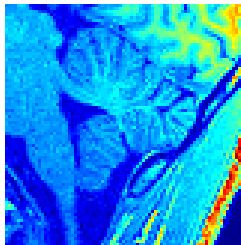
# Imagerie par Résonance Magnétique Parallèle

*(Collaboration avec le CEA Neurospin)*

- ★ Acquisition parallèle et échantillonnage compressé
  - ★ Signaux à valeurs complexes
- ⇒ Méthode primale-duale pour minimiser le critère pénalisé



Originale

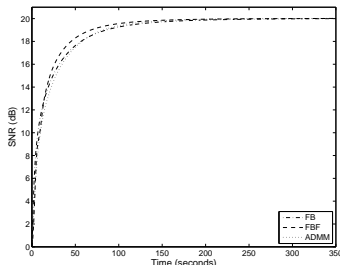
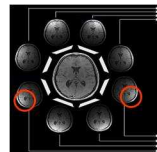


Reconstruite

# Imagerie par Résonance Magnétique Parallèle

*(Collaboration avec le CEA Neurospin)*

- ★ Acquisition parallèle et échantillonnage compressé
  - ★ Signaux à valeurs complexes
- ⇒ Méthode primale-duale pour minimiser le critère pénalisé



Vitesse de convergence de trois algorithmes proximaux primaux-duaux

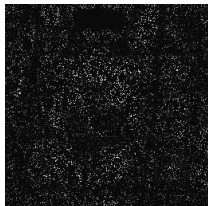
# Restauration d'images multispectrales

*(Collaboration avec l'ENS Lyon et Telecom ParisTech)*

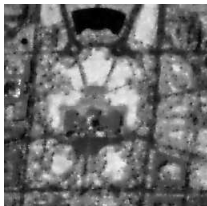
- ★ Choix des paramètres de régularisation  $\rightsquigarrow$  introduction de contraintes non locales
- ⇒ Approche primale-duale par épigraphes pour résoudre le problème contraint



Originale



Observée



TV

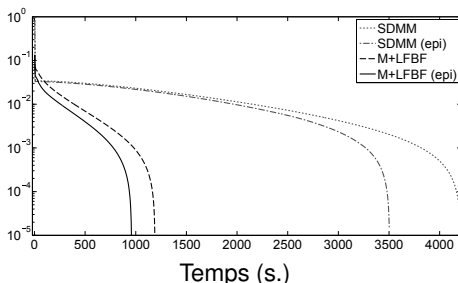


Restaurée

# Restauration d'images multispectrales

*(Collaboration avec l'ENS Lyon et Telecom ParisTech)*

- ★ Choix des paramètres de régularisation  $\rightsquigarrow$  introduction de contraintes non locales
- ⇒ Approche primale-duale par épigraphes pour résoudre le problème contraint



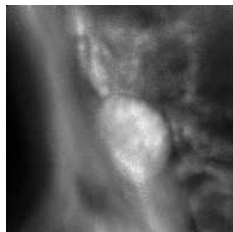
Formulation contrainte VS Formulation variationnelle



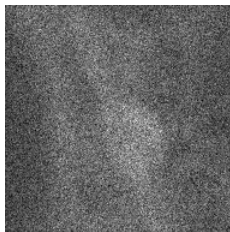
# Déconvolution sous un bruit Poisson-Gaussien

(ANR DIAMOND)

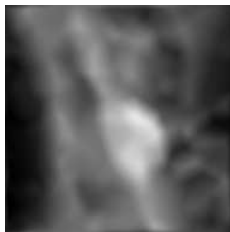
- ★ Analyse des propriétés de la log-vraisemblance associée au modèle de bruit
  - ★ Calcul approché du gradient  $\rightsquigarrow$  Robustesse aux erreurs nécessaire
- ⇒ Méthode primale-duale pour minimiser le critère pénalisé



Originale  
(190 × 190 pixels)



Observée  
SNR = 2.2 dB

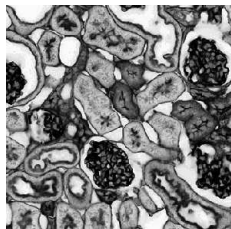


Restaurée  
SNR = 19.81 dB

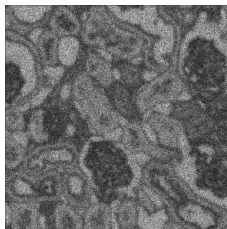
# Déconvolution sous un bruit Poisson-Gaussien

(ANR DIAMOND)

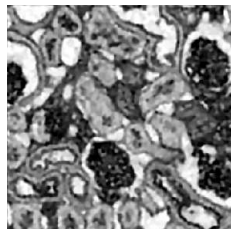
- ★ Analyse des propriétés de la log-vraisemblance associée au modèle de bruit
  - ★ Calcul approché du gradient  $\rightsquigarrow$  Robustesse aux erreurs nécessaire
- ⇒ Méthode primale-duale pour minimiser le critère pénalisé



Originale  
(350 × 350 pixels)



Observée  
SNR = 7.6 dB

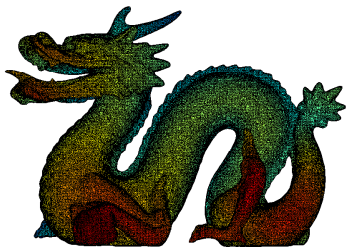


Restaurée  
SNR = 14.16 dB

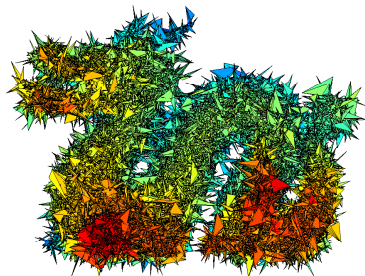
## Débruitage de maillages 3D

(ANR GRAPHSIP)

- ★ Données structurées sur un graphe
  - ★ Occupation mémoire limitée
- ⇒ Minimisation du critère pénalisé par une méthode primale-duale par blocs



Maillage original ( $M = 100250$  noeuds)

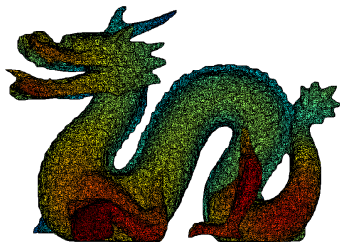


Maillage bruité,  $MSE = 2.89 \times 10^{-6}$

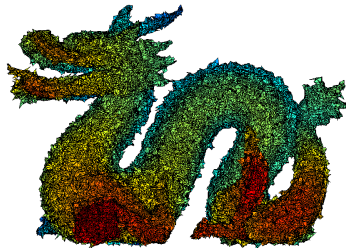
## Débruitage de maillages 3D

(ANR GRAPHSIP)

- ★ Données structurées sur un graphe
  - ★ Occupation mémoire limitée
- ⇒ Minimisation du critère pénalisé par une méthode primale-duale par blocs



Maillage restauré,  $MSE = 8.09 \times 10^{-8}$

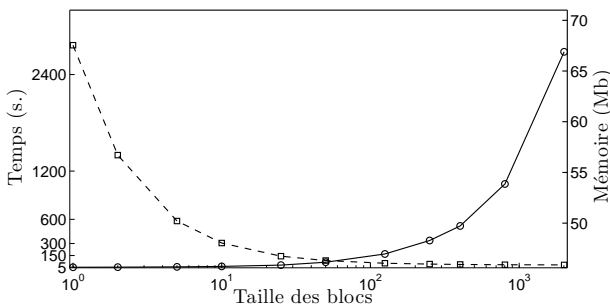


Maillage filtré (Laplacien),  $MSE = 5.23 \times 10^{-7}$

# Débruitage de maillages 3D

(ANR GRAPHSIP)

- ★ Données structurées sur un graphe
  - ★ Occupation mémoire limitée
- ⇒ Minimisation du critère pénalisé par une méthode primale-duale par blocs

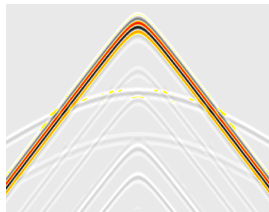
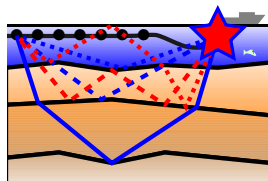


Mémoire nécessaire (—) VS Temps de calcul (— — —)

# Filtrage de multiples des signaux sismiques

*(Collaboration avec l'IFPEN)*

- ★ Formulation variationnelle du problème
  - ★ Nombreuses contraintes sur le signal primaire et le filtre
- ⇒ Méthode primale-duale pour gérer ces contraintes de façon parallèle

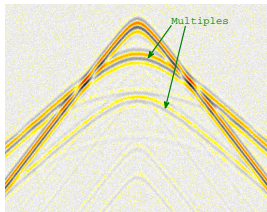
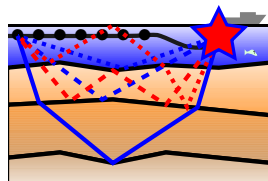


Signal primaire original (512 × 512)

# Filtrage de multiples des signaux sismiques

*(Collaboration avec l'IFPEN)*

- ★ Formulation variationnelle du problème
  - ★ Nombreuses contraintes sur le signal primaire et le filtre
- ⇒ Méthode primale-duale pour gérer ces contraintes de façon parallèle

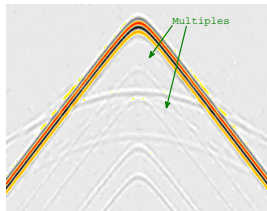
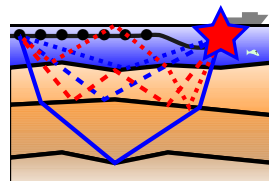


Signal observé (SNR = 1.13 dB)

# Filtrage de multiples des signaux sismiques

*(Collaboration avec l'IFPEN)*

- ★ Formulation variationnelle du problème
  - ★ Nombreuses contraintes sur le signal primaire et le filtre
- ⇒ Méthode primale-duale pour gérer ces contraintes de façon parallèle



Signal primaire restauré (SNR = 17.00 dB)



# Déconvolution aveugle et désentrelacement de vidéos

*(Collaboration avec l'INA)*

- ★ Problème aveugle  $\rightsquigarrow$  minimisation alternée
  - ★ Opérateurs linéaires complexes prenant en compte le mouvement inter-frames
- ⇒ Opérateur proximal lié à la vidéo calculé de façon approchée par une approche primale-duale



Trame n° 5 de la vidéo originale entrelacée (lignes paires)

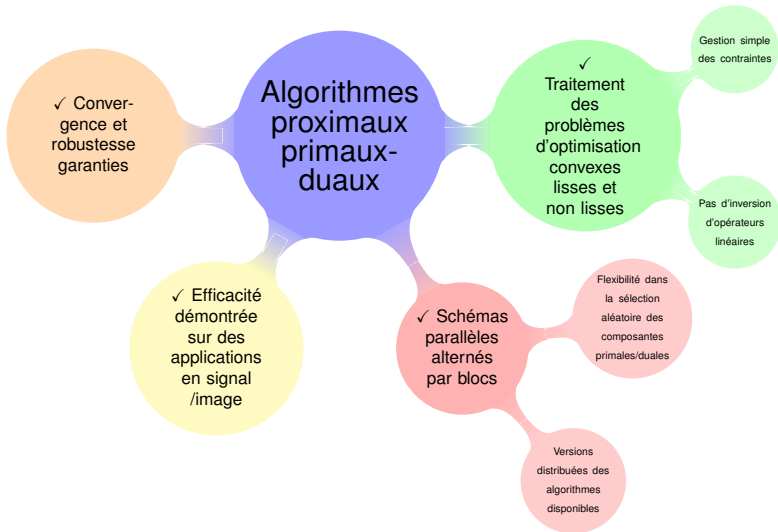
# Déconvolution aveugle et désentrelacement de vidéos

*(Collaboration avec l'INA)*

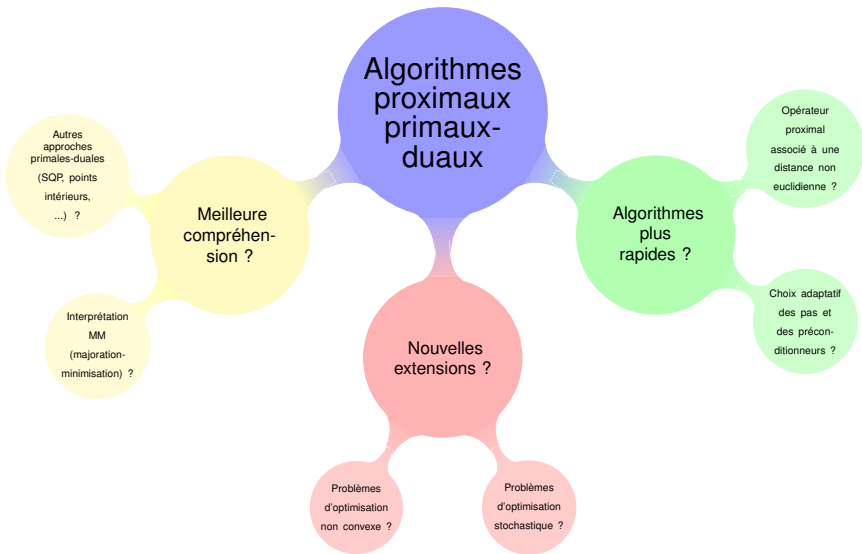


Trame n° 5 de la vidéo restaurée et désentrelacée  
(Temps de reconstruction  $\sim$  2 secondes par trames (code Matlab))

# Conclusion



# Perspectives



# Quelques références ...

- 
**F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot and L. Laborelli**  
**A Dual Block Coordinate Proximal Algorithm with Application to Deconvolution of Interlaced Video Sequences**  
*IEEE International Conference on Image Processing (ICIP 2015)*, 5 p., Québec, Canada, 27-30 Sep. 2015.
- 
**G. Chierchia, N. Pustelnik, B. Pesquet-Popescu and J.-C. Pesquet**  
**A Non-Local Structure Tensor Based Approach for Multicomponent Image Recovery Problems**  
*IEEE Transaction on Image Processing*, 23(12), pp. 5531 - 5544, Dec. 2014.
- 
**P. L. Combettes and J.-C. Pesquet**  
**Proximal Splitting methods in Signal Processing**  
*in Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz editors. Springer-Verlag, New York, pp. 185-212, 2011.
- 
**P. L. Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ**  
**A Forward-Backward View of Some Primal-Dual Optimization Methods in Image Recovery**  
*IEEE International Conference on Image Processing (ICIP 2014)*, 5 p., Paris, France, Oct. 27-30, 2014.
- 
**P. Combettes and J.-C. Pesquet**  
**Stochastic Quasi-Fejér Block-Coordinate Fixed Point Iterations with Random Sweeping**  
*SIAM Journal on Optimization*, 25(2), pp. 1221-1248, 2015.
- 
**A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu and S. Ciochina**  
**A Majorize-Minimize Memory Gradient Method for Complex-Valued Inverse Problems**  
*Signal Processing*, 103, pp. 285-295, 2014.
- 
**A. Jezierska, E. Chouzenoux, J.-C. Pesquet and H. Talbot**  
**A Convex Approach for Image Restoration with Exact Poisson-Gaussian Likelihood**  
 to appear in *SIAM Journal in Imaging Sciences*, 2015.
- 
**N. Komodakis and J.-C. Pesquet**  
**Playing with Duality: An Overview of Recent Primal-Dual Approaches for Solving Large-Scale Optimization Problems**  
 to appear in *IEEE Signal Processing Magazine*, 2015.
- 
**J.-C. Pesquet and A. Repetti**  
**A Class of Randomized Primal-Dual Algorithms for Distributed Optimization**  
 to appear in *Journal of Nonlinear and Convex Analysis*, 2015.
- 
**A. Repetti, E. Chouzenoux and J.-C. Pesquet**  
**A Random Block-Coordinate Primal-Dual Proximal Algorithm with Application to 3D Mesh Denoising**  
*IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015)*, 5 p., Brisbane, Australia, Apr. 19-24, 2015.