

Plan du cours

RAPPELS

- Problème d'optimisation
- Trames et parcimonie

UTILISATION DE TRAMES

- Objectif
- Approches

ALGORITHMES

- Définitions
- A propos de l'opérateur proximal
- Exemple

UN PEU PLUS LOIN

- Généralisation

CONCLUSIONS

- Conclusion
- Références

Problème direct

=

Problème direct

$z =$

- z : observations (e.g. signal 2D de taille $M = M_1 \times M_2$)

Problème direct

$$z = y$$

- z : observations (e.g. signal 2D de taille $M = M_1 \times M_2$)
- y : signal original (inconnu de taille N)

Problème direct

$$z = Ly$$

- z : observations (e.g. signal 2D de taille $M = M_1 \times M_2$)
- y : signal original (inconnu de taille N)
- L : opérateur linéaire (matrice de taille $M \times N$)

Problème direct

$$z = \mathcal{D}_\alpha(Ly)$$

- z : observations (e.g. signal 2D de taille $M = M_1 \times M_2$)
- y : signal original (inconnu de taille N)
- L : opérateur linéaire (matrice de taille $M \times N$)
- \mathcal{D}_α : perturbation de paramètre α

Problème direct

$$z = \mathcal{D}_\alpha(Ly)$$

- z : observations (e.g. signal 2D de taille $M = M_1 \times M_2$)
- y : signal original (inconnu de taille N)
- L : opérateur linéaire (matrice de taille $M \times N$)
- \mathcal{D}_α : perturbation de paramètre α

Objectif : résoudre le problème inverse

Trouver une estimation \tilde{y} de y à partir des observations z .

MAP

Problème de minimisation

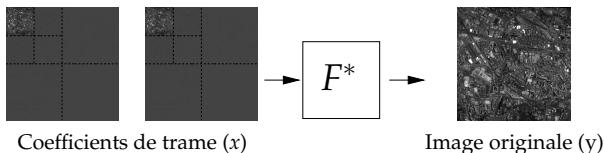
$$\min_{y \in \mathbb{R}^N} \underbrace{f_1(y)}_{\text{Terme de fidélité}} + \underbrace{f_2(y)}_{\text{Terme de régularisation}}$$

Question : que choisir pour f_1 et f_2 ?

Réponse rapide :

- ▶ $f_1(y) = -\log(\mu_{Z|Y=y}(z))$ est liée au bruit (anti-log vraisemblance).
- ▶ f_2 représente un *a priori* que l'on peut avoir sur y ou une représentation de y (on le détaillera dans la suite).

Représentations en ondelettes



- ▶ $x \in \mathbb{R}^K$: **Coefficients de trame** de l'image originale $y \in \mathbb{R}^N$
- ▶ $F^* : \mathbb{R}^K \rightarrow \mathbb{R}^N$: **Opérateur de synthèse de trame** tel que
 $\forall (\underline{\nu}, \bar{\nu}) \in]0, +\infty[^2, \underline{\nu}\text{Id} \leq F^* \circ F \leq \bar{\nu}\text{Id}$
 (la trame est dite ajustée lorsque $\underline{\nu} = \bar{\nu} = \nu$)

$$y = F^* x$$

[L. Jacques et al., 2011]

Exemples de trames

Exemples de trames :

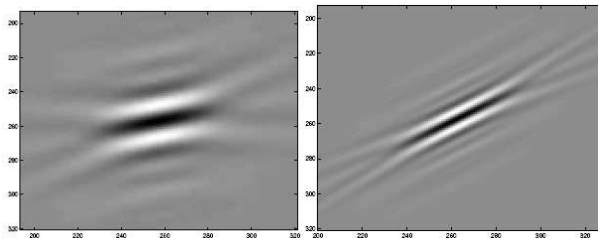
- ▶ Base d'ondelettes



Exemples de trames

Exemples de trames :

- ▶ Base d'ondelettes
- ▶ Curvelet [Candès, Donoho 02]



Exemples de trames

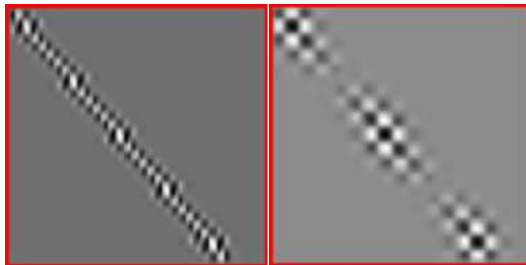
Exemples de trames :

- ▶ Base d'ondelettes
- ▶ Curvelet [Candès, Donoho 02]
- ▶ Contourlet [Do, Vetterli 05]

Exemples de trames

Exemples de trames :

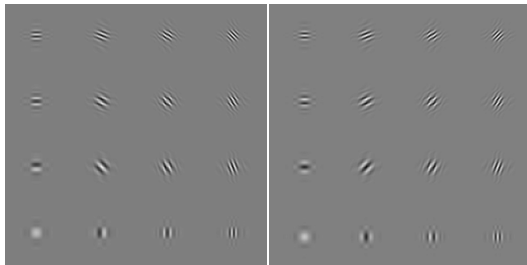
- ▶ Base d'ondelettes
- ▶ Curvelet [Candès, Donoho 02]
- ▶ Contourlet [Do, Vetterli 05]
- ▶ Bandlet [Le Pennec, Mallat 05]



Exemples de trames

Exemples de trames :

- ▶ Base d'ondelettes
- ▶ Curvelet [Candès, Donoho 02]
- ▶ Contourlet [Do, Vetterli 05]
- ▶ Bandlet [Le Pennec, Mallat 05]
- ▶ Transformée en arbre dual (*Dual-tree transform*) [Kingsbury 98], [Selesnick 01]



Exemples de trames

Exemples de trames :

- ▶ Base d'ondelettes
- ▶ Curvelet [Candès, Donoho 02]
- ▶ Contourlet [Do, Vetterli 05]
- ▶ Bandlet [Le Pennec, Mallat 05]
- ▶ Transformée en arbre dual (*Dual-tree transform*) [Kingsbury 98], [Selesnick 01]
- ▶ ...

Le problème peut-être reformuler comme :

Restaurer les coefficients de représentation sur une trame x (où $y = F^* x$) à partir de l'observation d'une image z .

A propos de parcimonie : définition

Définition

- Dire qu'un signal est parcimonieux, c'est dire qu'il n'a pas beaucoup de features.

A propos de parcimonie : définition

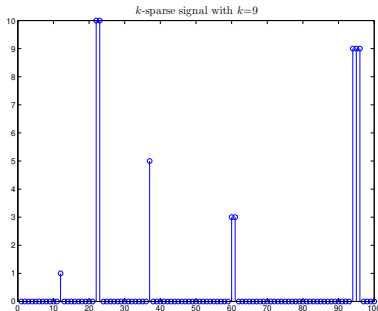
Définition

- Dire qu'un signal est parcimonieux, c'est dire qu'il n'a pas beaucoup de features.
- Autrement dit, un signal parcimonieux a beaucoup beaucoup de valeurs (approximativement) à zéro.

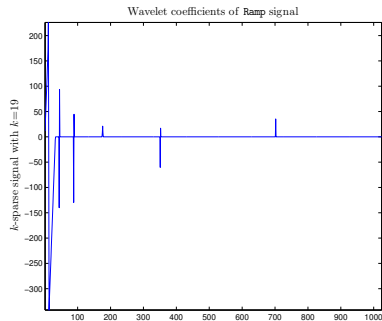
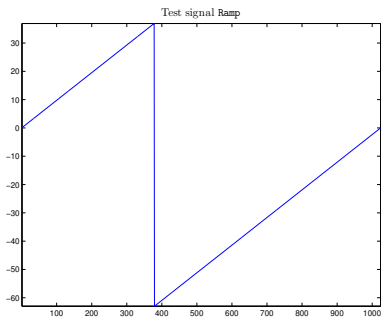
A propos de parcimonie : définition

Définition

- Dire qu'un signal est parcimonieux, c'est dire qu'il n'a pas beaucoup de features.
- Autrement dit, un signal parcimonieux a beaucoup beaucoup de valeurs (approximativement) à zéro.
- x est un signal k -parcimonieux : $\#\{i|x_i \neq 0\} = k$

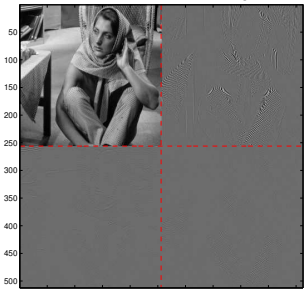


A propos de parcimonie : exemple jouet

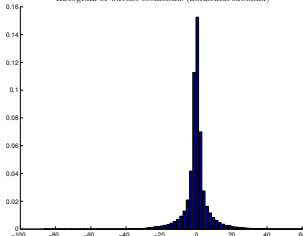


A propos de parcimonie : les images

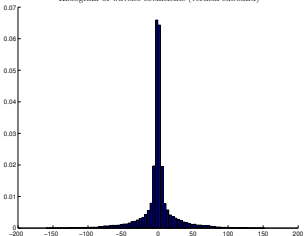
Wavelet coefficients of Barbara image



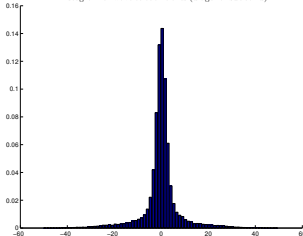
Histogram of wavelet coefficients (horizontal subband)



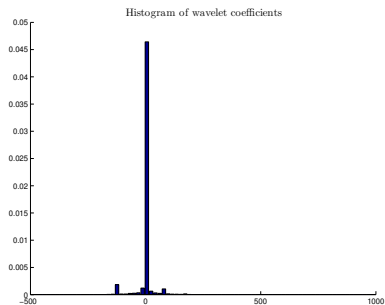
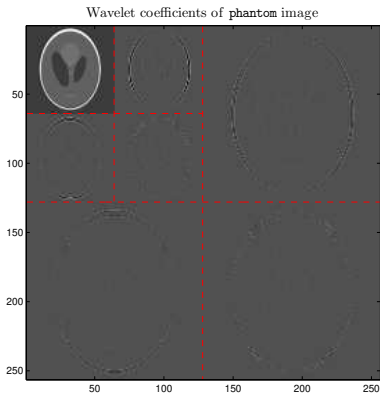
Histogram of wavelet coefficients (vertical subband)



Histogram of wavelet coefficients (diagonal subband)

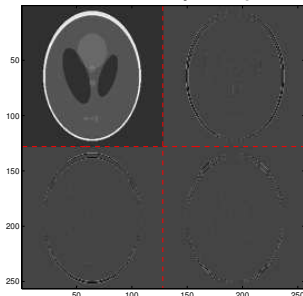


A propos de parcimonie : les images

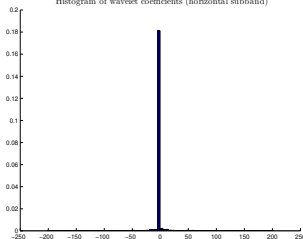


A propos de parcimonie : les images

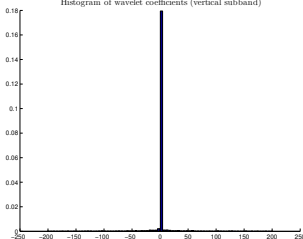
Wavelet coefficients of phantom image



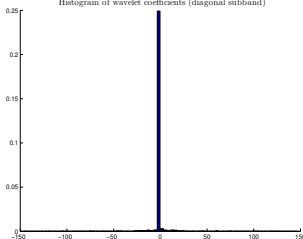
Histogram of wavelet coefficients (horizontal subband)



Histogram of wavelet coefficients (vertical subband)



Histogram of wavelet coefficients (diagonal subband)



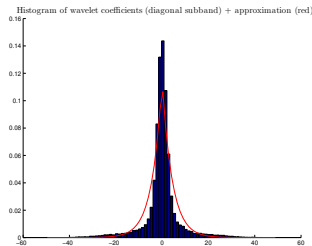
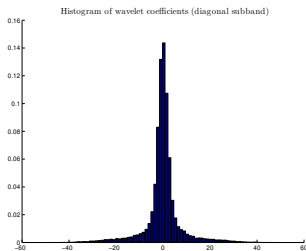
Représentation du signal

- ▶ Formuler un *a priori* directement sur le signal : pas toujours facile ;

Représentation du signal

- ▶ Formuler un *a priori* directement sur le signal : pas toujours facile ;
- ▶ Possibilité d'introduire de l'information *a priori* sur la représentation d'un signal, i.e. sur Fy plutôt que y .

Exemple : F : base d'ondelettes orthogonales ($K = N$).



Dans ce cas, $f_2(y) = \omega \sum_{k=1}^N |(Fy)^{(k)}|$ où $\omega \in]0, +\infty[$.

Approche à l'analyse vs. à la synthèse

Lorsque des **decompositions sur des trames** sont considérées, le problème peut être formulé soit :

Par abus de notation, on gardera les mêmes notations pour les fonctions f_1 et f_2 malgré la présence d'opérateurs linéaires.

Approche à l'analyse vs. à la synthèse

Lorsque des **decompositions sur des trames** sont considérées, le problème peut être formulé soit :
à la synthèse (SF) :

$$\underset{x \in \mathbb{R}^K}{\text{minimiser}} f_1(F^* x) + f_2(x)$$

Par abus de notation, on gardera les mêmes notations pour les fonctions f_1 et f_2 malgré la présence d'opérateurs linéaires.

Approche à l'analyse vs. à la synthèse

Lorsque des **decompositions sur des trames** sont considérées, le problème peut être formulé soit :

à la **synthèse** (SF) :

$$\underset{x \in \mathbb{R}^K}{\text{minimiser}} f_1(F^* x) + f_2(x)$$

à l'**analyse** (AF) :

$$\underset{y \in \mathbb{R}^N}{\text{minimiser}} f_1(y) + f_2(Fy).$$

Par abus de notation, on gardera les mêmes notations pour les fonctions f_1 et f_2 malgré la présence d'opérateurs linéaires.

Approche à l'analyse vs. à la synthèse

Lorsque des **decompositions sur des trames** sont considérées, le problème peut être formulé soit :

à la **synthèse** (SF) :

$$\underset{x \in \mathbb{R}^K}{\text{minimiser}} f_1(F^*x) + f_2(x)$$

à l'**analyse** (AF) :

$$\underset{y \in \mathbb{R}^N}{\text{minimiser}} f_1(y) + f_2(Fy).$$

Inclusion

AF est un cas particulier de SF [Chaâri et al., 2009].

Equivalence

Equivalence lorsque F modélise une transformée orthogonale.

Par abus de notation, on gardera les mêmes notations pour les fonctions f_1 et f_2 malgré la présence d'opérateurs linéaires.

Quelles régularisations ?

- ▶ Tikhonov [Tikhonov, 1963]

$$f_2(\mathbf{y}) = \|\Lambda \mathbf{y}\|^2$$

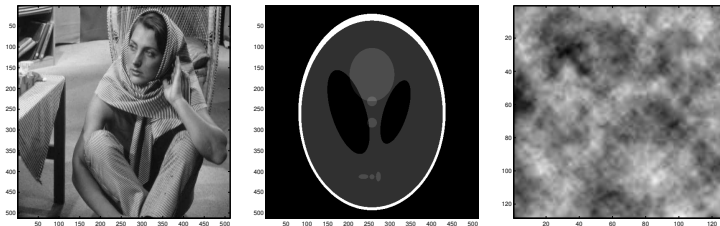
$$= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \{l * \mathbf{y}\}_{i,j}^2 \quad \text{avec} \quad l = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- ▶ Variation Totale (TV) [Rudin *et al.*, 1992]

$$f_2(\mathbf{y}) = \|\nabla \mathbf{y}\|_{1,2}$$

$$= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sqrt{(y_{i,j} - y_{i-1,j})^2 + (y_{i,j} - y_{i,j-1})^2}$$

Un choix important ...

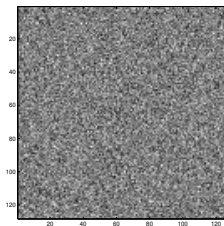
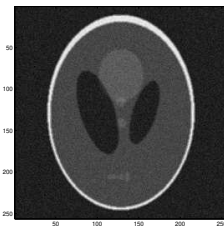
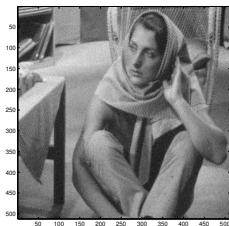


Grande diversité des images

Evaluation numérique des résultats :

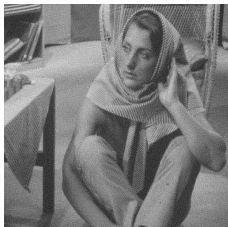
1. Rapport Signal à Bruit (SNR)
2. Indice de similarité (SSIM)
3. ...

Un choix important ...



Images observées

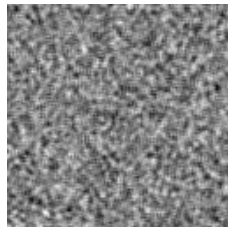
Un choix important ...



SNR=15.72 dB



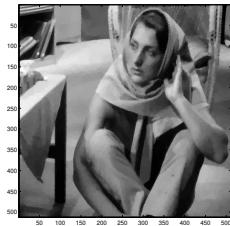
SNR=9.61 dB



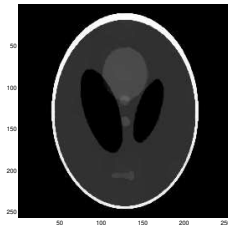
SNR=0.05 dB

Régularisation de Tikhonov

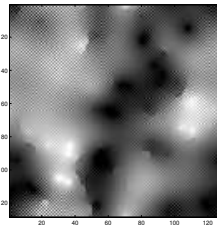
Un choix important ...



SNR=17.82 dB



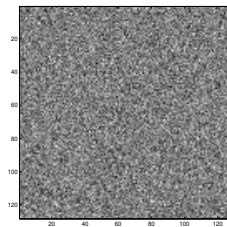
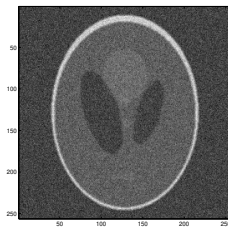
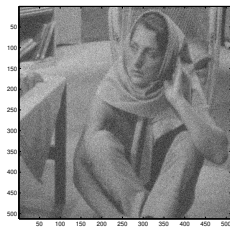
SNR=19.74 dB



SNR=5.38 dB

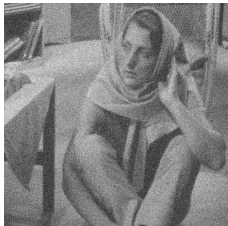
Régularisation par Variation Totale

Un choix important ...

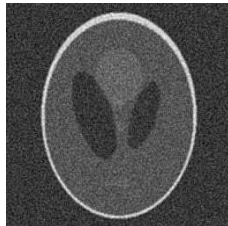


Images observées

Un choix important ...

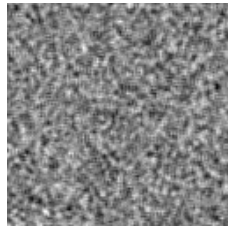


SNR=10.62 dB



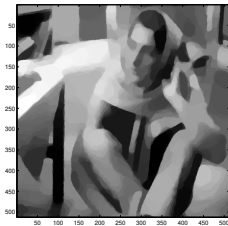
SNR=7.59 dB

Régularisation de Tikhonov

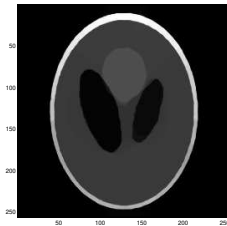


SNR=-0.13 dB

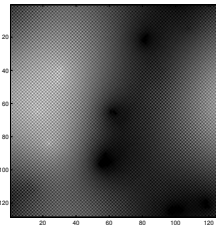
Un choix important ...



SNR=15.80 dB



SNR=9.93 dB



SNR=3.50 dB

Régularisation par Variation Totale

Définitions

Fonction propre

Soit $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$.

- ▶ Le domaine de f est $\text{dom} f = \{x \in \mathbb{R}^N \mid f(x) < +\infty\}$.
- ▶ f est **propre** si $\text{dom} f \neq \emptyset$.

Fonction semie-continue inférieurement

Soit $f : \mathbb{R}^N \rightarrow]-\infty, +\infty]$.

f est une **fonction semie-continue inférieurement** (s.c.i.) sur \mathbb{R}^N si, pour tout $x \in \mathbb{R}^N$ et pour toutes les suites $(x_i)_{i \in \mathbb{N}}$ de \mathbb{R}^N ,

$$\lim_{i \rightarrow +\infty} \|x_i - x\| = 0 \quad \Rightarrow \quad f(x) \leq \liminf f(x_i).$$

$\Gamma_0(\mathbb{R}^N)$ représente la classe des fonctions propres s.c.i.

Problème de minimisation simple

Objectif : minimiser une fonction $f \in \Gamma_0(\mathbb{R}^N)$

Problème de minimisation simple

Objectif : minimiser une fonction $f \in \Gamma_0(\mathbb{R}^N)$

- ▶ si f est de gradient β -Lipschitz avec $\beta \in]0, +\infty[$

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = y_n - \gamma_n \nabla f(y_n)$$

Algo. itératif de descente de gradient

où $0 < \inf_{n \in \mathbb{N}} \gamma_n$ et $\sup_{n \in \mathbb{N}} \gamma_n < 2/\beta$

Problème de minimisation simple

Objectif : minimiser une fonction $f \in \Gamma_0(\mathbb{R}^N)$

- ▶ si f est de gradient **β -Lipschitz** avec $\beta \in]0, +\infty[$

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = y_n - \gamma_n \nabla f(y_n)$$

Algo. itératif de descente de gradient

où $0 < \inf_{n \in \mathbb{N}} \gamma_n$ et $\sup_{n \in \mathbb{N}} \gamma_n < 2/\beta$

- ▶ Lorsque f est **non différentiable**, le gradient doit être remplacé par un sous-gradient

$$\partial f(x) = \left\{ t \in \mathbb{R}^N \mid (\forall y \in \mathbb{R}^N) \quad f(y) \geq f(x) + \langle t | y - x \rangle \right\}$$

$t \in \partial f(x)$: **sous-gradient** en $x \in \mathbb{R}^N$

$\partial f : \mathbb{R}^N \rightarrow 2^{\mathbb{R}^N}$: **sous-différentielle**

Sous-différentielle : exemples

- ▶ Si f est différentiable en $x \in \mathbb{R}^N$, alors $\partial f(x) = \{\nabla f(x)\}$.
- ▶ Si $f = |\cdot|$ alors

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

- ▶ Si $f = \iota_C$ où C est un ensemble convexe fermé non vide de \mathbb{R}^N alors

$$(\forall x \in \mathbb{R}^N), \partial \iota_C(x) = N_C(x)$$

où $N_C(x)$ est le cône normal à C en x ($N_C(x) = \emptyset$ si $x \notin C$).

Algorithme de sous-gradient [Shor, 1979]

► Forme explicite

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n t_n, t_n \in \partial f(x_n)$$

Algorithme de Shor.

où $(\forall n \in \mathbb{N}) \gamma_n \in]0, +\infty[$ tel que $\sum_{n=0}^{\infty} \gamma_n^2 < \infty$ et $\sum_{n=0}^{\infty} \gamma_n = \infty$.

Algorithme de sous-gradient [Shor, 1979]

- Forme explicite

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n t_n, t_n \in \partial f(x_n)$$

Algorithme de Shor.

où $(\forall n \in \mathbb{N}) \gamma_n \in]0, +\infty[$ tel que $\sum_{n=0}^{\infty} \gamma_n^2 < \infty$ et $\sum_{n=0}^{\infty} \gamma_n = \infty$.

- Forme implicite

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n t'_n, t'_n \in \partial f(x_{n+1})$$

$$\Leftrightarrow x_n - x_{n+1} \in \gamma_n \partial f(x_{n+1})$$

Algorithme de point proximal

où $(\forall n \in \mathbb{N}) \gamma_n \in]0, +\infty[$ tel que $\sum_{n=0}^{\infty} \gamma_n^2 < \infty$ et $\sum_{n=0}^{\infty} \gamma_n = \infty$.

Opérateur proximal

Proposition

Soit $\varphi \in \Gamma_0(\mathbb{R}^N)$. Pour tout $x \in \mathbb{R}^N$, il existe un unique vecteur $\hat{x} \in \mathbb{R}^N$ tel que $x - \hat{x} \in \partial\varphi(\hat{x})$.

- ▶ $\hat{x} = \text{prox}_\varphi(x)$
- ▶ $\text{prox}_\varphi : \mathbb{R}^N \rightarrow \mathbb{R}^N$: opérateur proximal

$$\begin{aligned}
 (\forall n \in \mathbb{N}) \quad x_n - x_{n+1} &\in \gamma_n \partial f(x_{n+1}) \\
 \Leftrightarrow x_{n+1} &= \text{prox}_{\gamma_n f} x_n
 \end{aligned}$$

Algorithme de point proximal

où $\inf_{n \in \mathbb{N}} \gamma_n > 0$ tel que $\sum_{n=0}^{\infty} \gamma_n = \infty$.

A propos de l'opérateur proximal

L'opérateur proximal de $\phi \in \Gamma_0(\mathbb{R}^N)$ est défini par

$$\text{prox}_\phi : \mathbb{R}^N \rightarrow \mathbb{R}^N : u \mapsto \arg \min_{v \in \mathbb{R}^N} \frac{1}{2} \|v - u\|^2 + \phi(v).$$

Remarque : si C est un ensemble convexe fermé non vide de \mathbb{R}^N , et ι_C représente la **fonction indicatrice** de C , i.e., $(\forall u \in \mathcal{H}) \iota_C(u) = 0$ si $u \in C$, $+\infty$ sinon, alors, prox_{ι_C} se réduit à la **projection** Π_C sur C .

- Soit $\psi = \phi(\cdot - v)$, où $v \in \mathbb{R}^N$. Alors $(\forall u \in \mathbb{R}^N)$

$$\text{prox}_\psi u = v + \text{prox}_\phi(u - v)$$

- Soit $\psi : v \mapsto \phi(-v)$. Alors $(\forall u \in \mathbb{R}^N)$

$$\text{prox}_\psi u = -\text{prox}_\phi(-u)$$

- Soit $\phi \in \Gamma_0(\mathbb{R}^M)$, $L : \mathbb{R}^N \rightarrow \mathbb{R}^M$ un opérateur linéaire borné. Supposons que $LL^* = \chi I$, pour $\chi \in]0, +\infty[$. Alors

$$\text{prox}_{\phi \circ L} = I + \chi^{-1} L^* (\text{prox}_{\chi\phi} - I) L.$$

A propos de l'opérateur proximal

- Soit $\phi \in \Gamma_0(\mathbb{R}^M)$, $L: \mathbb{R}^N \rightarrow \mathbb{R}^M$ un opérateur linéaire borné. Supposons que $LL^* = \chi I$, pour $\chi \in]0, +\infty[$. Alors

$$\text{prox}_{\phi \circ L} = I + \chi^{-1} L^* (\text{prox}_{\chi \phi} - I) L.$$

Exemples

$L = F^*$: Opérateur de synthèse de trame ajustée ($\underline{\nu} = \bar{\nu} = \nu$).

Cas particulier

$L = U$: Transformée en ondelettes (i.e. $UU^* = \text{Id}$)

$$\text{prox}_{f \circ U} = U^* \circ \text{prox}_f \circ U.$$

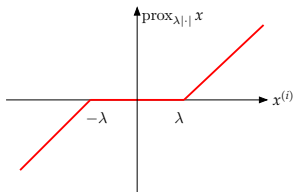
Opérateur proximal : un exemple

Si $f(x) = \|x\|_1 = \sum_{i=1}^N |x^{(i)}|$, alors

$$\text{prox}_{\lambda \|\cdot\|_1} x = (\text{prox}_{\lambda|\cdot|} x^{(i)})_{1 \leq i \leq N}$$

où

$$\text{prox}_{\lambda|\cdot|} x^{(i)} = \begin{cases} x^{(i)} - \lambda & \text{si } x^{(i)} > \lambda \\ 0 & \text{si } x^{(i)} \in [-\lambda, \lambda] \\ x^{(i)} + \lambda & \text{si } x^{(i)} < -\lambda. \end{cases}$$



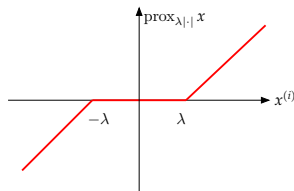
Opérateur proximal : un exemple

Si $f(x) = \|x\|_1 = \sum_{i=1}^N |x^{(i)}|$, alors

$$\text{prox}_{\lambda\|\cdot\|_1} x = (\text{prox}_{\lambda|\cdot|} x^{(i)})_{1 \leq i \leq N}$$

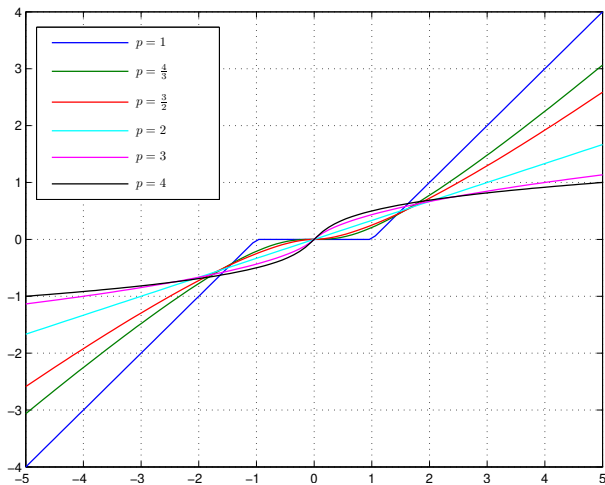
où

$$\text{prox}_{\lambda|\cdot|} x^{(i)} = \begin{cases} x^{(i)} - \lambda & \text{si } x^{(i)} > \lambda \\ 0 & \text{si } x^{(i)} \in [-\lambda, \lambda] \\ x^{(i)} + \lambda & \text{si } x^{(i)} < -\lambda. \end{cases}$$



L'opérateur proximal défini ici est le **seuillage doux**.

Opérateur proximal : un exemple



Tracé de $\text{prox}_{|\cdot|^p}$.

Qu'en est il de la minimisation de $f_1 + f_2$?

Objectif : minimisation de $f_1 + f_2$ où f_1 et f_2 appartiennent à $\Gamma_0(\mathbb{R}^N)$ et f_1 est de gradient β -Lipschitz.

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n(t'_n + \nabla f_1(x_n)), t'_n \in \partial f_2(x_{n+1})$$

$$\Leftrightarrow x_{n+1} = \text{prox}_{\gamma_n f_2}(x_n - \gamma_n \nabla f_1(x_n))$$

Algorithme Forward-Backward [Chen,Rockafellar,1997][Combettes,Wajs,2005]

Convergence de x_n si

1. $\text{Argmin}(f_1 + f_2) \neq \emptyset$
2. $0 < \inf_{n \in \mathbb{N}} \gamma_n$ et $\sup_{n \in \mathbb{N}} \gamma_n < 2/\beta$.

Autres algorithmes proches du FB

1. **Cas particuliers** : [Figuereido,Nowak, 2003][Bect et al., 2004][Daubechies,Defrise,De Mol, 2004]
2. **Méthodes accélérées** : [Nesterov, 2007][Bioucas-Dias,Figuereido, 2007][Beck,Teboulle, 2009]
3. **Plus général** : [Tseng, 2000][Raguet, Fadili, Peyré, 2013]

Algorithme de seuillage itératif rapide (FISTA)

Objectif : minimisation de $f_1 + f_2$ où f_1 et f_2 appartiennent à $\Gamma_0(\mathbb{R}^N)$ et f_1 est de gradient β -Lipschitz.

$$(\forall n \in \mathbb{N}) \quad \tilde{x}_n = \text{prox}_{f_2/\beta}(x_n - \frac{1}{\beta} \nabla f_1(x_n))$$

$$t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2}$$

$$x_{n+1} = \tilde{x}_n + \frac{t_n - 1}{t_{n+1}} (\tilde{x}_n - \tilde{x}_{n-1})$$

FISTA [Beck, Teboulle, 2009]

Exemple simple avec une approche à la synthèse

Soit le problème de **déconvolution** suivant :

$$z = L\bar{y} + b$$

où $b \sim \mathcal{N}(0, \sigma^2)$ et L est un opérateur de convolution.

Le problème peut se réécrire

$$z = LF^*\bar{x} + b$$

et on peut résoudre le problème inverse en résolvant

$$\underset{x \in \mathbb{R}^K}{\text{minimiser}} \sum_{i=1}^M \frac{1}{2\sigma^2} ((LF^*x)^{(i)} - z^{(i)})^2 + \lambda \sum_{k=1}^K |x^{(k)}|.$$

Exemple simple avec une approche à la synthèse

Soit le problème de **déconvolution** suivant :

$$z = L\bar{y} + b$$

où $b \sim \mathcal{N}(0, \sigma^2)$ et L est un opérateur de convolution.

Le problème peut se réécrire

$$z = LF^*\bar{x} + b$$

et on peut résoudre le problème inverse en résolvant

$$\underset{x \in \mathbb{R}^K}{\text{minimiser}} \sum_{i=1}^M \frac{1}{2} ((LF^*x)^{(i)} - z^{(i)})^2 + \lambda \sum_{k=1}^K |x^{(k)}|.$$

Kezako du pas de l'algorithme

Le terme de fidélité f_1 est donné par (norme ℓ_2)

$$f_1(x) = \sum_{i=1}^M \frac{1}{2} ((LF^*x)^{(i)} - z^{(i)})^2.$$

Il peut se réécrire comme

$$f_1 = \Psi \circ L \circ F^*$$

où $(\forall u = (u^{(i)})_{1 \leq i \leq M} \in \mathbb{R}^M)$ $\Psi(u) = \sum_{i=1}^M \psi_i(u^{(i)})$ et $\psi_i = \frac{1}{2}(\cdot - z^{(i)})^2$.

Kezako du pas de l'algorithme

$$f_1 = \Psi \circ L \circ F^*$$

$$\text{où } (\forall u = (u^{(i)})_{1 \leq i \leq M} \in \mathbb{R}^M) \quad \Psi(u) = \sum_{i=1}^M \psi_i(u^{(i)}) \text{ et } \psi_i = \frac{1}{2}(\cdot - z^{(i)})^2.$$

$$\forall x \in \mathbb{R}^K, \quad \nabla f_1(x) = FL^*(\nabla \Psi(LF^*x))$$

Pour tout x et x' de $(\mathbb{R}^K)^2$,

$$\begin{aligned} & \|FL^*(\nabla \Psi(LF^*x)) - FL^*(\nabla \Psi(LF^*x'))\| \\ & \leq \|F\| \|L^*\| \|\nabla \Psi(LF^*x) - \nabla \Psi(LF^*x')\| \\ & \leq \|F\| \|L^*\| \|LF^*x - LF^*x'\| \\ & \leq \|F\|^2 \|L\|^2 \|x - x'\| \\ & \leq \bar{\nu} \|L\|^2 \|x - x'\|. \end{aligned}$$

Kezako du pas de l'algorithme

$$f_1 = \Psi \circ L \circ F^*$$

$$\text{où } (\forall u = (u^{(i)})_{1 \leq i \leq M} \in \mathbb{R}^M) \quad \Psi(u) = \sum_{i=1}^M \psi_i(u^{(i)}) \text{ et } \psi_i = \frac{1}{2}(\cdot - z^{(i)})^2.$$

$$\forall x \in \mathbb{R}^K, \quad \nabla f_1(x) = FL^*(\nabla \Psi(LF^*x))$$

Pour tout x et x' de $(\mathbb{R}^K)^2$,

$$\begin{aligned} & \|FL^*(\nabla \Psi(LF^*x)) - FL^*(\nabla \Psi(LF^*x'))\| \\ & \leq \|F\| \|L^*\| \|\nabla \Psi(LF^*x) - \nabla \Psi(LF^*x')\| \\ & \leq \bar{\nu} \|L\|^2 \|x - x'\|. \end{aligned}$$

Si F est une trame ($F^* \circ F \leq \bar{\nu} \text{Id}$), alors $\gamma_n < 2(\bar{\nu} \|L\|^2)^{-1}$.

Exemple simple : algorithme

Soit $0 < \inf_{n \in \mathbb{N}} \gamma_n$ et $\sup_{n \in \mathbb{N}} \gamma_n < 2/\beta$ où β est la constante de Lipschitz

Initialiser x_0

For $n = 0, 1, \dots$ **do**

$$\left| \begin{array}{l} x_{n+1} = \text{soft}_{\gamma_n}(x_n - \gamma_n FL^*(LF^* x_n - z)) \text{ où } \text{soft}_{\gamma}(x) = \max(|x| - \\ \gamma, 0) \text{ sign}(x) \end{array} \right.$$

end For

Algorithm 1: Algorithme de seuillage itératif.

Exemple simple : résultats

Video : FB initialisé avec une matrice de zéros

Video : FB initialisé avec l'image bruitée (z)

Video : PPXA avec une régularisation TV

Video : PPXA avec une régularisation ℓ_1 (trame DTT)

Video : PPXA utilisant une régularisation hybride

Approche variationnelle

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^J f_j(L_j x)$$

où $(f_j)_{1 \leq j \leq J}$: fonctions de la classe $\Gamma_0(\mathcal{G}_j)$ (classe des fonctions convexes semi-continues inférieurement prenant leur valeur dans $] - \infty, +\infty]$) et où, pour tout $j \in \{1, \dots, J\}$, $L_j: \mathcal{H} \rightarrow \mathcal{G}_j$ est un opérateur linéaire borné (où $(\mathcal{G}_j)_{1 \leq j \leq J}$ représente des espaces de Hilbert).

Ce critère peut être **non différentiable**.

Approche variationnelle

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^J f_j(L_j x)$$

où $(f_j)_{1 \leq j \leq J}$: fonctions de la classe $\Gamma_0(\mathcal{G}_j)$ (classe des fonctions convexes semi-continues inférieurement prenant leur valeur dans $] -\infty, +\infty]$) et où, pour tout $j \in \{1, \dots, J\}$, $L_j: \mathcal{H} \rightarrow \mathcal{G}_j$ est un opérateur linéaire borné (où $(\mathcal{G}_j)_{1 \leq j \leq J}$ représente des espaces de Hilbert).

Ce critère peut être **non différentiable**.

- ▶ f_j peut être **lié au bruit** (e.g. un terme quadratique quand le bruit est Gaussien).

Approche variationnelle

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^J f_j(L_j x)$$

où $(f_j)_{1 \leq j \leq J}$: fonctions de la classe $\Gamma_0(\mathcal{G}_j)$ (classe des fonctions convexes semi-continues inférieurement prenant leur valeur dans $] -\infty, +\infty]$) et où, pour tout $j \in \{1, \dots, J\}$, $L_j: \mathcal{H} \rightarrow \mathcal{G}_j$ est un opérateur linéaire borné (où $(\mathcal{G}_j)_{1 \leq j \leq J}$ représente des espaces de Hilbert).

Ce critère peut être **non différentiable**.

- ▶ f_j peut être **lié au bruit** (e.g. un terme quadratique quand le bruit est Gaussien).
- ▶ f_j peut être lié à un **a priori** sur la solution cible (e.g. un a priori sur la distribution des coefs. d'ondelettes).

Approche variationnelle

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^J f_j(L_j x)$$

où $(f_j)_{1 \leq j \leq J}$: fonctions de la classe $\Gamma_0(\mathcal{G}_j)$ (classe des fonctions convexes semi-continues inférieurement prenant leur valeur dans $] -\infty, +\infty]$) et où, pour tout $j \in \{1, \dots, J\}$, $L_j: \mathcal{H} \rightarrow \mathcal{G}_j$ est un opérateur linéaire borné (où $(\mathcal{G}_j)_{1 \leq j \leq J}$ représente des espaces de Hilbert).

Ce critère peut être **non différentiable**.

- ▶ f_j peut être **lié au bruit** (e.g. un terme quadratique quand le bruit est Gaussien).
- ▶ f_j peut être lié à un **a priori** sur la solution cible (e.g. un a priori sur la distribution des coefs. d'ondelettes).
- ▶ f_j peut être lié à une **contrainte** (e.g. contrainte de support).

Approche variationnelle

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^J f_j(L_j x)$$

où $(f_j)_{1 \leq j \leq J}$: fonctions de la classe $\Gamma_0(\mathcal{G}_j)$ (classe des fonctions convexes semi-continues inférieurement prenant leur valeur dans $] -\infty, +\infty]$) et où, pour tout $j \in \{1, \dots, J\}$, $L_j: \mathcal{H} \rightarrow \mathcal{G}_j$ est un opérateur linéaire borné (où $(\mathcal{G}_j)_{1 \leq j \leq J}$ représente des espaces de Hilbert).

Ce critère peut être **non différentiable**.

- ▶ f_j peut être **lié au bruit** (e.g. un terme quadratique quand le bruit est Gaussien).
- ▶ f_j peut être lié à un **a priori** sur la solution cible (e.g. un a priori sur la distribution des coefs. d'ondelettes).
- ▶ f_j peut être lié à une **contrainte** (e.g. contrainte de support).
- ▶ L_j peut modéliser un **flou**.

Approche variationnelle

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^J f_j(L_j x)$$

où $(f_j)_{1 \leq j \leq J}$: fonctions de la classe $\Gamma_0(\mathcal{G}_j)$ (classe des fonctions convexes semi-continues inférieurement prenant leur valeur dans $] -\infty, +\infty]$) et où, pour tout $j \in \{1, \dots, J\}$, $L_j: \mathcal{H} \rightarrow \mathcal{G}_j$ est un opérateur linéaire borné (où $(\mathcal{G}_j)_{1 \leq j \leq J}$ représente des espaces de Hilbert).

Ce critère peut être **non différentiable**.

- ▶ f_j peut être **lié au bruit** (e.g. un terme quadratique quand le bruit est Gaussien).
- ▶ f_j peut être lié à un **a priori** sur la solution cible (e.g. un a priori sur la distribution des coefs. d'ondelettes).
- ▶ f_j peut être lié à une **contrainte** (e.g. contrainte de support).
- ▶ L_j peut modéliser un **flou**.
- ▶ L_j peut modéliser un gradient (e.g. **variation totale**).

Approche variationnelle

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^J f_j(L_j x)$$

où $(f_j)_{1 \leq j \leq J}$: fonctions de la classe $\Gamma_0(\mathcal{G}_j)$ (classe des fonctions convexes semi-continues inférieurement prenant leur valeur dans $] -\infty, +\infty]$) et où, pour tout $j \in \{1, \dots, J\}$, $L_j: \mathcal{H} \rightarrow \mathcal{G}_j$ est un opérateur linéaire borné (où $(\mathcal{G}_j)_{1 \leq j \leq J}$ représente des espaces de Hilbert).

Ce critère peut être **non différentiable**.

- ▶ f_j peut être **lié au bruit** (e.g. un terme quadratique quand le bruit est Gaussien).
- ▶ f_j peut être lié à un **a priori** sur la solution cible (e.g. un a priori sur la distribution des coefs. d'ondelettes).
- ▶ f_j peut être lié à une **contrainte** (e.g. contrainte de support).
- ▶ L_j peut modéliser un **flou**.
- ▶ L_j peut modéliser un gradient (e.g. **variation totale**).
- ▶ L_j peut modéliser un opérateur de **trame**.

Minimiser $\sum_j^J f_j(x)$

- ▶ Lorsque $J = 2$:
 - ▶ Algo. Forward-Backward [Figueiredo et Nowak, 2003][Bect et al., 2004][Daubechies et al., 2004][Combettes et Wajs, 2005][Chaux et al., 2007][Beck et Teboulle, 2009],
 - ▶ Algo. Douglas-Rachford [Lions et Mercier, 1979][Combettes et Pesquet, 2007]
- ▶ Lorsque $J > 2$: Algorithme Parallèle Proximal (PPXA) [Combettes et Pesquet, 2008].

Autres approches proximales : Minimiser $\sum_j f_j(L_j x)$

- ▶ *Algorithme Parallèle Proximal + (PPXA+)* [Pesquet, Pustelnik, 2012]
Même esprit que PPXA, requiert de calculer chaque prox_{f_j} .
Minimisations quadratiques à effectuer à l'initialisation et pour le calcul de variables intermédiaires \Leftrightarrow inversion d'un opérateur linéaire de grande taille.
- ▶ Forward-Backward généralisé [Raguet et al., 2012]
- ▶ *Approches primales-duales* :
 - ▶ M+SFBF [Briceño-Arias, Combettes, 2011]
Requiert de calculer chaque prox_{f_j} et le pas de l'algorithme dépend de $\|L_j\|$.
 - ▶ M+LFBF [Combettes, Pesquet, 2011]
Possibilité qu'une fonction f_{j_0} soit de gradient Lipschitz ; requiert de calculer le gradient de f_{j_0} et chaque prox_{f_j} for $j \neq j_0$. Le pas de l'algorithme dépend de $\|L_j\|$.
 - ▶ Algorithmes basés FB [Chambolle, Pock, 2011],[Vũ,2013],[Condat,2013]

PPXA+ : un cadre général

1. PPXA [Combettes, Pesquet, 2008, Algorithm 3.1] est un cas particulier de PPXA+ correspondant au cas où $\epsilon_1 = \dots = \epsilon_J = 0$, $\mathcal{G}_1 = \dots = \mathcal{G}_J = \mathcal{H}$, et $L_1 = \dots = L_J = \text{Id}$.
2. L'algorithme SDMM dérivé du DR [Setzer et al., 2010] est un cas particulier de PPXA+ correspondant au cas où $\epsilon_1 = \dots = \epsilon_J = 0$, $\omega_1 = \dots = \omega_J$, $\lambda_n \equiv 1$ et $(a_j^{[n]})_{1 \leq j \leq J} \equiv (0, \dots, 0)$.
3. L'algorithme introduit par [Attouch and Soueiyatt, 2009] est un cas particulier de PPXA+ correspondant au cas où $\epsilon_1 = \dots = \epsilon_J = \frac{\alpha}{1 + \alpha'}$, $(a_j^{[n]})_{1 \leq j \leq J} \equiv (0, \dots, 0)$.

Autres algorithmes

► Approches gloutonnes

Fondations : Matching Pursuit [Mallat, Zhang, 1993]

Avec un opérateur linéaire : Gradient Pursuit [Blumensath et.al., 2008], CoSaMP
[Needell, Tropp, 2009], Signal Space CoSaMP.

Accélération : OMP, random selection [Peel et. al., 2012.]

Terme de fidélité aux données général : GRASP [Bahmani, et. al., 2011.],
FOSP [Dupé, Anthoine, 2018.]

► Approches MM (Majoration-Minimisation)

Fondations : MM [Figueiredo, et. al. 2007], en astrophysique, imagerie
médicale.

Raffinements : non différentiabilité [Chouzenoux, et. al. 2011], multiples
dépendances [Villaron, et. al. 2011]

Conclusion

- ▶ Résoudre un problème inverse : formulation variationnelle régularisée
- ▶ Algorithmes performants, potentiellement parallèles

Quelques références

- G. Chierchia, N. Pustelnik, J.-C. Pesquet, B. Pesquet-Popescu, “Epigraphical Projection and Proximal Tools for Solving Constrained Convex Optimization Problems”, *Signal, Image and Video Processing*, Jul. 2014.
- F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, “Optimization with sparsity-inducing penalties”, *Foundations and Trends in Machine Learning*, vol. 4, no. 1, pp. 1-106, 2012.
- P. L. Combettes and J.-C. Pesquet, “Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators,” *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307-330, Jun. 2012.
- J.-C. Pesquet and N. Pustelnik, “A Parallel Inertial Proximal Optimization Method”, *Pacific Journal of Optimization*, Vol. 8, No. 2, pp. 273-305, Apr. 2012.
- H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer-Verlag, New York, 2011.
- 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. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Editors), pp. 185-212. Springer-Verlag, New York, 2011.
- A. Chambolle, T. Pock, “A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging”, *Journal of Mathematical Imaging and Vision*, Vol. 40, No. 1, pp. 120-145, May 2011.
- P. L. Combettes and J.-C. Pesquet, “A proximal decomposition method for solving convex variational inverse problems,” *Inverse Problems*, vol. 24, no. 6, article ID 065014, 27 pp., Dec. 2008.
- M. Figueiredo et. al. “Majorization-minimization algorithms for wavelet- based image restoration, *IEEE Trans. Im. Proc.*, 16 :12 (2007).
- E. Chouzenoux, J. Idier and S. Moussaoui. “A Majorize-Minimize Strategy for Subspace Optimization Applied to Image Restoration”. *IEEE Transactions on Image Processing*, Vol. 20, No. 18, pages 1517-1528, juin 2011.
- S. G. Mallat and Z. Zhang. “Matching pursuits with time-frequency dictionaries”. *IEEE Trans. Signal Process.*, 41(12) :3397-3415, December 1993.
- S. Bahmani, P. Boufounos, and B. Raj. “Greedy sparsity-constrained optimization”. In *Asilomar Conference on Signals, Systems, and Computers*, 2011.
- F.-X. Dupé, and S. Anthoine. “Generalized greedy alternatives”. *Applied and computational harmonic analysis*, to appear.
- D. Needell and J. A. Tropp. “Cosamp : Iterative signal recovery from incomplete and inaccurate samples”. *Applied and Computational Harmonic Analysis*, 26(3) :301-321, 2009.
- V. N. Temlyakov. “Greedy approximation”. *Acta Numerica*, 17 :235-409, 2008.
- L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” *Journal of Optimization Theory and Applications*, Aug. 2013.
- H. Raguet, J. Fadili, G. Peyré, “Generalized forward-backward splitting”, *SIAM J. Imaging Sci.*, 2013.

Quelques références de codes

- ▶ SSIM <https://ece.uwaterloo.ca/~z70wang/research/ssim/>
- ▶ CVX toolbox <http://cvxr.com/cvx/>
- ▶ UNLocBoX <http://unlocbox.sourceforge.net/>