Some applications of proximal methods

Caroline CHAUX

Joint work with P. L. Combettes, L. Duval, J.-C. Pesquet and N. Pustelnik

LATP - UMR CNRS 7353, Aix-Marseille Université, France

OSL 2013, Les Houches, 7-11 Jan. 2013

Direct problem

$$=$$

Direct problem

$$z =$$

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

Direct problem

$$z = \qquad \overline{y}$$

- z: observations (e.g. 2D signal of size $M = M_1 \times M_2$)
- $\overline{y}$: original signal (unknown of size N)

Direct problem

$$z = L\overline{y}$$

- z: observations (e.g. 2D signal of size $M = M_1 \times M_2$)
- $\overline{y}$: original signal (unknown of size N)
- L: linear operator (matrix of size $M \times N$)

Direct problem

$$z = \mathcal{D}_\alpha(L\overline{y})$$

- z: observations (e.g. 2D signal of size $M = M_1 \times M_2$)
- $\overline{y}$: original signal (unknown of size N)
- L: linear operator (matrix of size $M \times N$)
- $\mathcal{D}_\alpha$: perturbation of parameter α

Direct problem

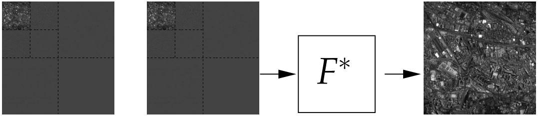
$$z = \mathcal{D}_\alpha(L\overline{y})$$

- z: observations (e.g. 2D signal of size $M = M_1 \times M_2$)
- $\overline{y}$: original signal (unknown of size N)
- L: linear operator (matrix of size $M \times N$)
- $\mathcal{D}_\alpha$: perturbation of parameter α

Objective: inverse problem

Find an estimation $\hat{y}$ of $\overline{y}$ from observations z.

Frame representation



Frame coefficients ($\overline{x}$) Original ($\overline{y}$)

- $\overline{x} \in \mathbb{R}^K$ **: Frame coefficients** of original image $\overline{y} \in \mathbb{R}^N$
- $F^* : \mathbb{R}^K \to \mathbb{R}^N$ **: Frame synthesis operator** such that $\exists(\underline{\nu}, \overline{\nu}) \in]0, +\infty[^2, \underline{\nu}\mathrm{Id} \leq F^* \circ F \leq \overline{\nu}\mathrm{Id}$
  (tight frame when $\underline{\nu} = \overline{\nu} = \nu$)

$$\boxed{\bar{y} = F^* \overline{x}}$$

[L. Jacques et al., 2011]

Variational approach

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

where $(f_j)_{1\le j\le J}$: functions in the class $\Gamma_0(\mathcal{G}_j)$ (class of l.s.c. proper convex functions on $\mathcal{G}_j$ taking their values in $]-\infty,+\infty]$) and where, for every $j\in\{1,\ldots,J\}$, $L_j\colon \mathcal{H}\to\mathcal{G}_j$ is a bounded linear operator (where $(\mathcal{G}_j)_{1\le j\le J}$ denote Hilbert spaces).

This criterion can be non differentiable.

Variational approach

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

where $(f_j)_{1\le j\le J}$: functions in the class $\Gamma_0(\mathcal{G}_j)$ (class of l.s.c. proper convex functions on $\mathcal{G}_j$ taking their values in $]-\infty,+\infty]$) and where, for every $j\in\{1,\ldots,J\}$, $L_j\colon\mathcal{H}\to\mathcal{G}_j$ is a bounded linear operator (where $(\mathcal{G}_j)_{1\le j\le J}$ denote Hilbert spaces).

This criterion can be non differentiable.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian)

Variational approach

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

where $(f_j)_{1\le j\le J}$: functions in the class $\Gamma_0(\mathcal{G}_j)$ (class of l.s.c. proper convex functions on $\mathcal{G}_j$ taking their values in $]-\infty,+\infty]$) and where, for every $j\in\{1,\ldots,J\}$, $L_j\colon \mathcal{H}\to\mathcal{G}_j$ is a bounded linear operator (where $(\mathcal{G}_j)_{1\le j\le J}$ denote Hilbert spaces).

This criterion can be non differentiable.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian)
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution)

Variational approach

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

where $(f_j)_{1\le j\le J}$: functions in the class $\Gamma_0(\mathcal{G}_j)$ (class of l.s.c. proper convex functions on $\mathcal{G}_j$ taking their values in $]-\infty,+\infty]$) and where, for every $j\in\{1,\ldots,J\}$, $L_j\colon \mathcal{H}\to\mathcal{G}_j$ is a bounded linear operator (where $(\mathcal{G}_j)_{1\le j\le J}$ denote Hilbert spaces).

This criterion can be non differentiable.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian)
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution)
- f_j can be related to a constraint (e.g. a support constraint)

Variational approach

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

where $(f_j)_{1\le j\le J}$: functions in the class $\Gamma_0(\mathcal{G}_j)$ (class of l.s.c. proper convex functions on $\mathcal{G}_j$ taking their values in $]-\infty,+\infty]$) and where, for every $j\in\{1,\ldots,J\}$, $L_j\colon \mathcal{H}\to\mathcal{G}_j$ is a bounded linear operator (where $(\mathcal{G}_j)_{1\le j\le J}$ denote Hilbert spaces).

This criterion can be non differentiable.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian)
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution)
- f_j can be related to a constraint (e.g. a support constraint)
- L_j can model a blur operator.

Variational approach

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

where $(f_j)_{1\le j\le J}$: functions in the class $\Gamma_0(\mathcal{G}_j)$ (class of l.s.c. proper convex functions on $\mathcal{G}_j$ taking their values in $]-\infty,+\infty]$) and where, for every $j\in\{1,\ldots,J\}$, $L_j\colon \mathcal{H}\to\mathcal{G}_j$ is a bounded linear operator (where $(\mathcal{G}_j)_{1\le j\le J}$ denote Hilbert spaces).

This criterion can be non differentiable.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian)
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution)
- f_j can be related to a constraint (e.g. a support constraint)
- L_j can model a blur operator.
- L_j can model a gradient operator (e.g. total variation).

Variational approach

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

where $(f_j)_{1\le j\le J}$: functions in the class $\Gamma_0(\mathcal{G}_j)$ (class of l.s.c. proper convex functions on $\mathcal{G}_j$ taking their values in $]-\infty,+\infty]$) and where, for every $j\in\{1,\ldots,J\}$, $L_j\colon \mathcal{H}\to\mathcal{G}_j$ is a bounded linear operator (where $(\mathcal{G}_j)_{1\le j\le J}$ denote Hilbert spaces).

This criterion can be non differentiable.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian)
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution)
- f_j can be related to a constraint (e.g. a support constraint)
- L_j can model a blur operator.
- L_j can model a gradient operator (e.g. total variation).
- L_j can model a frame operator.

Analysis approach vs. synthesis

When frame decompositions are considered, the problem can be formulated under a:

Analysis approach vs. synthesis

When frame decompositions are considered, the problem can be formulated under a:

Synthesis Form (SF):

$$\underset{x\in\mathbb{R}^K}{\text{minimize}} \ \sum_{r=1}^{R} f_r(L_r F^* x) + \sum_{s=1}^{S} g_s(x)$$

Analysis approach vs. synthesis

When frame decompositions are considered, the problem can be formulated under a:

Synthesis Form (SF):

$$\underset{x\in\mathbb{R}^K}{\text{minimize}} \sum_{r=1}^{R} f_r(L_r F^* x) + \sum_{s=1}^{S} g_s(x)$$

Analysis Form (AF):

$$\underset{y\in\mathbb{R}^N}{\text{minimize}} \sum_{r=1}^{R} f_r(L_r y) + \sum_{s=1}^{S} g_s(Fy).$$

ANALYSIS APPROACH VS. SYNTHESIS

When frame decompositions are considered, the problem can be formulated under a:

Synthesis Form (SF):

$$\underset{x \in \mathbb{R}^K}{\text{minimize}} \sum_{r=1}^{R} f_r(L_r F^* x) + \sum_{s=1}^{S} g_s(x)$$

Analysis Form (AF):

$$\underset{y \in \mathbb{R}^N}{\text{minimize}} \sum_{r=1}^{R} f_r(L_r y) + \sum_{s=1}^{S} g_s(Fy).$$

Inclusion

AF is a particular case of SF [Chaâri et al., 2009].

Equivalence

Equivalence when F is an orthonormal transform.

Proximal approaches

The proximity operator of $\phi \in \Gamma_0(\mathcal{H})$ is defined as

$$\operatorname{prox}_{\phi} \colon \mathcal{H} \to \mathcal{H} \colon u \mapsto \arg\min_{v \in \mathcal{H}} \frac{1}{2} \|v - u\|^2 + \phi(v).$$

Proximal approaches

The proximity operator of $\phi \in \Gamma_0(\mathcal{H})$ is defined as

$$\operatorname{prox}_{\phi} \colon \mathcal{H} \to \mathcal{H} \colon u \mapsto \arg\min_{v \in \mathcal{H}} \frac{1}{2} \|v - u\|^2 + \phi(v).$$

Remark: if C is a nonempty closed convex set of $\mathcal{H}$, and ι_C denotes the indicator function of C, i.e., $(\forall u \in \mathcal{H})\ \iota_C(u) = 0$ if $u \in C$, $+\infty$ otherwise, then, $\operatorname{prox}_{\iota_C}$ reduces to the projection Π_C onto C.

Proximal approaches

The proximity operator of $\phi \in \Gamma_0(\mathcal{H})$ is defined as

$$\operatorname{prox}_\phi \colon \mathcal{H} \to \mathcal{H} \colon u \mapsto \arg\min_{v \in \mathcal{H}} \frac{1}{2} \|v - u\|^2 + \phi(v).$$

Remark: if C is a nonempty closed convex set of $\mathcal{H}$, and ι_C denotes the indicator function of C, i.e., $(\forall u \in \mathcal{H})\ \iota_C(u) = 0$ if $u \in C$, $+\infty$ otherwise, then, $\operatorname{prox}_{\iota_C}$ reduces to the projection Π_C onto C.

• Let $\phi \in \Gamma_0(\mathcal{G})$, $L \colon \mathcal{H} \to \mathcal{G}$ a bounded linear operator. Suppose $LL^* = \chi\,\mathrm{I}$, for some $\chi \in \,]0, +\infty[$. Then

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

$\underset{x}{\text{Minimize}} \sum_j^J f_j(x)$

- When $J = 2$: Forward-Backward algorithm [Figueiredo and Nowak, 2003][Bect et al., 2004][Daubechies et al., 2004][Combettes and Wajs, 2005][Chaux et al., 2007][Beck and Teboulle, 2009], Douglas-Rachford algorithm [Lions and Mercier, 1979][Combettes and Pesquet, 2007]
- When $J > 2$: Parallel ProXimal Algorithm (PPXA) [Combettes and Pesquet, 2008]

PPXA+: $\underset{u \in \mathcal{H}}{\text{minimize}} \; \sum_{j=1}^{J} f_j(L_j u)$

Initialization

$(\epsilon_j)_{1 \le j \le J} \in [0,1[^J, (\omega_j)_{1 \le j \le J} \in]0,+\infty[^J,$

$(\lambda_n)_{n \in \mathbb{N}}$ a sequence of reals,

$(z_j^{[0]})_{1 \le j \le J} \in (\mathcal{G}_j)^J, (p_j^{[-1]})_{1 \le j \le J} \in (\mathcal{G}_j)^J,$

$u^{[0]} = \arg\min_{u \in \mathcal{H}} \sum_{j=1}^{J} \omega_j \|L_j u - z_j^{[0]}\|^2$

For every $j \in \{1,\dots,J\}$, $(a_j^{[n]})_{n \in \mathbb{N}}$ a sequence of reals,

For $n = 0, 1, \dots$

For $j = 1, \dots, J$

$$p_j^{[n]} = \mathrm{prox}_{\frac{(1-\epsilon_j) f_j}{\omega_j}} \left((1-\epsilon_j) z_j^{[n]} + \epsilon_j p_j^{[n-1]} \right) + a_j^{[n]}$$

$$c^{[n]} = \arg\min_{u \in \mathcal{H}} \sum_{j=1}^{J} \omega_j \|L_j u - p_j^{[n]}\|^2$$

For $j = 1, \dots, J$

$$z_j^{[n+1]} = z_j^{[n]} + \lambda_n \big(L_j (2c^{[n]} - u^{[n]}) - p_j^{[n]} \big)$$

$$u^{[n+1]} = u^{[n]} + \lambda_n (c^{[n]} - u^{[n]})$$

PPXA+: $\underset{u \in \mathcal{H}}{\text{minimize}} \sum_{j=1}^{J} f_j(L_j u)$

Initialization

$(\epsilon_j)_{1\leq j\leq J} \in [0,1[^J, (\omega_j)_{1\leq j\leq J} \in]0,+\infty[^J,$

$(\lambda_n)_{n\in\mathbb{N}}$ a sequence of reals,

$(z_j^{[0]})_{1\leq j\leq J} \in (\mathcal{G}_j)^J, (p_j^{[-1]})_{1\leq j\leq J} \in (\mathcal{G}_j)^J,$

$u^{[0]} = \arg\min_{u\in\mathcal{H}} \sum_{j=1}^{J} \omega_j \|L_j u - z_j^{[0]}\|^2$

For every $j \in \{1,\ldots,J\}$, $(a_j^{[n]})_{n\in\mathbb{N}}$ a sequence of reals,

For $n = 0, 1, \ldots$

For $j = 1, \ldots, J$

$p_j^{[n]} = \mathrm{prox}_{\frac{(1-\epsilon_j)f_j}{\omega_j}}\left((1-\epsilon_j)z_j^{[n]} + \epsilon_j p_j^{[n-1]}\right) + a_j^{[n]}$

$c^{[n]} = \arg\min_{u\in\mathcal{H}} \sum_{j=1}^{n} \omega_j \|L_j u - p_j^{[n]}\|^2$

For $j = 1, \ldots, J$

$z_j^{[n+1]} = z_j^{[n]} + \lambda_n\left(L_j(2c^{[n]} - u^{[n]}) - p_j^{[n]}\right)$

$u^{[n+1]} = u^{[n]} + \lambda_n(c^{[n]} - u^{[n]})$

PPXA+ CONVERGENCE

Proposition [Pesquet and Pustelnik, 2012]

The weak convergence of the sequence $(u^{[n]})_{n\in\mathbb{N}}$ to a minimizer of $\sum_{j=1}^{J} f_j \circ L_j$ is established under the following assumptions:

1. $\mathbf{0} \in \operatorname{sri}\{(L_1 v - w_1, \dots, L_J v - w_J) \mid v \in \mathcal{H}, w_1 \in \operatorname{dom} f_1, \dots, w_J \in \operatorname{dom} f_J\}$,
2. There exists $\underline{\lambda} \in]0,2[$ such that $(\forall n \in \mathbb{N})$, $\underline{\lambda} \leq \lambda_{n+1} \leq \lambda_n$,
3. For every $j \in \{1,\dots,J\}$, $a_j^{[n]}$ are absolutely summable sequences in $\mathcal{H}$.
4. $\sum_{j=1}^{J} \omega_j L_j^* L_j$ is an isomorphism. (PPXA+ iterations can be slightly modified to avoid this assumption)

PPXA+: a general framework

1. PPXA [Combettes, Pesquet, 2008, Algorithm 3.1] is a special case of PPXA+ corresponding to the case when $\epsilon_1 = \cdots = \epsilon_J = 0$, $\mathcal{G}_1 = \cdots = \mathcal{G}_J = \mathcal{H}$, and $L_1 = \cdots = L_J = \mathrm{Id}$.
2. The SDMM algorithm derived from DR in [Setzer et al., 2010] is a special case of PPXA+ corresponding to the case when $\epsilon_1 = \cdots = \epsilon_J = 0$, $\omega_1 = \cdots = \omega_J$, $\lambda_n \equiv 1$ and $(a_j^{[n]})_{1\leq j\leq J} \equiv (0, \cdots, 0)$.
3. Algorithm introduced in [Attouch and Soueycatt, 2009] is a special case of PPXA+ corresponding to the case when $\epsilon_1 = \cdots = \epsilon_J = \dfrac{\alpha}{1+\alpha}$, $(a_j^{[n]})_{1\leq j\leq J} \equiv (0, \cdots, 0)$.

OTHER PROXIMAL APPROACHES: $\underset{x}{\text{Minimize}} \sum_j^J f_j(L_j x)$

- *Parallel ProXimal Algorithm + (PPXA+)* [Pesquet, Pustelnik, 2012]
  In the same spirit as PPXA, requires to compute each prox_{f_j}. Quadratic minimizations need to be performed in the initialization step and in the computation of one intermediate variables $\Leftrightarrow$ invert a large-size linear operator.
- Generalized Forward-Backward [Raguet et al., 2012]
- *Primal-Dual approaches:*
  - M+SFBF [Briceño-Arias, Combettes, 2011]
    Requires to compute each prox_{f_j} and algorithm stepsize dependent on $\|L_j\|$.
  - M+LFBF [Combettes, Pesquet, 2011]
    Possibility that one function f_{j_0} is Lipschitz gradient; requires to compute the gradient of f_{j_0} and each prox_{f_j} for $j \neq j_0$. The algorithm stepsize is dependent on $\|L_j\|$.
  - FB based algorithms [Chambolle, Pock, 2011],[Vũ,2013],[Condat,2013]

Constrained formulation

$$\boxed{\underset{x\in\mathcal{H}}{\text{Minimize}} \sum_{r=1}^{R} g_r(T_r x) \quad \text{s.t.} \quad \begin{cases} H_1 x \in C_1, \\ \vdots \\ H_S x \in C_S, \end{cases}}$$

where

- $\mathcal{H}$: real Hilbert space,
- $\Gamma_0(\mathcal{H})$: class of proper, l.s.c, convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$,
- $(\forall s \in \{1,\ldots,S\})$, $H_s : \mathcal{H} \to \mathbb{R}^{Q_s}$ is a bounded linear operator,
- $(\forall s \in \{1,\ldots,S\})$, C_s is a nonempty closed convex subset of $\mathbb{R}^{Q_s}$,
- $(\forall r \in \{1,\ldots,R\})$, $T_r : \mathcal{H} \to \mathbb{R}^{N_r}$ is a bounded linear operator,
- $(\forall r \in \{1,\ldots,R\})$, $g_r \in \Gamma_0(\mathbb{R}^{N_r})$.

Constrained formulation

For $n = 0, 1, \ldots$

$$x^{[n]} = \sum_{r=1}^R \omega_r u_r^{[n]} + \sum_{s=1}^S \omega_s \overline{u}_s^{[n]}$$

⟵ Under technical assumptions, $(x^{[n]})_{n\in\mathbb{N}}$ generated by M+SFBF [Combettes,Briceño-Arias,2011] converge to $\hat{x}$

For $r = 1, \ldots, R$

$$w_{1,r}^{[n]} = u_r^{[n]} - \gamma_n T_r^* v_r^{[n]}$$
$$w_{2,r}^{[n]} = v_r^{[n]} + \gamma_n T_r u_r^{[n]}$$

For $s = 1, \ldots, S$

$$\overline{w}_{1,s}^{[n]} = \overline{u}_s^{[n]} - \gamma_n H_s^* \overline{v}_s^{[n]}$$
$$\overline{w}_{2,s}^{[n]} = \overline{u}_s^{[n]} + \gamma_n H_s \overline{u}_s^{[n]}$$

$$p_1^{[n]} = \sum_{r=1}^R \omega_r w_{1,r}^{[n]} + \sum_{s=1}^S \omega_s \overline{w}_{1,s}^{[n]}$$

For $r = 1, \ldots, R$

$$\boxed{p_{2,r}^{[n]} = w_{2,r}^{[n]} - \frac{\gamma_n}{\omega_r} \operatorname{prox}_{\frac{\omega_r}{\gamma_n} g_r}\left(\frac{\omega_r}{\gamma_n} w_{2,r}^{[n]}\right)}$$

⟵ Proximity operator computation

$$q_{1,r}^{[n]} = p_1^{[n]} - \gamma_n (T_r^* p_{2,r}^{[n]})$$
$$q_{2,r}^{[n]} = p_{2,r}^{[n]} + \gamma_n (T_r p_1^{[n]})$$

Update $u_1^{[n+1]}$ and $v_1^{[n+1]}$

For $s = 1, \ldots, S$

$$\boxed{\overline{p}_{2,s}^{[n]} = \overline{w}_{2,s}^{[n]} - \frac{\gamma_n}{\omega_s} \Pi_{C_s}\left(\frac{\omega_s}{\gamma_n} \overline{w}_{2,s}^{[n]}\right)}$$

⟵ Projection computation

$$\overline{q}_{1,s}^{[n]} = \overline{p}_1^{[n]} - \gamma_n (H_s^* \overline{p}_{2,s}^{[n]})$$
$$\overline{q}_{2,s}^{[n]} = \overline{p}_{2,s}^{[n]} + \gamma_n (H_s \overline{p}_1^{[n]})$$

Update $\overline{u}_1^{[n+1]}$ and $\overline{v}_1^{[n+1]}$

Constrained formulation

$$\boxed{(\forall x \in \mathcal{H}) \qquad H_s x \in C_s \quad \Leftrightarrow \quad h_s(H_s x) \leq \eta_s}$$

Constrained formulation

$$\boxed{(\forall x \in \mathcal{H}) \qquad H_s x \in C_s \quad \Leftrightarrow \quad h_s(H_s x) \leq \eta_s}$$

$$\vdots$$

$$\boxed{(\forall u \in \mathbb{R}^Q) \qquad u \in C \quad \Leftrightarrow \quad h(u) \leq \eta}$$

Constrained formulation

$$\boxed{(\forall x \in \mathcal{H}) \qquad H_s x \in C_s \quad \Leftrightarrow \quad h_s(H_s x) \leq \eta_s}$$

$$\vdots$$

$$\boxed{(\forall u \in \mathbb{R}^Q) \qquad u \in C \quad \Leftrightarrow \quad h(u) \leq \eta}$$

$$\vdots$$

$$\boxed{(\forall u = [\underbrace{(\mathsf{u}^{(1)})^\top}_{\text{size } Q^{(1)}}, \ldots, \underbrace{(\mathsf{u}^{(L)})^\top}_{\text{size } Q^{(L)}}]^\top \in \mathbb{R}^Q) \qquad u \in C \quad \Leftrightarrow \quad \sum_{\ell=1}^{L} h^{(\ell)}(\mathsf{u}^{(\ell)}) \leq \eta}$$

CONSTRAINED FORMULATION

$$\boxed{(\forall x \in \mathcal{H}) \qquad H_s x \in C_s \quad \Leftrightarrow \quad h_s(H_s x) \leq \eta_s}$$

$$\vdots$$

$$\boxed{(\forall u \in \mathbb{R}^Q) \qquad u \in C \quad \Leftrightarrow \quad h(u) \leq \eta}$$

$$\vdots$$

$$\boxed{(\forall u = [\underbrace{(\mathsf{u}^{(1)})^\top}_{\text{size } Q^{(1)}}, \ldots, \underbrace{(\mathsf{u}^{(L)})^\top}_{\text{size } Q^{(L)}}]^\top \in \mathbb{R}^Q) \qquad u \in C \quad \Leftrightarrow \quad \sum_{\ell=1}^{L} h^{(\ell)}(\mathsf{u}^{(\ell)}) \leq \eta}$$

$\rightarrow$ Any closed convex subset C can be expressed in this way by setting $\eta = 0$, $L = 1$ and $h = d_C$.

Constrained formulation

$$\boxed{(\forall x \in \mathcal{H}) \qquad H_s x \in C_s \quad \Leftrightarrow \quad h_s(H_s x) \leq \eta_s}$$

$$\vdots$$

$$\boxed{(\forall u \in \mathbb{R}^Q) \qquad u \in C \quad \Leftrightarrow \quad h(u) \leq \eta}$$

$$\vdots$$

$$\boxed{(\forall u = [\underbrace{(\mathsf{u}^{(1)})^\top}_{\text{size } Q^{(1)}}, \ldots, \underbrace{(\mathsf{u}^{(L)})^\top}_{\text{size } Q^{(L)}}]^\top \in \mathbb{R}^Q) \qquad u \in C \quad \Leftrightarrow \quad \sum_{\ell=1}^{L} h^{(\ell)}(\mathsf{u}^{(\ell)}) \leq \eta}$$

→ Any closed convex subset C can be expressed in this way by setting $\eta = 0$, $L = 1$ and $h = d_C$.

Question: What can we do if Π_C does not have a closed form ?

EPIGRAPHICAL PROJECTION

For every $u = [\underbrace{(\mathsf{u}^{(1)})^\top}_{\text{size } Q^{(1)}}, \ldots, \underbrace{(\mathsf{u}^{(L)})^\top}_{\text{size } Q^{(L)}}]^\top \in \mathbb{R}^Q$,

$$\boxed{u \in C \quad \Leftrightarrow \quad \sum_{\ell=1}^{L} h^{(\ell)}(\mathsf{u}^{(\ell)}) \leq \eta.}$$

By introducing now the auxiliary vector $\zeta = \left(\zeta^{(\ell)}\right)_{1 \leq \ell \leq L} \in \mathbb{R}^L$,

$$\boxed{u \in C \quad \Leftrightarrow \quad \begin{cases} \sum_{\ell=1}^{L} \zeta^{(\ell)} \leq \eta, \\ (\forall \ell \in \{1, \ldots, L\}) \qquad h^{(\ell)}(\mathsf{u}^{(\ell)}) \leq \zeta^{(\ell)}. \end{cases}}$$

Epigraphical projection

$$\boxed{u \in C \Leftrightarrow \quad \begin{cases} \zeta \in V \\ (u, \zeta) \in E \end{cases}}$$

where

- V denotes a closed half-space such that:

$$\boxed{V = \left\{ \zeta \in \mathbb{R}^L \mid \mathbf{1}_L^\top \zeta \leq \eta \right\}}$$

- E is the closed convex set associated to the epigraphical constraint:

$$\boxed{E = \left\{ (u, \zeta) \in \mathbb{R}^Q \times \mathbb{R}^L \mid (\forall \ell \in \{1, \ldots, L\}) \ (\mathrm{u}^{(\ell)}, \zeta^{(\ell)}) \in \operatorname{epi} h^{(\ell)} \right\}}$$

Epigraphical projection

$$\boxed{u \in C \Leftrightarrow \quad \begin{cases} \zeta \in V \\ (u, \zeta) \in E \end{cases}}$$

where

- V denotes a closed half-space such that:

$$\boxed{V = \left\{\zeta \in \mathbb{R}^L \mid \mathbf{1}_L^\top \zeta \leq \eta\right\}}$$

$\rightarrow \Pi_V$ has a closed form: projection onto an half-space.

- E is the closed convex set associated to the epigraphical constraint:

$$\boxed{E = \left\{(u, \zeta) \in \mathbb{R}^Q \times \mathbb{R}^L \mid (\forall \ell \in \{1, \ldots, L\})\ (\mathrm{u}^{(\ell)}, \zeta^{(\ell)}) \in \operatorname{epi} h^{(\ell)}\right\}}$$

$\rightarrow \Pi_E$ has a closed form for specific choice of $h^{(\ell)}$.

Epigraphical projection

- Euclidean norm functions defined as:

$$\boxed{\left(\forall \ell \in \{1,\ldots,L\}\right)\left(\forall \mathsf{u}^{(\ell)} \in \mathbb{R}^{Q^{(\ell)}}\right) \qquad h^{(\ell)}(\mathsf{u}^{(\ell)}) = \tau^{(\ell)}\|\mathsf{u}^{(\ell)}\|}$$

where $\tau^{(\ell)} \in \,]0,+\infty[$.

Epigraphical projection

- Euclidean norm functions defined as:

$$\boxed{(\forall \ell \in \{1,\dots,L\})\,(\forall \mathsf{u}^{(\ell)} \in \mathbb{R}^{Q^{(\ell)}}) \qquad h^{(\ell)}(\mathsf{u}^{(\ell)}) = \tau^{(\ell)} \|\mathsf{u}^{(\ell)}\|}$$

where $\tau^{(\ell)} \in \,]0,+\infty[$.

- Epigraphical projection: for every $(\mathsf{u}^{(\ell)}, \zeta^{(\ell)}) \in \mathbb{R}^{Q^{(\ell)}} \times \mathbb{R}$

$$\boxed{\Pi_{\operatorname{epi} h^{(\ell)}}(\mathsf{u}^{(\ell)}, \zeta^{(\ell)}) = \begin{cases} (\mathsf{u}^{(\ell)}, \zeta^{(\ell)}), & \text{if } \|\mathsf{u}^{(\ell)}\| < \frac{\zeta^{(\ell)}}{\tau^{(\ell)}}, \\ (0,0), & \text{if } \|\mathsf{u}^{(\ell)}\| < -\tau^{(\ell)}\zeta^{(\ell)}, \\ \alpha^{(\ell)}\left(\mathsf{u}^{(\ell)}, \tau^{(\ell)}\|\mathsf{u}^{(\ell)}\|\right), & \text{otherwise,} \end{cases}}$$

where $\alpha^{(\ell)} = \dfrac{1}{1+(\tau^{(\ell)})^2}\left(1 + \dfrac{\tau^{(\ell)}\zeta^{(\ell)}}{\|\mathsf{u}^{(\ell)}\|}\right)$.

Epigraphical projection

- Infinity norms defined as:

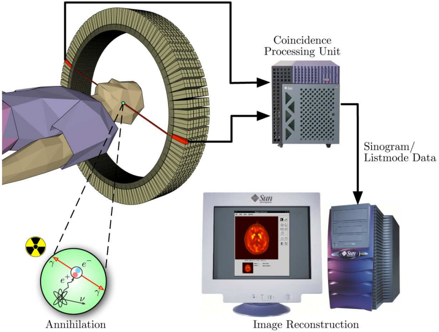
$$\Big(\forall \ell \in \{1,\dots,L\}\Big)\Big(\forall \mathsf{u}^{(\ell)} = (\mathsf{u}^{(\ell,m)})_{1\le m\le Q^{(\ell)}} \in \mathbb{R}^{Q^{(\ell)}}\Big)$$

$$\boxed{h^{(\ell)}(\mathsf{u}^{(\ell)}) = \max\left\{ \frac{|\mathsf{u}^{(\ell,m)}|}{\tau^{(\ell,m)}} \mid 1 \le m \le Q^{(\ell)} \right\}}$$

where $(\tau^{(\ell,m)})_{1\le m\le Q^{(\ell)}} \in \,]0,+\infty[^{Q^{(\ell)}}$.

$\Pi_{\operatorname{epi} h^{(\ell)}}(\mathsf{u}^{(\ell)}, \zeta^{(\ell)})$ has a closed form [G. Chierchia et al., 2012].

Reconstruction problem: PET



- High level of noise
- Large amount of data

Reconstruction problem

$$\boxed{z = \mathcal{P}_\alpha(A\overline{y})}$$

where

- $\mathcal{P}_\alpha$: Poisson noise of scale parameter α
- A: projection matrix

RECONSTRUCTION PROBLEM

Our objective is:

$$\boxed{\min_{x \in \mathbb{R}^K} \quad \sum_{t=1}^{T} D_{\mathrm{KL}}(AF_t^* x, z) + \kappa\, \mathrm{tv}(F_t^* x) + \iota_C(x) + \vartheta\, \|x\|_{\ell_1}}$$

$$\boxed{y = F^* x = (F_t^* x)_{1 \le t \le T}}$$

where $\kappa > 0$, $\vartheta > 0$ and

- D_{KL} is the Kullback-Leibler divergence
- tv represents a total variation term
- ι_C is the indicator function of a closed convex set C
- $\|x\|_{\ell_1}$ denotes the ℓ_1-norm.

Reconstruction problem

Our objective is:

$$\boxed{\min_{x\in\mathbb{R}^K} \quad \sum_{t=1}^{T} D_{\mathrm{KL}}(AF_t^* x, z) + \kappa\,\mathrm{tv}(F_t^* x) + \iota_C(x) + \vartheta\,\|x\|_{\ell_1}}$$

$$\boxed{y = F^* x = (F_t^* x)_{1\le t\le T}}$$

where $\kappa > 0$, $\vartheta > 0$ and

- D_{KL} is the Kullback-Leibler divergence $\Rightarrow$ split into several proximable functions
- tv represents a total variation term $\Rightarrow$ closed form in [Combettes and Pesquet, 2008]
- ι_C is the indicator function of a closed convex set C $\Rightarrow$ projection onto C
- $\|x\|_{\ell_1}$ denotes the ℓ_1-norm. $\Rightarrow$ soft thresholding [Chaux et al., 2007]

Reconstruction problem

Our objective is:

$$\boxed{\min_{x\in\mathbb{R}^K} \quad \sum_{t=1}^{T} D_{\mathrm{KL}}(AF_t^* x, z) + \kappa\, \mathrm{tv}(F_t^* x) + \iota_C(x) + \vartheta\, \|x\|_{\ell_1}}$$

$$\boxed{y = F^* x = (F_t^* x)_{1\le t\le T}}$$

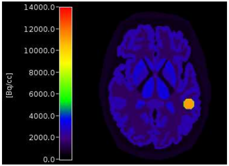
where $\kappa > 0$, $\vartheta > 0$ and

- D_{KL} is the Kullback-Leibler divergence
- tv represents a total variation term
- ι_C is the indicator function of a closed convex set C
- $\|x\|_{\ell_1}$ denotes the ℓ_1-norm.

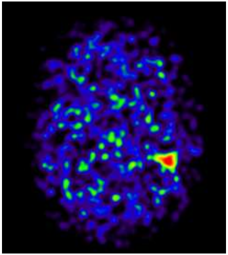
PET reconstruction results

Slice n 1

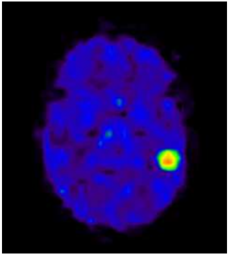
PET reconstruction results



Original

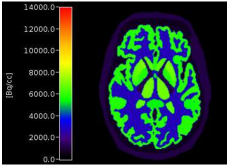


SIEVES

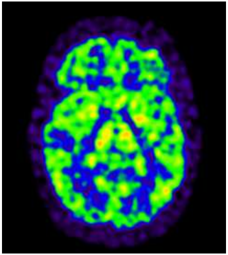


PPXA

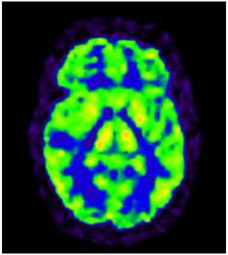
PET reconstruction results



Original

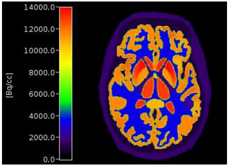


SIEVES

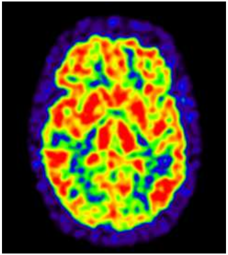


PPXA

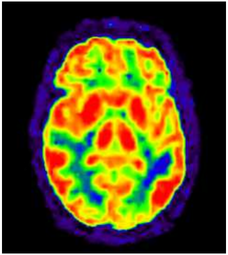
PET reconstruction results



Original

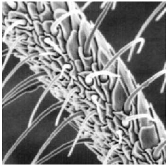
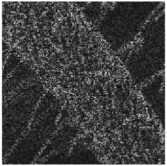


SIEVES



PPXA

Image restoration with missing samples

Original: $\overline{y} \in \mathbb{R}^N$ Degraded: $z \in \mathbb{R}^M$

$$\boxed{z = A\overline{y} + b}$$

- $\overline{y}$: original image in $[0, 255]^N$
  $\rightarrow$ assumed to be sparse after some appropriate transform,
- $A \in \mathbb{R}^{M \times N}$: randomly decimated convolution,
- $b \in \mathbb{R}^M$: realization of a zero-mean white Gaussian noise,
- z: degraded image of size M.

IMAGE RESTORATION WITH MISSING SAMPLES

$$\widehat{y} \in \underset{y \in [0,255]^N}{\text{Argmin}} \|Ay - z\|^2 \quad \text{s.t.} \quad \sum_{\ell=1}^{N} \|Y^{(\ell)}\|_p \leq \eta$$

where

- $Y^{(\ell)} = \left(\omega_{\ell,n}(y^{(\ell)} - y^{(n)})\right)_{n \in \mathcal{N}_\ell}$
- $p \geq 1$ and $\eta > 0$.

Particular cases:

- $\ell_2 - TV$: $p = 2$, $\omega_{\ell,n} = 1$, and $\mathcal{N}_\ell$ horizontal and vertical neighbours,
- $\ell_\infty - TV$: $p = \infty$, $\omega_{\ell,n} = 1$, and $\mathcal{N}_\ell$ horizontal and vertical neighbours,
- $\ell_2 - NLTV$: $p = 2$, $\omega_{\ell,n}$ as in [Foi, Boracchi, 2012] and $\mathcal{N}_\ell$ as in [Gilboa, Osher, 2007],
- $\ell_\infty - NLTV$: $p = \infty$, $\omega_{\ell,n}$ as in [Foi, Boracchi, 2012] and $\mathcal{N}_\ell$ as in [Gilboa, Osher, 2007].

Image restoration with missing samples

$$\underset{y}{\operatorname{Argmin}} \|Ay - z\|^2 \quad \text{s.t.} \quad \begin{cases} \sum_{\ell=1}^{N} \|Y^{(\ell)}\|_p \leq \eta \\ y \in [0, 255]^N \end{cases}$$

$$\vdots$$

$$\underset{y,\zeta}{\operatorname{Argmin}} \|Ay - z\|^2 \ \text{s.t.} \ \begin{cases} (\forall \ell \in \{1, \ldots, N\}) \ \ \|Y^{(\ell)}\|_p \leq \zeta^{(\ell)} \\ \sum_{\ell=1}^{N} \zeta^{(\ell)} \leq \eta \\ y \in [0, 255]^N \end{cases}$$

Image restoration with missing samples

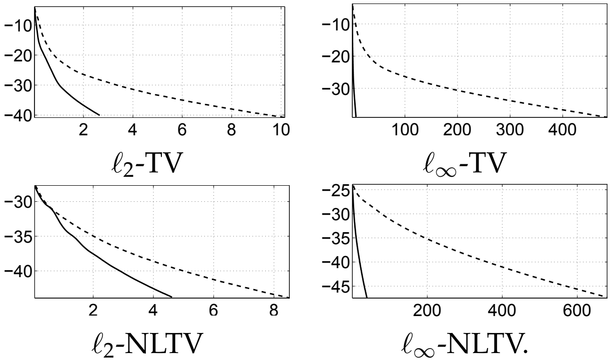
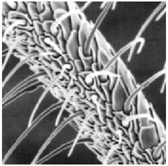
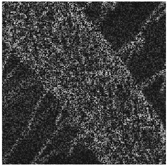
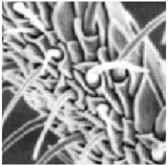


Figure: Comparison between epigraphical method (*solid line*) and direct method (*dashed line*): $\frac{\|y^{[n]}-y^{[\infty]}\|}{\|y^{[\infty]}\|}$ in dB vs time.

IMAGE RESTORATION WITH MISSING SAMPLES


Culicoidae

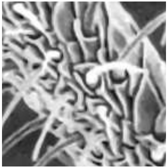

Degraded


Zoom


GPSR
SNR: 17.03 dB


ℓ_2-TV
SNR: 20.80 dB


ℓ_∞-TV
SNR: 20.25 dB


ℓ_2-NLTV
SNR: **22.62 dB**

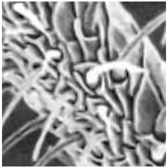

ℓ_∞-NLTV
SNR: 22.38 dB

Image restoration with missing samples

Culicoidae

Degraded

Zoom

GPSR
SNR: 20.26 dB

ℓ_2-TV
SNR: 23.18 dB

ℓ_∞-TV
SNR: 22.77 dB

ℓ_2-NLTV
SNR: **24.18 dB**

ℓ_∞-NLTV
SNR: 24.14 dB

Seismic data acquisition

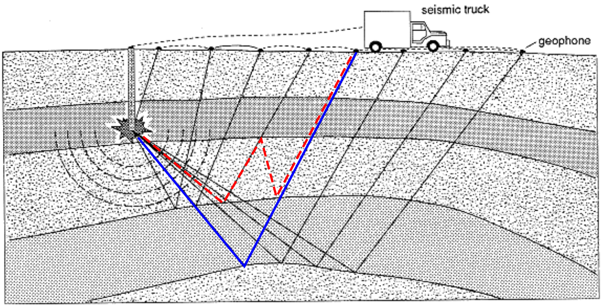


Figure: Principles of seismic wave propagation, with reflections on different layers, and data acquisition. Solid blue: primary; dashed red: multiple.

Observation model

$$\boxed{z^{(n)} = s^{(n)} + y^{(n)}}$$

where

- $n \in \{0, \cdots, N-1\}$: the time index
- $z = (z^{(n)})_{0 \le n < N}$: the observed data combining
  1. the primary $y = (y^{(n)})_{0 \le n < N}$ (signal of interest, unknown)
  2. the multiples $(s^{(n)})_{0 \le n < N}$ (sum of undesired reflected signals). We assume that a template $(r^{(n)})_{0 \le n < N}$ (for the disturbance signal) is available and that $s^{(n)} = \sum_{p=p'}^{p'+P-1} h^{(n)}(p) r^{(n-p)}$

We can rewrite the problem as

$$\boxed{z = Rh + y}$$

MAP ESTIMATION - FILTERS h

Assumptions:

1. $x = Fy$ (where $F \in \mathbb{R}^{N\times N}$ denotes the analysis operator) is a realization of a random vector, whose probability density function (pdf) is given by $(\forall x \in \mathbb{R}^N) \quad f_X(x) \propto \exp(-\varphi(x))$
2. h is a realization of a random vector, whose pdf is expressed as $(\forall h \in \mathbb{R}^{NP}) \quad f_H(h) \propto \exp(-\rho(h))$, and which is independent of x.

MAP estimation of h

$$\underset{h\in\mathbb{R}^{NP}}{\text{minimize}}\ \varphi\big(F(z - Rh)\big) + \rho(h).$$

- φ: data fidelity term taking into account the statistical properties of the basis coefficients
- ρ: prior informations that are available on h.

Convex constraints on the filters

Assumption: filters are varying along the time index n.

$$\big(\forall (n,p)\big) \qquad |h^{(n+1)}(p) - h^{(n)}(p)| \le \varepsilon_p$$

The associated closed convex set is defined as

$$C \quad = \quad \Big\{ h \in \mathbb{R}^{NP} \mid \forall (n,p) \;\; |h^{(n+1)}(p) - h^{(n)}(p)| \le \varepsilon_p \Big\}.$$

Convex constraints on the filters

Assumption: filters are varying along the time index n.

$$(\forall (n,p)) \qquad |h^{(n+1)}(p) - h^{(n)}(p)| \leq \varepsilon_p$$

The associated closed convex set is defined as

$$C \quad = \quad \left\{ h \in \mathbb{R}^{NP} \mid \forall (n,p) \;\; |h^{(n+1)}(p) - h^{(n)}(p)| \leq \varepsilon_p \right\}.$$

Minimization problem to be solved

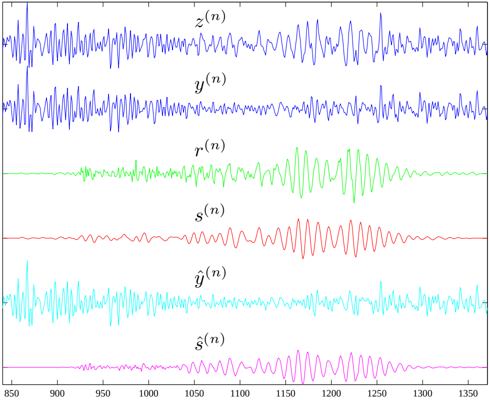
$$\underset{h \in \mathbb{R}^{NP}}{\text{minimize}} \; \varphi\big(F(z - Rh)\big) + \widetilde{\rho}(h) + \iota_{C_1}(h) + \iota_{C_2}(h).$$

Use of PPXA+ to perform the minimization.

Results: Context

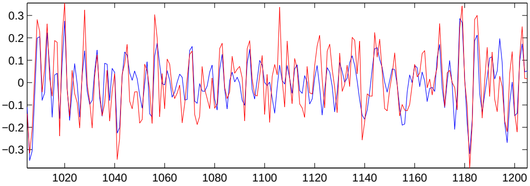
- $N = 2048$; filter length: $P = 14$ (noise-free case), $P = 10$ (noisy case)
- PPXA+ parameters: $\lambda_i \equiv 1.5$, $\omega_1 = 10000/N, \omega_2 = \omega_1/P, \omega_3 = \omega_4 = 10\,\omega_2$;
- Iteration number: 10000 (stopping criterion at iteration i if $\|h^{[i+1]} - h^{[i]}\| < 10^{-5}$);
- Functions choice: $\varphi_k \equiv |\cdot|$ and $\widetilde{\rho} = \mu\|\cdot\|^2$, $\mu = 0.01$;
- Basis choice: Symlet wavelets of length 8 over 3 resolution levels.

Results: Non noisy case

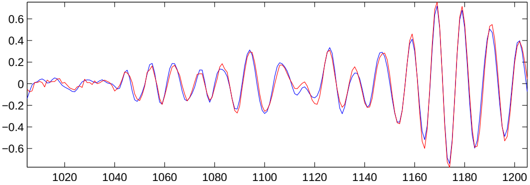


Observed signal z
Original signal y
Model r
Original multiple s
Estimated signal $\widehat{y}$
Estimated multiples $\widehat{s}$

Noisy case



Reference signal y and estimated signal $\widehat{y}$



Multiples s and estimated multiples $\widehat{s}$

CONCLUSION

- Proximity operators and proximal methods are shown to be very flexible tools for solving variational problems encountered in inverse problems.
- The convex criterion can be composed of various terms modelizing data fidelity (often linked to noise statictics) and also prior information, possibly formulated under convex (hard) constraints.
- Frames can be used to introduce prior information.
- Many other applications have been investigated (pMRI, compressive sensing, satellite imaging, stereovision, microcopy imaging,...).

Future work:

- Use of these methods in statistical learning.
- Extension to the non convex case.

Conclusion

- Proximity operators and proximal methods are shown to be very flexible tools for solving variational problems encountered in inverse problems.
- The convex criterion can be composed of various terms modelizing data fidelity (often linked to noise statictics) and also prior information, possibly formulated under convex (hard) constraints.
- Frames can be used to introduce prior information.
- Many other applications have been investigated (pMRI, compressive sensing, satellite imaging, stereovision, microcopy imaging,...).

Future work:

- Use of these methods in statistical learning.
- Extension to the non convex case.

Thank you !

SOME REFERENCES

L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” J. Optimization Theory and Applications, to appear, 2013.

G. Chierchia, N. Pustelnik, J.-C. Pesquet, B. Pesquet-Popescu, “Epigraphical Projection and Proximal Tools for Solving Constrained Convex Optimization Problems: Part I”, arXiv:1210.5844, 2012.

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.

N. Pustelnik, J.-C. Pesquet, C. Chaux, “Relaxing Tight Frame Condition in Parallel Proximal Methods for Signal Restoration,” IEEE Trans. on Sig. Proc., vol. 60 , no. 2, pp. 968 - 973, Feb. 2012.

H. H. Bauschke and P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer-Verlag, New York, 2011.

H. Raguet, J. Fadili, G. Peyré, “Generalized forward-backward splitting”, Preprint, arXiv:1108.4404, 2011.

B. C. Vũ, “A splitting algorithm for dual monotone inclusions involving cocoercive operators”, Advances in Computational Mathematics, Nov. 2011.

L. M. Briceño-Arias and P. L. Combettes, “A monotone+skew splitting model for composite monotone inclusions in duality,” SIAM Journal on Optimization, vol. 21, no. 4, pp. 1230-1250, Oct. 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.

P. L. Combettes and J.-C. Pesquet, “A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery,” IEEE Journal of Selected Topics in Signal Processing, vol. 1, no. 4, pp. 564-574, Dec. 2007.

C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, “A variational formulation for frame-based inverse problems,” Inverse Problems, vol. 23, no. 4, pp. 1495-1518, Aug. 2007.