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

\[ z = \bar{y} \]

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

\[ z = L \bar{y} \]

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

\[ z = \mathcal{D}_\alpha(L\bar{y}) \]

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

\[ z = D_\alpha(L\bar{y}) \]

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

Objective: inverse problem

Find an estimation \( \hat{y} \) of \( \bar{y} \) from observations \( z \).
Frame representation

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

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

[L. Jacques et al., 2011]
**Variational approach**

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

where \((f_j)_{1 \leq j \leq J}\): 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 \(j \in \{1, \ldots, J\}\), \(L_j: \mathcal{H} \to \mathcal{G}_j\) is a bounded linear operator (where \((\mathcal{G}_j)_{1 \leq j \leq J}\) denote Hilbert spaces).

This criterion can be non differentiable.
**Variational approach**

\[
\text{minimize}_{x \in \mathcal{H}} \quad \sum_{j=1}^{J} f_j(L_jx)
\]

where \((f_j)_{1 \leq j \leq J}\): 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 \(j \in \{1, \ldots, J\}\), \(L_j: \mathcal{H} \to \mathcal{G}_j\) is a bounded linear operator (where \((\mathcal{G}_j)_{1 \leq j \leq J}\) 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**

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

where \((f_j)_{1 \leq j \leq J}\): 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 \(j \in \{1, \ldots, J\}\), \(L_j: \mathcal{H} \to \mathcal{G}_j\) is a bounded linear operator (where \((\mathcal{G}_j)_{1 \leq j \leq J}\) 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)
VARATIONAL APPROACH

\[
\text{minimize}_{x \in \mathcal{H}} \quad \sum_{j=1}^{J} f_j(L_jx)
\]

where \((f_j)_{1 \leq j \leq J}\): 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 \(j \in \{1, \ldots, J\}\), \(L_j : \mathcal{H} \to \mathcal{G}_j\) is a bounded linear operator (where \((\mathcal{G}_j)_{1 \leq j \leq J}\) 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**

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

where \((f_j)_{1 \leq j \leq J}\): 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 \(j \in \{1, \ldots, J\}\), \(L_j: \mathcal{H} \rightarrow \mathcal{G}_j\) is a bounded linear operator (where \((\mathcal{G}_j)_{1 \leq j \leq J}\) 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**

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

where \((f_j)_{1 \leq j \leq J}\): 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 \(j \in \{1, \ldots, J\}\), \(L_j: \mathcal{H} \to \mathcal{G}_j\) is a bounded linear operator (where \((\mathcal{G}_j)_{1 \leq j \leq J}\) 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**

\[
\text{minimize}_{x \in \mathcal{H}} \sum_{j=1}^{J} f_j(L_jx)
\]

where \((f_j)_{1 \leq j \leq J}\): 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 \(j \in \{1, \ldots, J\}\), \(L_j: \mathcal{H} \rightarrow \mathcal{G}_j\) is a bounded linear operator (where \((\mathcal{G}_j)_{1 \leq j \leq J}\) 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):**

\[
\text{minimize } \sum_{r=1}^{R} f_r(L_rF^*x) + \sum_{s=1}^{S} g_s(x)
\]
ANALYSIS APPROACH VS. SYNTHESIS

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

Synthesis Form (SF):

\[
\begin{align*}
\text{minimize} & \quad \sum_{r=1}^{R} f_r(L_rF^*x) + \sum_{s=1}^{S} g_s(x) \\
& \quad \text{subject to:}
\end{align*}
\]

Analysis Form (AF):

\[
\begin{align*}
\text{minimize} & \quad \sum_{r=1}^{R} f_r(L_ry) + \sum_{s=1}^{S} g_s(Fy) \\
& \quad \text{subject to:}
\end{align*}
\]
**ANALYSIS APPROACH VS. SYNTHESIS**

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

**Synthesis Form (SF):**

\[
\begin{align*}
\text{minimize} & \quad \sum_{r=1}^{R} f_r(L_r F^* x) + \sum_{s=1}^{S} g_s(x) \\
\text{subject to} & \quad x \in \mathbb{R}^K
\end{align*}
\]

**Analysis Form (AF):**

\[
\begin{align*}
\text{minimize} & \quad \sum_{r=1}^{R} f_r(L_r y) + \sum_{s=1}^{S} g_s(F y) \\
\text{subject to} & \quad y \in \mathbb{R}^N
\end{align*}
\]

**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

$$
\text{prox}_\phi: \mathcal{H} \to \mathcal{H}: u \mapsto \arg \min_{v \in \mathcal{H}} \frac{1}{2} \|v - u\|^2 + \phi(v).
$$
The proximity operator of $\phi \in \Gamma_0(\mathcal{H})$ is defined as

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

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

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

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

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

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

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

- When $J > 2$: Parallel ProXimal Algorithm (PPXA) [Combettes and Pesquet, 2008]
PPXA+: minimize \[
\sum_{j=1}^{J} f_j(L_j u)
\]
\[
\text{Initialization}
\]
\[
\begin{align*}
\epsilon_j &\in [0,1]^J, \omega_j \in ]0, +\infty[^J, \\
\lambda_n &\in \mathbb{R} \quad \text{a sequence of reals}, \\
\mathbf{z}_j^{[0]} &\in (\mathcal{G}_j)^J, \mathbf{p}_j^{[-1]} \in (\mathcal{G}_j)^J, \\
u^{[0]} &= \arg\min_{u\in\mathcal{H}} \sum_{j=1}^{J} \omega_j \|L_j u - \mathbf{z}_j^{[0]}\|^2
\end{align*}
\]
\[
\text{For every } j \in \{1, \ldots, J\}, (a_j^{[n]})_{n\in\mathbb{N}} \text{ a sequence of reals,}
\]
\[
\text{For } n = 0, 1, \ldots
\]
\[
\begin{align*}
\text{For } j = 1, \ldots, J
\quad \mathbf{p}_j^{[n]} &= \text{prox}^{(1-\epsilon_j)f_j}_\omega ((1-\epsilon_j)\mathbf{z}_j^{[n]} + \epsilon_j \mathbf{p}_j^{[n-1]}) + a_j^{[n]} \\
c^{[n]} &= \arg\min_{u\in\mathcal{H}} \sum_{j=1}^{J} \omega_j \|L_j u - \mathbf{p}_j^{[n]}\|^2 \\
\text{For } j = 1, \ldots, J
\quad \mathbf{z}_j^{[n+1]} &= \mathbf{z}_j^{[n]} + \lambda_n (L_j (2c^{[n]} - u^{[n]}) - \mathbf{p}_j^{[n]}) \\
u^{[n+1]} &= u^{[n]} + \lambda_n (c^{[n]} - u^{[n]})
\end{align*}
\]
PPXA+: minimize $\sum_{j=1}^{J} f_j(L_j u)$

**Initialization**

\[
\begin{align*}
& (\epsilon_j)_{1 \leq j \leq J} \in [0, 1]^J, (\omega_j)_{1 \leq j \leq J} \in [0, +\infty]^J, \\
& (\lambda_n)_{n \in \mathbb{N}} \text{ a sequence of reals,} \\
& (z_j^{[0]})_{1 \leq j \leq J} \in (\mathcal{G}_j)^J, (p_j^{[-1]})_{1 \leq j \leq J} \in (\mathcal{G}_j)^J, \\
& u^{[0]} = \arg\min_{u \in \mathcal{H}} \sum_{j=1}^{J} \omega_j \|L_j u - z_j^{[0]}\|^2
\end{align*}
\]

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

**For** $n = 0, 1, \ldots$

**For** $j = 1, \ldots, J$

\[
\begin{align*}
& p_j^{[n]} = \operatorname{prox}_{\frac{(1-\epsilon_j)f_j}{\omega_j}}((1-\epsilon_j)z_j^{[n]} + \epsilon_j p_j^{[n-1]}) + a_j^{[n]} \\
& c^{[n]} = \arg\min_{u \in \mathcal{H}} \sum_{j=1}^{n} \omega_j \|L_j u - p_j^{[n]}\|^2
\end{align*}
\]

For every $j \in \{1, \ldots, J\}$

\[
\begin{align*}
& z_j^{[n+1]} = z_j^{[n]} + \lambda_n (L_j(2c^{[n]} - u^{[n]}) - p_j^{[n]}) \\
& u^{[n+1]} = u^{[n]} + \lambda_n (c^{[n]} - u^{[n]})
\end{align*}
\]
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. \(0 \in \text{sri} \{ (L_1 v - w_1, \ldots, L_J v - w_J) \mid v \in H, w_1 \in \text{dom } f_1, \ldots, w_J \in \text{dom } f_J \}\),
2. There exists \(\lambda \in ]0, 2[\) such that \((\forall n \in \mathbb{N})\), \(\lambda \leq \lambda_{n+1} \leq \lambda_n\),
3. For every \(j \in \{1, \ldots, J\}\), \(a_j^{[n]}\) are absolutely summable sequences in \(H\).
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 $\epsilon_1 = \cdots = \epsilon_J = 0$, $G_1 = \cdots = G_J = \mathcal{H}$, and $L_1 = \cdots = L_J = \text{Id}$.

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 $\epsilon_1 = \cdots = \epsilon_J = \frac{\alpha}{1 + \alpha}$, $(a_j^{[n]})_{1 \leq j \leq J} \equiv (0, \cdots, 0)$. 
Other proximal approaches: Minimize $\sum_{j=1}^{J} f_j(L_jx)$

- **Parallel ProXimal Algorithm + (PPXA+)** [Pesquet, Pustelnik, 2012]
  In the same spirit as PPXA, requires to compute each $\text{prox}_{f_j}$. 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 $\text{prox}_{f_j}$ 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 $\text{prox}_{f_j}$ for $j \neq j_0$. The algorithm stepsize is dependent on $\|L_j\|$.
  - **FB based algorithms** [Chambolle, Pock, 2011],[Vũ,2013],[Condat,2013]
CONSTRANDED FORMULATION

\[
\begin{align*}
\text{Minimize} & \quad \sum_{r=1}^{R} g_r(T_r x) \\
\text{s.t.} & \quad H_1 x \in C_1, \\
& \quad \vdots \\
& \quad H_S x \in C_S,
\end{align*}
\]

where

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

For \( n = 0, 1, \ldots \)

\[
    x[n] = \sum_{r=1}^{R} \omega_r u_r[n] + \sum_{s=1}^{S} \omega_s u_s[n]
\]

For \( r = 1, \ldots, R \)

\[
    w_1[r] = u_r[n] - \gamma_n T_r^* v_r[n] \\
    w_2[r] = v_r[n] + \gamma_n T_r u_r[n]
\]

For \( s = 1, \ldots, S \)

\[
    \overline{w}_1[s] = \overline{u}_s[n] - \gamma_n H_s^* \overline{v}_s[n] \\
    \overline{w}_2[s] = \overline{u}_s[n] + \gamma_n H_s \overline{v}_s[n]
\]

\[
    p_1[n] = \sum_{r=1}^{R} \omega_r w_1[n] + \sum_{s=1}^{S} \omega_s \overline{w}_1[n]
\]

For \( r = 1, \ldots, R \)

\[
    p_2[n] = w_2[r] - \frac{\gamma_n}{\omega_r} \operatorname{prox}_{\frac{\omega_r}{\gamma_n} g_r} \left( \frac{\omega_r w_1[n]}{\gamma_n} \right)
\]

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

Update \( u_1^{n+1} \) and \( v_1^{n+1} \)

For \( s = 1, \ldots, S \)

\[
    \overline{p}_2[s] = \overline{w}_2[s] - \frac{\gamma_n}{\omega_s} \Pi_{C_s} \left( \frac{\omega_s \overline{w}_1[n]}{\gamma_n} \right)
\]

\[
    q_{1,s}[n] = \overline{p}_{1,s}[n] - \gamma_n \left( H_s^* \overline{p}_{2,s}[n] \right) \\
    q_{2,s}[n] = \overline{p}_{2,s}[n] + \gamma_n \left( H_s \overline{p}_{1,s}[n] \right)
\]

Update \( \overline{u}_1^{n+1} \) and \( \overline{v}_1^{n+1} \)

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

Proximity operator computation

Projection computation
**Constrained Formulation**

\[(\forall x \in \mathcal{H}) \quad H_s x \in C_s \iff h_s(H_s x) \leq \eta_s\]
**Constrained Formulation**

\[
(\forall x \in \mathcal{H}) \quad H_s x \in C_s \iff h_s(H_s x) \leq \eta_s
\]

\[
\vdots
\]

\[
(\forall u \in \mathbb{R}^Q) \quad u \in C \iff h(u) \leq \eta
\]
CONSTRANGED FORMULATION

\[(\forall x \in \mathcal{H}) \quad H_s x \in C_s \iff h_s(H_s x) \leq \eta_s\]

\[\vdots\]

\[(\forall u \in \mathbb{R}^Q) \quad u \in C \iff h(u) \leq \eta\]

\[\vdots\]

\[(\forall u = [(u^{(1)})^\top, \ldots, (u^{(L)})^\top]^\top \in \mathbb{R}^Q) \quad u \in C \iff \sum_{\ell=1}^{L} h^{(\ell)}(u^{(\ell)}) \leq \eta\]
**Constrained Formulation**

\[
\forall x \in \mathcal{H} \quad H_s x \in C_s \iff h_s(H_s x) \leq \eta_s
\]

\[
\vdots
\]

\[
\forall u \in \mathbb{R}^Q \quad u \in C \iff h(u) \leq \eta
\]

\[
\vdots
\]

\[
\forall u = [(u^{(1)})^\top, \ldots, (u^{(L)})^\top]^\top \in \mathbb{R}^Q \quad u \in C \iff \sum_{\ell=1}^{L} h^{(\ell)}(u^{(\ell)}) \leq \eta
\]

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

\[(\forall x \in \mathcal{H}) \quad H_s x \in C_s \iff h_s(H_s x) \leq \eta_s\]

\[
\vdots
\]

\[(\forall u \in \mathbb{R}^Q): u \in C \iff h(u) \leq \eta\]

\[
\vdots
\]

\[(\forall u = [(u^{(1)})^\top, \ldots, (u^{(L)})^\top]^\top \in \mathbb{R}^Q): u \in C \iff \sum_{\ell=1}^{L} h^{(\ell)}(u^{(\ell)}) \leq \eta\]

→ 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 $\Pi_C$ does not have a closed form?
**EpiGraphical Projection**

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

\[
\begin{bmatrix}
\text{size } Q^{(1)} \\
\text{size } Q^{(L)}
\end{bmatrix}
\]

\[u \in C \iff \sum_{\ell=1}^{L} h^{(\ell)}(u^{(\ell)}) \leq \eta.
\]

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

\[u \in C \iff \left\{ \begin{array}{ll}
\sum_{\ell=1}^{L} \zeta^{(\ell)} \leq \eta, \\
(\forall \ell \in \{1, \ldots, L\}) h^{(\ell)}(u^{(\ell)}) \leq \zeta^{(\ell)}.
\end{array} \right\}
\]
**EPIGRAPHICAL PROJECTION**

\[ u \in C \iff \begin{cases} \zeta \in V \\ (u, \zeta) \in E \end{cases} \]

where

- \( V \) denotes a closed half-space such that:

\[ V = \{ \zeta \in \mathbb{R}^L \mid 1_L^T \zeta \leq \eta \} \]

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

\[ E = \{ (u, \zeta) \in \mathbb{R}^Q \times \mathbb{R}^L \mid (\forall \ell \in \{1, \ldots, L\}) (u^{(\ell)}, \zeta^{(\ell)}) \in \text{epi } h^{(\ell)} \} \]
EPIGRAPHICAL PROJECTION

\[ u \in C \Leftrightarrow \begin{cases} \zeta \in V \\ (u, \zeta) \in E \end{cases} \]

where

- \( V \) denotes a closed half-space such that:

\[ V = \{ \zeta \in \mathbb{R}^L \mid 1_L^\top \zeta \leq \eta \} \]

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

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

\[ E = \{ (u, \zeta) \in \mathbb{R}^Q \times \mathbb{R}^L \mid (\forall \ell \in \{1, \ldots, L\}) (u^{(\ell)}, \zeta^{(\ell)}) \in \text{epi} h^{(\ell)} \} \]

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

- **Euclidean norm** functions defined as:

\[
\left( \forall \ell \in \{1, \ldots, L\} \right) \left( \forall u^{(\ell)} \in \mathbb{R}^{Q^{(\ell)}} \right) \quad h^{(\ell)}(u^{(\ell)}) = \tau^{(\ell)} \|u^{(\ell)}\|
\]

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

- **Euclidean norm** functions defined as:

\[
(\forall \ell \in \{1, \ldots, L\}) (\forall u^{(\ell)} \in \mathbb{R}^{Q^{(\ell)}}) \quad h^{(\ell)}(u^{(\ell)}) = \tau^{(\ell)} \|u^{(\ell)}\|
\]

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

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

\[
\Pi_{\text{epi}} h^{(\ell)} (u^{(\ell)}, \zeta^{(\ell)}) = \begin{cases} 
(u^{(\ell)}, \zeta^{(\ell)}), & \text{if } \|u^{(\ell)}\| < \frac{\zeta^{(\ell)}}{\tau^{(\ell)}}, \\
(0, 0), & \text{if } \|u^{(\ell)}\| < -\tau^{(\ell)} \zeta^{(\ell)}, \\
\alpha^{(\ell)} (u^{(\ell)}, \tau^{(\ell)} \|u^{(\ell)}\|), & \text{otherwise,}
\end{cases}
\]

where \(\alpha^{(\ell)} = \frac{1}{1 + (\tau^{(\ell)})^2} \left(1 + \frac{\tau^{(\ell)} \zeta^{(\ell)}}{\|u^{(\ell)}\|}\right).\)
EPIGRAPHICAL PROJECTION

- **Infinity norms** defined as:

  \[(\forall \ell \in \{1, \ldots, L\}) \left( \forall u^{(\ell)} = (u^{(\ell,m)})_{1 \leq m \leq Q^{(\ell)}} \in \mathbb{R}^{Q^{(\ell)}} \right) \]

  \[
h^{(\ell)}(u^{(\ell)}) = \max \left\{ \frac{|u^{(\ell,m)}|}{\tau^{(\ell,m)}} | 1 \leq m \leq Q^{(\ell)} \right\}
\]

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

\[\Pi_{epi} h^{(\ell)}(u^{(\ell)}, \zeta^{(\ell)})\] has a closed form [G. Chierchia et al., 2012].
Reconstruction problem: PET

- High level of noise
- Large amount of data
RECONSTRUCTION PROBLEM

\[ z = \mathcal{P}_\alpha (A\bar{y}) \]

where

- \( \mathcal{P}_\alpha \): Poisson noise of scale parameter \( \alpha \)
- \( A \): projection matrix
RECONSTRUCTION PROBLEM

Our objective is:

\[
\min_{x \in \mathbb{R}^K} \sum_{t=1}^{T} D_{KL}(AF_t^* x, z) + \kappa \text{tv}(F_t^* x) + \iota C(x) + \vartheta \|x\|_{\ell_1}
\]

where \(\kappa > 0, \vartheta > 0\) and

- \(D_{KL}\) is the Kullback-Leibler divergence
- \(\text{tv}\) represents a total variation term
- \(\iota C\) is the indicator function of a closed convex set \(C\)
- \(\|x\|_{\ell_1}\) denotes the \(\ell_1\)-norm.
**RECONSTRUCTION PROBLEM**

Our objective is:

\[
\min_{x \in \mathbb{R}^K} \sum_{t=1}^{T} D_{KL}(A F_t^* x, z) + \kappa \text{tv}(F_t^* x) + \iota_C(x) + \vartheta \|x\|_{\ell_1}
\]

\[
y = F^* x = (F_t^* x)_{1 \leq t \leq T}
\]

where \(\kappa > 0, \vartheta > 0\) and

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

Our objective is:

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

\[
y = F^*x = (F_t^*x)_{1 \leq t \leq T}
\]

where \(\kappa > 0, \vartheta > 0\) and

- \(D_{KL}\) is the Kullback-Leibler divergence
- \(\text{tv}\) represents a total variation term
- \(\iota_C\) is the indicator function of a closed convex set \(C\)
- \(\|x\|_{\ell_1}\) denotes the \(\ell_1\)-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: $\bar{y} \in \mathbb{R}^N$  
Degraded: $z \in \mathbb{R}^M$

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

- $\bar{y}$: original image in $[0, 255]^N$
  - 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**

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

where

- \( Y^{(\ell)} = (\omega_{\ell,n}(y^{(\ell)} - y^{(n)}))_{n \in \mathcal{N}_\ell} \)
- \( p \geq 1 \) 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**

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

\[
\begin{cases}
\vdots
\end{cases}
\]

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

![Graphs showing comparison between epigraphical method (solid line) and direct method (dashed line): $\|y[n]-y[\infty]\|/\|y[\infty]\|$ in dB vs time.](image)

Figure: Comparison between epigraphical method (*solid line*) and direct method (*dashed line*): $\|y[n]-y[\infty]\|/\|y[\infty]\|$ in dB vs time.
IMAGE RESTORATION WITH MISSING SAMPLES

Culicoidae

Degraded

Zoom

GPSR
SNR: 17.03 dB

$\ell_2$-TV
SNR: 20.80 dB

$\ell_\infty$-TV
SNR: 20.25 dB

$\ell_2$-NLTV
SNR: 22.62 dB

$\ell_\infty$-NLTV
SNR: 22.38 dB
IMAGE RESTORATION WITH MISSING SAMPLES

Culicoidae

Degraded

Zoom

GPSR
SNR: 20.26 dB

$\ell_2$-TV
SNR: 23.18 dB

$\ell_\infty$-TV
SNR: 22.77 dB

$\ell_2$-NLTV
SNR: 24.18 dB

$\ell_\infty$-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

\[
  z(n) = s(n) + y(n)
\]

where

- \( n \in \{0, \cdots, N - 1\} \): the time index

- \( z = (z(n))_{0 \leq n < N} \): the observed data combining
  1. the primary \( y = (y(n))_{0 \leq n < N} \) (signal of interest, unknown)
  2. the multiples \( s(n) = (s(n))_{0 \leq n < N} \) (sum of undesired reflected signals). We assume that a template \( (r(n))_{0 \leq n < N} \) (for the disturbance signal) is available and that

\[
  s(n) = \sum_{p=0}^{p' + P - 1} h(n)(p)r(n-p)
\]

We can rewrite the problem as

\[
  z = Rh + y
\]
**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 \((\forall x \in \mathbb{R}^N) \ f_X(x) \propto \exp(-\varphi(x))\)

2. \( h \) is a realization of a random vector, whose pdf is expressed as \((\forall h \in \mathbb{R}^{NP}) \ f_H(h) \propto \exp(-\rho(h))\), and which is independent of \( x \).

**MAP estimation of \( h \)**

\[
\text{minimize} \quad \varphi(F(z - Rh)) + \rho(h).
\]

- \( \varphi \): *data fidelity* term taking into account the statistical properties of the basis coefficients
- \( \rho \): *prior* informations that are available on \( h \).
Convex Constraints on the Filters

**Assumption**: filters are varying along the time index $n$.

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

The associated closed convex set is defined as

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

**Assumption:** filters are varying along the time index $n$.

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

The associated closed convex set is defined as

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

Minimization problem to be solved

$$\min_{h \in \mathbb{R}^{NP}} \varphi(F(z - Rh)) + \tilde{\rho}(h) + \iota_{C_1}(h) + \iota_{C_2}(h).$$

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 $\tilde{\rho} = \mu \| \cdot \|^2,$ $\mu = 0.01$;
- 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 $\hat{y}$
Estimated multiples $\hat{s}$
**Noisy Case**

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

Multiples $s$ and estimated multiples $\hat{s}$
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 modeling data fidelity (often linked to noise statistics) 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, microscopy 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 statistics) 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, microscopy imaging,...).

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

Thank you!
SOME REFERENCES


