Some Aspects of the Interplay Between Convex Analysis and Monotone Operator Theory

Patrick L. Combettes

Laboratoire Jacques-Louis Lions
Faculté de Mathématiques
Université Pierre et Marie Curie – Paris 6
75005 Paris, France

Sestri Levante, 12 settembre 2014
Our objective is to discuss certain aspects of the importance of the theory of monotone operators and of non-expansive operators in the analysis and the numerical solution of problems in inverse problems and learning theory, even when those admit a purely variational formulation.

Special emphasis is placed on the role played by duality.
Duality in Hilbert spaces

Duality-closed classes of objects:

- Closed vector subspaces: $V \rightarrow V^\perp$ (Fréchet?)

- Closed convex cones: $K \rightarrow K^\ominus$ (Fenchel; $\uparrow$: $K = \partial f$)

- Lower semicontinuous proper convex functions: $f \rightarrow f^\ast$ (Fenchel, Moreau; $\uparrow$: $f = \iota_K$)

- Maximally monotone operators: $A \rightarrow A^\ominus$ (Moreau; $\uparrow$: $A = \partial f$)

- Firmly nonexpansive operators: $J \rightarrow \text{Id} - J^\uparrow$ (Minty; $\leftrightarrow J = (\text{Id} + A)^\ominus$)

- Nonexpansive operators: $R \rightarrow -R$ ($\leftrightarrow R = 2J - \text{Id}$)
Duality in Hilbert spaces

Duality-closed classes of objects:

- Closed vector subspaces: $V \to V^\perp$ (Fréchet?)
- Closed convex cones: $K \to K^\circ$ (Fenchel; $\uparrow$: $K = V$)
Duality in Hilbert spaces

Duality-closed classes of objects:
- Closed vector subspaces: \( V \to V^\perp \) (Fréchet?)
- Closed convex cones: \( K \to K^\ominus \) (Fenchel; ↑: \( K = V \))
- Lower semicontinuous proper convex functions: \( f \to f^* \) (Fenchel, Moreau; ↑: \( f = \nu_K \))
Duality in Hilbert spaces

Duality-closed classes of objects:

- Closed vector subspaces: $V \rightarrow V^\perp$ (Fréchet?)
- Closed convex cones: $K \rightarrow K^\ominus$ (Fenchel; $\uparrow$: $K = V$)
- Lower semicontinuous proper convex functions: $f \rightarrow f^*$ (Fenchel, Moreau; $\uparrow$: $f = \iota_K$)
- Maximally monotone operators: $A \rightarrow A^{-1}$ (Moreau; $\uparrow$: $A = \partial f$)
Duality in Hilbert spaces

Duality-closed classes of objects:

- Closed vector subspaces: $V \rightarrow V^\perp$ (Fréchet?)
- Closed convex cones: $K \rightarrow K^\circ$ (Fenchel; $\uparrow$: $K = V$)
- Lower semicontinuous proper convex functions: $f \rightarrow f^*$ (Fenchel, Moreau; $\uparrow$: $f = \iota_K$)
- Maximally monotone operators: $A \rightarrow A^{-1}$ (Moreau; $\uparrow$: $A = \partial f$)
- Firmly nonexpansive operators: $J \rightarrow \text{Id} - J$ $\uparrow$: (Minty; $\downarrow$: $J = (\text{Id} + A)^{-1}$)
Duality in Hilbert spaces

Duality-closed classes of objects:

- Closed vector subspaces: $V ightarrow V^\perp$ (Fréchet?)
- Closed convex cones: $K ightarrow K^\Theta$ (Fenchel; $\uparrow$: $K = V$)
- Lower semicontinuous proper convex functions:
  $f \rightarrow f^*$ (Fenchel, Moreau; $\uparrow$: $f = \iota_K$)
- Maximally monotone operators: $A \rightarrow A^{-1}$ (Moreau; $\uparrow$: $A = \partial f$)
- Firmly nonexpansive operators: $J \rightarrow \text{Id} - J$ (Minty; $\downarrow$: $J = (\text{Id} + A)^{-1}$)
- Nonexpansive operators: $R \rightarrow -R$ ($\uparrow$: $R = 2J - \text{Id}$)
Problems in duality

- Functional setting (Fenchel, Moreau): $f + g \rightarrow f^\vee + g^*$

- Composite functional setting (Rockafellar): $f + g \circ L \rightarrow f^\vee \circ L^* + g^*$

- Maximally monotone operator setting (Mosco 1972 (for V.I.), Mercier 1980, Attouch-Théra 1996):
  
  $A + B \leftrightarrow -A^{-1} \circ (-\text{Id}) + B^{-1}$.

- Composite maximally monotone operator setting (Robinson 1999, Pennanen 2000):
  
  $A + L^* B L \leftrightarrow -L A^{-1} (L^*)^{-1} + B^{-1}$.

Open problem 1: (Firmly) nonexpansive operator setting?
Problems in duality

- Functional setting (Fenchel, Moreau): $f + g \to f^\vee + g^*$
- Composite functional setting (Rockafellar):
  \[
  f + g \circ L \to f^\vee \circ L^* + g^*
  \]
Problems in duality

- Functional setting (Fenchel, Moreau): \( f + g \rightarrow f^\vee + g^* \)
- Composite functional setting (Rockafellar):

\[
f + g \circ L \rightarrow f^\vee \circ L^* + g^*
\]

- Maximally monotone operator setting (Mosco 1972 (for V.I.), Mercier 1980, Attouch-Théra 1996):

\[
A + B \leftrightarrow -A^{-1} \circ (-\text{Id}) + B^{-1}.
\]
Problems in duality

- Functional setting (Fenchel, Moreau): \( f + g \rightarrow f^\vee + g^* \)
- Composite functional setting (Rockafellar):

\[
f + g \circ L \rightarrow f^\vee \circ L^* + g^*
\]

- Maximally monotone operator setting (Mosco 1972 (for V.I.), Mercier 1980, Attouch-Théra 1996):

\[
A + B \leftrightarrow -A^{-1} \circ (-\text{Id}) + B^{-1}.
\]

- Composite maximally monotone operator setting (Robinson 1999, Pennanen 2000):

\[
A + L^*BL \leftrightarrow -LA^{-1}(-L^*) + B^{-1}
\]
Problems in duality

- Functional setting (Fenchel, Moreau): \( f + g \rightarrow f^* + g^* \)

- Composite functional setting (Rockafellar):
  \[
  f + g \circ L \rightarrow f^* \circ L^* + g^*
  \]

- Maximally monotone operator setting (Mosco 1972 (for V.I.), Mercier 1980, Attouch-Théra 1996):
  \[
  A + B \leftrightarrow -A^{-1} \circ (-\text{Id}) + B^{-1}
  \]

- Composite maximally monotone operator setting (Robinson 1999, Pennanen 2000):
  \[
  A + L^*BL \leftrightarrow -LA^{-1}(-L^*) + B^{-1}
  \]

- **Open problem 1**: (Firmly) nonexpansive operator setting?
Open problem 1: Duality for nonexpansive operators?

Consider the inclusion $0 \in Ax + Bx$, call $\mathcal{P}$ its solution set and $\mathcal{D}$ the dual solution set. Then:

- $\mathcal{P} = \{ x \in \mathcal{H} \mid (\exists u \in \mathcal{D}) - u \in Ax \text{ and } u \in Bx \}$ and $\mathcal{D} = \{ u \in \mathcal{H} \mid (\exists x \in \mathcal{P}) x \in A^{-1}(-u) \text{ and } x \in B^{-1}u \}$
- $\mathcal{P} = \text{dom} (A \cap (-B))$ and $\mathcal{D} = \text{dom} (A^{-1} \circ (-\text{Id}) \cap B^{-1})$
- Suppose that $A$ and $B$ are maximally monotone, set $T_{A,B} = J_A(2J_B - \text{Id}) + \text{Id} - J_B$, and let $\gamma > 0$. Then:
  - $T_{A,B}$ is firmly nonexpansive
  - $\mathcal{P} = J_{\gamma B}(\text{Fix } T_{\gamma A,\gamma B})$
  - $\mathcal{D} = \gamma B(\text{Fix } T_{\gamma A,\gamma B})$
  - $\mathcal{P} \neq \emptyset \iff \mathcal{D} \neq \emptyset \iff \text{Fix } T_{\gamma A,\gamma B} \neq \emptyset$
Suppose that $0 \in \text{sri}(\text{dom } f - \text{dom } g)$. Then

$$\inf (f + g)(\mathcal{H}) = - \min (f^* + g^*)(\mathcal{H})$$
Open problem 2: True Fenchel duality for monotone operators?

- Suppose that \( 0 \in \text{sri} (\text{dom } f - \text{dom } g) \). Then

\[
\inf (f + g)(\mathcal{H}) = - \min (f^* + g^*)(\mathcal{H})
\]

- We always have

\[
\text{zer}(A + B) \neq \emptyset \iff \text{zer}(-A^{-1} \circ (-\text{Id}) + B^{-1}) \neq \emptyset.
\]
Open problem 2: True Fenchel duality for monotone operators?

- Suppose that $0 \in \text{sri}(\text{dom } f - \text{dom } g)$. Then

$$\inf (f + g)(\mathcal{H}) = - \min (f^\vee + g^\vee)(\mathcal{H})$$

- We always have

$$\text{zer}(A + B) \neq \emptyset \iff \text{zer}(-A^{-1} \circ (-\text{Id}) + B^{-1}) \neq \emptyset.$$ 

- For suitable “closure” or “enlargement” operations, can

$$\text{zer} (A + B)^{\sharp} \neq \emptyset \implies \text{zer} (-A^{-1} \circ (-\text{Id}) + B^{-1})^{\flat} \neq \emptyset$$

recover Fenchel duality?
Splitting methods: Quick overview

Traditional splitting techniques were developed in the late 1970s for inclusions involving the sum of two maximally monotone operators:

$$0 \in Ax + Bx$$

Open problem 3: The continuous-time dynamics leading to FB are reasonably understood ($-x' \in Ax + Bx$); what about FBF and DR?

Main algorithms:
- Forward-backward method (Mercier, 1979)
- Douglas-Rachford method (Lions and Mercier, 1979)
- Forward-backward-forward-forward algorithm (Tseng, 2000).

Until recently, general splitting methods lacked for the composite problem

$$0 \in Ax + L^*BLx$$
Splitting methods: Composite problems

- \( L: \mathcal{H} \to \mathcal{G} \) linear and bounded, \( A: \mathcal{H} \to 2^\mathcal{H} \), \( B: \mathcal{G} \to 2^\mathcal{G} \) maximally monotone

- Solve
  \[ 0 \in Ax + L^* BLx \]

- Main issue: 3 objects \((A, B, L)\) to split... and a binary relation \(\in\) binding them

- We need to reduce a problem to a 2-object problem in a larger space

- Key: recast the problem in the primal dual space
Kuhn-Tucker set of a composite inclusion

- Primal solutions: \( \mathcal{P} = \{ x \in \mathcal{H} \mid 0 \in Ax + L^*BLx \} \)

- Dual solutions: \( \mathcal{D} = \{ v^* \in \mathcal{G} \mid 0 \in -L \circ A^{-1}(-L^*v^*) + B^{-1}v^* \} \)
Kuhn-Tucker set of a composite inclusion

- Primal solutions: \( \mathcal{P} = \{ x \in \mathcal{H} \mid 0 \in Ax + L^*BLx \} \)
- Dual solutions: \( \mathcal{D} = \{ v^* \in \mathcal{G} \mid 0 \in -L \circ A^{-1}(-L^*v^*) + B^{-1}v^* \} \)
- Kuhn-Tucker set

\[
\mathcal{Z} = \{(x, v^*) \in \mathcal{H} \oplus \mathcal{G} \mid -L^*v^* \in Ax \text{ and } Lx \in B^{-1}v^* \}
\]
Kuhn-Tucker set of a composite inclusion

- Primal solutions: \( \mathcal{P} = \{ x \in \mathcal{H} \mid 0 \in Ax + L^*BLx \} \)
- Dual solutions: \( \mathcal{D} = \{ v^* \in \mathcal{G} \mid 0 \in -L \circ A^{-1}(-L^*v^*) + B^{-1}v^* \} \)
- Kuhn-Tucker set

\[
\mathcal{Z} = \{(x, v^*) \in \mathcal{H} \oplus \mathcal{G} \mid -L^*v^* \in Ax \text{ and } Lx \in B^{-1}v^* \}
\]

- \( \mathcal{Z} \) is a closed convex set
- \( \mathcal{Z} \neq \emptyset \iff \mathcal{P} \neq \emptyset \iff \mathcal{D} \neq \emptyset \)
- \( \mathcal{Z} \subset \mathcal{P} \times \mathcal{D} \)
Kuhn-Tucker set of a composite inclusion

- Strategy: find a Kuhn-Tucker pair \((x, v^*)\) by applying a standard (e.g., Douglas-Rachford or forward-backward-forward) method to a monotone+skew decomposition of the problem in \(\mathcal{H} \oplus \mathcal{G}\).

\[
\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} A & 0 \\ 0 & B^{-1} \end{bmatrix} \begin{bmatrix} x \\ v^* \end{bmatrix} + \begin{bmatrix} 0 & L^* \\ -L & 0 \end{bmatrix} \begin{bmatrix} x \\ v^* \end{bmatrix}
\]

Kuhn-Tucker set of a composite inclusion

- Strategy: find a Kuhn-Tucker pair \((x, v^*)\) by applying a standard (e.g., Douglas-Rachford or forward-backward-forward) method to a monotone+skew decomposition of the problem in \(\mathcal{H} \oplus \mathcal{G}\).

\[
\begin{bmatrix}
0 \\
0
\end{bmatrix} \in \begin{bmatrix}
A & 0 \\
0 & B^{-1}
\end{bmatrix} \begin{bmatrix}
x \\
v^*
\end{bmatrix} + \begin{bmatrix}
0 & L^* \\
-L & 0
\end{bmatrix} \begin{bmatrix}
x \\
v^*
\end{bmatrix}
\]


- Even in minimization problems, such a framework cannot be reduced to a functional setting: monotone operator splitting is required!

- Possible limitation: linear inversions (in DR) or knowledge of \(\|L\|\) (in FBF) are necessary.
Revisiting the proximal point algorithm

- A maximally monotone, \((\gamma_n)_{n \in \mathbb{N}} \in ]0, +\infty[^{\mathbb{N}}, \text{zer } A \neq \emptyset, \]

\[x_{n+1} = J_{\gamma_n A} x_n\]

Classical results by Brézis&Lions (1978):
- If \(\sum_{n \in \mathbb{N}} \gamma_n^2 = +\infty\), then \(x_n \rightharpoonup x \in \text{zer } A\)
- If \(A = \partial f (f \in \Gamma_0(\mathcal{H}))\) and \(\sum_{n \in \mathbb{N}} \gamma_n = +\infty\), then \(x_n \rightharpoonup x \in \text{zer } A = \text{Argmin } f\)

- Is the proximal point algorithm of any use?
Revisiting the proximal point algorithm

- $A : \mathcal{H} \to 2^\mathcal{H}$ maximally monotone, $V$ be a closed vector subspace of $\mathcal{H}$
- $A_V$: partial inverse of $A$ w.r.t. $V$ (Spingarn, 1983)

\[
\text{gra } A_V = \{(P_V x + P_{V^\perp} u, P_V u + P_{V^\perp} x) \mid (x, u) \in \text{gra } A\}
\]

Then (Spingarn, 1983):
- $A_V$ is maximally monotone
- $z \in \text{zer } A_V \iff (P_V z, P_{V^\perp} z) \in \text{gra } A$
- $p = J_{A_V} z \iff P_V p + P_{V^\perp} (z - p) = J_A z$

Aside – Open problem 4:
Let $f \in \Gamma_0(H)$ and $V = \{0\}$. Then $(\partial f)_V = \partial f^*$, but how to define a partial conjugate of $f$? (in general $(\partial f)_V$ is not a subdifferential)
Revisiting the proximal point algorithm

- \( \mathbf{A} : \mathcal{H} \rightarrow 2^\mathcal{H} \) maximally monotone, \( \mathbf{V} \) be a closed vector subspace of \( \mathcal{H} \)

- \( \mathbf{A}_\mathbf{V} \): partial inverse of \( \mathbf{A} \) w.r.t. \( \mathbf{V} \) (Spingarn, 1983)

\[
\text{gra} \mathbf{A}_\mathbf{V} = \{(P_\mathbf{V}x + P_\mathbf{V}^\perp u, P_\mathbf{V}u + P_\mathbf{V}^\perp x) \mid (x, u) \in \text{gra} \mathbf{A}\}
\]

Then (Spingarn, 1983):
- \( \mathbf{A}_\mathbf{V} \) is maximally monotone
- \( z \in \text{zer} \mathbf{A}_\mathbf{V} \iff (P_\mathbf{V}z, P_\mathbf{V}^\perp z) \in \text{gra} \mathbf{A} \)
- \( p = J_{\mathbf{A}_\mathbf{V}}z \iff P_\mathbf{V}p + P_\mathbf{V}^\perp(z - p) = J_{\mathbf{A}}z \)

Aside – **Open problem 4**: Let \( f \in \Gamma_0(\mathcal{H}) \) and \( \mathbf{V} = \{0\} \). Then \((\partial f)_\mathbf{V} = \partial f^*\), but how to define a partial conjugate of \( f \)? (in general \((\partial f)_\mathbf{V} \) is not a subdifferential)
Back to the Kuhn-Tucker set

\[ Z = \{(x, v^*) \in \mathcal{H} \oplus \mathcal{G} \mid -L^*v^* \in Ax \text{ and } Lx \in B^{-1}v^*\} \]

Define \( \mathcal{H} = \mathcal{H} \oplus \mathcal{G} \), \( A: (x, y) \mapsto Ax \times By \), and \( V = \{(x, y) \in \mathcal{H} \oplus \mathcal{G} \mid Lx = y\} \)

Apply the (monotone operator) proximal point algorithm to \( A_V \)

Revisiting the proximal point algorithm

**Theorem**

Set $Q = (Id + L^*L)^{-1}$ and assume that $\text{zer}(A + L^*BL) \neq \emptyset$. Let $(\lambda_n)_{n \in \mathbb{N}} \in ]0,2[\mathbb{N}$, such that $\sum_{n \in \mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, let $x_0 \in \mathcal{H}$, $v_0^* \in \mathcal{G}$, and set $y_0 = Lx_0$, $u_0 = -L^*v_0^*$, and

$$(\forall n \in \mathbb{N})$$

\[
\begin{align*}
\rho_n &\approx J_A(x_n + u_n) \\
q_n &\approx J_B(y_n + v_n^*) \\
r_n &= x_n + u_n - \rho_n \\
s_n &= y_n + v_n - q_n \\
t_n &= Q(r_n + L^*s_n) \\
w_n &= Q(p_n + L^*q_n) \\
x_{n+1} &= x_n - \lambda_n t_n \\
y_{n+1} &= y_n - \lambda_n Lt_n \\
u_{n+1} &= u_n + \lambda_n(w_n - \rho_n) \\
v_{n+1}^* &= v_{n+1}^* + \lambda_n(Lw_n - q_n).
\end{align*}
\]

Then $x_n - w_n \rightarrow 0$, $y_n - Lw_n \rightarrow 0$, $u_n - r_n + t_n \rightarrow 0$, and $v_{n+1}^* - s_n + Lt_n \rightarrow 0$, and $(x_n, v_{n+1}^*) \rightarrow (\bar{x}, \bar{v}^*) \in K$. 

Patrick L. Combettes
A strongly convergent splitting method

- We have seen that the Kuhn-Tucker set
  \[ Z = \{(x, v^*) \in \mathcal{H} \oplus \mathcal{G} \mid -L^*v^* \in Ax \text{ and } Lx \in B^{-1}v^*\} \]
  is a closed and convex

- The previous methods require efficient linear inversions schemes or knowledge of \( \|L\| \), neither of which may be available in many situations (e.g., domain decomposition methods). In addition they require stringent additional hypotheses to guarantee strong convergence (to an unspecified Kuhn-Tucker point)

- We present a method which alleviates all these limitations.
We have seen that the Kuhn-Tucker set

\[ Z = \{ (x, v^*) \in H \oplus G \mid -L^*v^* \in Ax \text{ and } Lx \in B^{-1}v^* \} \]

is a closed and convex

The previous methods require efficient linear inversions schemes or knowledge of \( \|L\| \), neither of which may be available in many situations (e.g., domain decomposition methods). In addition they require stringent additional hypotheses to guarantee strong convergence (to an unspecified Kuhn-Tucker point).

We present a method which alleviates all these limitations.

Aside – **Open problem 5**: in infinite-dimensional problems is weak convergence relevant?
A strongly convergent splitting method

An abstract Haugazeau scheme:
Let $C$ be a nonempty closed convex subset of $\mathcal{H}$ and let $x_0 \in \mathcal{H}$. Iterate

$$x_{n+1} = P_{H(x_0, x_n) \cap H(x_n, x_{n+1/2})} x_0$$

for $n = 0, 1, \ldots$

Suppose that, for every $x \in \mathcal{H}$ and every strictly increasing sequence $(k_n)_{n \in \mathbb{N}}$ in $\mathbb{N}$, $x_{k_n} \rightharpoonup x \Rightarrow x \in C$. Then $(x_n)_{n \in \mathbb{N}}$ is well defined and $x_n \to P_C x_0$.

Strategy: Apply this principle to $C = Z$, construct $x_{n+1/2}$ by suitably choosing points in gra $A$ and gra $B$

A strongly convergent splitting method

**Theorem**

Fix \((x_0, v_0^*) \in \mathcal{H} \times \mathcal{G} and \varepsilon \in ]0, 1[, and iterate for every \(n \in \mathbb{N}\)

\[
(\gamma_n, \mu_n) \in [\varepsilon, 1/\varepsilon]^2
\]

\[
a_n = J_{\gamma_n A}(x_n - \gamma_n L^* v_n^*), \quad l_n = Lx_n, \quad b_n = J_{\mu_n B}(l_n + \mu_n v_n^*)
\]

\[
s_n^* = \gamma_n^{-1}(x_n - a_n) + \mu_n^{-1} L^*(l_n - b_n), \quad t_n = b_n - L a_n
\]

\[
\tau_n = \|s_n^*\|^2 + \|t_n\|^2, \quad \lambda_n \in [\varepsilon, 1]
\]

\[
\theta_n = \lambda_n \left(\gamma_n^{-1} \|x_n - a_n\|^2 + \mu_n^{-1} \|l_n - b_n\|^2\right) / \tau_n
\]

\[
x_{n+1/2} = x_n - \theta_n s_n^*, \quad v_{n+1/2}^* = v_n^* - \theta_n t_n
\]

\[
\chi_n = \langle x_0 - x_n \mid x_n - x_{n+1/2}\rangle + \langle v_0^* - v_n^* \mid v_n^* - v_{n+1/2}^*\rangle
\]

\[
\mu_n = \|x_0 - x_n\|^2 + \|v_0^* - v_n^*\|^2, \quad \nu_n = \|x_n - x_{n+1/2}\|^2 + \|v_n^* - v_{n+1/2}^*\|^2
\]

\[
\rho_n = \mu_n \nu_n - \chi_n^2
\]

if \(\rho_n > 0\) and \(\chi_n \nu_n \geq \rho_n\)

\[
x_{n+1} = x_0 + (1 + \chi_n / \nu_n)(x_{n+1/2} - x_n)
\]

\[
v_{n+1}^* = v_0^* + (1 + \chi_n / \nu_n)(v_{n+1/2}^* - v_n^*)
\]

if (ETC... easy steps)

Then \((x_n, v_n^*) \to P_K(x_0, v_0^*)\)