Between Subdifferentials and Monotone Operators

Patrick L. Combettes

Department of Mathematics
North Carolina State University
Raleigh, NC 27695, USA

Modern Maximal Monotone Operator Theory: From Nonsmooth Optimization to Differential Inclusions
Wien, 27 Februar 2019

Supported by the NSF under grant DMS-1818946
Nonlinear functional analysis

1950’s
Linear functional analysis

- Topological vector spaces
- Linear operators
- Duality
- Theory of distributions
- etc.
Nonlinear functional analysis → outgrowths of linear analysis

1950’s
Linear functional analysis
- Topological vector spaces
- Linear operators
- Duality
- Theory of distributions
- etc.

Early 1960’s

Monotone operators

Convex analysis

Nonexpansive operators

These new structured theories, which often revolve around turning equalities in classical linear analysis into inequalities, benefit from tight connections between each other.
Convex analysis (Moreau, Rockafellar, 1962+)

- $\Gamma_0(\mathcal{H})$: lower semicontinuous convex functions $f: \mathcal{H} \to ]-\infty, +\infty]$ such that $\text{dom } f = \{x \in \mathcal{H} \mid f(x) < +\infty\} \neq \emptyset$

- $f^*: x^* \mapsto \sup_{x \in \mathcal{H}} \langle x \mid x^* \rangle - f(x)$ is the conjugate of $f$; if $f \in \Gamma_0(\mathcal{H})$, then $f^* \in \Gamma_0(\mathcal{H})$ and $f^{**} = f$

- The subdifferential of $f$ at $x \in \mathcal{H}$ is

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

Infimal operations:

- $(f \square g): x \mapsto \inf_{y \in \mathcal{H}} f(y) + g(x - y)$
- $(L \triangleright g): x \mapsto \inf_{L y = x} g(y)$

Fermat’s rule:

$x$ minimizes $f \iff 0 \in \partial f(x)$
Nonexpansive operators (Browder, Minty)

- $T \in \mathcal{B}(\mathcal{H})$ is an isometry if $(\forall x \in \mathcal{H}) \|Tx\| = \|x\|$, i.e.,
  $$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \|T - T\| = \|x - y\|.$$  

- $T : \mathcal{H} \to \mathcal{H}$ is nonexpansive if
  $$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \|T - T\| \leq \|x - y\|,$$

  firmly nonexpansive if
  $$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \|T - T\|^2 + \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \|x - y\|^2.$$  

  and $\alpha$-averaged ($\alpha \in [0, 1]$), if
  $$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \|T - T\|^2 + \frac{1 - \alpha}{\alpha} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \|x - y\|^2.$$
Monotone operators (Kačurovskii, Minty, Zarantonello, 1960)

- A ∈ ℬ(ℋ) is skew if (∀x ∈ ℋ) ⟨x | Ax⟩ = 0 and it is positive if (∀x ∈ ℋ) ⟨x | Ax⟩ ≥ 0, i.e.,

\[
(∀x ∈ ℋ)(∀y ∈ ℋ) \quad ⟨x − y | Ax − Ay⟩ ≥ 0. \tag{1}
\]

- In 1960, Kačurovskii, Minty, and Zarantonello independently called monotone a nonlinear operator A: ℋ → ℋ that satisfies (1)

- More generally, a set-valued operator A: ℋ → 2ℋ with graph \( \text{gra } A = \{(x, x^*) ∈ ℋ × ℋ | x^* ∈ Ax\} \) is monotone if

\[
(∀(x, x^*) ∈ \text{gra } A)(∀(y, y^*) ∈ \text{gra } A) \quad ⟨x − y | x^* − y^*⟩ ≥ 0,
\]

and maximally monotone if there is no monotone operator B: ℋ → 2ℋ such that \( \text{gra } A ⊂ \text{gra } B ≠ \text{gra } A \)
Convexity/Nonexpansiveness/Monotonicity

- If $f \in \Gamma_0(H)$, $A = \partial f$ is maximally monotone
- (Minty) If $T: H \to H$ is firmly nonexpansive, then $T = J_A$ for some maximally monotone $A: H \to 2^H$ and $\text{Fix } T = \text{zer } A$
- (Minty) If $A: H \to 2^H$ is maximally monotone, the resolvent $J_A = (\text{Id} + A)^{-1}$ is firmly nonexpansive with $\text{dom } J_A = H$, and the reflected resolvent $R_A = 2J_A - \text{Id}$ is nonexpansive
- If $T: H \to H$ is nonexpansive, $A = \text{Id} - T$ is max. mon., $\text{Fix } T = \{x \in H \mid Tx = x\}$ is closed and convex, and $\text{Fix } T = \text{zer } A$
- If $A: H \to 2^H$ is max. mon., $(\forall x \in H) \ Ax$ is closed and convex; $\text{zer } A = A^{-1}(0)$ is closed and convex
- If $A: H \to 2^H$ is maximally monotone, $\text{int dom } A$, $\overline{\text{dom } A}$, $\text{int ran } A$, and $\overline{\text{ran } A}$ are convex
- If $T: H \to H$ is an $\alpha$-averaged ($\alpha \leq 1/2$) nonexpansive operator, it is maximally monotone
- If $A = \beta B$ is firmly nonexpansive (hence max. mon.), $0 < \gamma < 2\beta$, and $\alpha = \gamma/(2\beta)$, then $\text{Id} - \gamma B$ is an $\alpha$-averaged nonexpansive operator
What is a maximally monotone operator in general?

- Who knows? ...certainly a complicated object
- The Asplund decomposition

\[ A = \partial f + \text{something (acyclic)} \]

is not fully understood

- If \( \mathcal{H} = \mathbb{R} \), something = 0
- In the Borwein-Wiersma decomposition, “something” is the restriction of a skew operator
- Jon Borwein’s conjecture was that in general “something” is locally the restriction (localization) of a skew linear relation
Moreau’s proximity operator

- In 1962, Jean Jacques Moreau (1923–2014) introduced the proximity operator of $f \in \Gamma_0(\mathcal{H})$

$$\text{prox}_f : x \mapsto \arg\min_{y \in \mathcal{H}} f(y) + \frac{1}{2} \|x - y\|^2$$

and derived all its main properties

- Set $q = \| \cdot \|^2/2$. Then $f \Box q + f^* \Box q = q$ and

$$\text{prox}_f = \nabla (f + q)^* = \nabla (f^* \Box q) = \text{id} - \text{prox}_{f^*} = (\text{id} + \partial f)^{-1}$$

- $\text{prox}_f = J_{\partial f}$, hence
  - Fix $\text{prox}_f = \text{zer} \ \partial f = \text{Argmin} \ f$
  - $(\text{prox}_f x, x - \text{prox}_f x) \in \text{gra} \ \partial f$
  - $\|\text{prox}_f x - \text{prox}_f y\|^2 + \|\text{prox}_{f^*} x - \text{prox}_{f^*} y\|^2 \leq \|x - y\|^2$

- This suggests that (Martinet’s proximal point algorithm, 1970/72) $x_{n+1} = \text{prox}_f x_n \rightharpoonup x \in \text{Argmin} \ f$
Subdifferentials as maximally monotone ops. and proximity operators as firmly nonexpansive ops.

- Rockafellar (1966) has fully characterized subdifferentials as those maximally monotone operators which are cyclically maximally monotone.
- Moreau (1965) has fully characterized proximity operators as those (firmly) nonexpansive operators which are gradients of convex functions.
- Moreau (1963) showed that a convex average of proximity operator is again a proximity operator.
- Not all firm nonexpansiveness preserving operations are proximity preserving.

Set

\[
\begin{align*}
\mathcal{P}(\mathcal{H}) &= \left\{ T : \mathcal{H} \rightarrow \mathcal{H} \mid (\exists f \in \Gamma_0(\mathcal{H})) \ T = \text{prox}_f \right\} \\
A \boxdot B &= (A^{-1} + B^{-1})^{-1} \\
L \triangleright A &= (L \circ A^{-1} \circ L^*)^{-1}
\end{align*}
\]
Let $l$ be finite and put $q = \| \cdot \|^2_{\mathcal{H}} / 2$. For every $i \in l$, let $\omega_i \in [0, +\infty[$, put $q_i = \| \cdot \|^2_{\mathcal{G}_i} / 2$, let $L_i \in \mathcal{B}(\mathcal{H}, \mathcal{G}_i) \smallsetminus \{0\}$, let $M_i \in \mathcal{B}(\mathcal{K}_i, \mathcal{G}_i) \smallsetminus \{0\}$, let $f_i \in \Gamma_0(\mathcal{G}_i)$, let $g_i \in \Gamma_0(\mathcal{G}_i)$, and let $h_i \in \Gamma_0(\mathcal{K}_i)$. Suppose that $\sum_{i \in l} \omega_i \| L_i \|^2 \leq 1$ and that,

$$
(\forall i \in l) \begin{cases}
0 \in \mbox{sri} (\text{dom } h_i^* - M_i^* (\text{dom } f_i \cap \text{dom } g_i^*)) \\
0 \in \mbox{sri} (\text{dom } f_i - \text{dom } g_i^*).
\end{cases}
$$

Set

$$
T = \sum_{i \in l} \omega_i L_i^* \circ \left( \text{prox}_{f_i} \square (\partial g_i \square (M_i \triangleright \partial h_i)) \right) \circ L_i.
$$

Then $T \in \mathcal{P}(\mathcal{H})$. More specifically,

$$
T = \text{prox}_f, \quad \text{where} \quad f = \left( \sum_{i \in l} \omega_i \left( (f_i + g_i^* + h_i^* \circ M_i^*)^* \square q_i \right) \circ L_i \right)^* - q.
$$
Proximity-preserving transformations: Consequences

- Let \((T_i)_{i \in I}\) be a finite family in \(\mathcal{P}(\mathcal{H})\), \((\omega_i)_{i \in I}\) convex weights. Then \(\sum_{i \in I} \omega_i T_i \in \mathcal{P}(\mathcal{H})\) (Moreau, 1963).

- Auslender’s barycentric projection method

\[ x_{n+1} = \sum_{i \in I} \omega_i \text{proj}_{C_i} x_n \]

(and under-relaxations thereof) is a proximal algorithm.

- Let \(T_1\) and \(T_2\) be in \(\mathcal{P}(\mathcal{H})\). Then \((T_1 - T_2 + \text{Id})/2 \in \mathcal{P}(\mathcal{H})\).

- Let \(T \in \mathcal{P}(\mathcal{H})\) and let \(V\) be a closed vector subspace of \(\mathcal{H}\). Then \(\text{proj}_V \circ T \circ \text{proj}_V \in \mathcal{P}(\mathcal{H})\).

- Let \(T_1\) and \(T_2\) be in \(\mathcal{P}(\mathcal{H})\). Then \(T_1 \Box T_2 \in \mathcal{P}(\mathcal{H})\).
Proximity-preserving transformations: Consequences

- $K$ a closed convex cone in $\mathcal{H}$ with polar cone $K^\ominus$, $V$ a closed vector subspace of $\mathcal{H}$.
- Set

$$f = \left( \frac{1}{2} a_{K^\ominus}^2 \circ \text{proj}_V \right)^* - \frac{\| \cdot \|^2}{2} \quad \text{and} \quad T = \text{proj}_V \circ \text{proj}_K \circ \text{proj}_V.$$

- Then $T = \text{prox}_f$.
- Let $x_0 \in V$ and $(\forall n \in \mathbb{N}) \ x_{n+1} = \text{prox}_f x_n$.
- $(x_n)_{n \in \mathbb{N}}$ is identical to the alternating projection sequence $x_{n+1} = (\text{proj}_V \circ \text{proj}_K) x_n$.
- Hundal (2004) constructed a special $V$ and $K$ so that convergence of alternating projections is only weak and not strong. We thus obtain a new instance of the weak but not strong convergence of the proximal point algorithm.
Proximity-preserving transformations: Compositions and sums

- Take $T_1 = \text{prox}_{f_1} \in \mathcal{P}(\mathcal{H})$ and $T_2 = \text{prox}_{f_2} \in \mathcal{P}(\mathcal{H})$. Then $T_1 \circ T_2 \notin \mathcal{P}(\mathcal{H})$ (unless $\mathcal{H} = \mathbb{R}$) and $T_1 + T_2 \notin \mathcal{P}(\mathcal{H})$.

- The formula $T_1 \circ T_2 = \text{prox}_{f_1 + f_2}$ has been characterized. An interesting instance is (Briceño-Arias/PLC, 2009)

\[
\text{prox}_{\phi \circ \| \cdot \| + \sigma_C} = \text{prox}_{\phi \circ \| \cdot \|} \circ \text{prox}_{\sigma_C} : x \mapsto \\
\begin{cases} 
\frac{\text{prox}_\phi d_C(x)}{d_C(x)} (x - \text{proj}_C x), & \text{if } d_C(x) > \max \text{Argmin } \phi; \\
 x - \text{proj}_C x, & \text{if } d_C(x) \leq \max \text{Argmin } \phi.
\end{cases}
\]

- **Example:** $K$ a closed convex cone, $\phi = \gamma | \cdot |$. Then

\[
\text{prox}_{\gamma \| \cdot \| + \iota_K} x = \begin{cases} 
\frac{\|\text{proj}_K x\| - \gamma}{\|\text{proj}_K x\|} \text{proj}_K x, & \text{if } \|\text{proj}_K x\| > \gamma; \\
0, & \text{if } \|\text{proj}_K x\| \leq \gamma.
\end{cases}
\]
Proximity-preserving transformations: Compositions and sums

Example: $K$ a closed convex cone, $\phi = \nu_{[-\gamma, \gamma]}$. Then

$$\text{proj}_{B(0; \gamma) \cap K} x = \begin{cases} \gamma \frac{\text{proj}_K x}{\|\text{proj}_K x\|}, & \text{if } \|\text{proj}_K x\| > \gamma; \\ \text{proj}_K x, & \text{if } \|\text{proj}_K x\| \leq \gamma. \end{cases}$$

Suppose that $0 \in \text{sri} (\text{dom } f_1^* - \text{dom } f_2^*)$ and that

$$(f_1^* + f_2^*) \square q = f_1^* \square q + f_2^* \square q.$$

Then $T_1 + T_2 = \text{prox}_{f_1 \square f_2} \in \mathcal{P}(\mathcal{H})$. 
Self-dual classes: $T \in \mathcal{T}(\mathcal{H}) \iff \text{Id} - T \in \mathcal{T}(\mathcal{H})$
The need for monotone operators in optimization

- They offer a synthetic framework to formulate, analyze, and solve optimization problems but, more importantly,...
The need for monotone operators in optimization

- They offer a synthetic framework to formulate, analyze, and solve optimization problems but, more importantly,...
- ... some key maximal monotone operators arising in the analysis and the numerical solution of convex minimization problems are not subdifferentials, for instance:
  - (Rockafellar, 1970) The saddle operator
    \[ A: (x_1, x_2) \mapsto \partial L(\cdot, x_2)(x_1) \times \partial(-L(x_1, \cdot))(x_2) \]
    associated with a closed convex-concave function \( L \)
  - (Spingarn, 1983) The partial inverse of a maximally monotone operator (and even of a subdifferential)
  - Some operators which arise in the perturbation of optimization problems are no longer subdifferentials
  - Skew linear operators arising in composite primal-dual minimization problems (PLC et al., 2011+)
Interplay: The proximal point algorithm

- First derived by Martinet (1970/72) for \( f \in \Gamma_0(\mathcal{H}) \) with constant proximal parameters, and then by Brézis-Lions (1978)

\[
x_{n+1} = \text{prox}_{\gamma_n f} x_n \rightharpoonup x \in \text{Argmin } f \quad \text{if} \quad \sum_{n \in \mathbb{N}} \gamma_n = +\infty \quad (2)
\]

- Then extended to a maximally monotone operator \( A \) by Rockafellar (1976) and Brézis-Lions (1978)

\[
x_{n+1} = J_{\gamma_n A} x_n \rightharpoonup x \in \text{zer } A \quad \text{if} \quad \sum_{n \in \mathbb{N}} \gamma_n^2 = +\infty \quad (3)
\]

- Note that (2) has more general parameters. However (3) is considerably more useful to optimization than (2)
Interplay: The proximal point algorithm

- (Rockafellar, 1976) Applying the general proximal point algorithm (3) to the saddle operator leads to various minimization algorithms (e.g., the proximal method of multipliers in the case of the ordinary Lagrangian).

- It was noted by Eckstein/Bersekas (1992) that the Douglas-Rachford splitting algorithm is implicitly driven by a proximal iteration for a maximally monotone operator. The same is true for the forward-backward algorithm!

- Applying the general proximal point algorithm (3) to the partial inverse of a suitably constructed partial inverse makes it possible to solve the convex composite problem (Alghamdi, Alotaibi, PLC, Shahzad, 2014)

\[
\text{minimize } \sum_{i \in I} \left( f_i(x_i) - \langle x_i | z_i \rangle \right) + g \left( \sum_{i \in I} L_i x_i - r \right)
\]
The need for monotone operators in optimization

- They offer a synthetic framework to formulate, analyze, and solve optimization problems but, more importantly,...

- ... some key maximal monotone operators arising in the analysis and the numerical solution of convex minimization problems are **not** subdifferentials, for instance

  
  (Rockafellar, 1970) The saddle operator

  \[ A: (x_1, x_2) \mapsto \partial L(\cdot, x_2)(x_1) \times \partial (-L(x_1, \cdot))(x_2) \]

  associated with a closed convex concave function \( L \)

  (Spingarn, 1983) The partial inverse of a maximally monotone operator (and even of a subdifferential)

  Some operators which arise in the perturbation of optimization problems are no longer subdifferentials

  Skew linear operators arising in composite primal-dual minimization problems (PLC et al., 2011+)
Periodic projection methods: inconsistent case

- Basic feasibility problem: find a common point of nonempty closed convex sets \((C_i)_{1 \leq i \leq m}\) by the method of periodic projections \(x_{mn+1} = \text{proj}_1 \cdots \text{proj}_m x_{mn}\)

- If the sets turn out not to intersect, the method produces a cycle \((\bar{y}_1, \bar{y}_2, \bar{y}_3)\)
Denote by $\text{cyc}(C_1, \ldots, C_m)$ is the set of cycles of $(C_1, \ldots, C_m)$, i.e.,

$$
\text{cyc}(C_1, \ldots, C_m) = \{ (\overline{y}_1, \ldots, \overline{y}_m) \in \mathcal{H}^m \mid \overline{y}_1 = \text{proj}_1 \overline{y}_2, \ldots, \\
\overline{y}_{m-1} = \text{proj}_{m-1} \overline{y}_m, \overline{y}_m = \text{proj}_m \overline{y}_1 \}.
$$

**Question (Gurin-Polyak-Raik, 1967):** Let $m \geq 3$ be an integer. Does there exist a function $\Phi: \mathcal{H}^m \to \mathbb{R}$ such that, for every ordered family of nonempty closed convex subsets $(C_1, \ldots, C_m)$ of $\mathcal{H}$, $\text{cyc}(C_1, \ldots, C_m)$ is the set of solutions to

$$
\text{minimize} \quad \Phi(y_1, \ldots, y_m) \quad \text{subject to} \quad y_1 \in C_1, \ldots, y_m \in C_m
$$
Cyclic projection methods

- **Theorem (Baillon/PLC/Cominetti, 2012):** Suppose that $\dim \mathcal{H} \geq 2$ and let $\mathbb{N} \ni m \geq 3$. There exists no function $\Phi: \mathcal{H}^m \to \mathbb{R}$ such that, for every ordered family of nonempty closed convex subsets $(C_1, \ldots, C_m)$ of $\mathcal{H}$, $\text{cyc}(C_1, \ldots, C_m)$ is the set of solutions to the variational problem

$$\begin{align*}
\text{minimize} \quad & \Phi(y_1, \ldots, y_m) \\
\text{subject to} \quad & y_1 \in C_1, \ldots, y_m \in C_m
\end{align*}$$
Theorem (Baillon/PLC/Cominetti, 2012): Suppose that \( \dim \mathcal{H} \geq 2 \) and let \( \mathbb{N} \ni m \geq 3 \). There exists no function \( \Phi : \mathcal{H}^m \rightarrow \mathbb{R} \) such that, for every ordered family of nonempty closed convex subsets \( (C_1, \ldots, C_m) \) of \( \mathcal{H} \), \( \text{cyc}(C_1, \ldots, C_m) \) is the set of solutions to the variational problem

\[
\text{minimize} \quad \Phi(y_1, \ldots, y_m).
\]

\( y_1 \in C_1, \ldots, y_m \in C_m \)

However, cycles do have a meaning: if we denote by \( L \) the circular left shift, they solve the inclusion

\[
(0, \ldots, 0) \in \sum_{\text{subdifferential}} N_{C_1 \times \cdots \times C_m}(y_1, \ldots, y_m) + (\text{Id} - L)(y_1, \ldots, y_m),
\]

which involves two maximally monotone operators
The need for monotone operators in optimization

- They offer a synthetic framework to formulate, analyze, and solve optimization problems but, more importantly,...
- ... some key maximal monotone operators arising in the analysis and the numerical solution of convex minimization problems are not subdifferentials, for instance
  - (Rockafellar, 1970) The saddle operator
    \[ A: (x_1, x_2) \mapsto \partial L(\cdot, x_2)(x_1) \times \partial (-L(x_1, \cdot))(x_2) \]
  associated with a closed convex concave function \( L \)
  - (Spingarn, 1983) The partial inverse of a maximally monotone operator (and even of a subdifferential)
  - Some operators which arise in the perturbation of optimization problems are no longer subdifferentials
  - Skew linear operators arising in composite primal-dual minimization problems (PLC et al., 2011+)
Splitting structured problems: 3 basic methods

\[ A, B : \mathcal{H} \rightarrow 2^\mathcal{H} \text{ maximally monotone, solve } 0 \in A\overline{x} + B\overline{x}. \]

- **Douglas-Rachford splitting (1979)**

\[
\begin{aligned}
  y_n &= J_{\gamma B} x_n \\
  z_n &= J_{\gamma A} (2y_n - x_n) \\
  x_{n+1} &= x_n + z_n - y_n
\end{aligned}
\]

- **B : \mathcal{H} \rightarrow \mathcal{H} 1/\beta\text{-cocoercive: forward-backward splitting (1979+)}**

\[
\begin{aligned}
  0 < \gamma_n < 2/\beta \\
  y_n &= x_n - \gamma_n Bx_n \\
  x_{n+1} &= J_{\gamma_n A} y_n
\end{aligned}
\]

- **B : \mathcal{H} \rightarrow \mathcal{H} \mu\text{-Lipschitzian: forward-backward-forward splitting (2000)}**

\[
\begin{aligned}
  0 < \gamma_n < 1/\mu \\
  y_n &= x_n - \gamma_n Bx_n \\
  z_n &= J_{\gamma_n A} y_n \\
  r_n &= z_n - \gamma_n Bz_n \\
  x_{n+1} &= x_n - y_n + r_n
\end{aligned}
\]
Splitting structured problems: 3 basic methods

- A large number of minimization methods are special cases of these **monotone operator** splitting methods in a suitable setting that may involve
  - product spaces
  - dual spaces
  - primal-dual spaces
  - renormed spaces
  - or a combination thereof

- The simplifying reformulations typically involve monotone operators which are **not** subdifferentials. For instance, the primal-dual minimization of $f + g \circ L$ leads to the monotone+skew model (Briceño-Arias/PLC, 2011)

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \partial f & 0 \\ 0 & \partial g^* \end{bmatrix} \begin{bmatrix} x \\ x^* \end{bmatrix} + \begin{bmatrix} 0 & L^* \\ -L & 0 \end{bmatrix} \begin{bmatrix} x \\ x^* \end{bmatrix}$$
Proximal splitting methods in convex optimization

- \( f \in \Gamma_0(\mathcal{H}), \varphi_k \in \Gamma_0(\mathcal{G}_k), \ell_k \in \Gamma_0(\mathcal{G}_k) \) strongly convex, \( L_k : \mathcal{H} \to \mathcal{G}_k \) linear bounded, \( \|L_k\| = 1 \), \( h : \mathcal{H} \to \mathbb{R} \) convex and smooth:

  \[
  \text{minimize}_{x \in \mathcal{H}} \quad f(x) + \sum_{k=1}^{p} (\varphi_k \square \ell_k)(L_k x - r_k) + h(x)
  \]

- A splitting algorithm activates each function and each linear operator individually
Proximal splitting methods in convex optimization

- $A = \partial f$, $C = \nabla h$, $B_k = \partial g_k$, and $D_k = \partial \ell_k$
- $\mathcal{K} = \mathcal{H} \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$
- Subdifferential: $M: \mathcal{K} \to 2^\mathcal{K}: (x, v_1, \ldots, v_p) \mapsto (-z + Ax) \times (r_1 + B_1^{-1}v_1) \times \cdots \times (r_p + B_p^{-1}v_p)$
- Not a subdifferential: $Q: \mathcal{K} \to \mathcal{K}: (x, v_1, \ldots, v_p) \mapsto (Cx + \sum_{k=1}^{p} L_k^* v_k, -L_1 x + D_1^{-1} v_1, \ldots, -L_p x + D_p^{-1} v_p)$
- $M$ and $Q$ are maximally monotone, $Q$ is Lipschitzian, the zeros of $M + Q$ are primal-dual solutions
- Solve $0 \in Mx + Qx$, where $x = (x, v_1, \ldots, v_p)$ via Tseng’s forward-backward-forward-forward splitting algorithm

\[
\begin{align*}
    y_n &= x_n - Qx_n \\
    p_n &= (\text{Id} + M)^{-1} y_n \\
    q_n &= p_n - Qp_n \\
    x_{n+1} &= x_n - y_n + q_n
\end{align*}
\]

in $\mathcal{K}$ to get...
Proximal splitting methods in convex optimization

- Algorithm:
  
  for $n = 0, 1, \ldots$
  
  $y_{1,n} = x_n - (\nabla h(x_n) + \sum_{k=1}^{m} L^*_k v_{k,n})$
  
  $p_{1,n} = \text{prox}_f y_{1,n}$
  
  For $k = 1, \ldots, p$
  
  $y_{2,k,n} = v_{k,n} + (L_k x_n - \nabla \ell^*_k(v_{k,n}))$
  
  $p_{2,k,n} = \text{prox}_{g_k^*}(y_{2,k,n} - r_k)$
  
  $q_{2,k,n} = p_{2,k,n} + (L_k p_{1,n} - \nabla \ell^*_k(p_{2,k,n}))$
  
  $v_{k,n+1} = v_{k,n} - y_{2,k,n} + q_{2,k,n}$
  
  $q_{1,n} = p_{1,n} - (\nabla h(p_{1,n}) + \sum_{k=1}^{m} L^*_k p_{2,k,n})$
  
  $x_{n+1} = x_n - y_{1,n} + q_{1,n}$

- $(x_n)_{n \in \mathbb{N}}$ converges weakly to a solution and $(v_{k,n})_{1 \leq k \leq p} (n \in \mathbb{N})$ converges weakly to a solution and to a dual solution (PLC/Pesquet, 2012; PLC, 2013)
Some limitations of the state-of-the-art

We present a new framework that circumvents simultaneously the limitations of current methods, which require:

- inversions of linear operators or knowledge of bounds on norms of all the $L_{ki}$
- the proximal parameters must be the same for all the subdifferential operators
- activation of the proximal operators of all the functions: impossible in huge-scale problems
- synchronicity: all proximity operator evaluations must be computed and used during the current iteration

and, in general,

- converge only weakly
Let $F$ be the set of solutions to the problem

$$\min_{x_i \in \mathcal{H}_i, i \in I} \sum_{i \in I} (f_i(x_i) - \langle x_i \mid z_i^* \rangle) + \sum_{k \in K} g_k \left( \sum_{i \in I} L_{ki} x_i - r_k \right)$$

where $f_i \in \Gamma_0(\mathcal{H}_i)$, $g_k \in \Gamma_0(\mathcal{G}_k)$, $L_{ki} \in \mathcal{B}(\mathcal{H}_i, \mathcal{G}_k)$

Let $F^*$ be the set of solutions to the dual problem

$$\min_{\nabla_k^* \in \mathcal{G}_k, k \in K} \sum_{i \in I} f_i^* \left( z_i^* - \sum_{k \in K} L_{ki}^* \nabla_k^* \right) + \sum_{k \in K} (g_k^*(\nabla_k^*) + \langle \nabla_k^* \mid r_k \rangle)$$

Associated Kuhn-Tucker set (set of zeros a maximally monotone operator which is not a subdifferential)

$$Z = \left\{ \left( (\overline{x}_i)_{i \in I}, (\overline{\nabla}_k^*)_{k \in K} \right) \bigg| \overline{x}_i \in \mathcal{H}_i \text{ and } z_i^* - \sum_{k \in K} L_{ki}^* \overline{\nabla}_k^* \in \partial f_i(\overline{x}_i), \overline{\nabla}_k^* \in \mathcal{G}_k \text{ and } \sum_{i \in I} L_{ki} \overline{x}_i - r_k \in \partial g_k^*(\overline{\nabla}_k^*) \right\}$$
Underlying geometry: The Kuhn-Tucker set

\[ \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p \]

\[ \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_m \]

\[ F^* \]

\[ Z \]
Underlying geometry: The Kuhn-Tucker set

Choose suitable points in the graphs of \((\partial f_i)_{i \in I}\) and \((\partial g_k)_{k \in K}\) to construct a half-space \(H_n\) containing \(Z\).

Algorithm: \((x_{n+1}, v_{n+1}^*) = P_{H_n}(x_n, v_n^*) \rightarrow (x, v^*) \in Z \subset F \times F^*\)
Asynchronous block-iterative proximal splitting (PLC/Eckstein, 2018)

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

for every \( i \in I_n \)

\[
\begin{align*}
   l_{i,n}^* &= \sum_{k \in K} L_{ki}^* v_{k, i}(n) \\
   (a_{i,n}, a_{i,n}^*) &= \left( \operatorname{prox}_{\gamma_{i,c_i}(n)} f_i(x_{i,c_i}(n) + \gamma_{i,c_i}(n)(z_i - l_{i,n}^*)), \gamma_{i,c_i}(n)(x_{i,c_i}(n) - a_{i,n}) - l_{i,n}^* \right)
\end{align*}
\]

for every \( i \in I \setminus I_n \)

\[
(a_{i,n}, a_{i,n}^*) = (a_{i,n-1}, a_{i,n-1}^*)
\]

for every \( k \in K_n \)

\[
\begin{align*}
   l_{k,n} &= \sum_{i \in I} L_{ki} x_{i,k}(n) \\
   (b_{k,n}, b_{k,n}^*) &= \left( r_k + \operatorname{prox}_{\mu_k, d_k(n)} g_k(l_k, n + \mu_k, d_k(n) v_{k, n}(k) - r_k), v_{k, n}(k) + \mu_k, d_k(n)(l_k - b_{k,n}) \right)
\end{align*}
\]

for every \( k \in K \setminus K_n \)

\[
(b_{k,n}, b_{k,n}^*) = (b_{k,n-1}, b_{k,n-1}^*)
\]

\[
((t_{i,n}^*)_{i \in I}, (t_{k,n})_{k \in K}) = \left( (a_{i,n}^* + \sum_{k \in K} L_{ki}^* b_{k,n}^*)_{i \in I}, (b_{k,n} - \sum_{i \in I} L_{ki} a_{i,n})_{k \in K} \right)
\]

\[\tau_n = \sum_{i \in I} \|t_{i,n}\|^2 + \sum_{k \in K} \|t_{k,n}\|^2\]

if \( \tau_n > 0 \)

\[\theta_n = \frac{\lambda_n}{\tau_n} \max \left\{ 0, \sum_{i \in I} \left( \langle x_{i,n} \mid t_{i,n}^* \rangle - \langle a_{i,n} \mid a_{i,n}^* \rangle \right) + \sum_{k \in K} \left( \langle t_{k,n} \mid v_{k,n}^* \rangle - \langle b_{k,n} \mid b_{k,n}^* \rangle \right) \right\}\]

else \( \theta_n = 0 \)

for every \( i \in I \)

\[x_{i,n+1} = x_{i,n} - \theta_n t_{i,n}^*\]

for every \( k \in K \)

\[v_{k,n+1} = v_{k,n} - \theta_n t_{k,n}\]
Asynchronous block-iterative proximal splitting II

- Construct $H_n$ as before
- The half-space $D_n$ satisfies $(x_n, v_n^*) = P_{D_n}(x_0, v_0^*)$
- Algorithm: $(x_{n+1}, v_{n+1}^*) = P_{H_n \cap D_n}(x_0, v_0^*) \rightarrow P_Z(x_0, v_0^*) \in F \times F^*$
Just like in the early 1960s the frontier separating linear from nonlinear problems was not a useful one, the current dichotomy between the class of convex/monotone problems and its complement (“everything else”) is not pertinent.

One must define a structured extension of the remarkably efficient convexity/nonexpansiveness/monotonicity trio that would ideally enjoy similar rich connections. This is an extremely challenging task.
References