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

Functional analysis: Linear versus nonlinear

Nonlinear functional analysis

1950's

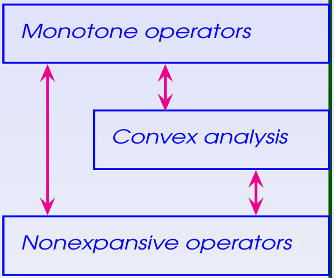
Linear functional analysis

- *Topological vector spaces*
- *Linear operators*
- *Duality*
- *Theory of distributions*
- *etc.*

Functional analysis: Linear versus nonlinear

Early 1960's

Nonlinear functional analysis → outgrowths of linear analysis



1950's

Linear functional analysis

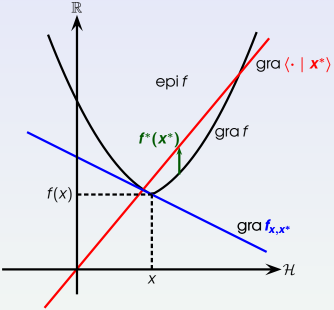
- *Topological vector spaces*
- *Linear operators*
- *Duality*
- *Theory of distributions*
- *etc.*

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 $f\colon \mathcal{H} \to]-\infty, +\infty]$ such that $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\} \neq \varnothing$
- $f^*\colon x^* \mapsto \sup_{x\in\mathcal{H}} \langle x \mid x^*\rangle - f(x)$ 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

$$\partial f(x) = \Big\{x^* \in \mathcal{H} \;\Big|\; (\forall y \in \mathcal{H})\; \underbrace{\langle y - x \mid x^*\rangle + f(x)}_{f_{x,x^*}(y)} \leqslant f(y)\Big\}$$



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

Fermat's rule:
x minimizes $f \Leftrightarrow 0 \in \partial f(x)$

Nonexpansive operators (Browder, Minty)

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

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \ \|Tx - Ty\| = \|x - y\|.$$

- $T\colon \mathcal{H} \to \mathcal{H}$ is *nonexpansive* if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \ \|Tx - Ty\| \leqslant \|x - y\|,$$

  firmly nonexpansive if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \ \|Tx-Ty\|^2+\|(\mathrm{Id}-T)x-(\mathrm{Id}-T)y\|^2 \leqslant \|x-y\|^2.$$

  and α-*averaged* ($\alpha \in \,]0,1[$), if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \ \|Tx-Ty\|^2+\frac{1-\alpha}{\alpha}\|(\mathrm{Id}-T)x-(\mathrm{Id}-T)y\|^2 \leqslant \|x-y\|^2$$

Monotone operators (Kačurovskiĭ, Minty, Zarantonello, 1960)

- $A \in \mathscr{B}(\mathcal{H})$ is skew if $(\forall x \in \mathcal{H})\ \langle x \mid Ax\rangle = 0$ and it is positive if $(\forall x \in \mathcal{H})\ \langle x \mid Ax\rangle \geqslant 0$, i.e.,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Ax - Ay\rangle \geqslant 0. \tag{1}$$

- In 1960, Kačurovskiĭ, Minty, and Zarantonello independently called *monotone* a nonlinear operator $A\colon \mathcal{H} \to \mathcal{H}$ that satisfies (1)
- More generally, a set-valued operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ with graph $\operatorname{gra} A = \big\{(x, x^*) \in \mathcal{H} \times \mathcal{H} \mid x^* \in Ax\big\}$ is monotone if

$$(\forall (x, x^*) \in \operatorname{gra} A)(\forall (y, y^*) \in \operatorname{gra} A) \quad \langle x - y \mid x^* - y^*\rangle \geqslant 0,$$

and *maximally monotone* if there is no monotone operator $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} A \subset \operatorname{gra} B \neq \operatorname{gra} A$

Convexity/Nonexpansiveness/Monotonicity

- If $f \in \Gamma_0(\mathcal{H})$, $A = \partial f$ is maximally monotone
- (Minty) If $T\colon \mathcal{H} \to \mathcal{H}$ is firmly nonexpansive, then $T = J_A$ for some maximally monotone $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $\operatorname{Fix} T = \operatorname{zer} A$
- (Minty) If $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone, the resolvent $J_A = (\mathrm{Id} + A)^{-1}$ is firmly nonexpansive with $\operatorname{dom} J_A = \mathcal{H}$, and the reflected resolvent $R_A = 2J_A - \mathrm{Id}$ is nonexpansive
- If $T\colon \mathcal{H} \to \mathcal{H}$ is nonexpansive, $A = \mathrm{Id} - T$ is max. mon., $\operatorname{Fix} T = \{x \in \mathcal{H} \mid Tx = x\}$ is closed and convex, and $\operatorname{Fix} T = \operatorname{zer} A$
- If $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is max. mon., $(\forall x \in \mathcal{H})$ Ax is closed and convex; $\operatorname{zer} A = A^{-1}(0)$ is closed and convex
- If $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone, $\operatorname{int}\operatorname{dom} A$, $\overline{\operatorname{dom}}\, A$, $\operatorname{int}\operatorname{ran} A$, and $\overline{\operatorname{ran}}\, A$ are convex
- If $T\colon \mathcal{H} \to \mathcal{H}$ is an α-averaged ($\alpha \leqslant 1/2$) 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 $\mathrm{Id} - \gamma B$ is an α-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 $f \in \Gamma_0(\mathcal{H})$

$$\operatorname{prox}_f \colon x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}}\, f(y) + \frac{1}{2}\|x - y\|^2$$

  and derived all its main properties
- Set $q = \|\cdot\|^2/2$. Then $f \,\Box\, q + f^* \,\Box\, q = q$ and

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

- $\operatorname{prox}_f = J_{\partial f}$, hence
  - $\operatorname{Fix} \operatorname{prox}_f = \operatorname{zer} \partial f = \operatorname{Argmin} f$
  - $(\operatorname{prox}_f x, x - \operatorname{prox}_f x) \in \operatorname{gra} \partial f$
  - $\|\operatorname{prox}_f x - \operatorname{prox}_f y\|^2 + \|\operatorname{prox}_{f^*} x - \operatorname{prox}_{f^*} y\|^2 \leqslant \|x - y\|^2$
- This suggests that (Martinet's proximal point algorithm, 1970/72) $x_{n+1} = \operatorname{prox}_f x_n \rightharpoonup x \in \operatorname{Argmin} f$

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

$$\begin{cases} \mathcal{P}(\mathcal{H}) = \big\{T\colon \mathcal{H}\to\mathcal{H} \;\big|\; (\exists\, f\in\Gamma_0(\mathcal{H}))\; T=\mathrm{prox}_f\big\}\\ A\,\square\, B = (A^{-1}+B^{-1})^{-1}\\ L\rhd A = \big(L\circ A^{-1}\circ L^*\big)^{-1}\end{cases}$$

Proximity-preserving transformations

Let I be finite and put $q = \|\cdot\|_{\mathcal{H}}^2/2$. For every $i \in I$, let $\omega_i \in \left]0,+\infty\right[$, put $q_i = \|\cdot\|_{\mathcal{G}_i}^2/2$, let $L_i \in \mathscr{B}(\mathcal{H},\mathcal{G}_i) \smallsetminus \{0\}$, let $M_i \in \mathscr{B}(\mathcal{K}_i,\mathcal{G}_i) \smallsetminus \{0\}$, 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 $\sum_{i\in I} \omega_i \|L_i\|^2 \leqslant 1$ and that,

$$(\forall i \in I) \quad \begin{cases} 0 \in \operatorname{sri}\big(\operatorname{dom} h_i^* - M_i^*(\operatorname{dom} f_i \cap \operatorname{dom} g_i^*)\big) \\ 0 \in \operatorname{sri}\big(\operatorname{dom} f_i - \operatorname{dom} g_i^*\big). \end{cases}$$

Set

$$T = \sum_{i\in I} \omega_i L_i^* \circ \Big(\operatorname{prox}_{f_i} \,\square\, \big(\partial g_i \,\square\, (M_i \triangleright \partial h_i)\big)\Big) \circ L_i.$$

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

$$T = \operatorname{prox}_f, \quad \text{where} \quad f = \left(\sum_{i\in I} \omega_i \Big(\big(f_i + g_i^* + h_i^* \circ M_i^*\big)^* \,\square\, q_i\Big) \circ L_i\right)^* - q.$$

Proximity-preserving transformations: Consequences

- $(T_i)_{i \in I}$ 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

$$x_{n+1} = \sum_{i \in I} \omega_i \mathrm{proj}_{C_i} x_n$$

  (and under-relaxations thereof) is a proximal algorithm.
- Let T_1 and T_2 be in $\mathcal{P}(\mathcal{H})$. Then $(T_1 - T_2 + \mathrm{Id})/2 \in \mathcal{P}(\mathcal{H})$.
- Let $T \in \mathcal{P}(\mathcal{H})$ and let V be a closed vector subspace of $\mathcal{H}$. Then $\mathrm{proj}_V \circ T \circ \mathrm{proj}_V \in \mathcal{P}(\mathcal{H})$.
- Let T_1 and T_2 be in $\mathcal{P}(\mathcal{H})$. Then $T_1 \,\square\, T_2 \in \mathcal{P}(\mathcal{H})$.

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

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

- Then $T = \mathrm{prox}_f$.
- Let $x_0 \in V$ and $(\forall n \in \mathbb{N})\ x_{n+1} = \mathrm{prox}_f x_n$.
- $(x_n)_{n\in\mathbb{N}}$ is identical to the alternating projection sequence $x_{n+1} = (\mathrm{proj}_V \circ \mathrm{proj}_K)x_n$.
- 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 $T_1 = \operatorname{prox}_{f_1} \in \mathcal{P}(\mathcal{H})$ and $T_2 = \operatorname{prox}_{f_2} \in \mathcal{P}(\mathcal{H})$. 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 $T_1 \circ T_2 = \operatorname{prox}_{f_1+f_2}$ has been characterized. An interesting instance is (Briceño-Arias/PLC, 2009)

$$\operatorname{prox}_{\phi\circ\|\cdot\|+\sigma_C} = \operatorname{prox}_{\phi\circ\|\cdot\|} \circ \operatorname{prox}_{\sigma_C} : x \mapsto$$
$$\begin{cases} \dfrac{\operatorname{prox}_\phi d_C(x)}{d_C(x)}(x - \operatorname{proj}_C x), & \text{if } d_C(x) > \max \operatorname{Argmin} \phi; \\ x - \operatorname{proj}_C x, & \text{if } d_C(x) \leqslant \max \operatorname{Argmin} \phi \end{cases}$$

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

$$\operatorname{prox}_{\gamma\|\cdot\|+\iota_K} x = \begin{cases} \dfrac{\|\operatorname{proj}_K x\| - \gamma}{\|\operatorname{proj}_K x\|}\operatorname{proj}_K x, & \text{if } \|\operatorname{proj}_K x\| > \gamma; \\ 0, & \text{if } \|\operatorname{proj}_K x\| \leqslant \gamma. \end{cases}$$

Proximity-preserving transformations: Compositions and sums

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

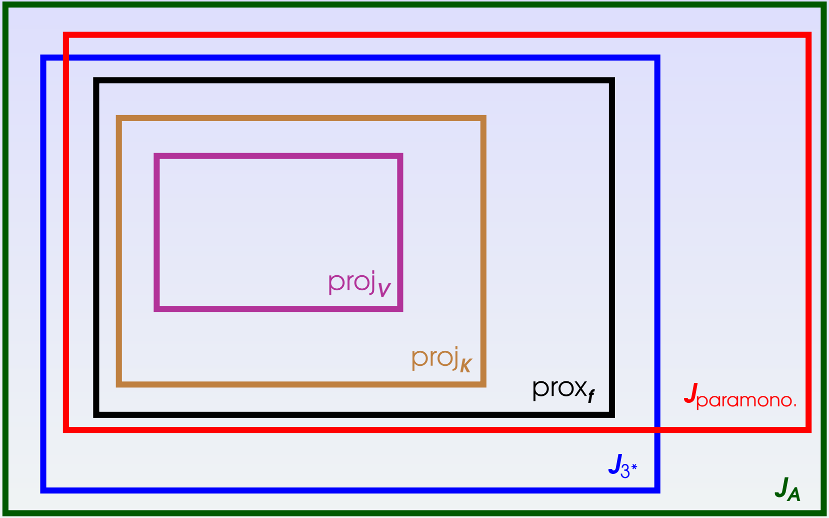
$$\operatorname{proj}_{B(0;\gamma)\cap K} x = \begin{cases} \dfrac{\gamma}{\|\operatorname{proj}_K x\|}\operatorname{proj}_K x, & \text{if } \|\operatorname{proj}_K x\| > \gamma; \\ \operatorname{proj}_K x, & \text{if } \|\operatorname{proj}_K x\| \leqslant \gamma. \end{cases}$$

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

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

Then $T_1 + T_2 = \operatorname{prox}_{f_1 \,\square\, f_2} \in \mathcal{P}(\mathcal{H})$.

Self-dual classes: $T \in \mathcal{T}(\mathcal{H}) \Leftrightarrow \mathrm{Id} - T \in \mathcal{T}(\mathcal{H})$



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

    $$A\colon (x_1, x_2) \mapsto \partial\mathcal{L}(\cdot, x_2)(x_1) \times \partial(-\mathcal{L}(x_1, \cdot))(x_2)$$

    associated with a closed convex-concave function $\mathcal{L}$
  - (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)

$$x_{n+1} = \operatorname{prox}_{\gamma_n f} x_n \rightharpoonup x \in \operatorname{Argmin} f \quad \text{if} \quad \sum_{n\in\mathbb{N}} \gamma_n = +\infty \tag{2}$$

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

$$x_{n+1} = J_{\gamma_n A} x_n \rightharpoonup x \in \operatorname{zer} A \quad \text{if} \quad \sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty \tag{3}$$

- 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)

$$\underset{(\forall i\in I)\ x_i\in\mathcal{H}_i}{\text{minimize}} \sum_{i\in I}\big(f_i(x_i)-\langle x_i \mid z_i\rangle\big)+g\left(\sum_{i\in I}L_i x_i - r\right)$$

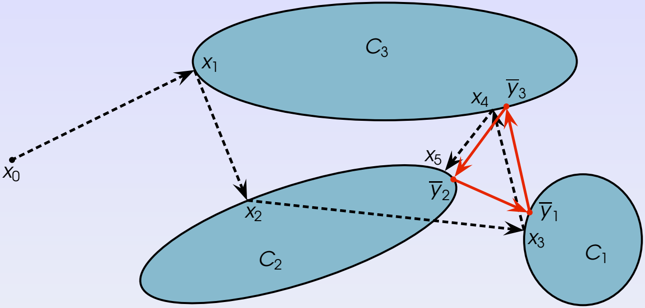
The need for monotone operators in optimization

- 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

$$A\colon (x_1, x_2) \mapsto \partial\mathcal{L}(\cdot, x_2)(x_1) \times \partial(-\mathcal{L}(x_1, \cdot))(x_2)$$

    associated with a closed convex concave function $\mathcal{L}$
  - (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 $(C_i)_{1\leqslant i\leqslant m}$ 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 $(\overline{y}_1, \overline{y}_2, \overline{y}_3)$

Periodic projection methods: inconsistent case

- Denote by $\operatorname{cyc}(C_1,\ldots,C_m)$ is the set of cycles of $(C_1,\ldots,C_m)$, i.e.,

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

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

$$\underset{y_1\in C_1,\ldots,\,y_m\in C_m}{\text{minimize}}\quad \Phi(y_1,\ldots,y_m)\ ?$$

Cyclic projection methods

- **Theorem (Baillon/PLC/Cominetti, 2012):** Suppose that $\dim \mathcal{H} \geqslant 2$ and let $\mathbb{N} \ni m \geqslant 3$. There exists ***no*** function $\Phi\colon \mathcal{H}^m \to \mathbb{R}$ 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

$$\underset{y_1 \in C_1, \ldots, y_m \in C_m}{\text{minimize}} \quad \Phi(y_1, \ldots, y_m).$$

Cyclic projection methods

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

$$\underset{y_1 \in C_1, \ldots, y_m \in C_m}{\text{minimize}} \quad \Phi(y_1, \ldots, y_m).$$

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

$$(0, \ldots, 0) \in \underbrace{N_{C_1 \times \cdots \times C_m}}_{\text{subdifferential}}(y_1, \ldots, y_m) + \underbrace{(\mathrm{Id} - L)}_{\text{not a subdifferential}}(y_1, \ldots, y_m),$$

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

$$A\colon (x_1,x_2) \mapsto \partial\mathcal{L}(\cdot,x_2)(x_1) \times \partial(-\mathcal{L}(x_1,\cdot))(x_2)$$

    associated with a closed convex concave function $\mathcal{L}$
  - (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

$A, B\colon \mathcal{H} \to 2^{\mathcal{H}}$ maximally monotone, solve $0 \in A\overline{x} + B\overline{x}$.

- Douglas-Rachford splitting (1979)

$$\left\lfloor \begin{array}{l} y_n = J_{\gamma B} x_n \\ z_n = J_{\gamma A}(2y_n - x_n) \\ x_{n+1} = x_n + z_n - y_n \end{array} \right.$$

- $B\colon \mathcal{H} \to \mathcal{H}$ $1/\beta$-cocoercive: forward-backward splitting (1979+)

$$\left\lfloor \begin{array}{l} 0 < \gamma_n < 2/\beta \\ y_n = x_n - \gamma_n B x_n \\ x_{n+1} = J_{\gamma_n A} y_n \end{array} \right.$$

- $B\colon \mathcal{H} \to \mathcal{H}$ μ-Lipschitzian: forward-backward-forward splitting (2000)

$$\left\lfloor \begin{array}{l} 0 < \gamma_n < 1/\mu \\ y_n = x_n - \gamma_n B x_n \\ z_n = J_{\gamma_n A} y_n \\ r_n = z_n - \gamma_n B z_n \\ x_{n+1} = x_n - y_n + r_n \end{array} \right.$$

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, $L_k\colon \mathcal{H} \to \mathcal{G}_k$ linear bounded, $\|L_k\| = 1$, $h\colon \mathcal{H} \to \mathbb{R}$ convex and smooth:

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{k=1}^{p} (\varphi_k \,\Box\, \ell_k)(L_k x - r_k) + h(x)$$

- 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: $\boldsymbol{M}\colon \mathcal{K} \to 2^{\mathcal{K}}\colon (x, v_1, \ldots, v_p) \mapsto$ $(-z + Ax) \times (r_1 + B_1^{-1} v_1) \times \cdots \times (r_p + B_p^{-1} v_p)$
- Not a subdifferential: $\boldsymbol{Q}\colon \mathcal{K} \to \mathcal{K}\colon (x, v_1, \ldots, v_p) \mapsto$ $\left(Cx + \sum_{k=1}^{p} L_k^* v_k, -L_1 x + D_1^{-1} v_1, \ldots, -L_p x + D_p^{-1} v_p\right)$
- $\boldsymbol{M}$ and $\boldsymbol{Q}$ are maximally monotone, $\boldsymbol{Q}$ is Lipschitzian, the zeros of $\boldsymbol{M} + \boldsymbol{Q}$ are primal-dual solutions
- Solve $\boldsymbol{0} \in \boldsymbol{M}\boldsymbol{x} + \boldsymbol{Q}\boldsymbol{x}$, where $\boldsymbol{x} = (x, v_1, \ldots, v_p)$ via Tseng's forward-backward-forward splitting algorithm

$$
\left\lfloor
\begin{array}{l}
\boldsymbol{y}_n = \boldsymbol{x}_n - \boldsymbol{Q}\boldsymbol{x}_n \\
\boldsymbol{p}_n = (\mathrm{Id} + \boldsymbol{M})^{-1} \boldsymbol{y}_n \\
\boldsymbol{q}_n = \boldsymbol{p}_n - \boldsymbol{Q}\boldsymbol{p}_n \\
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \boldsymbol{y}_n + \boldsymbol{q}_n
\end{array}
\right.
$$

in $\mathcal{K}$ to get...

Proximal splitting methods in convex optimization

- Algorithm:

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
y_{1,n} = x_n - \left(\nabla h(x_n) + \sum_{k=1}^{m} L_k^* v_{k,n}\right) \\
p_{1,n} = \operatorname{prox}_f y_{1,n} \\
\text{For } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
y_{2,k,n} = v_{k,n} + (L_k x_n - \nabla \ell_k^*(v_{k,n})) \\
p_{2,k,n} = \operatorname{prox}_{g_k^*}(y_{2,k,n} - r_k) \\
q_{2,k,n} = p_{2,k,n} + (L_k p_{1,n} - \nabla \ell_k^*(p_{2,k,n})) \\
v_{k,n+1} = v_{k,n} - y_{2,k,n} + q_{2,k,n}
\end{array}
\right. \\
q_{1,n} = p_{1,n} - \left(\nabla h(p_{1,n}) + \sum_{k=1}^{m} L_k^* p_{2,k,n}\right) \\
x_{n+1} = x_n - y_{1,n} + q_{1,n}
\end{array}
\right.
\end{array}
$$

- $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution and $((v_{k,n})_{1\leqslant k\leqslant p})_{n\in\mathbb{N}}$ 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

Composite convex optimization problem

- Let $\boldsymbol{F}$ be the set of solutions to the problem

$$\underset{x_i\in\mathcal{H}_i,\, i\in I}{\text{minimize}} \; \sum_{i\in I}\big(f_i(x_i) - \langle x_i \mid z_i^*\rangle\big) + \sum_{k\in K} g_k\Big(\sum_{i\in I} L_{ki}x_i - r_k\Big)$$

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

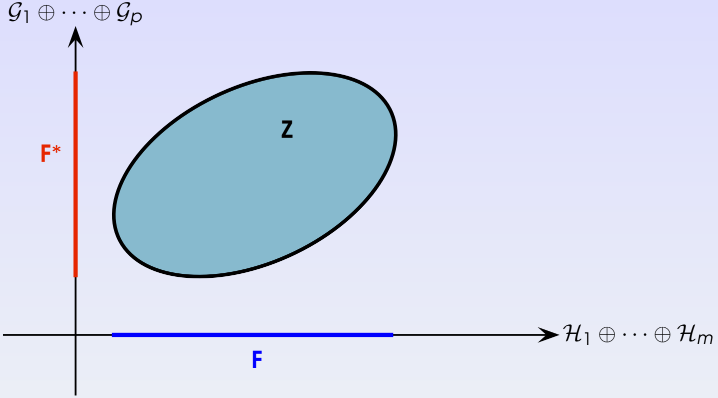
- Let $\boldsymbol{F}^*$ be the set of solutions to the dual problem

$$\underset{v_k^*\in\mathcal{G}_k,\, k\in K}{\text{minimize}} \; \sum_{i\in I} f_i^*\Big(z_i^* - \sum_{k\in K} L_{ki}^* v_k^*\Big) + \sum_{k\in K}\big(g_k^*(v_k^*) + \langle v_k^* \mid r_k\rangle\big)$$

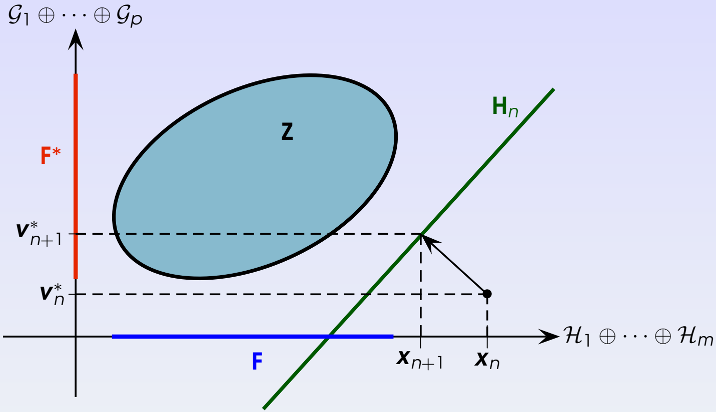
- Associated Kuhn-Tucker set (set of zeros a maximally monotone operator which is **not** a subdifferential)

$$\boldsymbol{Z} = \Big\{ \big((\overline{x}_i)_{i\in I}, (\overline{v}_k^*)_{k\in K}\big) \;\Big|\; \overline{x}_i \in \mathcal{H}_i \;\text{ and }\; z_i^* - \sum_{k\in K} L_{ki}^*\overline{v}_k^* \in \partial f_i(\overline{x}_i), \quad \overline{v}_k^* \in \mathcal{G}_k \;\text{ and }\; \sum_{i\in I} L_{ki}\overline{x}_i - r_k \in \partial g_k^*(\overline{v}_k^*) \Big\}$$

Underlying geometry: The Kuhn-Tucker set



Underlying geometry: The Kuhn-Tucker set

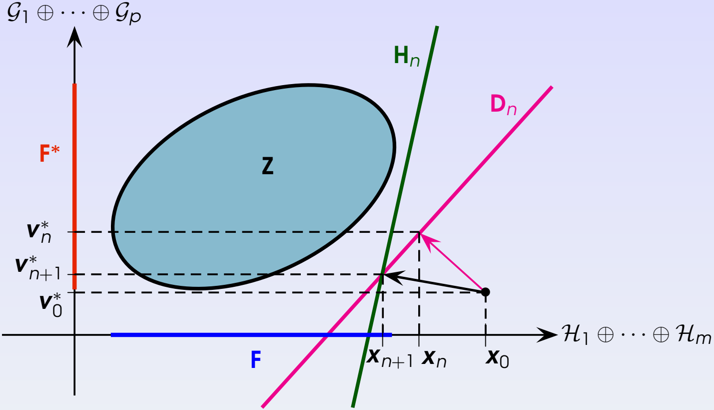


- Choose suitable points in the graphs of $(\partial f_i)_{i\in I}$ and $(\partial g_k)_{k\in K}$ to construct a half-space $\mathbf{H}_n$ containing $\mathbf{Z}$
- Algorithm: $(\boldsymbol{x}_{n+1}, \boldsymbol{v}^*_{n+1}) = P_{\mathbf{H}_n}(\boldsymbol{x}_n, \boldsymbol{v}^*_n) \rightharpoonup (\boldsymbol{x}, \boldsymbol{v}^*) \in \mathbf{Z} \subset \mathbf{F} \times \mathbf{F}^*$

Asynchronous block-iterative proximal splitting (PLC/Eckstein, 2018)

for $n = 0, 1, \ldots$
- for every $i \in I_n$
  - $l_{i,n}^* = \sum_{k\in K} L_{ki}^* v_{k,c_i(n)}^*$
  - $(a_{i,n}, a_{i,n}^*) = \big(\mathrm{prox}_{\gamma_{i,c_i(n)} f_i}(x_{i,c_i(n)} + \gamma_{i,c_i(n)}(z_i - l_{i,n}^*)), \gamma_{i,c_i(n)}^{-1}(x_{i,c_i(n)} - a_{i,n}) - l_{i,n}^*\big)$
- for every $i \in I \smallsetminus I_n$
  - $(a_{i,n}, a_{i,n}^*) = (a_{i,n-1}, a_{i,n-1}^*)$
- for every $k \in K_n$
  - $l_{k,n} = \sum_{i\in I} L_{ki} x_{i,d_k(n)}$
  - $(b_{k,n}, b_{k,n}^*) = \big(r_k + \mathrm{prox}_{\mu_{k,d_k(n)} g_k}(l_{k,n} + \mu_{k,d_k(n)} v_{k,d_k(n)}^* - r_k), v_{k,d_k(n)}^* + \mu_{k,d_k(n)}^{-1}(l_{k,n} - b_{k,n})\big)$
- for every $k \in K \smallsetminus K_n$
  - $(b_{k,n}, b_{k,n}^*) = (b_{k,n-1}, b_{k,n-1}^*)$
- $((t_{i,n}^*)_{i\in I}, (t_{k,n})_{k\in K}) = ((a_{i,n}^* + \sum_{k\in K} L_{ki}^* b_{k,n}^*)_{i\in I}, (b_{k,n} - \sum_{i\in I} L_{ki} a_{i,n})_{k\in K})$
- $\tau_n = \sum_{i\in I} \|t_{i,n}^*\|^2 + \sum_{k\in K} \|t_{k,n}\|^2$
- if $\tau_n > 0$
  - $$\theta_n = \frac{\lambda_n}{\tau_n} \max\Big\{0, \sum_{i\in I} (\langle x_{i,n} \mid t_{i,n}^*\rangle - \langle a_{i,n} \mid a_{i,n}^*\rangle) + \sum_{k\in K} (\langle t_{k,n} \mid v_{k,n}^*\rangle - \langle b_{k,n} \mid b_{k,n}^*\rangle)\Big\}$$
- 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 $\mathbf{H}_n$ as before
- The half-space $\mathbf{D}_n$ satisfies $(\boldsymbol{x}_n, \boldsymbol{v}_n^*) = P_{\mathbf{D}_n}(\boldsymbol{x}_0, \boldsymbol{v}_0^*)$
- Algorithm: $(\boldsymbol{x}_{n+1}, \boldsymbol{v}_{n+1}^*) = P_{\mathbf{H}_n \cap \mathbf{D}_n}(\boldsymbol{x}_0, \boldsymbol{v}_0^*) \to P_{\mathbf{Z}}(\boldsymbol{x}_0, \boldsymbol{v}_0^*) \in \mathbf{F} \times \mathbf{F}^*$

Outlook

- Just like in the early 1960s the frontier separating linear from noninear 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 extrememely challenging task.

References

- PLC, Monotone operator theory in convex optimization, *Math. Program.*, vol. B170, 2018.
- H. H. Bauschke and PLC, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. Springer, New York, 2017.
- PLC and J. Eckstein, Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions, *Math. Program.*, vol. B168, 2018.
- PLC and L. E. Glaudin, Proximal activation of smooth functions in splitting algorithms for convex image recovery, arxiv, 2018.
- L. M. Briceño-Arias and PLC, A monotone+skew splitting model for composite monotone inclusions in duality, *SIAM J. Optim.*, vol. 21, 2011.
- Chierchia, Chouzenoux, PLC, Pesquet, *Proximity Operator Repository*, `http://proximity-operator.net/`