Monotone Operator Theory in Convex Optimization

Patrick L. Combettes

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

International Symposium on Mathematical Programming
Bordeaux, July 4, 2018
Supported by the NSF under grant CCF-1715671

Notation

- $\mathcal{H}$, $\mathcal{G}$, etc are real Hilbert spaces
- $\mathscr{B}(\mathcal{H},\mathcal{G})$ is the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$; $\mathscr{B}(\mathcal{H}) = \mathscr{B}(\mathcal{H},\mathcal{H})$
- Synthetic problem: given $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$,

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) \tag{1}$$

Notation

- $\mathcal{H}$, $\mathcal{G}$, etc are real Hilbert spaces
- $\mathscr{B}(\mathcal{H},\mathcal{G})$ is the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$; $\mathscr{B}(\mathcal{H}) = \mathscr{B}(\mathcal{H},\mathcal{H})$
- Synthetic problem: given $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$,

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) \tag{1}$$

- *Convex optimization* refers to the case when f in (1) is proper, lower semicontinuous, and convex, which we denote by $f \in \Gamma_0(\mathcal{H})$

Notation

- $\mathcal{H}, \mathcal{G}$, etc are real Hilbert spaces
- $\mathscr{B}(\mathcal{H}, \mathcal{G})$ is the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$; $\mathscr{B}(\mathcal{H}) = \mathscr{B}(\mathcal{H}, \mathcal{H})$
- Synthetic problem: given $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$,

$$\underset{x \in \mathcal{H}}{\text{minimize}} \; f(x) \tag{1}$$

- *Convex optimization* refers to the case when f in (1) is proper, lower semicontinuous, and convex, which we denote by $f \in \Gamma_0(\mathcal{H})$
- We interpret (1) in the strict sense of producing a point in $\operatorname{Argmin} f$, not in the looser sense of making f small
  - Minimizing sequences may have little to do with actually approaching a point in $\operatorname{Argmin} f$ as we can have for $p > 2$ (even in $\mathcal{H} = \mathbb{R}^2$): $f(x_n) - \min f(\mathcal{H}) = 1/(n+1)^p \downarrow 0$ and $d_{\operatorname{Argmin} f}(x_n) \uparrow +\infty$

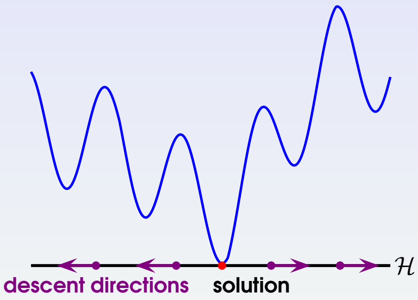
A few words on nonconvex minimization

1. Nonconvex optimization is an unstructured corpus of results, not a constructive theory

A few words on nonconvex minimization

1. Nonconvex optimization is an unstructured corpus of results, not a constructive theory

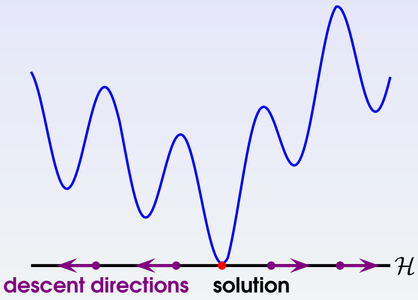
2. Moving permanently *away* from solutions in descent methods:



A few words on nonconvex minimization

1. Nonconvex optimization is an unstructured corpus of results, not a constructive theory

2. Moving permanently *away* from solutions in descent methods:

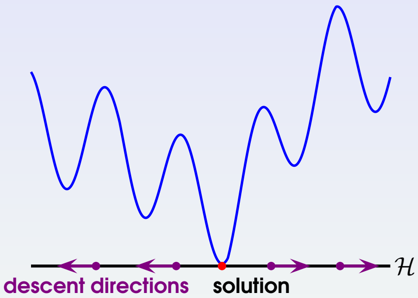


3. Loose connections with other branches of nonlinear analysis

A few words on nonconvex minimization

1. Nonconvex optimization is an unstructured corpus of results, not a constructive theory

2. Moving permanently *away* from solutions in descent methods:



3. Loose connections with other branches of nonlinear analysis

4. Algorithms may yield trivial solutions:

Let $f: \mathcal{H} \to \{0, \ldots, p\}$ be l.s.c. (e.g., rank etc.), let $C \neq \varnothing$. Then ***any*** point in C is a local minimizer of:

$$\underset{x \in C}{\text{minimize}} \;\; f(x)$$

J.-B. Hiriart-Urruty, When only global optimization matters, *J. Global Optim.*, vol. 56, pp. 761–763, 2013

Convex minimization

- Synthetic problem: given $f \in \Gamma_0(\mathcal{H})$,

$$\underset{x \in \mathcal{H}}{\text{minimize}} \; f(x)$$

Convex minimization

- Synthetic problem: given $f \in \Gamma_0(\mathcal{H})$,

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x)$$

- A more structured version

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

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

Convex minimization

- Synthetic problem: given $f \in \Gamma_0(\mathcal{H})$,

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x)$$

- A more structured version

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

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

- To analyze and solve such complex minimization problem, one must borrow tools from functional and numerical analysis

Convex minimization

- Synthetic problem: given $f \in \Gamma_0(\mathcal{H})$,

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x)$$

- A more structured version

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

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

- To analyze and solve such complex minimization problem, one must borrow tools from functional and numerical analysis

- Our main message is that *monotone operator theory* plays an increasingly central role in convex optimization and that both fields maintain a tight and productive interplay

Functional analysis: Historical overview

1950's

Linear functional analysis

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

Functional analysis: Historical overview

Nonlinear functional analysis: "anything not linear"

1950's

Linear functional analysis

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

Functional analysis: Historical overview

Early 1960's

Nonlinear functional analysis: outgrowths of linear analysis

Monotone operators

Convex analysis

Nonexpansive operators

1950's

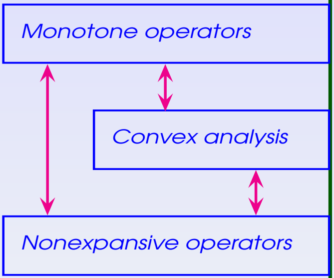
Linear functional analysis

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

Functional analysis: Historical overview

Early 1960's

Nonlinear functional analysis: outgrowths of linear analysis



1950's

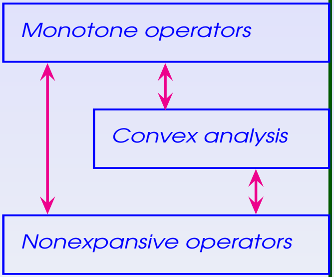
Linear functional analysis

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

Functional analysis: Historical overview

Early 1960's

Nonlinear functional analysis: outgrowths of linear analysis



1950's

Linear functional analysis

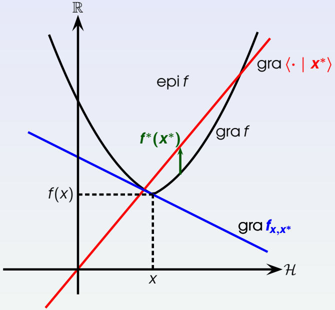
- *Topological vector spaces*
- *Linear operators*
- *Duality*
- *Theory of distributions*
- *etc.*

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

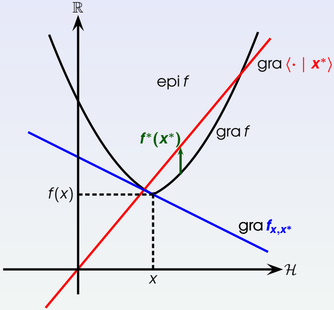
$$\partial f(x) = \Big\{x^* \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \underbrace{\langle y - x \mid x^*\rangle + f(x)}_{f_{x,x^*}(y)} \leqslant f(y)\Big\}$$



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

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



Infimal convolution of $f, g\colon \mathcal{H} \to]-\infty,+\infty]$:
$(f \,\square\, g)\colon x \mapsto \inf_{y\in\mathcal{H}} f(y) + g(x-y)$

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)

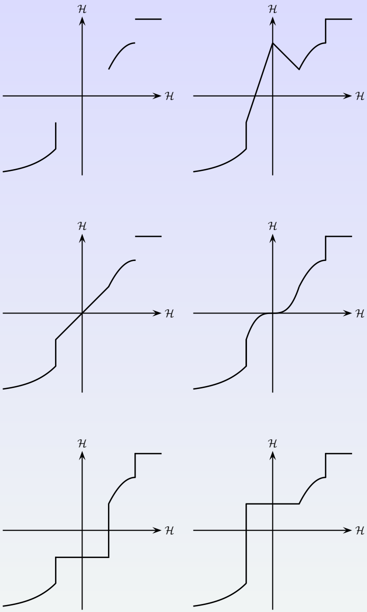
- $A \in \mathscr{B}(\mathcal{H})$ is skew if $(\forall x \in \mathcal{H})\ \langle x \mid Ax \rangle = 0$ and it is positive if $(\forall x \in \mathcal{H})\ \langle x \mid Ax \rangle \geqslant 0$, i.e.,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Ax - Ay \rangle \geqslant 0. \tag{2}$$

- In 1960, Kačurovskiĭ, Minty, and Zarantonello independently called *monotone* a nonlinear operator $A\colon \mathcal{H} \to \mathcal{H}$ that satisfies (2)
- 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 $\operatorname{gra} A \subset \operatorname{gra} B \neq \operatorname{gra} A$



monotone, not monotone

monotone, max. monotone

max. monotone, max. monotone

Minty's theorem: A monotone is max. monotone $\Leftrightarrow$ $\operatorname{ran}(\operatorname{Id} + A) = \mathcal{H}$

First examples of maximally monotone operators

- $A\colon \mathbb{R} \to \mathbb{R}$ is a continuous increasing function
- $A \in \mathscr{B}(\mathcal{H})$ is a skew operator
- (Moreau) $f \in \Gamma_0(\mathcal{H})$ and $A = \partial f$
- C is a nonempty closed convex subset of $\mathcal{H}$ and

$$(\forall x \in \mathcal{H}) \;\; Ax = \begin{cases} \big\{u \in \mathcal{H} \mid (\forall y \in C) \;\; \langle y - x \mid u \rangle \leqslant 0\big\} & \text{if } \; x \in C \\ \varnothing & \text{if } \; x \notin C \end{cases}$$

  A is the normal cone operator of C

- V is a closed vector subspace of $\mathcal{H}$ and

$$(\forall x \in \mathcal{H}) \;\; Ax = \begin{cases} V^{\perp} & \text{if } x \in V \\ \varnothing & \text{if } x \notin V \end{cases}$$

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
- 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

Convexity/Nonexpansiveness/Monotonicity

- If $f \in \Gamma_0(\mathcal{H})$, $A = \partial f$ is maximally monotone
- If $T\colon \mathcal{H} \to \mathcal{H}$ is nonexpansive, $A = \mathrm{Id} - T$ is max. mon. and $\operatorname{Fix} T = \{x \in \mathcal{H} \mid Tx = x\}$ 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
- (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 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

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 Operators

- 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. A number of additional "proximity preserving" transformations are identified in the accompanying paper (PLC, 2018), which lead to:
  - A new example of weakly but not strongly convergent proximal iteration
  - New explicit expressions for proximity operators of certain composite functions
  - A study of self-dual classes of firmly nonexpansive operators

The need for monotone operators in optimization

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

The need for monotone operators in optimization

- They of course 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)

$$x_{n+1} = \operatorname{prox}_{\gamma_n f} x_n \rightharpoonup x \in \operatorname{Argmin} f \quad \text{if} \quad \sum_{n\in\mathbb{N}} \gamma_n = +\infty \tag{3}$$

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

$$x_{n+1} = J_{\gamma_n A} x_n \rightharpoonup x \in \operatorname{zer} A \quad \text{if} \quad \sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty \tag{4}$$

- Note that (3) has more general parameters. However (4) is considerably more useful **to optimization** than (3)

Interplay: The proximal point algorithm

- (Rockafellar, 1976) Applying the general proximal point algorithm (4) to the saddle operator leads to various minimization algorithms (e.g., the proximal method of multipliers in the case of the ordinary Lagrangian)
- Applying the general proximal point algorithm (4) to the partial inverse of a suitably constructed partial inverse makes it possible to solve (Alghamdi, Alotaibi, PLC, Shahzad, 2014)

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

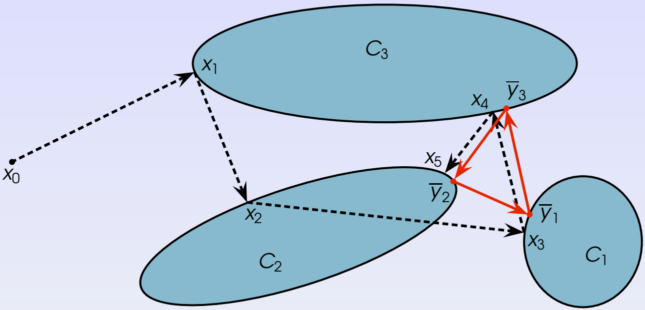
The need for monotone operators in optimization

- They of course 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 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

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

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 of course 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

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:

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

where: $\varphi_k \,\square\, \ell_k \colon x \mapsto \inf_{y \in \mathcal{H}} \big(\varphi_k(y) + \ell_k(x - y)\big)$

- Example: multiview total variation image recovery from observations $r_k = L_k \overline{x} + w_k$:

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; \sum_{k \in \mathbb{N}} \phi_k(\langle x \mid e_k \rangle) + \sum_{k=1}^{p-1} \alpha_k \underbrace{d_{C_k}}_{\iota_C \,\square\, \|\cdot\|} (L_k x - r_k) + \beta \|\nabla x\|_{1,2}$$

- 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: $\boldsymbol{Q}\colon \mathcal{K} \to \mathcal{K}\colon (x, v_1, \ldots, v_p) \mapsto$
$$\Big(Cx + \textstyle\sum_{k=1}^{p} L_k^* v_k, -L_1x+D_1^{-1}v_1, \ldots, -L_px+D_p^{-1}v_p\Big)$$
- $\boldsymbol{M}$ and $\boldsymbol{Q}$ are maximally monotone, $\boldsymbol{Q}$ is Lipschitzian, the zeros of $\boldsymbol{M}+\boldsymbol{Q}$ are primal-dual solutions
- Solve $\mathbf{0} \in \boldsymbol{M}\boldsymbol{x} + \boldsymbol{Q}\boldsymbol{x}$, 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:

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

- $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution (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 $f_i \in \Gamma_0(\mathcal{H}_i)$, $g_k \in \Gamma_0(\mathcal{G}_k)$, $L_{ki} \in \mathscr{B}(\mathcal{H}_i, \mathcal{G}_k)$

Composite convex optimization problem

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

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

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

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

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

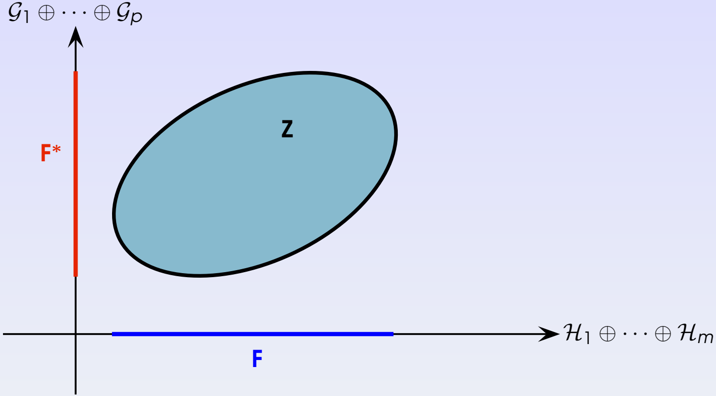
- Let $\boldsymbol{F}^*$ be the set of solutions to the dual problem

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

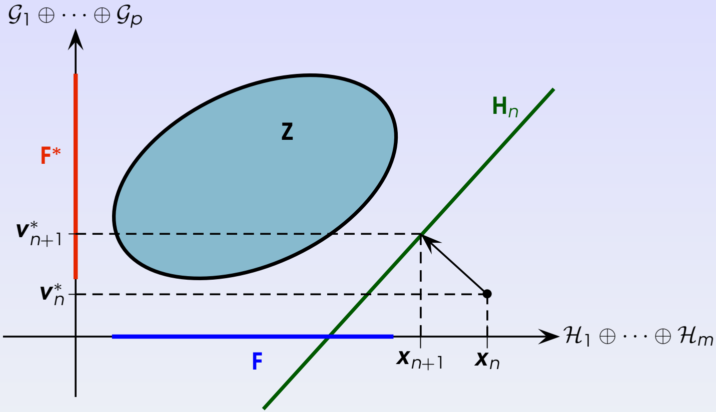
- Associated Kuhn-Tucker set

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

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 H_n containing Z
- Algorithm: $(x_{n+1}, v^*_{n+1}) = P_{\mathsf{H}_n}(x_n, v^*_n) \rightharpoonup (x, v^*) \in \mathsf{Z} \subset \mathsf{F} \times \mathsf{F}^*$

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

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\quad \text{for every } i \in I_n \\
\quad\quad l_{i,n}^* = \sum_{k\in K} L_{ki}^* v_{k,c_i(n)}^* \\
\quad\quad (a_{i,n}, a_{i,n}^*) = \Big(\operatorname{prox}_{\gamma_{i,c_i(n)} f_i}\big(x_{i,c_i(n)} + \gamma_{i,c_i(n)}(z_i - l_{i,n}^*)\big), \gamma_{i,c_i(n)}^{-1}(x_{i,c_i(n)} - a_{i,n}) - l_{i,n}^*\Big) \\
\quad \text{for every } i \in I \smallsetminus I_n \\
\quad\quad (a_{i,n}, a_{i,n}^*) = (a_{i,n-1}, a_{i,n-1}^*) \\
\quad \text{for every } k \in K_n \\
\quad\quad l_{k,n} = \sum_{i\in I} L_{ki} x_{i,d_k(n)} \\
\quad\quad (b_{k,n}, b_{k,n}^*) = \Big(r_k + \operatorname{prox}_{\mu_{k,d_k(n)} g_k}\big(l_{k,n} + \mu_{k,d_k(n)} v_{k,d_k(n)}^* - r_k\big), v_{k,d_k(n)}^* + \mu_{k,d_k(n)}^{-1}(l_{k,n} - b_{k,n})\Big) \\
\quad \text{for every } k \in K \smallsetminus K_n \\
\quad\quad (b_{k,n}, b_{k,n}^*) = (b_{k,n-1}, b_{k,n-1}^*) \\
\quad \big((t_{i,n}^*)_{i\in I}, (t_{k,n})_{k\in K}\big) = \big((a_{i,n}^* + \sum_{k\in K} L_{ki}^* b_{k,n}^*)_{i\in I}, (b_{k,n} - \sum_{i\in I} L_{ki} a_{i,n})_{k\in K}\big) \\
\quad \tau_n = \sum_{i\in I} \|t_{i,n}^*\|^2 + \sum_{k\in K} \|t_{k,n}\|^2 \\
\quad \text{if } \tau_n > 0 \\
\quad\quad \theta_n = \dfrac{\lambda_n}{\tau_n} \max\Big\{0, \sum_{i\in I} \big(\langle x_{i,n} \mid t_{i,n}^*\rangle - \langle a_{i,n} \mid a_{i,n}^*\rangle\big) + \sum_{k\in K} \big(\langle t_{k,n} \mid v_{k,n}^*\rangle - \langle b_{k,n} \mid b_{k,n}^*\rangle\big)\Big\} \\
\quad \text{else } \theta_n = 0 \\
\quad \text{for every } i \in I \\
\quad\quad x_{i,n+1} = x_{i,n} - \theta_n t_{i,n}^* \\
\quad \text{for every } k \in K \\
\quad\quad v_{k,n+1}^* = v_{k,n}^* - \theta_n t_{k,n}
\end{array}
$$

Numerical example

The problem is to

$$\underset{x\in C}{\text{minimize}}\ \ 6\|\nabla x\|_{1,2} + 5d_D^2(x) + 10\|H_1x - y_1\|_2^2 + 10\|H_2x - y_2\|_2^2,$$

where

- $C = [0, 255]^N$, $N = 128 \times 128$
- $\|\nabla\|_{1,2} \colon \mathbb{R}^N \to \mathbb{R}$ is the total variation
- $D = \big\{x \in \mathbb{R}^N \mid \widehat{x}1_K = \widehat{\overline{x}}1_K\big\}$ where the set K contains the frequencies in $\{0, \ldots, \sqrt{N}/8 - 1\}^2$ (+ symmetries)
- H_1 and H_2 model convolution blurs of size 3×11 and 7×5, y_1 and y_2 are noisy observations

Fully proximal implementations of splitting algorithms

- The problem

$$\underset{x \in C}{\text{minimize}} \; 6\|\nabla x\|_{1,2} + 5d_D^2(x) + 10\|H_1 x - y_1\|^2 + 10\|H_2 x - y_2\|_2^2$$

  contains 3 smooth terms

- However each prox_{g_k} in the $g_k \circ L_k$ terms has an explicit prox
- Although some of the primal dual FB and FBF (see below) algorithms can exploit smoothness, a fully proximal implementation turned out to be faster

We apply these algorithms with

- $f = \iota_C$
- $g_1 = 6\|\cdot\|_{1,2}$ and $L_1 = \nabla$
- $g_2 = 5d_D^2$ and $L_2 = \mathrm{Id}$
- $g_3 = 10\|H_1 \cdot - y_1\|_2^2$ and $L_3 = \mathrm{Id}$
- $g_4 = 10\|H_2 \cdot - y_2\|_2^2$ and $L_4 = \mathrm{Id}$

and

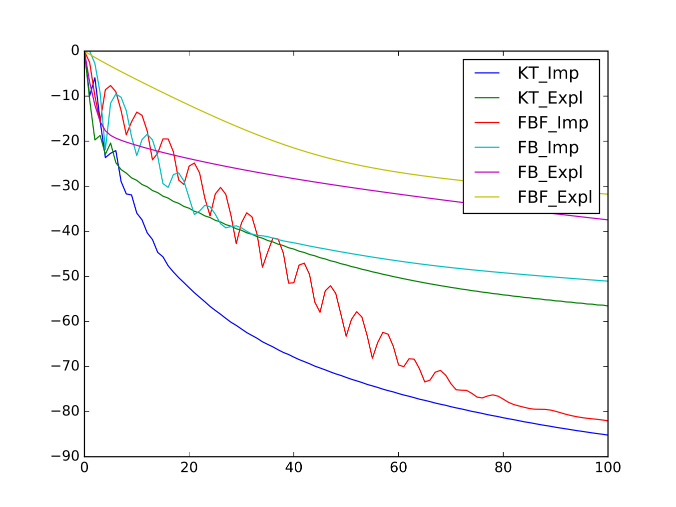
- FBF_Imp, FB_Imp, and KT_Imp: $I = \{1, 2, 3, 4\}$ and $J = \varnothing$
- FBF_Expl, FB_Expl, and KT_Expl: $I = \{1\}$ and $J = \{2, 3, 4\}$

Parameters

Set $\beta = \sqrt{\sum_{k\in K} \|L_k\|^2}$ and:

- FBF_Imp: $\gamma_n \equiv 0.99\beta$
- FBF_Expl: $\gamma_n \equiv 0.99\beta$
- FB_Imp: $\sigma_{1,n} \equiv 3/(2\beta)$, $\sigma_{2,n} \equiv 3/(2\beta)$, $\sigma_{3,n} \equiv 1/(10\beta)$, $\sigma_{4,n} \equiv 1/(10\beta)$, $\tau_n \equiv 1/\beta$, and $\lambda_n \equiv 1$
- FB_Expl: $\sigma_{1,n} \equiv 3/(2\beta)$, $\tau_n \equiv 1/(10\beta)$, and $\lambda_n \equiv 1$
- KT_Imp : $\gamma_n \equiv 0.4$, $\mu_{1,n} \equiv 1$, $\mu_{2,n} \equiv 1$, $\mu_{3,n} \equiv 1$,$\mu_{4,n} \equiv 3/2$, and $\lambda_n \equiv 1$
- KT_Expl : $\gamma_n \equiv 1.5$, $\mu_{1,n} \equiv 0.04$, $\mu_{2,n} \equiv 0.04$, $\mu_{3,n} \equiv 0.09$, $\mu_{4,n} \equiv 0.5$, and $\lambda_n \equiv 1$

Numerical results



Distance to respective solution

Numerical results

Original

Degraded 1

Degraded 2

Restored

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 minimization, arxiv, 2018.
- Chierchia, Chouzenoux, PLC, Pesquet, *Proximity Operator Repository*, `http://proximity-operator.net/`