Back to Single-Resolvent Iterations, with Warping

Patrick L. Combettes

– joint work with Minh N. Bùi –

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

One World Optimization Seminar, May 25, 2020

NC STATE UNIVERSITY

Outline

- Part 1: Background
- Part 2: The warped resolvent
- Part 3: Warped proximal iterations in Hilbert spaces
- Part 4: Warped proximal iterations with Bregman kernels

Background

PART 1:

Background

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$.

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$.
- Considerable range of applications: optimization,
  - Subdifferential: $M = \partial f$ (Fermat's rule)
  - Kuhn-Tucker operator: $M = \begin{bmatrix} \partial f & L^* \\ -L & \partial g^* \end{bmatrix}$.

    (Rockafellar 1967)
  - etc. (Eckstein 1994, PLC 2018, Bùi/PLC 2020).

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$.
- Considerable range of applications: optimization, variational inequalities, statistics, mechanics, neural networks, finance, partial differential equations, optimal transportation, signal and image processing, control, game theory, machine learning, economics, mean fields games, etc.

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$.
- The proximal point algorithm (Bellman 1966, Martinet 1970, Rockafellar 1976):

$$x_{n+1} = J_M x_n, \text{ where } J_M = (\mathrm{Id} + M)^{-1} \text{ is the resolvent of } M.$$

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$. The proximal point algorithm (Bellman 1966, Martinet 1970, Rockafellar 1976):

$$x_{n+1} = J_M x_n, \text{ where } J_M = (\mathrm{Id} + M)^{-1} \text{ is the resolvent of } M.$$

- Acknowledging the fact that J_M may be hard to implement, *splitting methods* have been developed: the goal is to express M as a combination of operators, and devise an algorithm that uses these operators individually.

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$. The proximal point algorithm (Bellman 1966, Martinet 1970, Rockafellar 1976):

$$x_{n+1} = J_M x_n, \quad \text{where } J_M = (\mathrm{Id} + M)^{-1} \text{ is the resolvent of } M.$$

- *Splitting methods*: express M as a combination of operators, and devise an algorithm that uses these operators individually.
- The following structures have been considered:

$$M = A + B$$

(Mercier 1979, Lions/Mercier 1979, Tseng 2000)

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$. The proximal point algorithm (Bellman 1966, Martinet 1970, Rockafellar 1976):

$$x_{n+1} = J_M x_n, \text{ where } J_M = (\mathrm{Id} + M)^{-1} \text{ is the resolvent of } M.$$

- *Splitting methods*: express M as a combination of operators, and devise an algorithm that uses these operators individually.
- The following structures have been considered:

$$M = \sum_{k=1}^{p} A_k$$

(Spingarn 1983, Gol'stein 1985, Eckstein/Svaiter 2009, PLC 2009)

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$. The proximal point algorithm (Bellman 1966, Martinet 1970, Rockafellar 1976):

$$x_{n+1} = J_M x_n, \text{ where } J_M = (\mathrm{Id} + M)^{-1} \text{ is the resolvent of } M.$$

- *Splitting methods*: express M as a combination of operators, and devise an algorithm that uses these operators individually.
- The following structures have been considered:

$$M = \sum_{k=1}^{p} L_k^* \circ B_k \circ L_k$$

(Briceño-Arias/PLC 2011)

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$. The proximal point algorithm (Bellman 1966, Martinet 1970, Rockafellar 1976):

$$x_{n+1} = J_M x_n, \text{ where } J_M = (\mathrm{Id} + M)^{-1} \text{ is the resolvent of } M.$$

- *Splitting methods*: express M as a combination of operators, and devise an algorithm that uses these operators individually.
- The following structures have been considered:

$$M = A + \sum_{k=1}^{p} L_k^* \circ (B_k \,\square\, D_k) \circ L_k + C$$

(PLC/Pesquet 2012, Vũ 2013, Condat 2013, Boţ/Hendrich 2013)

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$. The proximal point algorithm (Bellman 1966, Martinet 1970, Rockafellar 1976):

$$x_{n+1} = J_M x_n, \text{ where } J_M = (\mathrm{Id} + M)^{-1} \text{ is the resolvent of } M.$$

- *Splitting methods*: express M as a combination of operators, and devise an algorithm that uses these operators individually.
- The following structures have been considered:

$$M = A + \sum_{k=1}^{p} B_k + C$$

(Raguet/Fadili/Peyré 2013, Briceño-Arias 2015, Davis/Yin 2017, Raguet 2019)

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$. The proximal point algorithm (Bellman 1966, Martinet 1970, Rockafellar 1976):

$$x_{n+1} = J_M x_n, \text{ where } J_M = (\mathrm{Id} + M)^{-1} \text{ is the resolvent of } M.$$

- *Splitting methods*: express M as a combination of operators, and devise an algorithm that uses these operators individually.
- The following structures have been considered:

$$M\colon (x_1, \ldots, x_m) \mapsto \bigtimes_{i=1}^{m} \Bigg(A_i x_i + C_i x_i + Q_i x_i + \sum_{k=1}^{p} L_{ki}^{*}\bigg(\Big(\big(B_k^m + B_k^c + B_k^l\big) \,\square\, \big(D_k^m + D_k^c + D_k^l\big)\Big)\Big(\sum_{j=1}^{m} L_{kj} x_j\Big)\bigg)\Bigg)$$

(Bùi/PLC 2020)

Monotone operator splitting in Hilbert spaces

- Basic problem: Given a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, find $x \in \mathcal{X}$ such that $0 \in Mx$. The proximal point algorithm (Bellman 1966, Martinet 1970, Rockafellar 1976):

$$x_{n+1} = J_M x_n, \quad \text{where } J_M = (\mathrm{Id} + M)^{-1} \text{ is the resolvent of } M.$$

- *Splitting methods*: express M as a combination of operators, and devise an algorithm that uses these operators individually.
- ... which models in particular

$$\underset{x_1 \in \mathcal{X}_1, \ldots, x_m \in \mathcal{X}_m}{\text{minimize}} \quad \sum_{i=1}^{m} \big(f_i(x_i) + \varphi_i(x_i)\big) + \sum_{k=1}^{p} \big((g_k + \psi_k) \,\square\, h_k\big)\Bigg(\sum_{j \in I} L_{kj} x_j\Bigg).$$

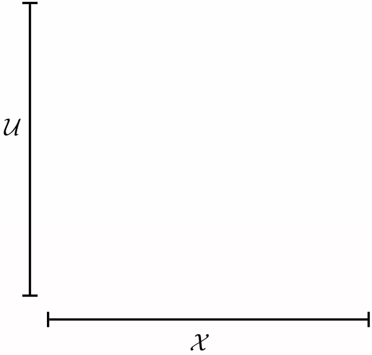
(Bùi/PLC 2020)

Monotone operator splitting

- The field has evolved in many exciting directions and various algorithms are now available for complex structured problems, together with block-coordinate, block-iterative, and asynchronous implementations.
- A common feature of these developments is to move away from single-resolvent iterations such as the proximal point algorithm.
- We introduce an extended notion of a resolvent, called warped resolvent, and show that considering the warped resolvent iterations of a single operator provides a surprisingly broad platform to not only recover existing schemes in a synthetic framework, but also design new ones.

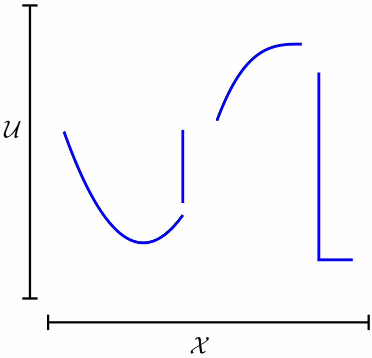
Set-valued operators

- $\mathcal{X}$ and $\mathcal{U}$ nonempty sets, $2^{\mathcal{U}}$ the power set of $\mathcal{U}$.



Set-valued operators

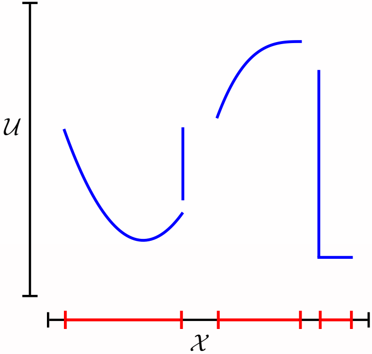
- $\mathcal{X}$ and $\mathcal{U}$ nonempty sets, $2^{\mathcal{U}}$ the power set of $\mathcal{U}$.
- $M\colon \mathcal{X} \to 2^{\mathcal{U}} : x \mapsto Mx \subset \mathcal{U}$ a set-valued operator.



graph of M: gra $M = \big\{(x,u) \in \mathcal{X} \times \mathcal{U} \mid u \in Mx\big\}$.

Set-valued operators

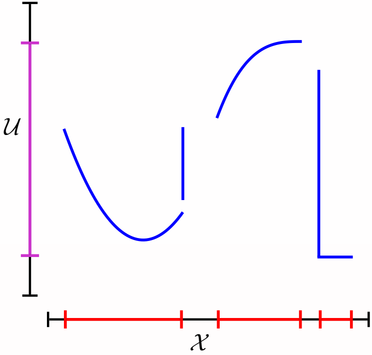
- $\mathcal{X}$ and $\mathcal{U}$ nonempty sets, $2^{\mathcal{U}}$ the power set of $\mathcal{U}$.
- $M\colon \mathcal{X} \to 2^{\mathcal{U}} : x \mapsto Mx \subset \mathcal{U}$ a set-valued operator.



domain of M: dom $M = \{x \in \mathcal{X} \mid Mx \neq \varnothing\}$.

Set-valued operators

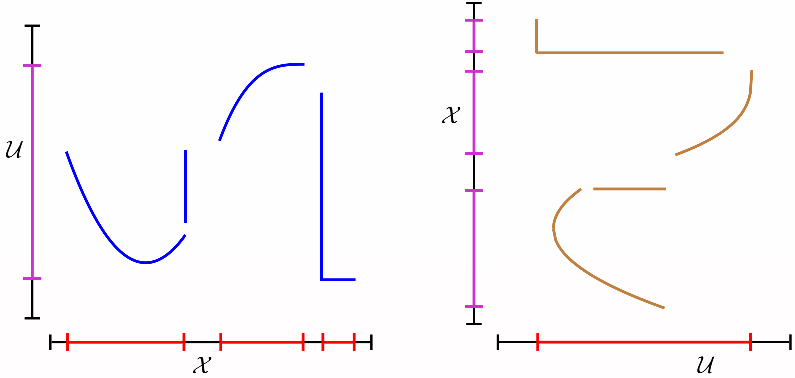
- $\mathcal{X}$ and $\mathcal{U}$ nonempty sets, $2^{\mathcal{U}}$ the power set of $\mathcal{U}$.
- $M\colon \mathcal{X} \to 2^{\mathcal{U}} : x \mapsto Mx \subset \mathcal{U}$ a set-valued operator.



range of M: ran $M = \bigcup_{x\in \operatorname{dom} M} Mx$.

Set-valued operators

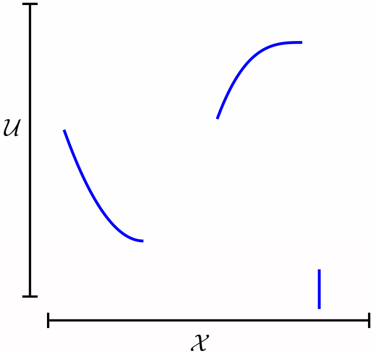
- $\mathcal{X}$ and $\mathcal{U}$ nonempty sets, $2^{\mathcal{U}}$ the power set of $\mathcal{U}$.
- $M\colon \mathcal{X} \to 2^{\mathcal{U}} : x \mapsto Mx \subset \mathcal{U}$ a set-valued operator.



inverse of M: gra $M^{-1} = \{(u,x) \in \mathcal{U} \times \mathcal{X} \mid u \in Mx\}$.

Set-valued operators

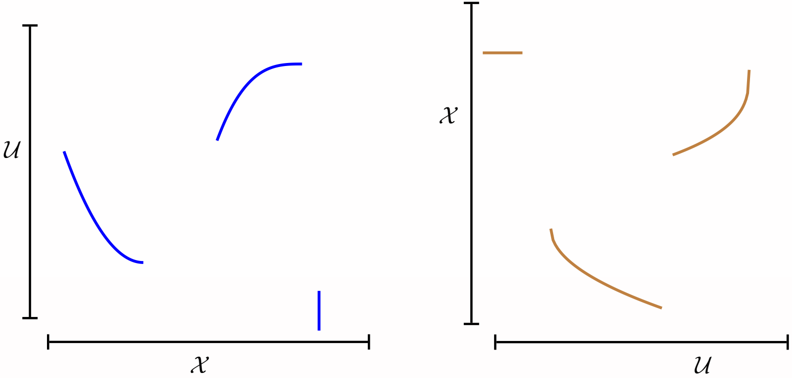
- $\mathcal{X}$ and $\mathcal{U}$ nonempty sets, $2^{\mathcal{U}}$ the power set of $\mathcal{U}$.
- $M\colon \mathcal{X} \to 2^{\mathcal{U}}$ a set-valued operator.



M is injective if : $(\forall x \in \mathcal{X})(\forall y \in \mathcal{X})\ \ Mx \cap My \neq \varnothing \ \ \Rightarrow \ \ x = y.$

Set-valued operators

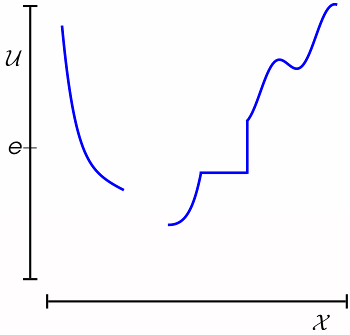
- $\mathcal{X}$ and $\mathcal{U}$ nonempty sets, $2^{\mathcal{U}}$ the power set of $\mathcal{U}$.
- $M\colon \mathcal{X} \to 2^{\mathcal{U}}$ a set-valued operator.



M is injective if : $(\forall x \in \mathcal{X})(\forall y \in \mathcal{X})\;\; Mx \cap My \neq \varnothing \;\;\Rightarrow\;\; x = y$.
This implies that M^{-1} is at most single-valued.

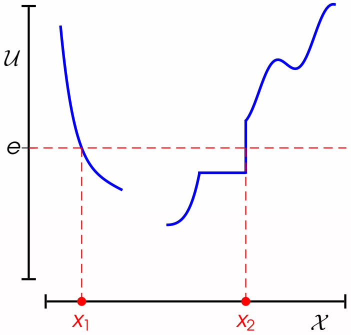
Problem model: Solving set-valued inclusions

- $\mathcal{X}$ a set, $(\mathcal{U}, \boxplus)$ a group with identity e, $M\colon \mathcal{X} \to 2^{\mathcal{U}}$.



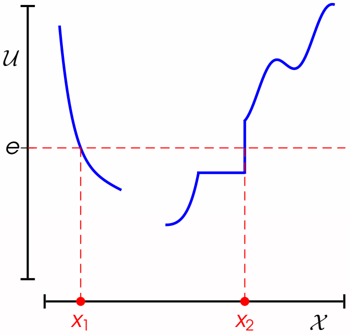
Problem model: Solving set-valued inclusions

- $\mathcal{X}$ a set, $(\mathcal{U}, \boxplus)$ a group with identity e, $M\colon \mathcal{X} \to 2^{\mathcal{U}}$.
- **Objective**: Find a point in $Z = \{x \in \mathcal{X} \mid e \in Mx\}$.



Problem model: Solving set-valued inclusions

- $\mathcal{X}$ a set, $(\mathcal{U}, \boxplus)$ a group with identity e, $M\colon \mathcal{X} \to 2^{\mathcal{U}}$.
- **Objective:** Find a point in $Z = \{x \in \mathcal{X} \mid e \in Mx\}$.
- Take $K\colon \mathcal{X} \to \mathcal{U}$ such that $K \boxplus M\colon x \mapsto \{Kx \boxplus u \mid u \in Mx\}$ is injective.

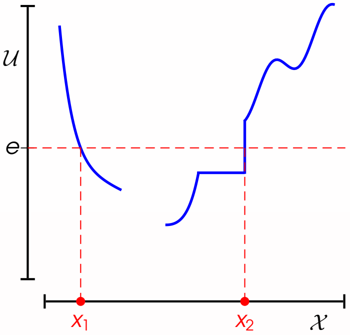


Problem model: Solving set-valued inclusions

- $\mathcal{X}$ a set, $(\mathcal{U}, \boxplus)$ a group with identity e, $M\colon \mathcal{X} \to 2^{\mathcal{U}}$.
- **Objective:** Find a point in $Z = \{x \in \mathcal{X} \mid e \in Mx\}$.
- Take $K\colon \mathcal{X} \to \mathcal{U}$ such that $K \boxplus M\colon x \mapsto \{Kx \boxplus u \mid u \in Mx\}$ is injective.

- Clearly,

$$\begin{aligned} x \in Z &\Leftrightarrow e \in Mx \\ &\Leftrightarrow Kx \in Kx \boxplus Mx \\ &\Leftrightarrow x = (K \boxplus M)^{-1}(Kx). \end{aligned}$$

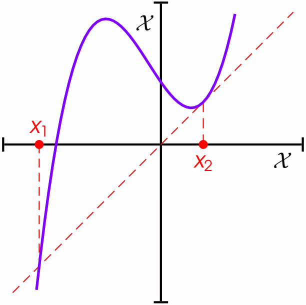


Problem model: Solving set-valued inclusions

- $\mathcal{X}$ a set, $(\mathcal{U}, \boxplus)$ a group with identity e, $M\colon \mathcal{X} \to 2^{\mathcal{U}}$.
- **Objective:** Find a point in $Z = \{x \in \mathcal{X} \mid e \in Mx\}$.
- Take $K\colon \mathcal{X} \to \mathcal{U}$ such that $K \boxplus M\colon x \mapsto \{Kx \boxplus u \mid u \in Mx\}$ is injective.

- Clearly,

$$\begin{aligned} x \in Z &\Leftrightarrow e \in Mx \\ &\Leftrightarrow Kx \in Kx \boxplus Mx \\ &\Leftrightarrow x = (K \boxplus M)^{-1}(Kx). \end{aligned}$$



- Thus $Z = \operatorname{Fix} J_M^K$, where

$$J_M^K = (K \boxplus M)^{-1} \circ K$$

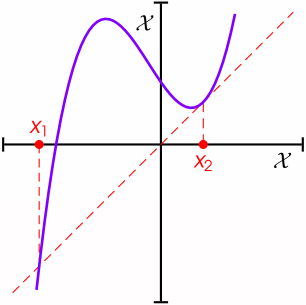
is the *warped resolvent* of M with *kernel* K.

Problem model: Solving set-valued inclusions

- $\mathcal{X}$ a set, $(\mathcal{U}, \boxplus)$ a group with identity e, $M\colon \mathcal{X} \to 2^{\mathcal{U}}$.
- **Objective:** Find a point in $Z = \{x \in \mathcal{X} \mid e \in Mx\}$.
- Take $K\colon \mathcal{X} \to \mathcal{U}$ such that $K \boxplus M\colon x \mapsto \{Kx \boxplus u \mid u \in Mx\}$ is injective.

- Clearly,

$$\begin{aligned} x \in Z &\Leftrightarrow e \in Mx \\ &\Leftrightarrow Kx \in Kx \boxplus Mx \\ &\Leftrightarrow x = (K \boxplus M)^{-1}(Kx). \end{aligned}$$



- Thus $Z = \operatorname{Fix} J_M^K$, where

$$J_M^K = (K \boxplus M)^{-1} \circ K$$

is the *warped resolvent* of M with *kernel* K.

- $p = J_M^K x \Leftrightarrow (p, Kx \boxminus Kp) \in \operatorname{gra} M$.

The warped resolvent

PART 2:

The warped resolvent

The warped resolvent: Definition

- $\mathcal{X}$ is a reflexive real Banach space with topological dual $\mathcal{X}^*$.
- An operator $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ is *monotone* if

$$(\forall (x_1, x_1^*) \in \operatorname{gra} M)\,(\forall (x_2, x_2^*) \in \operatorname{gra} M) \quad \langle x_1 - x_2, x_1^* - x_2^* \rangle \geqslant 0,$$

and *maximally monotone* if, furthermore, no point can be added to $\operatorname{gra} M$ without compromising monotonicity.

Definition

Let $\varnothing \neq D \subset \mathcal{X}$, let $K\colon D \to \mathcal{X}^*$, and let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be such that $\operatorname{ran} K \subset \operatorname{ran}(K + M)$ and $K + M$ is injective. The warped resolvent of M with kernel K is $J_M^K = (K + M)^{-1} \circ K$.

The warped resolvent: Properties

- Sufficient conditions for ran $K \subset \operatorname{ran}(K+M)$ and $K+M$ is injective are given in (Bùi/PLC, 2019).
- $J_M^K\colon D \to D$.
- $\operatorname{Fix} J_M^K = D \cap \operatorname{zer} M$.
- $p = J_M^K x \Leftrightarrow (p, Kx - Kp) \in \operatorname{gra} M$.
- Suppose that M is monotone. Let $x \in D$, and set $y = J_M^K x$ and $y^* = Kx - Ky$. Then

$$\operatorname{zer} M \subset \big\{ z \in \mathcal{X} \mid \langle z - y, y^* \rangle \leqslant 0 \big\}.$$

- Suppose that M is monotone. Set $p = J_M^K x$ and $q = J_M^K y$. Then

$$\langle p - q, Kx - Ky \rangle \geqslant \langle p - q, Kp - Kq \rangle.$$

The warped resolvent: Examples

$M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ is maximally monotone.

- If $\mathcal{X}$ is Hilbertian and $K = \text{Id}$, J_M^K is the classical resolvent.

The warped resolvent: Examples

$M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ is maximally monotone.

- If $\mathcal{X}$ is Hilbertian and $K = \mathrm{Id}$, J_M^K is the classical resolvent.
- If $\mathcal{X}$ is strictly convex with normalized duality mapping K, then J_M^K is the extended resolvent of (Kassay, 1985).

The warped resolvent: Examples

$M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ is maximally monotone.

- If $\mathcal{X}$ is Hilbertian and $K = \mathrm{Id}$, J_M^K is the classical resolvent.
- If $\mathcal{X}$ is strictly convex with normalized duality mapping K, then J_M^K is the extended resolvent of (Kassay, 1985).
- Let $f\colon \mathcal{X} \to \left]-\infty, +\infty\right]$ be a Legendre function such that $\mathrm{dom}\, M \subset \mathrm{int}\,\mathrm{dom}\, f$, and set $K = \nabla f$. Then J_M^K is the D-resolvent of (Bauschke/Borwein/PLC, 2003).

The warped resolvent: Examples

$M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ is maximally monotone.

- If $\mathcal{X}$ is Hilbertian and $K = \mathrm{Id}$, J_M^K is the classical resolvent.
- If $\mathcal{X}$ is strictly convex with normalized duality mapping K, then J_M^K is the extended resolvent of (Kassay, 1985).
- Let $f\colon \mathcal{X} \to \left]-\infty, +\infty\right]$ be a Legendre function such that $\operatorname{dom} M \subset \operatorname{int}\operatorname{dom} f$, and set $K = \nabla f$. Then J_M^K is the D-resolvent of (Bauschke/Borwein/PLC, 2003).
- $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ and $B\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ are maximally monotone, and $f\colon \mathcal{X} \to \left]-\infty, +\infty\right]$ is a suitable convex function. Set

$$M = A + B \quad \text{and} \quad K\colon \operatorname{int}\operatorname{dom} f \to \mathcal{X}^*\colon x \mapsto \nabla f(x) - Bx.$$

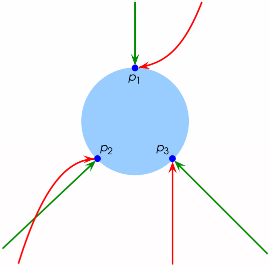
Then $J_M^K = (\nabla f + A)^{-1} \circ (\nabla f - B)$ is the Bregman forward-backward operator to be investigated in Part 4.

The warped resolvent: Examples

- Let $K\colon \mathcal{X} \to \mathcal{X}^*$ be strictly monotone, 3^* monotone, and surjective. Then J_M^K is the K-resolvent of (Bauschke/Wang/Yao, 2010).

The warped resolvent: Examples

- Let $K\colon \mathcal{X} \to \mathcal{X}^*$ be strictly monotone, 3^* monotone, and surjective. Then J_M^K is the K-resolvent of (Bauschke/Wang/Yao, 2010).
- Let $\varnothing \neq C \subset \mathcal{X}$ be closed and convex, with normal cone operator N_C. The warped projection operator is $\operatorname{proj}_C^K = J_{N_C}^K = (K+N_C)^{-1}\circ K$.



Left: Warped projections onto $B(0;1)$. Sets of points projecting onto p_1, p_2, and p_3 for $K_1 = \mathrm{Id}$ and

$$K_2\colon (\xi_1,\xi_2) \mapsto \left(\frac{\xi_1^3}{2} + \frac{\xi_1}{5} - \xi_2, \xi_1 + \xi_2\right)$$

Note that K_2 is not a gradient.

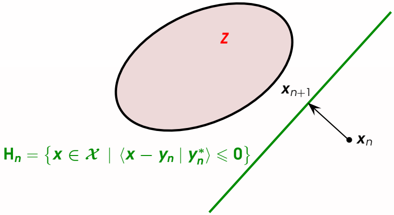
Warped proximal iterations in Hilbert space

PART 3:

Warped proximal iterations in Hilbert spaces

Finding zeros of monotone operators: Geometry

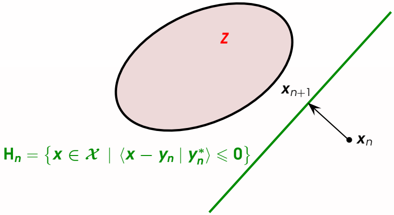
- M maximally monotone with $Z = \operatorname{zer} M \neq \varnothing$.



Finding zeros of monotone operators: Geometry

- M maximally monotone with $Z = \operatorname{zer} M \neq \varnothing$.
- Iterate

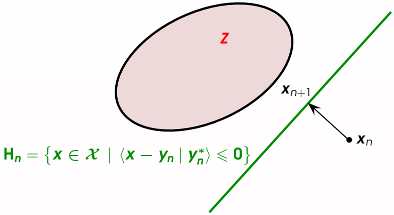
$$
\left\lfloor
\begin{array}{l}
(y_n, y_n^*) \in \operatorname{gra} M \\
\lambda_n \in [\varepsilon, 2-\varepsilon] \\
\text{if } \langle y_n - x_n \mid y_n^* \rangle < 0 \\
\left\lfloor\; x_{n+1} = x_n + \lambda_n \langle y_n - x_n \mid y_n^* \rangle y_n^* / \|y_n^*\|^2 \right. \\
\text{else} \\
\left\lfloor\; x_{n+1} = x_n. \right.
\end{array}
\right.
$$



Finding zeros of monotone operators: Geometry

- M maximally monotone with $Z = \operatorname{zer} M \neq \varnothing$.
- Iterate

$$
\left\lfloor
\begin{array}{l}
(y_n, y_n^*) \in \operatorname{gra} M \\
\lambda_n \in [\varepsilon, 2-\varepsilon] \\
\text{if } \langle y_n - x_n \mid y_n^* \rangle < 0 \\
\left\lfloor\; x_{n+1} = x_n + \lambda_n \langle y_n - x_n \mid y_n^* \rangle y_n^* / \|y_n^*\|^2 \right. \\
\text{else} \\
\left\lfloor\; x_{n+1} = x_n. \right.
\end{array}
\right.
$$

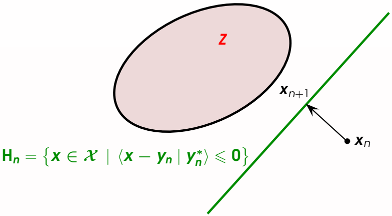


- Weak convergence to a point in Z if weak cluster points are in Z.
- The weak–to–strong convergence principle (Bauschke/PLC, 2001) gives strong convergence of a 2 half-spaces variant.
- How to choose $(y_n, y_n^*) \in \operatorname{gra} M$?

Finding zeros of monotone operators: Geometry

- M maximally monotone with $Z = \operatorname{zer} M \neq \varnothing$.
- Iterate

$$
\left\lfloor
\begin{array}{l}
y_n = J_{\gamma_n M}^{K_n} \widetilde{x}_n \\
y_n^* = \gamma_n^{-1}(K_n \widetilde{x}_n - K_n y_n) \\
\lambda_n \in [\varepsilon, 2-\varepsilon] \\
\text{if } \langle y_n - x_n \mid y_n^* \rangle < 0 \\
\left\lfloor \; x_{n+1} = x_n + \lambda_n \langle y_n - x_n \mid y_n^* \rangle y_n^* / \|y_n^*\|^2 \right. \\
\text{else} \\
\left\lfloor \; x_{n+1} = x_n. \right.
\end{array}
\right.
$$



- **Key:** Move beyond Minty's parametrization of $\operatorname{gra} M$ and use a warped resolvent to pick $(y_n, y_n^*) \in \operatorname{gra} M$.
- Simply evaluate a warped resolvent at some point $\widetilde{x}_n$.

Convergence

Notation: $(y^*)^\sharp = y^*/\|y^*\|$ if $y^* \neq 0$; $= 0$ otherwise.

Theorem

Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, +\infty[$. For every $n \in \mathbb{N}$, let $\widetilde{x}_n \in \mathcal{X}$ and let $K_n \colon \mathcal{X} \to \mathcal{X}$ be a monotone operator such that $\operatorname{ran} K_n \subset \operatorname{ran}(K_n + \gamma_n M)$ and $K_n + \gamma_n M$ is injective. Suppose that:

- $\widetilde{x}_n - x_n \to 0$.
- $\left\langle \widetilde{x}_n - y_n \mid (K_n\widetilde{x}_n - K_n y_n)^\sharp \right\rangle \to 0 \quad \Rightarrow \quad \begin{cases} \widetilde{x}_n - y_n \rightharpoonup 0 \\ K_n\widetilde{x}_n - K_n y_n \to 0. \end{cases}$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Convergence

Notation: $(y^*)^\sharp = y^*/\|y^*\|$ if $y^* \neq 0$; $= 0$ otherwise.

Theorem

Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, +\infty[$. For every $n \in \mathbb{N}$, let $\widetilde{x}_n \in \mathcal{X}$ and let $K_n\colon \mathcal{X} \to \mathcal{X}$ be a monotone operator such that $\operatorname{ran} K_n \subset \operatorname{ran}(K_n + \gamma_n M)$ and $K_n + \gamma_n M$ is injective. Suppose that:

- $\widetilde{x}_n - x_n \to 0$.
- $\langle \widetilde{x}_n - y_n \mid (K_n\widetilde{x}_n - K_n y_n)^\sharp \rangle \to 0 \quad \Rightarrow \quad \begin{cases} \widetilde{x}_n - y_n \rightharpoonup 0 \\ K_n\widetilde{x}_n - K_n y_n \to 0. \end{cases}$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

- We also have a strongly convergent version.

Choosing the evaluation points $(\widetilde{x}_n)_{n\in\mathbb{N}}$

The auxiliary sequence $(\widetilde{x}_n)_{n\in\mathbb{N}}$ can serve several purposes:

- $\widetilde{x}_n$ can model an additive perturbation of x_n, say $\widetilde{x}_n = x_n + e_n$, where we require only $\|e_n\| \to 0$.

Choosing the evaluation points $(\widetilde{x}_n)_{n\in\mathbb{N}}$

The auxiliary sequence $(\widetilde{x}_n)_{n\in\mathbb{N}}$ can serve several purposes:

- $\widetilde{x}_n$ can model an additive perturbation of x_n, say $\widetilde{x}_n = x_n + e_n$, where we require only $\|e_n\| \to 0$.
- Modeling inertia: let $(\alpha_n)_{n\in\mathbb{N}}$ be **any** bounded sequence in $\mathbb{R}$ and set $\widetilde{x}_n = x_n + \alpha_n(x_n - x_{n-1})$.

Choosing the evaluation points $(\widetilde{x}_n)_{n\in\mathbb{N}}$

The auxiliary sequence $(\widetilde{x}_n)_{n\in\mathbb{N}}$ can serve several purposes:

- $\widetilde{x}_n$ can model an additive perturbation of x_n, say $\widetilde{x}_n = x_n + e_n$, where we require only $\|e_n\| \to 0$.
- Modeling inertia: let $(\alpha_n)_{n\in\mathbb{N}}$ be **any** bounded sequence in $\mathbb{R}$ and set $\widetilde{x}_n = x_n + \alpha_n(x_n - x_{n-1})$.
- More generally,

$$(\forall n \in \mathbb{N}) \quad \widetilde{x}_n = \sum_{j=0}^{n} \mu_{n,j} x_j.$$

with $\sum_{j=0}^{n} \mu_{n,j} = 1$ and $(1 - \mu_{n,n})x_n - \sum_{j=0}^{n-1} \mu_{n,j} x_j \to 0$.

Choosing the evaluation points $(\widetilde{x}_n)_{n\in\mathbb{N}}$

The auxiliary sequence $(\widetilde{x}_n)_{n\in\mathbb{N}}$ can serve several purposes:

- $\widetilde{x}_n$ can model an additive perturbation of x_n, say $\widetilde{x}_n = x_n + e_n$, where we require only $\|e_n\| \to 0$.
- Modeling inertia: let $(\alpha_n)_{n\in\mathbb{N}}$ be **any** bounded sequence in $\mathbb{R}$ and set $\widetilde{x}_n = x_n + \alpha_n(x_n - x_{n-1})$.
- More generally,

$$(\forall n \in \mathbb{N}) \quad \widetilde{x}_n = \sum_{j=0}^{n} \mu_{n,j} x_j.$$

  with $\sum_{j=0}^{n} \mu_{n,j} = 1$ and $(1 - \mu_{n,n})x_n - \sum_{j=0}^{n-1} \mu_{n,j} x_j \to 0$.
- Nonlinear perturbations can also be considered. For instance, at iteration n, $\widetilde{x}_n = \mathrm{proj}_{C_n} x_n$ is an approximation to x_n from some suitable closed convex set $C_n \subset \mathcal{X}$.

Corollary 1

Corollary

Let $A\colon \mathcal{X} \to 2^{\mathcal{X}}$ *be maximally monotone, and let* $B\colon \mathcal{X} \to \mathcal{X}$ *be monotone and* β*-Lipschitzian, with* $\operatorname{zer}(A+B) \neq \varnothing$*. Let* $W_n\colon \mathcal{X} \to \mathcal{X}$ *be* α*-strongly monotone and* χ*-Lipschitzian, and let* $\gamma_n \in [\varepsilon, (\alpha-\varepsilon)/\beta]$*, let* $\lambda_n \in [\varepsilon, 2-\varepsilon]$*, and let* $\mathcal{X} \ni e_n \to 0$*. Furthermore, let* $m > 0$ *and let* $(\mu_{n,j})_{n\in\mathbb{N}, 0\leqslant j\leqslant n}$ *be bounded and satisfy*

- *For every* $n > m$ *and every integer* $j \in [0, n-m-1]$*,* $\mu_{n,j} = 0$*.*
- *For every* $n \in \mathbb{N}$*,* $\sum_{j=0}^{n} \mu_{n,j} = 1$*.*

Iterate

$$
\left\lfloor
\begin{array}{l}
\widetilde{x}_n = e_n + \sum_{j=0}^{n} \mu_{n,j} x_j \\
v_n^* = W_n\widetilde{x}_n - \gamma_n B\widetilde{x}_n \\
y_n = (W_n + \gamma_n A)^{-1} v_n^* \\
y_n^* = \gamma_n^{-1}(v_n^* - W_n y_n) + B y_n \\
\text{if } \langle y_n - x_n \mid y_n^* \rangle < 0 \\
\left\lfloor\; x_{n+1} = x_n + \dfrac{\lambda_n \langle y_n - x_n \mid y_n^* \rangle}{\|y_n^*\|^2}\, y_n^* \right. \\
\text{else } x_{n+1} = x_n.
\end{array}
\right.
$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{zer}(A+B)$*.*

Proof: $M = A + B$ and $K_n = W_n - \gamma_n B$. **Special case:** Tseng's algorithm.

Corollary 2: Multivariate inclusions

- **Problem:** find $(x_i)_{i\in I} \in \bigtimes_{i\in I} \mathcal{X}_i$ such that

$$(\forall i \in I) \quad 0 \in A_i x_i + \sum_{j\in J} L_{ji}^*\left((B_j + D_j)\left(\sum_{k\in I} L_{jk} x_k\right)\right) + C_i x_i$$

- **Warping:** Apply Theorem 2 to

$$M\colon \; ((x_i)_{i\in I}, (y_j)_{j\in J}, (v_j^*)_{j\in J}) \mapsto \left(\bigtimes_{i\in I}\left(A_i x_i + C_i x_i + \sum_{j\in J} L_{ji}^* v_j^*\right),\right.$$
$$\left.\bigtimes_{j\in J}(B_j y_j + D_j y_j - v_j^*), \bigtimes_{j\in J}\left\{y_j - \sum_{i\in I} L_{ji} x_i\right\}\right)$$

and $K_n\colon (x, y, v^*) \mapsto$

$$\left(\left(\gamma_{i,n}^{-1} F_{i,n} x_i - C_i x_i - \sum_{j\in J} L_{ji}^* v_j^*\right)_{i\in I}, (\tau_{j,n}^{-1} W_{j,n} y_j - D_j y_j + v_j^*)_{j\in J},\right.$$
$$\left.\left(-y_j + v_j^* + \sum_{i\in I} L_{ji} x_i\right)_{j\in J}\right),$$

where $F_{i,n}$ and $W_{j,n}$ are strongly monotone and Lipschitzian.

Corollary 2: Multivariate inclusions

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\quad \text{for every } i \in I \\
\quad\quad l_{i,n}^* = F_{i,n}\widetilde{x}_{i,n} - \gamma_{i,n} C_i \widetilde{x}_{i,n} - \gamma_{i,n} \sum_{j \in J} L_{ji}^* \widetilde{v}_{j,n}^* \\
\quad\quad a_{i,n} = \big(F_{i,n} + \gamma_{i,n} A_i\big)^{-1} (l_{i,n}^* + \gamma_{i,n} s_i^*) \\
\quad\quad a_{i,n}^* = \gamma_{i,n}^{-1} (l_{i,n}^* - F_{i,n} a_{i,n}) + C_i a_{i,n} \\
\quad \text{for every } j \in J \\
\quad\quad t_{j,n}^* = W_{j,n} \widetilde{y}_{j,n} - \tau_{j,n} D_j \widetilde{y}_{j,n} + \tau_{j,n} \widetilde{v}_{j,n}^* \\
\quad\quad b_{j,n} = \big(W_{j,n} + \tau_{j,n} B_j\big)^{-1} t_{j,n}^* \\
\quad\quad f_{j,n}^* = \tau_{j,n}^{-1} (t_{j,n}^* - W_{j,n} b_{j,n}) + D_j b_{j,n} \\
\quad\quad c_{j,n} = \sum_{i \in I} L_{ji} \widetilde{x}_{i,n} - \widetilde{y}_{j,n} + \widetilde{v}_{j,n}^* - r_j \\
\quad \text{for every } i \in I \\
\quad\quad a_{i,n}^* = o_{i,n}^* + \sum_{j \in J} L_{ji}^* c_{j,n} \\
\quad \text{for every } j \in J \\
\quad\quad b_{j,n}^* = f_{j,n}^* - c_{j,n} \\
\quad\quad c_{j,n}^* = r_j + b_{j,n} - \sum_{i \in I} L_{ji} a_{i,n} \\
\quad \sigma_n = \sum_{i \in I} \|a_{i,n}^*\|^2 + \sum_{j \in J} \big(\|b_{j,n}^*\|^2 + \|c_{j,n}^*\|^2\big) \\
\quad \theta_n = \sum_{i \in I} \langle a_{i,n} - x_{i,n} \mid a_{i,n}^* \rangle + \sum_{j \in J} \big(\langle b_{j,n} - y_{j,n} \mid b_{j,n}^* \rangle + \langle c_{j,n} - v_{j,n}^* \mid c_{j,n}^* \rangle\big) \\
\quad \text{if } \theta_n < 0 \\
\quad\quad \rho_n = \lambda_n \theta_n / \sigma_n \\
\quad \text{else} \\
\quad\quad \rho_n = 0 \\
\quad \text{for every } i \in I \\
\quad\quad x_{i,n+1} = x_{i,n} + \rho_n a_{i,n}^* \\
\quad \text{for every } j \in J \\
\quad\quad y_{j,n+1} = y_{j,n} + \rho_n b_{j,n}^* \\
\quad\quad v_{j,n+1}^* = v_{j,n}^* + \rho_n c_{j,n}^*.
\end{array}
$$

Further connections

- Primal-dual splitting.
  - Consider the inclusion $0 \in Ax + L^*(B(Lx))$ and the associated Kuhn–Tucker operator

$$M\colon \mathcal{X} \times \mathcal{Y} \to 2^{\mathcal{X} \times \mathcal{Y}} \colon (x, y^*) \mapsto (Ax + L^* y^*) \times (-Lx + B^{-1} y^*).$$

  - The cutting plane method of (Alotaibi/PLC/Shahzad, 2014) and (PLC/Eckstein, 2018) generate points $(a_n, a_n^*) \in \operatorname{gra} A$ and $(b_n, b_n^*) \in \operatorname{gra} B$. This implicitly provides

$$(y_n, y_n^*) = \big((a_n, b_n^*), (a_n^* + L^* b_n^*, -La_n + b_n)\big) \in \operatorname{gra} M$$

    to construct $H_n \supset \operatorname{zer} M$.
  - The primal-dual framework of (Alotaibi/PLC/Shahzad, 2014) is therefore an instance of Theorem 2 with

$$K_n\colon (x, y^*) \mapsto \big(\gamma_n^{-1} x - L^* y^*, Lx + \mu_n y^*\big).$$

Further connections

- Primal-dual splitting.
  - Consider the inclusion $0 \in Ax + L^*(B(Lx))$ and the associated Kuhn–Tucker operator

$$M\colon \mathcal{X} \times \mathcal{Y} \to 2^{\mathcal{X} \times \mathcal{Y}} \colon (x, y^*) \mapsto (Ax + L^*y^*) \times (-Lx + B^{-1}y^*).$$

  - The cutting plane method of (Alotaibi/PLC/Shahzad, 2014) and (PLC/Eckstein, 2018) generate points $(a_n, a_n^*) \in \operatorname{gra} A$ and $(b_n, b_n^*) \in \operatorname{gra} B$. This implicitly provides

$$(y_n, y_n^*) = \big((a_n, b_n^*), (a_n^* + L^*b_n^*, -La_n + b_n)\big) \in \operatorname{gra} M$$

    to construct $H_n \supset \operatorname{zer} M$.
  - The primal-dual framework of (Alotaibi/PLC/Shahzad, 2014) is therefore an instance of Theorem 2 with

$$K_n\colon (x, y^*) \mapsto \big(\gamma_n^{-1}x - L^*y^*, Lx + \mu_n y^*\big).$$

- An alternate cutting plane strategy was independently investigated in (Giselsson, arXiv 2019), where an instance of a warped resolvent (in our sense) was used.

Warped proximal iterations with Bregman kernels

PART 4:

Warped proximal iterations with Bregman kernels

Bregman forward-backward splitting

- $\mathcal{X}$ a reflexive real Banach space, $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ and $B\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ maximally monotone, and $f \in \Gamma_0(\mathcal{X})$ essentially smooth.
- $C = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A \subset \operatorname{int}\operatorname{dom} B$ and B is single-valued on $\operatorname{int}\operatorname{dom} B$.
- $(\forall x \in C)(\forall y \in C)(\forall z \in \mathscr{S})(\forall y^* \in Ay)(\forall z^* \in Az)$
$$\langle y - x, By - Bz\rangle \leqslant \kappa D_f(x, y) + \big\langle y - z, \delta_1(y^* - z^*) + \delta_2\big(By - Bz\big)\big\rangle.$$
- The objective is to
$$\text{find } x \in \mathscr{S} = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{zer}(A + B) \neq \varnothing.$$

Bregman forward-backward splitting

- $\mathcal{X}$ a reflexive real Banach space, $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ and $B\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ maximally monotone, and $f \in \Gamma_0(\mathcal{X})$ essentially smooth.
- $C = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A \subset \operatorname{int}\operatorname{dom} B$ and B is single-valued on $\operatorname{int}\operatorname{dom} B$.
- $(\forall x \in C)(\forall y \in C)(\forall z \in \mathscr{S})(\forall y^* \in Ay)(\forall z^* \in Az)$
$$\langle y - x, By - Bz \rangle \leqslant \kappa D_f(x, y) + \big\langle y - z, \delta_1(y^* - z^*) + \delta_2\big(By - Bz\big)\big\rangle.$$
- The objective is to
$$\text{find } x \in \mathscr{S} = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{zer}(A + B) \neq \varnothing.$$
- Apply the warped proximal point algorithm
$$x_{n+1} = J_M^{K_n} x_n$$
to $M = A + B$ with kernel $K_n = \gamma_n^{-1}\nabla f_n - B$ for a suitable essentially smooth function f_n.

Bregman forward-backward splitting

- $\mathcal{X}$ a reflexive real Banach space, $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ and $B\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ maximally monotone, and $f \in \Gamma_0(\mathcal{X})$ essentially smooth.
- $C = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A \subset \operatorname{int}\operatorname{dom} B$ and B is single-valued on $\operatorname{int}\operatorname{dom} B$.
- $(\forall x \in C)(\forall y \in C)(\forall z \in \mathscr{S})(\forall y^* \in Ay)(\forall z^* \in Az)$
$$\langle y - x, By - Bz\rangle \leqslant \kappa D_f(x, y) + \big\langle y - z, \delta_1(y^* - z^*) + \delta_2(By - Bz)\big\rangle.$$
- The objective is to
$$\text{find } x \in \mathscr{S} = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{zer}(A + B) \neq \varnothing.$$
- Apply the warped proximal point algorithm
$$x_{n+1} = J_M^{K_n} x_n$$
to $M = A + B$ with kernel $K_n = \gamma_n^{-1}\nabla f_n - B$ for a suitable essentially smooth function f_n.
- We obtain the Bregman forward-backward splitting algorithm
$$x_{n+1} = \big(\nabla f_n + \gamma_n A\big)^{-1}\big(\nabla f_n(x_n) - \gamma_n B x_n\big).$$

Convergence

Theorem

"Under suitable assumptions,"

$$x_{n+1} = \big(\nabla f_n + \gamma_n A\big)^{-1}\big(\nabla f_n(x_n) - \gamma_n B x_n\big) \rightharpoonup x \in \mathscr{S}.$$

- This result provides, for instance, the convergence of the basic Bregman forward-backward splitting method

$$\big(\nabla f + \gamma A\big)^{-1}\big(\nabla f(x_n) - \gamma B x_n\big),$$

  which is new even in Euclidean spaces.
- It also allows us to recover and extend 4, so far unrelated, splitting frameworks.

$x_{n+1} = (\nabla f_n + \gamma_n A)^{-1}(\nabla f_n(x_n) - \gamma_n B x_n)$: Instantiations

- The iteration $x_{n+1} = \big(\nabla f + \gamma_n A\big)^{-1}\big(\nabla f(x_n)\big)$ for finding a zero of A in a reflexive Banach space (Bauschke/Borwein/PLC, 2003).
- The iteration $x_{n+1} = \big(U_n + \gamma_n A\big)^{-1}\big(U_n x_n - \gamma_n B x_n\big)$ for finding a zero of $A + B$ in a Hilbert space, where U_n is a strongly positive Hermitian bounded linear operator (PLC/Vũ, 2014).
- The iteration

$$x_{n+1} = \big(\nabla f + \gamma A\big)^{-1}\big(\nabla f(x_n) - \gamma B x_n\big)$$

  for finding a zero of $A + B$ in a Hilbert space, where f is real-valued and strongly convex (Renaud/Cohen, 1997).
- The iteration

$$x_{n+1} = \big(\nabla f_n + \gamma_n \partial\varphi\big)^{-1}\big(\nabla f_n(x_n) - \gamma_n \nabla\psi(x_n)\big)$$

  for minimizing $\varphi + \psi$ in a reflexive Banach space (Nguyen, 2017; see also Bauschke/Bolte/Teboulle, 2017).

Illustration: The minimization setting

Let $\varphi \in \Gamma_0(\mathcal{X})$, $\psi \in \Gamma_0(\mathcal{X})$, and $f \in \Gamma_0(\mathcal{X})$ be essentially smooth. Set $C = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom}\partial\varphi$ and $\mathscr{S} = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{Argmin}(\varphi+\psi)$. Suppose that $C \neq \varnothing$, $\varphi+\psi$ is coercive, $C \subset \operatorname{int}\operatorname{dom}\psi$, $\mathscr{S} \neq \varnothing$, ψ is Gâteaux differentiable on $\operatorname{int}\operatorname{dom}\psi$, and $D_f \geqslant \beta D_\psi$.

Corollary

Take $x_0 \in C$ and set

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (\nabla f_n + \gamma_n \partial\varphi)^{-1}(\nabla f_n(x_n) - \gamma_n \nabla\psi(x_n)).$$

Then:

- $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathscr{S}$.
- $(\varphi+\psi)(x_n) - \min(\varphi+\psi)(\mathcal{X}) = o(1/n)$.
- $\sum_{n\in\mathbb{N}} n(D_{f_n}(x_{n+1}, x_n) + D_{f_n}(x_n, x_{n+1})) < +\infty$.

- Weak convergence was obtained in (Nguyen, 2017) under more restrictive assumptions.
- The rates are new, even in Euclidean spaces.

References

- **Bùi/PLC, Warped proximal iterations for monotone inclusions, arXiv, 2019.**
- **Bùi/PLC, Bregman forward-backward operator splitting, arXiv, 2019.**
- Bùi/PLC, Multivariate monotone inclusions in saddle form, arXiv, 2020.
- PLC, Monotone operator theory in convex optimization, *Math. Programming*, 2018.
- Bauschke/PLC, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. corrected printing, Springer, 2019.

Bregman distance

- $f \in \Gamma_0(\mathcal{X})$ is a Legendre function if it is both (Bauschke/Borwein/PLC, 2001):
  - Essentially smooth: ∂f is both locally bounded and single-valued on its domain.
  - Essentially strictly convex: ∂f^* is locally bounded on its domain and f is strictly convex on every convex subset of $\operatorname{dom} \partial f$.
- Take $f \in \Gamma_0(\mathcal{X})$, Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f \neq \varnothing$. The associated Bregman distance is

$$D_f \colon \mathcal{X} \times \mathcal{X} \to [0, +\infty]$$
$$(x, y) \mapsto \begin{cases} f(x) - f(y) - \langle x - y, \nabla f(y) \rangle, & \text{if } y \in \operatorname{int} \operatorname{dom} f; \\ +\infty, & \text{otherwise.} \end{cases}$$