Proximal and resolvent averages

Heinz H. Bauschke

Mathematics, University of British Columbia
Kelowna, B.C., Canada

Research supported by NSERC and by the CRC program

Optimization, Games, and Dynamics

Institut Henri Poincaré, Paris, France

November 28, 2011
9:50 – 10:40

Table of contents

Introduction

The feasibility problem and projection methods

Let $C_1, C_2, \ldots, C_m$ be sets in a Hilbert space X, which we assume to be closed, convex, $\neq \varnothing$. The convex feasibility problem asks to

$$\text{find} \qquad x \in C := C_1 \cap C_2 \cap \cdots \cap C_m.$$

We assume that the sets C_i are "simple" in the sense that the nearest point mappings (projection operators) P_i or

$$P_{C_i} : x \mapsto \operatorname*{argmin}_{c_i \in C_i} \|x - c_i\|$$

are easy to compute.

A projection method combines the projectors in some algorithmic fashion to generate a sequence converging to a solution of the feasibility problem.

Cyclic/alternating projections

The method of cyclic projections generates a sequence $(x_n)_{n\in\mathbb{N}}$ via

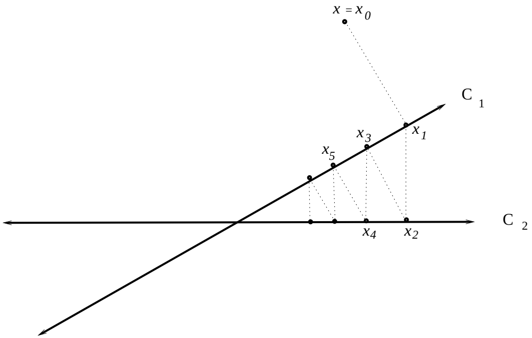
$$x_0 \xmapsto{P_1} x_1 \xmapsto{P_2} x_2 \cdots x_{m-1} \xmapsto{P_m} x_m \xmapsto{P_1} x_{m+1} \xmapsto{P_2} x_{m+2} \cdots$$

Cyclic/alternating projections

The method of cyclic projections generates a sequence $(x_n)_{n\in\mathbb{N}}$ via

$$x_0 \xmapsto{P_1} x_1 \xmapsto{P_2} x_2 \cdots x_{m-1} \xmapsto{P_m} x_m \xmapsto{P_1} x_{m+1} \xmapsto{P_2} x_{m+2} \cdots$$

When $m = 2$, this is also called alternating projections:



Method of alternating projections (for $m = 2$ subspaces)

von Neumann's result for subspaces

Theorem. (von Neumann, 1935)
Suppose that C_1 and C_2 are subspaces. The sequence generated by the method of alternating projections converges strongly to the projection of the starting point onto the intersection.

von Neumann's result for subspaces

Theorem. (von Neumann, 1935)
Suppose that C_1 and C_2 are subspaces. The sequence generated by the method of alternating projections converges strongly to the projection of the starting point onto the intersection.

Remark. (Aronszajn, 1950)
If the angle

$$\arccos \sup_{c_i \in C_i \cap (C_1 \cap C_2)^\perp, \|c_i\| \leq 1} \langle c_1, c_2 \rangle$$

between the subspaces is positive, then the rate of convergence is linear.

Bregman's weak convergence result for convex sets

Theorem. (Bregman, 1965)
Given a starting point $x_0 \in X$, define $(x_n)_{n\in\mathbb{N}}$, the sequence of alternating projections, by

$$x_0 \xrightarrow{P_1} x_1 \xrightarrow{P_2} x_2 \xrightarrow{P_1} x_3 \xrightarrow{P_2} x_4 \xrightarrow{P_1} \cdots .$$

Then

$$x_n \rightharpoonup \bar{c} \in C.$$

Regularity

Remark. (Gubin-Polyak-Raik, 1967)
If $\big(C_1 \cap \mathrm{int}(C_2)\big) \cup \big(C_2 \cap \mathrm{int}(C_1)\big) \neq \varnothing$, then $x_n \to \bar{c} \in C$ strongly (even linearly).

Remark. The results by Aronszajn and by Gubin-Polyak-Raik can be unified: indeed, either assumption implies the Attouch-Brezis constraint qualification

$$\bigcup_{\rho > 0} \rho(C_1 - C_2) \text{ is a closed subspace,}$$

which in turn yields linear convergence (B-Borwein).

Hundal's counterexample

Hundal's counterexample, 2004.
In $X = \ell_2$, there exist two closed convex sets H and K, a vector $f \in X$, and a starting point $y_0 \in K$ so that:

- $\|f\| = 1$;
- H is the hyperplane $\{f\}^\perp$;
- K is a closed convex cone with $\sup\langle f, K\rangle = 0$;
- $H \cap K = \{0\}$.

Then the sequence of alternating projections

converges weakly to 0, but not strongly.

Random projections (for the consistent case)

Rather than projecting cyclically let us "roll a die" instead: let

$$r : \mathbb{N} \to I = \{1, \ldots, m\}$$

be a random map, i.e., $r^{-1}(i)$ is infinite for every $i \in I$, and $x_0 \in X$. Consider the sequence of random projections

$$x_{n+1} = P_{C_{r(n)}} x_n.$$

Random projections (for the consistent case)

Rather than projecting cyclically let us "roll a die" instead: let

$$r\colon \mathbb{N} \to I = \{1, \ldots, m\}$$

be a random map, i.e., $r^{-1}(i)$ is infinite for every $i \in I$, and $x_0 \in X$. Consider the sequence of random projections

$$x_{n+1} = P_{C_{r(n)}} x_n.$$

Open Problem:
If each C_i is a subspace, must $(x_n)_{n\in\mathbb{N}}$ converge strongly?

Random projections (for the consistent case)

Rather than projecting cyclically let us "roll a die" instead: let

$$r \colon \mathbb{N} \to I = \{1, \ldots, m\}$$

be a random map, i.e., $r^{-1}(i)$ is infinite for every $i \in I$, and $x_0 \in X$. Consider the sequence of random projections

$$x_{n+1} = P_{C_{r(n)}} x_n.$$

Open Problem:
If each C_i is a subspace, must $(x_n)_{n \in \mathbb{N}}$ converge strongly?
(Weak convergence to $P_C x_0$ is due to Amemiya and Ando, 1965.
Works by Baillon and Bruck strongly suggest this is true.)

Random projections (for the consistent case)

Rather than projecting cyclically let us "roll a die" instead: let

$$r\colon \mathbb{N} \to I = \{1, \ldots, m\}$$

be a random map, i.e., $r^{-1}(i)$ is infinite for every $i \in I$, and $x_0 \in X$. Consider the sequence of random projections

$$x_{n+1} = P_{C_{r(n)}} x_n.$$

Open Problem:
If each C_i is a subspace, must $(x_n)_{n\in\mathbb{N}}$ converge strongly?
(Weak convergence to $P_C x_0$ is due to Amemiya and Ando, 1965. Works by Baillon and Bruck strongly suggest this is true.)

Open Problem:
In the convex case, must $(x_n)_{n\in\mathbb{N}}$ converge weakly to a point in C?

Random projections (for the consistent case)

Rather than projecting cyclically let us "roll a die" instead: let

$$r\colon \mathbb{N} \to I = \{1, \ldots, m\}$$

be a random map, i.e., $r^{-1}(i)$ is infinite for every $i \in I$, and $x_0 \in X$. Consider the sequence of random projections

$$x_{n+1} = P_{C_{r(n)}} x_n.$$

Open Problem:
If each C_i is a subspace, must $(x_n)_{n\in\mathbb{N}}$ converge strongly?
(Weak convergence to $P_C x_0$ is due to Amemiya and Ando, 1965.
Works by Baillon and Bruck strongly suggest this is true.)

Open Problem:
In the convex case, must $(x_n)_{n\in\mathbb{N}}$ converge weakly to a point in C?
(OK if $m = 2$; also OK if $m = 3$ by Dye and Reich, 1992.)

The inconsistent case when $m = 2$

Define the gap vector

$$v := P_{\overline{C_2 - C_1}} 0,$$

and the "generalized solution sets"

$$E_1 := C_1 \cap (C_2 - v) \text{ and } E_2 := (C_1 + v) \cap C_2.$$

(If $C_1 \cap C_2 \neq \varnothing$, then $v = 0$ and $E_1 = E_2 = C_1 \cap C_2$.)
Then $E_1 = \mathrm{Fix}(P_1 \circ P_2)$, $E_2 = \mathrm{Fix}(P_2 \circ P_1)$, and

$$x_{2n+2} - x_{2n+1} \to v, \quad x_{2n+1} - x_{2n} \to -v.$$

Furthermore: Either: $E_1 = E_2 = \varnothing$ and $\|x_n\| \to +\infty$;
Or: $x_{2n+1} \rightharpoonup e_1 \in E_1$ and $x_{2n} \rightharpoonup e_2 \in E_2$,

$$(e_1, e_2) \text{ is a minimizer for } \min_{(y_1, y_2) \in C_1 \times C_2} \|y_1 - y_2\|$$

as well as a cycle: $e_2 = P_2 e_1$ and $e_1 = P_1 e_2$.

The inconsistent case when $m \geq 3$

In striking contrast, Baillon-Combettes-Cominetti (2011) proved:

There exists no function F on X^m such that cycles $(e_1, \ldots, e_m)$ correspond to minimizers for the problem

$$\min_{(y_1, y_2, \ldots, y_m) \in C_1 \times C_2 \times \cdots \times C_m} F(y_1, \ldots, y_m).$$

Underrelaxed projections for the general case

For $\lambda \in \,]0,1]$, consider the composition of underrelaxed projections:

$$Q_\lambda := \big((1-\lambda)\operatorname{Id} + \lambda P_m\big) \circ \cdots \circ \big((1-\lambda)\operatorname{Id} + \lambda P_1\big).$$

Suppose that each $\operatorname{Fix} Q_\lambda \neq \varnothing$, and let

$$\mathcal{L} := \operatorname{Fix}\Big(\sum_{i=1}^{m} \tfrac{1}{m} P_i\Big)$$

be the set of least squares solutions, i.e., the minimizers of the function

$$x \mapsto \sum_{i=1}^{m} d_{C_i}^2(x).$$

De Pierro's Conjecture

Theory of strongly/averaged nonexpansive mappings implies that

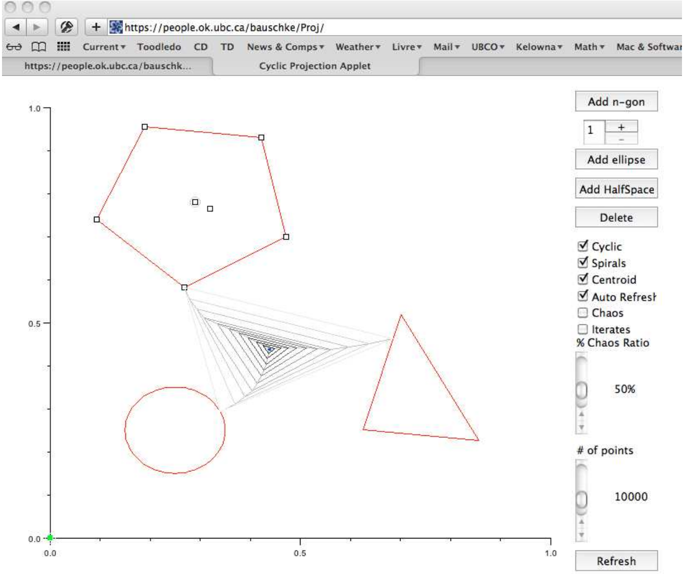
$$x_\lambda := \text{weak} \lim_{n \to +\infty} Q_\lambda^n x$$

exists, for every $x \in X$.

Open Problem: De Pierro's Conjecture
Does the curve $(x_\lambda)_{\lambda \in]0,1]}$ converge to $P_{\mathcal{L}} x$?

Remark. Censor-Eggermont-Gordon (1984) proved this for subspaces in Euclidean space; for further supporting results of this conjecture, see De Pierro (2001) (and also B-Edwards).

Experimental evidence for De Pierro's Conjecture



https://people.ok.ubc.ca/bauschke/Proj/

The arithmetic average

$$\frac{P_1 + P_2 + \cdots + P_m}{m}$$

is a much better behaved object than the composition

$$P_m \circ \cdots \circ P_2 \circ P_1.$$

$\ominus$ The composition is not firmly nonexpansive.
$\oplus$ The average is not a projection; however, it is still a proximal map (Moreau).

In the following, I will advocate the proximal average in the proximal mapping setting and the resolvent average in the general firmly nonexpansive setting (via Minty's correspondence).

The proximal average

Monotone operators

Recall that a set-valued operator $A\colon X \rightrightarrows X$ is *monotone* if

$$\left.\begin{array}{l}(x,u) \in \operatorname{gra} A\\ (y,v) \in \operatorname{gra} A\end{array}\right\} \quad \Rightarrow \quad \langle x-y, u-v\rangle \geq 0,$$

where $\operatorname{gra} A$ is the *graph* of A, and that A is *maximally monotone* if A cannot be properly extended without destroying monotonicity.

Basic examples are the *subdifferential operator* ∂f of $f\colon X \to \left]-\infty,+\infty\right]$, where f is convex, lower semicontinuous, and proper; any *bounded linear operator* $A\colon X \to X$ with a positive symmetric part.

Firmly nonexpansive mappings

Recall that $T\colon X \to X$ is *firmly nonexpansive* if

$$(\forall x \in X)(\forall y \in X) \quad \|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty\rangle.$$

Thanks to work by Minty (1962), Reich (1977), and Eckstein and Bertsekas (1992), we have for $T\colon X \to X$ and $A\colon X \rightrightarrows X$:

Firmly nonexpansive mappings

Recall that $T\colon X \to X$ is *firmly nonexpansive* if

$$(\forall x \in X)(\forall y \in X) \quad \|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty\rangle.$$

Thanks to work by Minty (1962), Reich (1977), and Eckstein and Bertsekas (1992), we have for $T\colon X \to X$ and $A\colon X \rightrightarrows X$:

- T is firmly nonexpansive
  $\Leftrightarrow$ $T^{-1} - \mathrm{Id}$ is maximally monotone;

Firmly nonexpansive mappings

Recall that $T\colon X \to X$ is *firmly nonexpansive* if

$$(\forall x \in X)(\forall y \in X) \quad \|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty\rangle.$$

Thanks to work by Minty (1962), Reich (1977), and Eckstein and Bertsekas (1992), we have for $T\colon X \to X$ and $A\colon X \rightrightarrows X$:

- T is firmly nonexpansive
  $\Leftrightarrow$ $T^{-1} - \operatorname{Id}$ is maximally monotone;
- A is maximally monotone
  $\Leftrightarrow$ $\operatorname{Id} + A$ is onto and the *resolvent*

  $$J_A := (\operatorname{Id} + A)^{-1}$$

  is firmly nonexpansive.

Firmly nonexpansive mappings

Recall that $T\colon X \to X$ is *firmly nonexpansive* if

$$(\forall x \in X)(\forall y \in X) \quad \|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty \rangle.$$

Thanks to work by Minty (1962), Reich (1977), and Eckstein and Bertsekas (1992), we have for $T\colon X \to X$ and $A\colon X \rightrightarrows X$:

- T is firmly nonexpansive
  $\Leftrightarrow$ $T^{-1} - \operatorname{Id}$ is maximally monotone;
- A is maximally monotone
  $\Leftrightarrow$ $\operatorname{Id} + A$ is onto and the *resolvent*

  $$J_A := (\operatorname{Id} + A)^{-1}$$

  is firmly nonexpansive.
- Critical vs fixed points: $0 \in Ax \Leftrightarrow x = J_A x$, i.e., $x \in \operatorname{Fix} J_A$.

Proximal mappings

We also have the equivalence

$$\begin{aligned} & T \text{ is firmly nonexpansive} \\ \Leftrightarrow\ & 2T - \mathrm{Id} \text{ is nonexpansive (Lipschitz-1).} \end{aligned} \qquad (*)$$

Moreau's *proximal map* (or *proximity operator*, early 1960s) is

$$J_{\partial f} x = \operatorname{Prox}_f x;$$

in fact, $\operatorname{Prox}_f x$ is the unique minimizer of $y \mapsto f(y) + \frac{1}{2}\|x - y\|^2$.

Proximal mappings

We also have the equivalence

$$\begin{aligned} &T \text{ is firmly nonexpansive} \\ &\Leftrightarrow 2T - \mathrm{Id} \text{ is nonexpansive (Lipschitz-1).} \end{aligned} \qquad (*)$$

Moreau's *proximal map* (or *proximity operator*, early 1960s) is

$$J_{\partial f} x = \mathrm{Prox}_f\, x;$$

in fact, $\mathrm{Prox}_f\, x$ is the unique minimizer of $y \mapsto f(y) + \frac{1}{2}\|x - y\|^2$.
(If A is the rotator by $\pi/2$ in $\mathbb{R}^2$, then J_A is not a proximal map.)

Proximal mappings

We also have the equivalence

$$\begin{aligned} &T \text{ is firmly nonexpansive} \\ &\Leftrightarrow 2T - \mathrm{Id} \text{ is nonexpansive (Lipschitz-1).} \end{aligned} \qquad (*)$$

Moreau's *proximal map* (or *proximity operator*, early 1960s) is

$$J_{\partial f} x = \mathrm{Prox}_f\, x;$$

in fact, $\mathrm{Prox}_f\, x$ is the unique minimizer of $y \mapsto f(y) + \frac{1}{2}\|x - y\|^2$.
(If A is the rotator by $\pi/2$ in $\mathbb{R}^2$, then J_A is not a proximal map.)

Suppose each $(T_i)_{i \in I}$ is firmly nonexpansive and $(\lambda_i)_{i \in I}$ are convex coefficients (weights): each $\lambda_i > 0$ and $\sum_{i \in I} \lambda_i = 1$. Set

$$T := \sum_{i \in I} \lambda_i T_i.$$

Then T is firmly nonexpansive

Proximal mappings

We also have the equivalence

$$\begin{aligned} &T \text{ is firmly nonexpansive} \\ &\Leftrightarrow 2T - \mathrm{Id} \text{ is nonexpansive (Lipschitz-1).} \end{aligned} \qquad (*)$$

Moreau's *proximal map* (or *proximity operator*, early 1960s) is

$$J_{\partial f} x = \mathrm{Prox}_f\, x;$$

in fact, $\mathrm{Prox}_f\, x$ is the unique minimizer of $y \mapsto f(y) + \frac{1}{2}\|x - y\|^2$.
(If A is the rotator by $\pi/2$ in $\mathbb{R}^2$, then J_A is not a proximal map.)

Suppose each $(T_i)_{i \in I}$ is firmly nonexpansive and $(\lambda_i)_{i \in I}$ are convex coefficients (weights): each $\lambda_i > 0$ and $\sum_{i \in I} \lambda_i = 1$. Set

$$T := \sum_{i \in I} \lambda_i T_i.$$

Then T is firmly nonexpansive since $2T - \mathrm{Id}$ is nonexpansive:

$$2T - \mathrm{Id} = 2\sum_{i \in I} \lambda_i T_i - \mathrm{Id} = \sum_{i \in I} \lambda_i (2T_i - \mathrm{Id}).$$

Proximal mappings form a convex set

Moreau showed that if each T_i is even a proximal map, then so is the average T.

Proximal mappings form a convex set

Moreau showed that if each T_i is even a proximal map, then so is the average T.

Put differently, given functions $(f_i)_{i \in I}$ in Γ, there exists $f \in \Gamma$ such that

$$\mathrm{Prox}_f = \sum_{i \in I} \lambda_i \, \mathrm{Prox}_{f_i} .$$

The function f is unique up to an additive constant; among all these functions, the proximal average that we shall formally define does have beautiful and useful properties.

Handy notation

- $\mathfrak{q}\colon x \mapsto \frac{1}{2}\langle x, x\rangle$ quadratic energy function
- $\Gamma =$ functions from X to $]-\infty, +\infty]$ that are convex, lower semicontinuous, and proper
- $\mathbf{f} = (f_1, \ldots, f_m) \in \Gamma^m$
- $\mathbf{f}^* = (f_1^*, \ldots, f_m^*)$
- $\mathbf{f}^{**} = (f_1^{**}, \ldots, f_m^{**}) = \mathbf{f}$
- $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}_+^m$
- $\lambda_1 + \cdots + \lambda_m = 1$
- $\mu > 0$

Definition of the proximal average

Definition. (B-Goebel-Lucet-Wang) The $\boldsymbol{\lambda}$*-weighted proximal average of* $\mathbf{f}$ *with parameter* μ is defined by

$$p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \lambda_1 \bullet (f_1 + \mu \bullet \mathfrak{q}) \,\Box \cdots \Box\, \lambda_m \bullet (f_m + \mu \bullet \mathfrak{q}) - \mu \bullet \mathfrak{q},$$

where *epi-addition* and *epi-multiplication* are

$$\big(f \,\Box\, g\big)(x) = \inf_{y+z=x} \big(f(y) + g(z)\big);$$

and $\alpha \bullet f = \alpha f(\cdot/\alpha)$, if $\alpha > 0$; $\alpha \bullet f = \iota_{\{0\}}$, if $\alpha = 0$.

Reformulations

If $I = \big\{ i \in \{1, \ldots, m\} \;\big|\; \lambda_i > 0 \big\}$, then

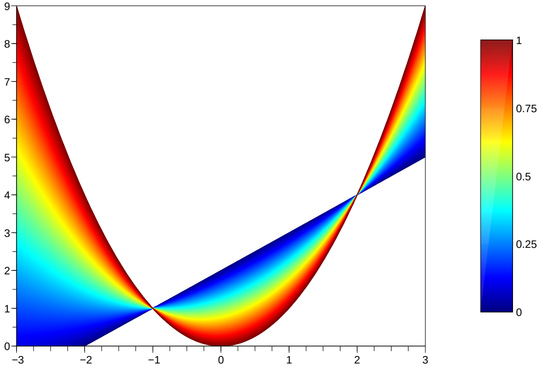
$$p_\mu(\mathbf{f}, \boldsymbol{\lambda})(x) = \frac{1}{\mu}\Big(-\tfrac{1}{2}\|x\|^2 + \inf_{\sum_{i\in I} x_i = x} \sum_{i\in I} \lambda_i\big(\mu f_i(x_i/\lambda_i) + \tfrac{1}{2}\|x_i/\lambda_i\|^2\big)\Big).$$

Furthermore,

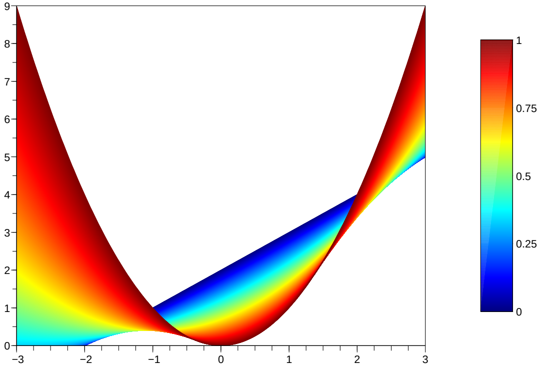
$$\begin{aligned} p_\mu(\mathbf{f}, \boldsymbol{\lambda}) &= \big(\lambda_1(f_1^* \,\square\, \mu\,\mathfrak{q}) + \cdots + \lambda_m(f_m^* \,\square\, \mu\,\mathfrak{q})\big)^* - \mu^{-1}\,\mathfrak{q} \\ &= \big(\lambda_1(f_1 + \mu^{-1}\,\mathfrak{q})^* + \cdots + \lambda_m(f_m + \mu^{-1}\,\mathfrak{q})^*\big)^* - \mu^{-1}\,\mathfrak{q}. \end{aligned}$$

Remark. This was first studied explicitly (for $m = 2$ and $\mu = 1$) by B-Matoušková-Reich to obtain a Güler-like counterexample for the proximal point algorithm — based on Hundal's counterexample!

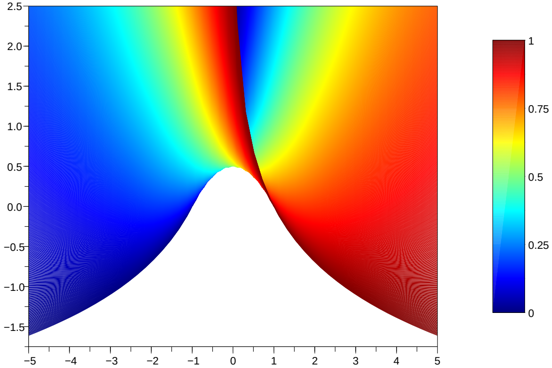
Visualizing the arithmetic average from $2x + 2$ to x^2



Visualizing the proximal average from $2x + 2$ to x^2



From $-\ln(-x)$ *to* $-\ln(x)$



Basic results — a selection

Theorem. $\operatorname{dom} p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_m \operatorname{dom} f_m$, and the epi-sum for $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$ is always exact (i.e., the infimum is attained).

Basic results — a selection

Theorem. $\operatorname{dom} p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_m \operatorname{dom} f_m$, and the epi-sum for $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$ is always exact (i.e., the infimum is attained).

Corollary. If some f_i has full domain and $\lambda_i > 0$, then $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$ has full domain as well.

Basic results — a selection

Theorem. $\operatorname{dom} p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_m \operatorname{dom} f_m$, and the epi-sum for $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$ is always exact (i.e., the infimum is attained).

Corollary. If some f_i has full domain and $\lambda_i > 0$, then $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$ has full domain as well.

Theorem. $\big(p_\mu(\mathbf{f}, \boldsymbol{\lambda})\big)^* = p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda})$.

Corollary. $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$ is convex, lower semicontinuous, and proper.

Proof. Applying the last theorem twice, we deduce that

$$\begin{aligned}\big(p_\mu(\mathbf{f}, \boldsymbol{\lambda})\big)^{**} &= \big(p_{\mu^{-1}}(\mathbf{f}^*, \boldsymbol{\lambda})\big)^* = p_{(\mu^{-1})^{-1}}(\mathbf{f}^{**}, \boldsymbol{\lambda}) \\ &= p_\mu(\mathbf{f}, \boldsymbol{\lambda}).\end{aligned}$$

∎

Basic results — a selection

Theorem. $\operatorname{dom} p_\mu(\mathbf{f},\boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_m \operatorname{dom} f_m$, and the epi-sum for $p_\mu(\mathbf{f},\boldsymbol{\lambda})$ is always exact (i.e., the infimum is attained).

Corollary. If some f_i has full domain and $\lambda_i > 0$, then $p_\mu(\mathbf{f},\boldsymbol{\lambda})$ has full domain as well.

Theorem. $\big(p_\mu(\mathbf{f},\boldsymbol{\lambda})\big)^* = p_{\mu^{-1}}(\mathbf{f}^*,\boldsymbol{\lambda})$.

Corollary. $p_\mu(\mathbf{f},\boldsymbol{\lambda})$ is convex, lower semicontinuous, and proper.

Proof. Applying the last theorem twice, we deduce that

$$\begin{aligned}\big(p_\mu(\mathbf{f},\boldsymbol{\lambda})\big)^{**} &= \big(p_{\mu^{-1}}(\mathbf{f}^*,\boldsymbol{\lambda})\big)^* = p_{(\mu^{-1})^{-1}}(\mathbf{f}^{**},\boldsymbol{\lambda}) \\ &= p_\mu(\mathbf{f},\boldsymbol{\lambda}). \qquad \blacksquare\end{aligned}$$

Example. $p_1\big(\mathbf{f},\mathbf{f}^*,1/(2m)\big) = \mathfrak{q}$.

Moreau envelope

Recall that the *Moreau envelope* of f with parameter μ is

$$e_\mu f = f \,\square\, \mu \bullet \mathfrak{q} = (f^* + \mu\,\mathfrak{q})^*.$$

Theorem. $e_\mu p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \lambda_1 e_\mu f_1 + \cdots + \lambda_m e_\mu f_m.$

Moreau envelope

Recall that the *Moreau envelope* of f with parameter μ is

$$e_\mu f = f \,\square\, \mu \bullet \mathfrak{q} = (f^* + \mu\,\mathfrak{q})^*.$$

Theorem. $e_\mu p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \lambda_1 e_\mu f_1 + \cdots + \lambda_m e_\mu f_m.$

Corollary.

$$\operatorname{argmin} p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \operatorname{argmin}\big(\lambda_1 e_\mu f_1 + \cdots + \lambda_m e_\mu f_m\big).$$

Example. (*least squares solutions* revisited)
If each $f_i = \iota_{C_i}$, where C_i is closed, convex, nonempty, then

$$\operatorname{argmin} p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \operatorname{argmin}(\lambda_1 d_{C_1}^2 + \cdots + \lambda_m d_{C_m}^2).$$

Proximal mapping

Recall that the *proximal mapping* of f with parameter μ is

$$P_\mu f := \mathrm{Prox}_{\mu f} = \big(\,\mathrm{Id} + \mu\partial f\big)^{-1};$$

it satisfies

$$(P_\mu f) \circ (\mu\,\mathrm{Id}) = \nabla(e_{\mu^{-1}}(f^*)).$$

Finally, we are able to motivate the term "proximal average":

Theorem.

$$P_\mu\big(p_\mu(\mathbf{f}, \boldsymbol{\lambda})\big) = \lambda_1 P_\mu f_1 + \cdots + \lambda_m P_\mu f_m.$$

Proof.

We have

$$\begin{aligned} e_{\mu^{-1}}\big((p_\mu(\mathbf{f},\boldsymbol{\lambda}))^*\big) &= e_{\mu^{-1}}\big(p_{\mu^{-1}}(\mathbf{f}^*,\boldsymbol{\lambda})\big) \\ &= \lambda_1 e_{\mu^{-1}}(f_1^*) + \cdots + \lambda_m e_{\mu^{-1}}(f_m^*). \end{aligned}$$

Taking gradients yields

$$\nabla\big(e_{\mu^{-1}}\big((p_\mu(\mathbf{f},\boldsymbol{\lambda}))^*\big)\big) = \lambda_1 \nabla(e_{\mu^{-1}}(f_1^*)) + \cdots + \lambda_m \nabla\big(e_{\mu^{-1}}(f_m^*)\big);$$

in turn, this is equivalent to

$$\big(P_\mu\big(p_\mu(\mathbf{f},\boldsymbol{\lambda})\big)\big)\circ(\mu\,\mathrm{Id}) = \lambda_1(P_\mu f_1)\circ(\mu\,\mathrm{Id}) + \cdots + \lambda_m(P_\mu f_m)\circ(\mu\,\mathrm{Id}),$$

i.e., to

$$P_\mu\big(p_\mu(\mathbf{f},\boldsymbol{\lambda})\big) = \lambda_1(P_\mu f_1) + \cdots + \lambda_m(P_\mu f_m). \qquad \blacksquare$$

Cones

Example. Let $K_1, \ldots, K_m$ be closed subspaces that are pairwise orthogonal and such that $K_1 \oplus \cdots \oplus K_m = X$, and suppose that each $f_i = \iota_{K_i}$ and $\lambda_i > 0$. Then

$$p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \mu^{-1} \sum_{i=1}^{m} (\lambda_i^{-1} - 1)(\mathfrak{q} \circ P_{K_i}).$$

Example. Let K be a nonempty closed convex cone in X and let $\lambda \in]0,1[$. Then

$$p_1\big((\iota_K, \iota_{K^\ominus}), (1-\lambda, \lambda)\big)(x) = \frac{\lambda^2 \|P_K x\|^2 + (1-\lambda)^2 \|P_{K^\ominus} x\|^2}{2(1-\lambda)\lambda},$$

where $K^\ominus = \big\{u \in X \;\big|\; \sup\langle u, K\rangle \le 0\big\}$ is the *polar cone* of K.

Legendre functions

Let $g \in \Gamma$. The following generalizes classical notions in $\mathbb{R}^N$:

- g is *essentially smooth* if ∂g is at most single-valued and $\operatorname{int} \operatorname{dom} g$ is nonempty;
- g is *essentially strictly convex* if g^* is essentially smooth;
- g is *Legendre* if g is both essentially smooth and essentially strictly convex.

Legendre functions

Let $g \in \Gamma$. The following generalizes classical notions in $\mathbb{R}^N$:

- g is *essentially smooth* if ∂g is at most single-valued and $\operatorname{int}\operatorname{dom} g$ is nonempty;
- g is *essentially strictly convex* if g^* is essentially smooth;
- g is *Legendre* if g is both essentially smooth and essentially strictly convex.

Corollary. (Inheritance) Suppose each $\lambda_i > 0$.

- If some f_i is essentially smooth, then so is $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$.
- If some f_j is essentially strictly convex, then so is $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$.
- If some f_i is essentially smooth and some f_j is essentially strictly convex (where not necessarily $i = j$), then $p_\mu(\mathbf{f}, \boldsymbol{\lambda})$ is Legendre.

Varying the parameter μ

Theorem. (pointwise limits) Let $x \in X$. Then the function

$$\mathbb{R}_{++} \to \,]-\infty, +\infty] : \mu \mapsto p_\mu(\mathbf{f}, \boldsymbol{\lambda})(x) \quad \text{is decreasing.}$$

In fact,

$$\lim_{\mu \to 0^+} p_\mu(\mathbf{f}, \boldsymbol{\lambda})(x) = \big(\lambda_1 f_1 + \cdots + \lambda_m f_m\big)(x)$$

and

$$\lim_{\mu \to +\infty} p_\mu(\mathbf{f}, \boldsymbol{\lambda})(x) = \big(\lambda_1 \bullet f_1 \,\Box \cdots \Box\, \lambda_m \bullet f_m\big)(x).$$

Antiderivatives

Recall that $f \in \Gamma$ is an antiderivative of A if $\operatorname{gra} A \subseteq \operatorname{gra} \partial f$.

Fact. (Rockafellar, 1970) Let A be cyclically monotone, i.e., $\sum_{i=1}^{n} \langle a_{i+1} - a_i, a_i^* \rangle \leq 0$ for $n \geq 2$, $(a_i, a_i^*) \in \operatorname{gra} A$ and $a_{n+1} = a_1$. Then the following hold:

- The *Rockafellar functions* $R_{A,(a,a^*)}(x)$ defined by

$$\sup_{2 \leq n,\, (a_i, a_i^*) \in \operatorname{gra} A} \left(\sum_{i=1}^{n-2} \langle a_{i+1} - a_i, a_i^* \rangle + \langle x - a_{n-1}, a_{n-1}^* \rangle \right)$$

  (with $(a, a^*) = (a_1, a_1^*) \in \operatorname{gra} A$ fixed) are antiderivatives of A.
- Maximally cyclically monotone operators are precisely subdifferential operators of functions in Γ.
- If A is maximally cyclically monotone, then antiderivatives of A differ only by constants.

Rockafellar's question

In 2005, R.T. Rockafellar asked the following.

> *Given a cyclically monotone operator with finite graph, find a method that produces an antiderivative of A that preserves the "natural symmetry" induced by convex duality.*

Neither Rockafellar's antiderivatives $R_{A,(a,a^*)}$ nor their pointwise maximum

$$m_A(x) := \sup_{(a,a^*)\in \operatorname{gra} A} R_{A,(a,a^*)}(x)$$

have this property.

An answer

Let $\mathcal{A}$ be the set of cyclically monotone operators with finite graph.

Theorem. (B-Lucet-Wang) The method

$$\mathfrak{m}\colon \mathcal{A} \to \Gamma\colon A \mapsto p_1\big(m_A, m^*_{A^{-1}}, \tfrac{1}{2}, \tfrac{1}{2}\big)$$

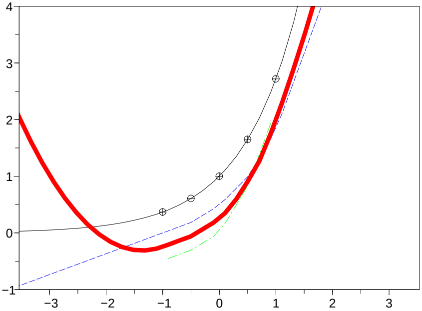
produces *primal-dual symmetric antiderivatives* in the sense that

$$(\mathfrak{m}_A)^* = \mathfrak{m}_{A^{-1}}.$$

In other words, the following is a commutative diagram.

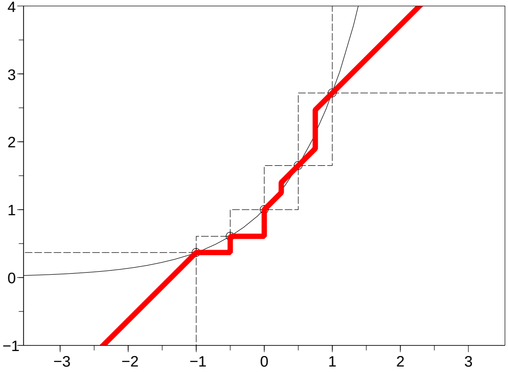
$$\begin{array}{ccc} A & \xrightarrow{\;*\;} & A^{-1} \\ {\scriptstyle\mathfrak{m}}\big\downarrow & & {\scriptstyle\mathfrak{m}}\big\downarrow \\ \mathfrak{m}_A & \xrightarrow{\;*\;} & \mathfrak{m}_{A^{-1}} \end{array}$$

An example - 5 points sampled from exp



thick black — exp; five circled points — the sample;
dashed blue — m_A; dashed-dotted green — $m^*_{A^{-1}}$;
thick red — $\mathfrak{m}_A$.

∂(primal-dual symmetric extension)



Note the "slope one" property of $\partial \mathfrak{m}_A$ outside the rectangle $\operatorname{conv} \operatorname{dom} A \times \operatorname{conv} \operatorname{ran} A$.

The resolvent average

Near equality and near convexity

We now assume for a while that

$$\boxed{X \text{ is finite-dimensional}} \quad \text{and that } I = \{1, 2, \ldots, m\},$$

because then the "relative interior" calculus works particularly well.

Definition. Let A and B be subsets of X. We say that A and B are *nearly equal* if

$$A \approx B \quad :\Leftrightarrow \quad \overline{A} = \overline{B} \quad \text{and} \quad \operatorname{ri} A = \operatorname{ri} B.$$

Proposition. Let $A \subseteq X$. Then A is *nearly convex* (in the sense of Rockafellar and Wets), i.e., there exists a convex subset C of X such that $C \subseteq A \subseteq \overline{C}$ if and only if

$$A \approx \operatorname{conv} A.$$

Calculus

Proposition. Assume that $A, A_1, \ldots, A_m$ are nearly convex subsets of X and that $B, B_1, \ldots, B_m$ are just subsets of X, all $\neq \varnothing$. Then:

- $A \approx \operatorname{conv} A \approx \overline{A} \approx \operatorname{ri} A$.
- If $A \approx B$, then B is nearly convex.
- If $(\forall i \in I)\ A_i \approx B_i$, then $\sum_{i \in I} A_i$ is nearly convex and

$$\sum_{i \in I} A_i \approx \sum_{i \in I} B_i.$$

- If B is compact and $A_1 + B \approx A_2 + B$, then $A_1 \approx A_2$.

Relevance for maximally monotone operators

Fact. Let $A\colon X \rightrightarrows X$ be maximally monotone. Then

$$\operatorname{dom} A \quad \text{and} \quad \operatorname{ran} A \quad \text{are nearly convex.}$$

Relevance for maximally monotone operators

Fact. Let $A\colon X \rightrightarrows X$ be maximally monotone. Then

$$\operatorname{dom} A \quad \text{and} \quad \operatorname{ran} A \quad \text{are nearly convex.}$$

Recall that $A\colon X \rightrightarrows X$ is *rectangular* (a.k.a. 3^* or star monotone; Brezis-Haraux 1976) if the *Fitzpatrick function* satisfies

$$\big(\forall x \in \operatorname{dom} A\big)\big(\forall x^* \in \operatorname{ran} A\big)$$
$$F_A(x,x^*) := \sup_{(a,a^*)\in \operatorname{gra} A} \big(\langle x, a^*\rangle + \langle a, x^*\rangle - \langle a, a^*\rangle\big) < +\infty.$$

Relevance for maximally monotone operators

Fact. Let $A\colon X \rightrightarrows X$ be maximally monotone. Then

$$\operatorname{dom} A \quad \text{and} \quad \operatorname{ran} A \quad \text{are nearly convex.}$$

Recall that $A\colon X \rightrightarrows X$ is *rectangular* (a.k.a. 3^* or star monotone; Brezis-Haraux 1976) if the *Fitzpatrick function* satisfies

$$\big(\forall x \in \operatorname{dom} A\big)\big(\forall x^* \in \operatorname{ran} A\big)$$
$$F_A(x, x^*) := \sup_{(a,a^*)\in \operatorname{gra} A} \big(\langle x, a^*\rangle + \langle a, x^*\rangle - \langle a, a^*\rangle\big) < +\infty.$$

Examples.

Relevance for maximally monotone operators

Fact. Let $A\colon X \rightrightarrows X$ be maximally monotone. Then

$$\operatorname{dom} A \quad \text{and} \quad \operatorname{ran} A \quad \text{are nearly convex.}$$

Recall that $A\colon X \rightrightarrows X$ is *rectangular* (a.k.a. 3^* or star monotone; Brezis-Haraux 1976) if the *Fitzpatrick function* satisfies

$$\big(\forall x \in \operatorname{dom} A\big)\big(\forall x^* \in \operatorname{ran} A\big)$$
$$F_A(x,x^*) := \sup_{(a,a^*)\in \operatorname{gra} A} \big(\langle x, a^*\rangle + \langle a, x^*\rangle - \langle a, a^*\rangle\big) < +\infty.$$

Examples.

- The skew rotator by $\pi/2$ in the plane is *not rectangular*.

Relevance for maximally monotone operators

Fact. Let $A\colon X \rightrightarrows X$ be maximally monotone. Then

$$\operatorname{dom} A \quad \text{and} \quad \operatorname{ran} A \quad \text{are nearly convex.}$$

Recall that $A\colon X \rightrightarrows X$ is *rectangular* (a.k.a. 3^* or star monotone; Brezis-Haraux 1976) if the *Fitzpatrick function* satisfies

$$\big(\forall x \in \operatorname{dom} A\big)\big(\forall x^* \in \operatorname{ran} A\big)$$
$$F_A(x,x^*) := \sup_{(a,a^*)\in \operatorname{gra} A} \big(\langle x, a^*\rangle + \langle a, x^*\rangle - \langle a, a^*\rangle\big) < +\infty.$$

Examples.

- The skew rotator by $\pi/2$ in the plane is *not rectangular*.
- ∂f is rectangular.

Relevance for maximally monotone operators

Fact. Let $A\colon X \rightrightarrows X$ be maximally monotone. Then

$$\operatorname{dom} A \quad \text{and} \quad \operatorname{ran} A \quad \text{are nearly convex.}$$

Recall that $A\colon X \rightrightarrows X$ is *rectangular* (a.k.a. 3^* or star monotone; Brezis-Haraux 1976) if the *Fitzpatrick function* satisfies

$$\big(\forall x \in \operatorname{dom} A\big)\big(\forall x^* \in \operatorname{ran} A\big)$$
$$F_A(x,x^*) := \sup_{(a,a^*)\in \operatorname{gra} A} \big(\langle x, a^*\rangle + \langle a, x^*\rangle - \langle a, a^*\rangle\big) < +\infty.$$

Examples.

- The skew rotator by $\pi/2$ in the plane is *not rectangular*.
- ∂f is rectangular.
- $J_A = (\operatorname{Id} + A)^{-1}$ is rectangular.

On the range

Theorem. Let $(A_i)_{i\in I}$ be a family of maximally monotone rectangular operators such that $\bigcap_{i\in I} \operatorname{ri}\operatorname{dom} A_i \neq \varnothing$, let $(\lambda_i)_{i\in I}$ be a family in $\mathbb{R}_{++}$, and let $j \in I$. Then $A := \sum_{i\in I} \lambda_i A_i$ is maximally monotone, rectangular,

$$\operatorname{ran} A = \operatorname{ran} \sum_{i\in I} \lambda_i A_i \approx \sum_{i\in I} \lambda_i \operatorname{ran} A_i \quad \text{is nearly convex,}$$

and the following hold:

- If $\sum_{i\in I} \lambda_i \operatorname{ran} A_i = X$, then A is surjective.
- If A_j is surjective, then A is surjective.
- If $0 \in \bigcap_{i\in I} \overline{\operatorname{ran}}\, A_i$, then $0 \in \overline{\operatorname{ran}}\, A$.
- If $0 \in (\operatorname{int}\operatorname{ran} A_j) \cap \bigcap_{i\in I\smallsetminus\{j\}} \overline{\operatorname{ran}}\, A_i$, then $0 \in \operatorname{int}\operatorname{ran} A$.

Application to firmly nonexpansive mappings

Corollary. Let $(T_i)_{i\in I}$ be a family of firmly nonexpansive mappings on X, let $(\lambda_i)_{i\in I}$ be a family in $\mathbb{R}_{++}$ such that $\sum_{i\in I}\lambda_i = 1$, and let $j \in I$. Set

$$T := \sum_{i\in I} \lambda_i T_i.$$

Then the following hold.

Application to firmly nonexpansive mappings

Corollary. Let $(T_i)_{i\in I}$ be a family of firmly nonexpansive mappings on X, let $(\lambda_i)_{i\in I}$ be a family in $\mathbb{R}_{++}$ such that $\sum_{i\in I}\lambda_i = 1$, and let $j \in I$. Set

$$T := \sum_{i\in I} \lambda_i T_i.$$

Then the following hold.

- T is firmly nonexpansive and $\operatorname{ran} T \approx \sum_{i\in I}\lambda_i \operatorname{ran} T_i$ is nearly convex.
- If T_j is surjective, then T is surjective.
- If $0 \in \bigcap_{i\in I} \overline{\operatorname{ran}}\, T_i$, then $0 \in \overline{\operatorname{ran}}\, T$.
- If $0 \in (\operatorname{int}\operatorname{ran} T_j) \cap \bigcap_{i\in I\smallsetminus\{j\}} \overline{\operatorname{ran}}\, T_i$, then $0 \in \operatorname{int}\operatorname{ran} T$.

Application to firmly nonexpansive mappings

Corollary. Let $(T_i)_{i\in I}$ be a family of firmly nonexpansive mappings on X, let $(\lambda_i)_{i\in I}$ be a family in $\mathbb{R}_{++}$ such that $\sum_{i\in I}\lambda_i = 1$, and let $j \in I$. Set

$$T := \sum_{i\in I}\lambda_i T_i.$$

Then the following hold.

- T is firmly nonexpansive and $\operatorname{ran} T \approx \sum_{i\in I}\lambda_i \operatorname{ran} T_i$ is nearly convex.
- If T_j is surjective, then T is surjective.
- If $0 \in \bigcap_{i\in I}\overline{\operatorname{ran}}\, T_i$, then $0 \in \overline{\operatorname{ran}}\, T$.
- If $0 \in (\operatorname{int}\operatorname{ran} T_j) \cap \bigcap_{i\in I\smallsetminus\{j\}}\overline{\operatorname{ran}}\, T_i$, then $0 \in \operatorname{int}\operatorname{ran} T$.

Proof. As a resolvent, each T_i is rectangular. Now apply the last result. ■

Back to projections

Example. Let $(C_i)_{i\in I}$ be a family of nonempty closed convex subsets of X with associated projection operators P_i, and let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers such that $\sum_{i\in I}\lambda_i = 1$. Then

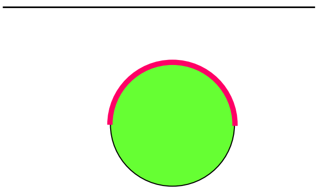
$$\operatorname{ran}\sum_{i\in I}\lambda_i P_i \approx \sum_{i\in I}\lambda_i C_i.$$

Back to projections

Example. Let $(C_i)_{i\in I}$ be a family of nonempty closed convex subsets of X with associated projection operators P_i, and let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers such that $\sum_{i\in I}\lambda_i = 1$. Then

$$\operatorname{ran}\sum_{i\in I}\lambda_i P_i \approx \sum_{i\in I}\lambda_i C_i.$$

Example. Suppose that $X = \mathbb{R}^2$, $m = 2$, $C_1 = \mathbb{R}\times\{2\}$, and $C_2 =$ unit ball centered at 0 of radius 1. The composition $P_2 \circ P_1$ is nonexpansive but ran $P_2 \circ P_1$ is not even nearly convex:



Asymptotic regularity

Theorem. Let $(T_i)_{i \in I}$ be a family of firmly nonexpansive mappings on X, and let $(\lambda_i)_{i \in I}$ be a family of strictly positive real numbers such that $\sum_{i \in I} \lambda_i = 1$. Suppose that each T_i is asymptotically regular, i.e., $0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_i)}$, i.e., T_i has—or "almost" has—a fixed point.

Asymptotic regularity

Theorem. Let $(T_i)_{i\in I}$ be a family of firmly nonexpansive mappings on X, and let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers such that $\sum_{i\in I}\lambda_i = 1$. Suppose that each T_i is asymptotically regular, i.e., $0 \in \overline{\operatorname{ran}(\operatorname{Id} - T_i)}$, i.e., T_i has—or "almost" has—a fixed point. Then

$$\sum_{i\in I}\lambda_i T_i$$

is asymptotically regular as well.

Asymptotic regularity

Theorem. Let $(T_i)_{i\in I}$ be a family of firmly nonexpansive mappings on X, and let $(\lambda_i)_{i\in I}$ be a family of strictly positive real numbers such that $\sum_{i\in I}\lambda_i = 1$. Suppose that each T_i is asymptotically regular, i.e., $0 \in \overline{\operatorname{ran}(\mathrm{Id} - T_i)}$, i.e., T_i has—or "almost" has—a fixed point. Then

$$\sum_{i\in I}\lambda_i T_i$$

is asymptotically regular as well.

Example. Suppose that $X = \mathbb{R}^2$, $m = 2$, $C_1 = \mathbb{R}\times\{0\}$ and $C_2 = \operatorname{epi}\exp$, with corresponding projectors P_1 and P_2. Then each $\operatorname{Fix} P_i = C_i \neq \varnothing$, yet $\operatorname{Fix}\left(\frac{1}{2}P_1 + \frac{1}{2}P_2\right) = \varnothing$.

The resolvent average

Theorem. Let $(A_i)_{i\in I}$ be a family of maximally monotone—not necessarily rectangular—operators, let $(\lambda_i)_{i\in I}$ be in $\mathbb{R}_{++}$ such that $\sum_{i\in I}\lambda_i = 1$, let $j \in I$, and define the resolvent average by

$$A := \Big(\sum_{i\in I}\lambda_i\big(\operatorname{Id} + A_i\big)^{-1}\Big)^{-1} - \operatorname{Id}.$$

Then the following hold.

The resolvent average

(i) A is maximally monotone and

$$J_A = \sum_{i \in I} \lambda_i J_{A_i}$$

The resolvent average

(i) A is maximally monotone and

$$J_A = \sum_{i \in I} \lambda_i J_{A_i}$$

(ii) $\operatorname{dom} A \approx \sum_{i \in I} \lambda_i \operatorname{dom} A_i$ and $\operatorname{ran} A \approx \sum_{i \in I} \lambda_i \operatorname{ran} A_i$.

(iii) If $\operatorname{dom} A_j = X$, then $\operatorname{dom} A = X$.

(iv) If $\operatorname{ran} A_j = X$, then $\operatorname{ran} A = X$.

(v) If $0 \in \bigcap_{i \in I} \overline{\operatorname{ran} A_i}$, then $0 \in \overline{\operatorname{ran} A}$.

(vi) If $0 \in (\operatorname{int} \operatorname{ran} A_j) \cap \bigcap_{i \in I \smallsetminus \{j\}} \overline{\operatorname{ran} A_i}$, then $0 \in \operatorname{int} \operatorname{ran} A$.

The resolvent average

(i) A is maximally monotone and

$$J_A = \sum_{i \in I} \lambda_i J_{A_i}$$

(ii) $\operatorname{dom} A \approx \sum_{i \in I} \lambda_i \operatorname{dom} A_i$ and $\operatorname{ran} A \approx \sum_{i \in I} \lambda_i \operatorname{ran} A_i$.

(iii) If $\operatorname{dom} A_j = X$, then $\operatorname{dom} A = X$.

(iv) If $\operatorname{ran} A_j = X$, then $\operatorname{ran} A = X$.

(v) If $0 \in \bigcap_{i \in I} \overline{\operatorname{ran} A_i}$, then $0 \in \overline{\operatorname{ran} A}$.

(vi) If $0 \in (\operatorname{int} \operatorname{ran} A_j) \cap \bigcap_{i \in I \smallsetminus \{j\}} \overline{\operatorname{ran} A_i}$, then $0 \in \operatorname{int} \operatorname{ran} A$.

Remark. If each $A_i = \partial f_i$, then $\rightsquigarrow$ proximal average of $(f_i)_{i \in I}$;

The resolvent average

(i) A is maximally monotone and

$$J_A = \sum_{i \in I} \lambda_i J_{A_i}$$

(ii) $\operatorname{dom} A \approx \sum_{i \in I} \lambda_i \operatorname{dom} A_i$ and $\operatorname{ran} A \approx \sum_{i \in I} \lambda_i \operatorname{ran} A_i$.

(iii) If $\operatorname{dom} A_j = X$, then $\operatorname{dom} A = X$.

(iv) If $\operatorname{ran} A_j = X$, then $\operatorname{ran} A = X$.

(v) If $0 \in \bigcap_{i \in I} \overline{\operatorname{ran} A_i}$, then $0 \in \overline{\operatorname{ran} A}$.

(vi) If $0 \in (\operatorname{int} \operatorname{ran} A_j) \cap \bigcap_{i \in I \smallsetminus \{j\}} \overline{\operatorname{ran} A_i}$, then $0 \in \operatorname{int} \operatorname{ran} A$.

Remark. If each $A_i = \partial f_i$, then $\rightsquigarrow$ proximal average of $(f_i)_{i \in I}$; item (iv) is abstract supercoercivity;

The resolvent average

(i) A is maximally monotone and

$$J_A = \sum_{i \in I} \lambda_i J_{A_i}$$

(ii) $\operatorname{dom} A \approx \sum_{i \in I} \lambda_i \operatorname{dom} A_i$ and $\operatorname{ran} A \approx \sum_{i \in I} \lambda_i \operatorname{ran} A_i$.

(iii) If $\operatorname{dom} A_j = X$, then $\operatorname{dom} A = X$.

(iv) If $\operatorname{ran} A_j = X$, then $\operatorname{ran} A = X$.

(v) If $0 \in \bigcap_{i \in I} \overline{\operatorname{ran} A_i}$, then $0 \in \overline{\operatorname{ran} A}$.

(vi) If $0 \in (\operatorname{int} \operatorname{ran} A_j) \cap \bigcap_{i \in I \smallsetminus \{j\}} \overline{\operatorname{ran} A_i}$, then $0 \in \operatorname{int} \operatorname{ran} A$.

Remark. If each $A_i = \partial f_i$, then $\rightsquigarrow$ proximal average of $(f_i)_{i \in I}$;
item (iv) is abstract supercoercivity;
item (vi) is abstract coercivity;

The resolvent average

(i) A is maximally monotone and

$$J_A = \sum_{i \in I} \lambda_i J_{A_i}$$

(ii) $\operatorname{dom} A \approx \sum_{i \in I} \lambda_i \operatorname{dom} A_i$ and $\operatorname{ran} A \approx \sum_{i \in I} \lambda_i \operatorname{ran} A_i$.

(iii) If $\operatorname{dom} A_j = X$, then $\operatorname{dom} A = X$.

(iv) If $\operatorname{ran} A_j = X$, then $\operatorname{ran} A = X$.

(v) If $0 \in \bigcap_{i \in I} \overline{\operatorname{ran} A_i}$, then $0 \in \overline{\operatorname{ran} A}$.

(vi) If $0 \in (\operatorname{int} \operatorname{ran} A_j) \cap \bigcap_{i \in I \smallsetminus \{j\}} \overline{\operatorname{ran} A_i}$, then $0 \in \operatorname{int} \operatorname{ran} A$.

Remark. If each $A_i = \partial f_i$, then $\rightsquigarrow$ proximal average of $(f_i)_{i \in I}$;
item (iv) is abstract supercoercivity;
item (vi) is abstract coercivity;
"if some A_j is good, then so is A" (B-Moffat-Wang, forthcoming).

Positive semidefinite matrices

As an illustration of both the proximal and the resolvent average, consider the following set up:

- $\mathbf{A} = (A_1, \dots, A_m) \in (\mathbb{S}^N_+)^m$
- $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m_+$
- $\lambda_1 + \cdots + \lambda_m = 1$
- $\mu > 0$

Now set

$$\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}) := \\ \left(\lambda_1(A_1 + \mu^{-1}\operatorname{Id})^{-1} + \cdots + \lambda_m(A_m + \mu^{-1}\operatorname{Id})^{-1}\right)^{-1} - \mu^{-1}\operatorname{Id},$$

so that

$$J_{\mu\mathcal{R}_\mu(\mathbf{A},\boldsymbol{\lambda})} = \lambda_1 J_{\mu A_1} + \cdots + \lambda_m J_{\mu A_m}.$$

The bridge to the proximal average

For $B \in \mathbb{S}^N$, set

$$q_B \colon x \mapsto \tfrac{1}{2}\langle x, Bx \rangle.$$

If each $f_i = q_{A_i}$, then

$$p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = q_{\mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda})}$$

and hence

$$\nabla p_\mu(\mathbf{f}, \boldsymbol{\lambda}) = \mathcal{R}_\mu(\mathbf{A}, \boldsymbol{\lambda}).$$

Thus the results on the proximal average are applicable!

Averages: harmonic vs resolvent vs arithmetic

Recall that the harmonic and arithmetic averages are defined by

$$\mathcal{H}(\mathbf{A}, \boldsymbol{\lambda}) = \left(\lambda_1 A_1^{-1} + \cdots + \lambda_m A_m^{-1}\right)^{-1},$$

and

$$\mathcal{A}(\mathbf{A}, \boldsymbol{\lambda}) = \lambda_1 A_1 + \cdots + \lambda_m A_m,$$

respectively.

Averages: harmonic vs resolvent vs arithmetic

Recall that the harmonic and arithmetic averages are defined by

$$\mathcal{H}(\mathbf{A},\boldsymbol{\lambda}) = \left(\lambda_1 A_1^{-1} + \cdots + \lambda_m A_m^{-1}\right)^{-1},$$

and

$$\mathcal{A}(\mathbf{A},\boldsymbol{\lambda}) = \lambda_1 A_1 + \cdots + \lambda_m A_m,$$

respectively.

Theorem. We have

$$\mathcal{H}(\mathbf{A},\boldsymbol{\lambda}) \preceq \mathcal{R}_\mu(\mathbf{A},\boldsymbol{\lambda}) \preceq \mathcal{A}(\mathbf{A},\boldsymbol{\lambda}),$$

$$\lim_{\mu\to 0^+} \mathcal{R}_\mu(\mathbf{A},\boldsymbol{\lambda}) = \mathcal{A}(\mathbf{A},\boldsymbol{\lambda}), \quad \lim_{\mu\to +\infty} \mathcal{R}_\mu(\mathbf{A},\boldsymbol{\lambda}) = \mathcal{H}(\mathbf{A},\boldsymbol{\lambda}),$$

and

$$\left(\mathcal{R}_\mu(\mathbf{A},\boldsymbol{\lambda})\right)^{-1} = \mathcal{R}_{\mu^{-1}}(\mathbf{A}^{-1},\boldsymbol{\lambda}).$$

Back to the general setting

We now let X be possibly infinite-dimensional again. Let

$$A$$

be a—not necessarily maximally—monotone operator on X.
Our aim is to find an **explicit** maximally monotone extension of A.
Recall the corresponding Fitzpatrick function is

$$F_A\colon (x,x^*) \mapsto \sup_{(a,a^*)\in \operatorname{gra} A} \big(\langle x,a^*\rangle + \langle a,x^*\rangle - \langle a,a^*\rangle\big).$$

Given $F \in \Gamma(X \times X)$, it will be convenient to define

$$F^\intercal(x^*,x) = F(x,x^*)$$

and also to define $G(F)\colon X \rightrightarrows X$ via

$$x^* \in G(F)x \quad \Leftrightarrow \quad (x^*,x) \in \partial F(x,x^*).$$

Explicit maximally monotone extension

Theorem. Let $A\colon X \rightrightarrows X$ be monotone and set

$$E_A := p_1\big(F_A, F_A^{*\intercal}, \tfrac{1}{2}, \tfrac{1}{2}\big).$$

Then $E_A^* = E_A^\intercal$, $G(E_A)$ is a *maximally monotone extension* of A that is primal-dual symmetric in the sense that

$$\big(G(E_A)\big)^{-1} = G(E_{A^{-1}}).$$

Explicit maximally monotone extension

Theorem. Let $A\colon X \rightrightarrows X$ be monotone and set

$$E_A := p_1\big(F_A, F_A^{*\intercal}, \tfrac{1}{2}, \tfrac{1}{2}\big).$$

Then $E_A^* = E_A^\intercal$, $G(E_A)$ is a *maximally monotone extension* of A that is primal-dual symmetric in the sense that

$$\big(G(E_A)\big)^{-1} = G(E_{A^{-1}}).$$

Remark.
This provided an answer to a problem of Fitzpatrick from 1988 (and it also works in reflexive spaces).
Note that this construction does not require Zorn's Lemma!
Similarly, via Minty, we also obtain Zorn's-Lemma-free extensions of (firmly) nonexpansive mappings in the spirit of Kirszbraun-Valentine!

Current/future work and open problems

Current/future work

- More basic theory for the resolvent average ✓
- Asymptotic regularity of compositions of resolvents ✓
- Extend resolvent average to nonreflexive Banach spaces and Bregman-distance like settings ?
- **Numerical convex analysis** (Lucet et al., on-going) ...
- **Numerical monotone operator theory** ?

Open problems

- Strong convergence of random projections for subspaces ?
- Weak convergence of random projections ?
- De Pierro's conjecture ?

Bibliographical starting points

For further information...

- Please email me at heinz.bauschke@ubc.ca if you wish to obtain detailed pointers to specific results.
- The interplay of maximally monotone operators and firmly nonexpansive mappings is a central theme in

Some classical references. . .

- H. Brézis and A. Haraux: "Image d'une somme d'opérateurs monotones et applications", *Israel J. Math.* 23 (1976), 165–186.
- G.J. Minty, "Monotone (nonlinear) operators in Hilbert spaces", *Duke Math. J.* 29 (1962), 341–346.
- J.-J. Moreau, "Proximité et dualité dan un espace hilbertien", *Bull. Soc. Math. France* 93 (1965), 273–299.
- R.T. Rockafellar, "On the maximal monotonicity of subdifferential mappings", *Pacific J. Math.* 33 (1970), 209–216.
- R.T. Rockafellar and R.J-B Wets, *Variational Analysis*, Springer-Verlag, 1998.

Some recent starting points...

- HHB, R. Goebel, Y. Lucet, and X. Wang: "The proximal average: basic theory", *SIAM J. Optim.* 19 (2008), 766–785.
- HHB, Y. Lucet, and X. Wang: "Primal-dual symmetric intrinsic methods for finding antiderivatives for cyclically monotone operators", *SIAM J. Control Optim.* 46 (2007), 2031–2051.
- HHB, S.M. Moffat, and X. Wang: "Near equality, near convexity, sums of maximally monotone operators, and averages of firmly nonexpansive mappings", *Math. Programming* in press. http://arxiv.org/abs/1105.0029
- HHB, S.M. Moffat, and X. Wang: "The resolvent average for positive semidefinite matrices", *Linear Algebra App.* 432 (2010), 1757–1771.
- HHB and X. Wang: "The kernel average for two convex functions and its applications to the extension and representation of monotone operators", *Trans. Amer. Math. Soc.* 361 (2009), 5947–5965.

Merci beaucoup!