Asymptotic behaviour of the composition of two prox operators

Heinz H. Bauschke

Mathematics and Statistics, University of Guelph

`hbauschk@uoguelph.ca`

Classical Analysis in honour of David Borwein's 80th Birthday

CMS Meeting, June 14, 2004, 5:00PM – 5:30PM

Arts Centre 409, Dalhousie University

An Analysis Problem

Let r be a positive constant and $c_0 \geq 0$. Consider the iteration

$$c_{n+1} = c_n + r - \frac{c_n}{\sqrt{1 + c_n^2}}.$$

(a) For which values of r does the sequence (c_n) converge?

(b) In case of convergence to c with $c \neq c_0$, prove that $\lim (c_{n+1} - c)/(c_n - c)$ exists and determine its value.

(c) In case of divergence, find an asymptotic expression for c_n.

An Analysis Problem

- This is D. Borwein and J. Borwein's Problem 10335 in *American Mathematical Monthly*, Vol. 100, 1993.
- Solution was used in a paper by HB and J. Borwein from 1994. In fact, the Acknowledgment of this paper reads:

  > *The authors wish to thank David Borwein for discussion of Example 5.3, Judith Borwein for preparing the manuscript, and two anonymous referees for helpful suggestions.*

  And our affiliations were Dalhousie and Waterloo!

Overview

1. MOTIVATION
2. BREGMAN OBJECTS
3. BREGMAN RESULTS
4. REFERENCES

Collaborators

Based on joint works with:

- *Patrick L. Combettes* (Paris 6, France),
- *Dominikus Noll* (Toulouse, France).

1. MOTIVATION

Set up

Throughout,

$$X = \mathbb{R}^J$$

with

$$\text{inner product } \langle x, y \rangle \text{ and norm } \|x\| = \sqrt{\langle x, x \rangle},$$

for x and y in X. Also, the proper lower semicontinuous convex functions on X are denoted by

$$\Gamma_0(X).$$

Alternating projections

Suppose

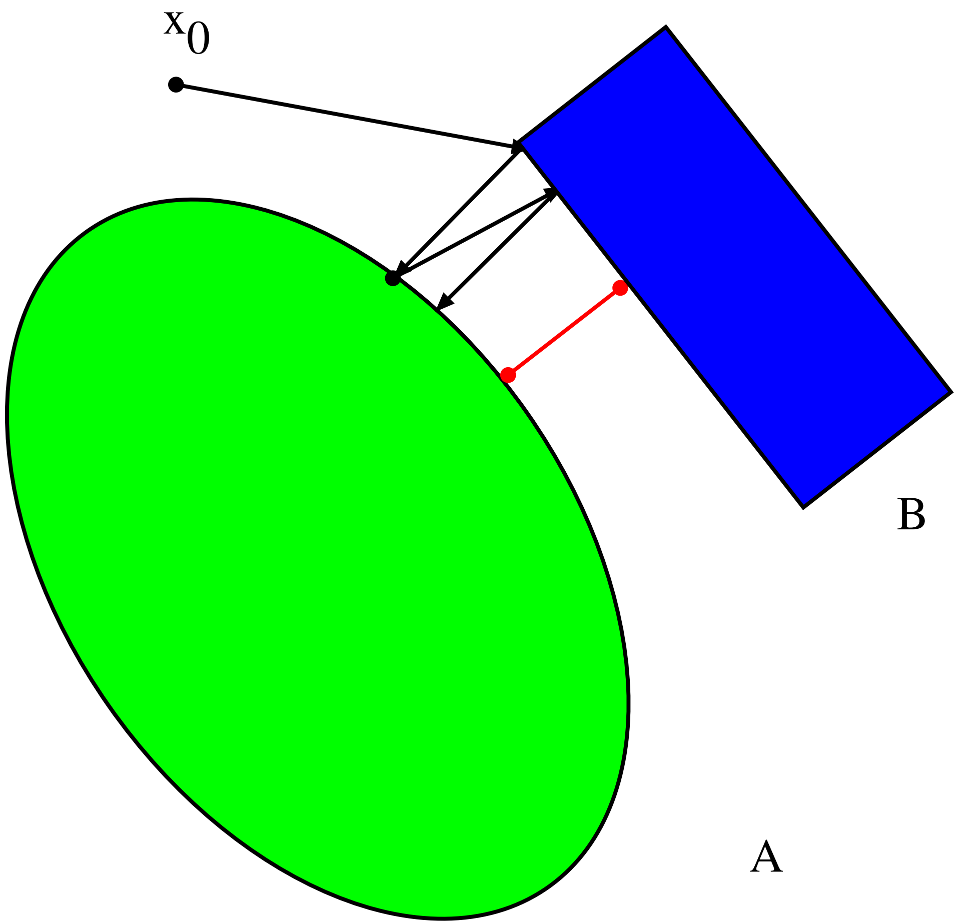
$$A \text{ and } B \text{ are nonempty closed convex sets in } X,$$

with corresponding **projectors** (nearest-point mappings)

$$P_A \text{ and } P_B.$$

Given a starting point $x_0 \in X$, the method of **alternating projections** generates sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ by

$$(\forall n \in \mathbb{N}) \quad y_n = P_B(x_n) \;\; x_{n+1} = P_A(y_n).$$


x_0
B
A

Basic convergence result

Theorem. (Cheney-Goldstein 1959, ...)
Suppose the **gap**

$$\gamma := \inf \|A - B\|$$

between A and B is attained. Then:
$(x_n)_{n\in\mathbb{N}}$ converges to $\bar{x} \in A$, $(y_n)_{n\in\mathbb{N}}$ converges to $\bar{y} \in B$, and $\|\bar{x} - \bar{y}\| = \gamma$.

Remark. True in Hilbert space with weak convergence — but not norm convergence, thanks to Hundal.

Observations on the limits

- Fixed point characterization: $\bar{x}$ and $\bar{y}$ satisfy $\bar{x} = P_A P_B \bar{x}$ and $\bar{y} = P_B \bar{x} = P_B P_A \bar{y}$
- The dual solution is

$$v := P_{\overline{B-A}}(0) \equiv \bar{y} - \bar{x},$$

i.e., the nearest point to 0 in the closure of the Minkowski difference $B - A$. Note that $\|v\|$ is exactly the gap γ!

Observations on the limits

- $(\bar{x}, \bar{y})$ solves the **optimization problem**

$$\text{minimize } (x, y) \mapsto \iota_A(x) + \iota_B(y) + \tfrac{1}{2}\|x - y\|^2.$$

Here ι_A and ι_B are **indicator functions**, defined by

$$\iota_C(x) := \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases}$$

Moreau envelope

Let $\theta \in \Gamma_0(X)$. The **Moreau envelope** of θ at z is

$$\mathrm{env}_\theta(z) := \big(\theta \Box \tfrac{1}{2}\|\cdot\|^2\big)(z) := \inf_{w \in X} \theta(w) + \tfrac{1}{2}\|z - w\|^2.$$

This operation regularizes θ. For instance, if $\theta = \iota_C$, then

$$\mathrm{env}_{\iota_C}(z) = \inf_{w \in X} \iota_C(w) + \tfrac{1}{2}\|z - w\|^2 = \inf_{c \in C} \tfrac{1}{2}\|z - c\|^2$$

is $\frac{1}{2} \cdot$ the square of the *distance* of z to C.

Proximity operator

The infimum in the definition of $\operatorname{env}_\theta(z)$, i.e.,

$$\operatorname{env}_\theta(z) = \inf_{w \in X} \theta(w) + \tfrac{1}{2}\|z - w\|^2,$$

is *always uniquely* attained! The induced map

$$\operatorname{prox}_\theta \colon X \to X \colon z \mapsto w_z := \operatorname*{argmin}_{w \in X} \theta(w) + \tfrac{1}{2}\|z - w\|^2$$

is called the **proximity operator** or **proximal map** of θ.

Note that $0 \in \partial\theta(w_z) - (z - w_z) \Leftrightarrow z \in (\mathsf{Id} + \partial\theta)(w_z)$; equivalently,

$$w_z = \operatorname{prox}_\theta(z) = (\mathsf{Id} + \partial\theta)^{-1}(z).$$

Since $\partial\theta$ is *maximal monotone*, the operator

$\operatorname{prox}_\theta$ is *firmly nonexpansive*

and everywhere defined.

If $\theta = \iota_C$, then $\operatorname{prox}_{\iota_C} = P_C$ (projector onto the set C).

Alternating prox operators

Let φ, ψ be in $\Gamma_0(X)$ and $x_0 \in X$. Consider the method of *alternating prox operators*:

$$(\forall n \in \mathbb{N}) \quad y_n := \operatorname{prox}_\psi(x_n) \;\; x_{n+1} := \operatorname{prox}_\varphi(y_n).$$

Theorem. (Acker-Prestel 1980)
$x_n \xrightarrow[\text{weak}]{} \bar{x} \in X$ and $y_n \xrightarrow[\text{weak}]{} \bar{y} \in X$, where $(\bar{x}, \bar{y})$ is a solution of the optimization problem

$$\text{minimize } (x,y) \mapsto \varphi(x) + \psi(y) + \tfrac{1}{2}\|x-y\|^2.$$

Remark. $\varphi = \iota_A$, $\psi = \iota_B$ yields alternating projections.

Purpose of this talk

- Consider the objective function

$$X \to \left]-\infty, +\infty\right] : (x, y) \mapsto \varphi(x) + \psi(y) + \tfrac{1}{2}\|x - y\|^2.$$

What happens if we replace

$$\tfrac{1}{2}\|x - y\|^2$$

by some other “distance-like” term?

2. Bregman objects

Suppose that

$$X = \mathbb{R}^J$$

and

$$f \in \Gamma_0(X) \text{ is differentiable on } U := \operatorname{int}\operatorname{dom} f \neq \varnothing.$$

Then the **Bregman distance** $D = D_f : X \times X \to [0, +\infty]$ corresponding to f is defined by

$$(x, y) \mapsto \begin{cases} f(x) - f(y) - \langle f'(y), x - y \rangle, & \text{if } y \in U; \\ +\infty, & \text{otherwise.} \end{cases}$$

Remark. D is not a distance in the sense of topology.

Further assumptions on f

We assume that f satisfies the following:

A1 f is of *Legendre* type;

A2 f'' exists and is continuous on U;

A3 D is *jointly convex*, i.e., it is convex on $X \times X$;

A4 $(\forall x \in U)\ D(x, \cdot)$ is strictly convex on U;

A5 $(\forall x \in U)\ D(x, \cdot)$ is *coercive*, i.e., it has bounded lower level sets.

Examples

Write $x = (\xi_j)$ and $y = (\eta_j)$. Then the following functions satisfy all assumptions on f:

(i) If f is the **energy** $x \mapsto \frac{1}{2}\|x\|^2$, then $U = X$ and

$$D(x, y) = \tfrac{1}{2}\|x - y\|^2.$$

(ii) If f is the **negative entropy** $x \mapsto \sum_j \xi_j \ln(\xi_j) - \xi_j$, then $U = \{x \in X : x > 0\}$ and

$$D(x, y) = \begin{cases} \sum_j \xi_j \ln(\xi_j/\eta_j) - \xi_j + \eta_j, & \text{if } x \geq 0 \text{ and } y > 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

Remarks

- Note that the setting considered earlier is covered, since $D(x,y) = \frac{1}{2}\|x-y\|^2$ when f is the energy.
- When f is the negative entropy, the term $D(x,y)$ is known as the **Kullback-Leibler information divergence** in statistics and information theory.
- Other examples are:
  (iii) the *Fermi-Dirac entropy* and
  (iv) the *log-quad function*.
- The function $f = -\ln$ has many good properties, but it does *not* satisfy all our assumptions.

Exploiting joint convexity

Since D is jointly convex (**A3**), its Bregman distance D_D is nonnegative.

Fact. (B-Noll 2002).
Take $\{x, y, u, v\} \subset U$. Then:

$$0 \leq D_D\big((x, y), (u, v)\big) = D(x, y) + D(x, u) - D(x, v) + \langle f''(v)(u - v), y - v \rangle.$$

Moreover:

(i) If f is the *energy*, then

$$D_D\big((x,y),(u,v)\big) = D\big(x, y + (u - v)\big).$$

(ii) If f is the *negative entropy*, then

$$D_D\big((x,y),(u,v)\big) = D\big(x, yu/v\big),$$

where the product and quotient is taken coordinate-wise.

In general, D is *not* symmetric; consequently, we expect **two** envelopes for a given function $\theta \in \Gamma_0(X)$.

The **backward Bregman envelope** of θ is

$$\overleftarrow{\mathrm{env}}_\theta \colon X \to [-\infty, +\infty] : z \mapsto \inf_{w \in X} \theta(w) + D(w, z),$$

and the **forward Bregman envelope** of θ is

$$\overrightarrow{\mathrm{env}}_\theta \colon X \to [-\infty, +\infty] : z \mapsto \inf_{w \in X} \theta(w) + D(z, w).$$

Examples

- If f is the *energy*, then backward & forward Bregman envelope coincide with the Moreau envelope.
- If $\theta = \iota_C$ for some closed convex set C, then we obtain the **backward Bregman distance**

$$\overleftarrow{D}_C := \overleftarrow{\mathrm{env}}_{\iota_C} : z \mapsto \inf_{c \in C} D(c, z)$$

and the **forward Bregman distance**

$$\overrightarrow{D}_C := \overrightarrow{\mathrm{env}}_{\iota_C} : z \mapsto \inf_{c \in C} D(z, c).$$

Definitions

Let $\theta \in \Gamma_0(X)$ such that $\operatorname{dom}\theta \cap U \neq \varnothing$. Under reasonable assumptions, we have:

(i) The **backward proximity operator** is well-defined by

$$\overleftarrow{\operatorname{prox}}_\theta \colon U \to U \colon y \mapsto \operatorname*{argmin}_{x \in X} \theta(x) + D(x, y).$$

(ii) The **forward proximity operator** is well-defined by

$$\overrightarrow{\operatorname{prox}}_\theta \colon U \to U \colon x \mapsto \operatorname*{argmin}_{y \in X} \theta(y) + D(x, y).$$

Proposition. ("the backward prox is very nice")
Suppose θ is nice and $(x, y) \in U \times U$. Then TFAE:

- $x = \overleftarrow{\operatorname{prox}}_\theta(y)$;
- $0 \in \partial\theta(x) + f'(x) - f'(y)$;
- $(\forall z \in X) \quad \langle f'(y) - f'(x), z - x\rangle + \theta(x) \leq \theta(z)$.

Moreover,

$$\overleftarrow{\operatorname{prox}}_\theta = (f' + \partial\theta)^{-1} \circ f'$$

is continuous on U, and

$$\nabla \overleftarrow{\operatorname{env}}_\theta(y) = f''(y)(y - \overleftarrow{\operatorname{prox}}_\theta(y)).$$

Proposition. ("the forward prox is just nice")
Suppose θ is nice and $(x, y) \in U \times U$. Then TFAE:

- $y = \overrightarrow{\mathrm{prox}}_\theta(x)$;
- $0 \in \partial\theta(y) + f''(y)(y - x)$;
- $(\forall z \in X) \quad \langle f''(y)(x - y), z - y\rangle + \theta(y) \le \theta(z)$.

Moreover, $\overrightarrow{\mathrm{prox}}_\theta$ is continuous on U, and

$$\nabla\, \overrightarrow{\mathrm{env}}_\theta(x) = f'(x) - f'(\overrightarrow{\mathrm{prox}}_\theta(x)).$$

Remark. Both propositions extend Moreau's results.

3. Bregman results

Optimization problem

Throughout, let φ, ψ in $\Gamma_0(X)$ be sufficiently "nice", and consider the *optimization problem*

minimize $\Lambda : (x, y) \mapsto \varphi(x) + \psi(y) + D(x, y)$ over $U \times U$.

Denote the *optimal value* and the set of *solutions* by

$p := \inf \Lambda(U \times U)$ and $S := \{(x, y) \in U \times U : \Lambda(x, y) = p\}$,

respectively. We assume that

$$p \in \mathbb{R}.$$

Alternating prox operators

In view of the characterization

$$(x,y) \in S \;\Leftrightarrow\; \big(x = \overleftarrow{\operatorname{prox}}_{\varphi}(y) \text{ and } y = \overrightarrow{\operatorname{prox}}_{\psi}(x)\big),$$

for any $(x,y) \in U \times U$, we propose to find a solution in S via the **method of alternating prox operators** with starting point $x_0 \in X$:

$$(\forall n \in \mathbb{N}) \quad y_n := \overrightarrow{\operatorname{prox}}_{\psi}(x_n), \;\; x_{n+1} := \overleftarrow{\operatorname{prox}}_{\varphi}(y_n). \qquad \text{(APO)}$$

Some inequalities

Suppose $\big((x_n, y_n)\big)_{n\in\mathbb{N}}$ is generated by (APO), $n \in \mathbb{N}$, and $\{x, y\} \subset U$. Then

$$\Lambda(x_{n+1}, y_{n+1}) \leq \Lambda(x_{n+1}, y_n) \leq \Lambda(x_n, y_n) \to \lambda, \tag{1}$$

and

$$\begin{aligned} D(x, x_{n+1}) \leq D(x, x_n) - D_D\big((x,y),(x_n,y_n)\big) \\ - \big(\Lambda(x_{n+1}, y_n) - \Lambda(x,y)\big). \end{aligned} \tag{2}$$

(*Proof.* Combine prox characterizations with Fact on D_D.)

Convergence in value

Corollary.

$$\lambda = \lim \Lambda(x_n, y_n) = \lim \Lambda(x_{n+1}, y_n) = p.$$

Proof. Clearly,

$$\lambda = \inf_{n \in \mathbb{N}} \Lambda(x_n, y_n) = \inf_{n \in \mathbb{N}} \Lambda(x_{n+1}, y_n) \geq \inf \Lambda(U \times U) = p.$$

Assume $\lambda > p$. Then obtain $(x, y) \in U \times U$ such that $\lambda = \Lambda(x, y) + \epsilon$, where $\epsilon > 0$. Now (2) implies

$$(\forall n \in \mathbb{N}) \quad \epsilon = \lambda - \Lambda(x, y) \leq D(x, x_n) - D(x, x_{n+1}).$$

Telescoping this yields a contradiction. Hence $\lambda = p$. ■

Bregman convergence result

Theorem. (B-Combettes-Noll 2004).
Suppose $(x, y) \in S \neq$ ø. Then:

$$\sum_{n\in\mathbb{N}} \big(\Lambda(x_{n+1}, y_n) - p\big) < +\infty,$$

$$\sum_{n\in\mathbb{N}} D_D\big((x, y), (x_n, y_n)\big) < +\infty,$$

and $\big((x_n, y_n)\big)_{n\in\mathbb{N}}$ converges to some point in S.

Geometry of solutions

Let (x, y), $(\widetilde{x}, \widetilde{y})$ be in S, and $x_0 = \widetilde{x}$. Then $(x_n)_{n\in\mathbb{N}} = (\widetilde{x})_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}} = (\widetilde{y})_{n\in\mathbb{N}}$. Thus the Theorem yields the invariance

$$D_D\big((x, y), (\widetilde{x}, \widetilde{y})\big) = 0$$

(This does *not* imply $(x, y) = (\widetilde{x}, \widetilde{y})$.) In particular:

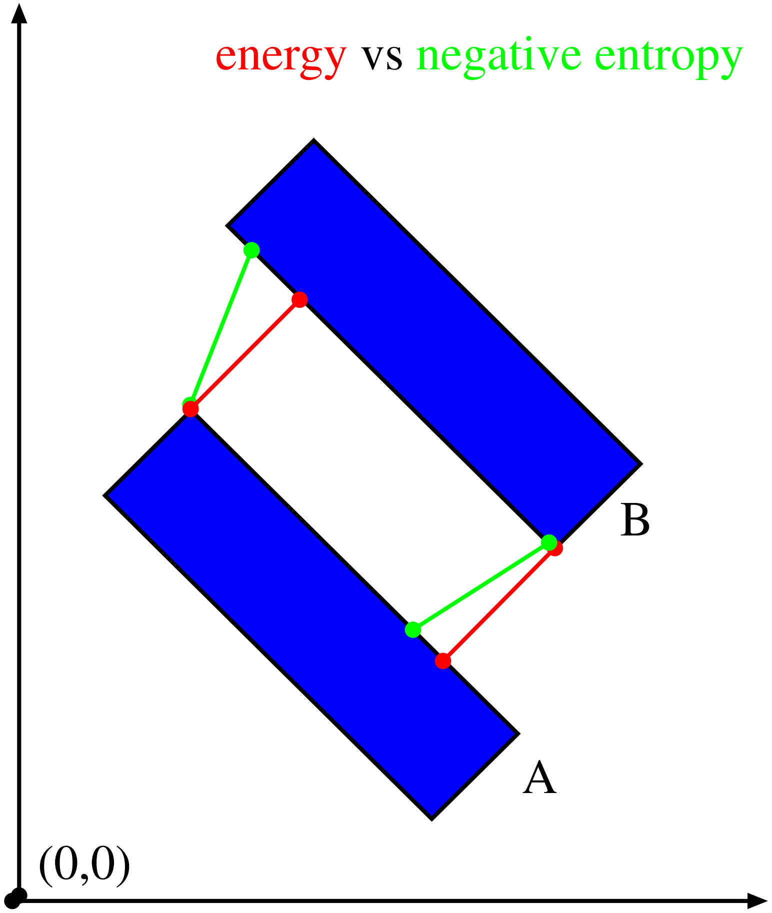
(i) If f is the energy, then $D(x, y + (\widetilde{x} - \widetilde{y})) = 0$, i.e.,

$$x - y = \widetilde{x} - \widetilde{y}.$$

(ii) If f is the negative entropy, then $D(x, y\widetilde{x}/\widetilde{y}) = 0$, i.e.,

$$x/y = \widetilde{x}/\widetilde{y}.$$

The invariance visualized



Applications

Corollary. (Acker-Prestel 1980)
Alternating (regular) prox operators ....
(*Proof.* Let f be the energy. ■)

Corollary. (Csiszár-Tusnády 1984)
Alternating "entropic projections": $x_0 > 0$ and

$$(\forall n \in \mathbb{N}) \quad y_n := \overrightarrow{P}_B(x_n), \;\; x_{n+1} := \overleftarrow{P}_A(y_n).$$

Then $(x_n, y_n) \to (\widetilde{x}, \widetilde{y})$, a Kullback-Leibler gap pair.
(*Proof.* Let f be the negative entropy, $\varphi = \iota_A$, $\psi = \iota_B$. ■)

Remark. Related to *Expectation-Maximization method*.

Applications

Suppose $\theta \in \Gamma_0(X)$ “nice” and assume

$$\varnothing \neq M := \text{minimizers of } \theta \text{ over } U.$$

Corollary. (Censor-Zenios 1992)
The sequence $(z_n)_{n\in\mathbb{N}}$ generated by the **backward proximal point iteration**

$$z_0 \in U, \quad (\forall n \in \mathbb{N}) \quad z_{n+1} = \overleftarrow{\operatorname{prox}}_\theta(z_n)$$

converges to a point in M.

(*Proof*. Set $\varphi = \theta$ and $\psi = 0$, then $\overrightarrow{\operatorname{prox}}_\psi = \mathrm{Id}$. $\blacksquare$)

Applications

Corollary. (B-Combettes-Noll 2004)
The sequence $(z_n)_{n\in\mathbb{N}}$ generated by the **forward proximal point iteration**

$$z_0 \in U, \quad (\forall n \in \mathbb{N}) \quad z_{n+1} = \overrightarrow{\mathrm{prox}}_{\theta}(z_n)$$

converges to a point in M.

(*Proof*. Set $\varphi = 0$ and $\psi = \theta$, then $\overleftarrow{\mathrm{prox}}_{\varphi} = \mathrm{Id}$. ■)

Remark. New parallel applications arise via a product space technique!

4. REFERENCES

- F. Acker and M. A. Prestel: Convergence d’un schéma de minimisation alternée, *Ann. Fac. Sci. Toulouse V. Sér. Math.* **2** (1980), 1–9.
- H. H. Bauschke and J. M. Borwein: Dykstra’s alternating projection algorithm for two sets, *J. Approx. Theory* **79** (1994), 418–443.
- H. H. Bauschke, P. L. Combettes, and D. Noll: Joint minimization with alternating Bregman proximity operators, submitted.