Bregman Monotone Optimization Methods
& Related Convex Functions

Prepared for
IV Brazilian Workshop on
Continuous Optimization Problems

IMPA, Rio de Janeiro, July 15-20, 2002

... ... ...

Jonathan M. Borwein, FRSC
Canada Research Chair

CECM
Centre for Experimental & Constructive Mathematics

Simon Fraser University, Burnaby, BC Canada

URL: www.cecm.sfu.ca/~jborwein/talks.html
Revised: July 9, 2002
ABSTRACT A broad class of optimization algorithms based on Bregman distances in Banach spaces may be unified around the notion of Bregman monotonicity. A systematic investigation of this notion leads to a simplified analysis of numerous algorithms and to the development of a new class of parallel block-iterative surrogate Bregman projection schemes.

Another key contribution is the introduction of a class of operators intrinsically tied to the notion of Bregman monotonicity and including those operators commonly needed in Bregman optimization methods.

Special emphasis is placed on the viability of the algorithms and the importance of Legendre functions in this regard. Various related classes of convex (barrier) functions and applications are discussed.

- This is joint work with Heinz Bauschke (Guelph), Patrick Combettes (Paris) (and Jon Vanderwerff, (La Sierra))
“If my teachers had begun by telling me that mathematics was pure play with presuppositions, and wholly in the air, I might have become a good mathematician. But they were overworked drudges, and I was largely inattentive, and inclined lazily to attribute to incapacity in myself or to a literary temperament that dullness which perhaps was due simply to lack of initiation.”

(George Santayana)

“Persons and Places,” 1945, pp. 238-9
TWO HEADS ARE BETTER THAN ONE

The case for intelligent parallelism
RESOURCES

◊ The transparencies, and other resources, for this presentation are available at
www.cecm.sfu.ca/personal/jborwein/talks.html.

◊ CECM preprint 2002-:184 is available at

◊ Earlier, closely related papers are lodged at
1995:034 (all published).

••••••••

SETTING: Real reflexive (or finite-dimensional)
normed spaces.

MOTIVATION: Produce a broad theory and
capture many concrete applications.

• I shall touch upon* the following topics:

*More is an unrealistic task!
0. OUTLINE

MATHEMATICAL CONTEXT

1. Functions of Legendre type in finite and infinite dimensions.

2. The Bregman distance, $\mathcal{D}$, and Bregman /Legendre functions.


ALGORITHMIC CONSEQUENCES

4. $\mathcal{D}$-viable operators, $\mathcal{D}$-resolvents and $\mathcal{D}$-projections.

5. A prototype algorithm.

6. Related examples and applications.
1a. **LEGENDRE FUNCTIONS**

- $\mathcal{X}$ is a reflexive Banach space with dual $\mathcal{X}^*$ and $f: \mathcal{X} \to [-\infty, +\infty]$ a lsc function which is Gâteaux differentiable on int dom $f \neq \emptyset$.

Then $f$ is **Legendre** [BBC], if it satisfies both:

1. $\partial f$ is both *locally bounded* and single-valued on its domain (*essential smoothness*);

2. $(\partial f)^{-1}$ is locally bounded on its domain and $f$ is strictly convex on every convex subset of dom $\partial f$ (*essential strict convexity*).

- $f$ is a *barrier function* for $C$ if $f(x_n) \to \infty$ as $d(x_n, C^c) \to 0$.

- Most ‘reasonable’ convex functions are Legendre (and *zone-consistent*). In $\mathbb{R}^n$ local boundedness is automatic. Essential strict convexity, where possible, is easy to impose.
1b. LEGENDRE FUNCTION EXAMPLES

1. $\sum_k \lambda_k \log(1 + \lambda_k)$ on $\ell_2(\mathbb{N})$ (or $\mathcal{H}S$) and $\frac{1}{p} \| \cdot \|_p$ on $\ell_p(\mathbb{N})$, $(1 < p < \infty)$ are Legendre.

2. (Strengthened concordancy) For $x \in A \subset \mathbb{R}^N$, open and convex, set

$$F(x) := \lambda_N((A - x)^\circ).$$

(i.) $F$ is an essentially smooth, log-convex, barrier function for $A$, & (ii.) $-F^{-p}$ is convex, essentially smooth, vanishes on $\text{bnd}(A)$ with domain $\overline{A}$, $(0 < p < 1/N)$.*

3. There are essentially smooth functions $f$ on bounded closed convex sets $C$ ($\subset c_0$) but as $d(x_n, \text{bnd}(C)) \to 0$, $\|f'(x_n)\| \not\to \infty$.

4. $\sum_n x_n^{2n}/n$ on $\ell_2$ is strictly convex but not essentially so!

*.$\lambda_N$ is Lebesgue measure and $(A - x)^\circ$ is the polar set.
1c. LEGENDRE FUNCTIONS THEOREMS

**Theorem.** [BV] (a) Suppose $\mathcal{X}$ admits an essentially $\beta$-smooth lsc convex function on a bounded open convex set $C$ with $0 \in C$. Then $\mathcal{X}$ admits a $\beta$-differentiable norm.

(b) Suppose $\mathcal{X}$ admits a strictly convex lsc function, continuous at one point. Then $X$ admits a strictly convex norm.

(c) Suppose $\mathcal{X}$ admits an essentially $\beta$-smooth lsc convex function $f$ on a bounded convex set $C$ with $0 \in \text{int}(C)$ that additionally satisfies $\|f'(x_n)\| \to \infty$ if $d(x_n, \text{bnd}(C)) \to 0$. Then $\mathcal{X}$ admits a $\beta$-differentiable norm.

(d) If $\mathcal{X}$ admits a Legendre function on a bounded open convex set, then $\mathcal{X}$ admits an equivalent strictly convex Gâteaux differentiable norm.
(e) If $\mathcal{X}$ admits an equivalent strictly convex norm, and if every open convex set is the domain of an essentially $\beta$-smooth function, then every open convex set is the domain of some $\beta$-Legendre function.

(f) **Suppose** $\mathcal{X}$ **admits an equivalent strictly convex norm whose dual is strictly convex (resp. LUR).** Then every open convex set is the domain of some Legendre function (resp. Fréchet-Legendre function).

- In particular (f) is true in all weakly compactly generated (resp. reflexive) spaces.
2a. **BREGMAN DISTANCES**

- The **Bregman distance** associated with $f$ is

$$D: \mathcal{X} \times \mathcal{X} \to [0, +\infty]$$

$$(x, y) \mapsto \begin{cases} f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ +\infty, \quad y \notin \text{int} \text{ dom } f. \end{cases}$$

- Let $x_0 \in \mathcal{X}$ and $(S_i)_{i \in I}$ a countable family of closed and convex subsets of $\mathcal{X}$ such that

$$x_0 \in \text{int dom } f, \quad (\text{int dom } f) \cap \bigcap_{i \in I} S_i \neq \emptyset,$$

(1) and $S = \overline{\text{dom } f} \cap \bigcap_{i \in I} S_i.$

- Our goal is to present methods for finding good or best Bregman approximation ($D$-approximations) to $x_0$ from $S$, i.e., for solving the problem: Find

(2)

$$\bar{x} \in S \text{ such that } (\forall x \in S) \; D(\bar{x}, x_0) \leq D(x, x_0).$$
2a. **BREGMAN DISTANCES – EXAMPLES**

- \( \mathcal{D} \) measures the error between a function and its tangent; so there are arbitrarily many such ‘distances’*. The two most central are the energy and the divergence:

- \( \mathcal{D}(x, y) = \| x - y \|^2_2 \) in Hilbert space
  \[
  (f := \| \cdot \|^2);
  \]

- \( \mathcal{D}(x, y) = \sum x_k \log(x_k/y_k) - (x_k - y_k) \) in \( \mathbb{R}^N \)
  \[
  (f(x) := \sum_k x_k \log(x_k) - x_k).
  \]

- Both are jointly convex, which is useful for iterative methods.

**Theorem.** [BaB00] In \( \mathbb{R}^N \), \( \mathcal{D}_f \) is **jointly convex** iff the inverse Hessian \( H_f^{-1} \) is Loewner-concave. In \( \mathbb{R} \) this holds iff \( 1/f'' \) is concave.

*And so more properties need to be imposed ...
This is also true of the spectral function

\[ D(\lambda, \mu) = \sum_k \lambda_k \log(\lambda_k/\mu_k) - (\lambda_k - \mu_k), \]

defined on the symmetric matrices.

While induced spectral functions inherit convexity and differentiability properties, they typically do not inherit Bregman properties.
2b. BREGMAN/LEGENDRE FUNCTIONS

- We introduce a verifiable subclass of Bregman/Legendre functions in $\mathbb{R}^N$ with well-behaved associated $\mathcal{D}$’s [BaB97], and $\mathcal{D}$-projections. We require $\text{dom} \ f^*$ to be open and also that

1. $\mathcal{D}_f(x, \cdot)$ is coercive $\forall x$ in $\text{dom} \ f / \text{int} \ \text{dom}(f)$;

2. $x \in \text{dom} \ f / \text{int} \ \text{dom}(f)$, $y_n \in \text{int} \ \text{dom}(f)$, $y_n \rightarrow y \in \text{dom} \ f / \text{int} \ \text{dom}(f)$, $\mathcal{D}_f(x, y_n)$ bounded
   $\Rightarrow \mathcal{D}_f(y, y_n) \rightarrow 0$ and hence $y \in \text{dom} \ f$;

3. $x_n, y_n \in \text{int} \ \text{dom}(f)$, $x_n, y_n \rightarrow x, y$
   $\in \text{dom} \ f / \text{int} \ \text{dom}(f)$, and $\mathcal{D}_f(x_n, y_n) \rightarrow 0$
   $\Rightarrow x = y$. 

14
**Proposition.** [BaB97] When dom $f$ open, then $f$ is Bregman/Legendre iff dom $f^*$ is open.

- For separable functions life is even better, since the case when $f : \mathbb{R} \mapsto \mathbb{R}$ is quite completely analyzable!

- *Re 3 above.* Solodov-Svaiter show that if $f$ is Bregman [à la Censor-Lent], and if $D(x_n, y_n) \to 0$, then the sequences co-converge to a common limit (“sequential consistency,” in the language of Burachik, Butnariu, Iusem et al.).
3. BREGMAN MONOTONICITY

A sequence \((x_n)_{n \in \mathbb{N}}\) in \(\mathcal{X}\) is Bregman monotone (\(\mathcal{D}_f\)-monotone) with respect to \(S \subset \mathcal{X}\) if:

1. (i) \(S \cap \text{dom } f \neq \emptyset\) and (ii) \(x_0 \in \text{int dom } f\).

2. \((\forall x \in S \cap \text{dom } f, \forall n \in \mathbb{N})\) \(\mathcal{D}(x, x_{n+1}) \leq \mathcal{D}(x, x_n)\).

- This captures the classical Fejér monotonicity property of the Euclidean norm.

Let \(\mathcal{D}_S(x) := \inf_{z \in S} \mathcal{D}_f(z, x)\) be the \(\mathcal{D}\)-distance to a set \(S\).
BASIC PROPERTIES.

1. $\forall x \in S \cap \text{dom } f, \quad \left( D(x, x_n) \right)_{n \in \mathbb{N}}$ converges;

2. $\forall n \in \mathbb{N}, \quad \mathcal{D}_S(x_{n+1}) \leq \mathcal{D}_S(x_n)$;

3. $\left( \mathcal{D}_S(x_n) \right)_{n \in \mathbb{N}}$ converges;

4. $\forall (x, x') \in (S \cap \text{dom } f)^2, \quad \left( \langle x - x', \nabla f(x_n) \rangle \right)_{n \in \mathbb{N}}$ converges;

5. $(x_n)_{n \in \mathbb{N}}$ is bounded if, for some $z \in S \cap \text{dom } f$, the set $\text{lev}_{\leq \mathcal{D}(z, x_0)} \mathcal{D}(z, \cdot)$ is bounded. This is true in particular if $S \cap \text{int dom } f \neq \emptyset$, $\mathcal{X}$ is reflexive, and either of the following is satisfied:

   (a) $f$ is super-coercive (i.e., super-linearly);

   (b) $\dim \mathcal{X}$ is finite and dom $f^*$ is open.
The prior example of \( \sum_{k \in \mathbb{N}} x_k - \ln(1 + x_k) \) shows the conclusion 5. may hold even though properties (a) and (b) are not satisfied.

The next two assumptions help the convergence analysis of \( \mathcal{D} \)-monotone sequences.

**A1.** Given \( S \subset \mathcal{X} \), for every bounded sequence \( (x_n)_{n \in \mathbb{N}} \) in \( \text{int dom } f \) that is \( \mathcal{D} \)-monotone with respect to \( S \), one has
\[
x, x' \in \mathcal{W}(x_n) \cap S \quad \Rightarrow \quad x = x'.
\]

**A2.** For all bounded sequences \( (x_n)_{n \in \mathbb{N}} \) and \( (y_n)_{n \in \mathbb{N}} \) in \( \text{int dom } f \), one has
\[
(3) \quad \mathcal{D}(x_n, y_n) \to 0 \quad \Rightarrow \quad x_n - y_n \to 0.
\]

[Here \( \mathcal{W}(x_n)_{n \in \mathbb{N}} \) and \( \mathcal{S}(x_n)_{n \in \mathbb{N}} \) denote the weak and strong cluster points respectively.]
• These assumptions capture familiar ground:

**Example 1.** A1 is satisfied if:

(a) $S$ is a subset of $\mathcal{X}$ such that $S \cap \overline{\text{dom} f}$ is a singleton; or

(b) $S \subset \text{int} \; \text{dom} f$ is convex, $f|_S$ is strictly convex, and $\nabla f$ is **sequentially weak-to-weak* continuous** at every point in $S$; or

(c) $\mathcal{X} = \mathbb{R}^N$, $f$ is Bregman/Legendre, and $S$ is a subset of $\mathcal{X}$ such that $S \cap \text{dom} f \neq \emptyset$.

**Example 2.** A2 is satisfied if:

(a) $f$ is uniformly convex (**totally convex**, for Butnariu, Iusem-Zalinescu) on bounded sets;

(b) $\mathcal{X} = \mathbb{R}^N$, $\text{dom} f$ is closed, and $f|_{\text{dom} f}$ is strictly convex and continuous; or

(c) $\mathcal{X} = \mathbb{R}$ and $f|_{\text{dom} f}$ is strictly convex.
A1 and A2 lead to remarkably simple weak and strong convergence criteria for $\mathcal{D}$-monotone sequences. When $\mathcal{X}$ is Hilbert and $f = w := \frac{1}{2} \| \cdot \|_2^2$, A1 and A2 are satisfied and these criteria can essentially be found in Gubin-Polyak-Raik (1967).

**Theorem.** *(Convergence)* Let $(x_n)_{n \in \mathbb{N}}$ be a bounded sequence in $\mathcal{X}$ which is $\mathcal{D}$-monotone with respect to a set $S \subset \mathcal{X}$. Suppose that $\mathcal{X}$ is reflexive and A1 is satisfied. Then

1. $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $S \cap \overline{\text{dom} f}$ if and only if $\mathcal{W}(x_n)_{n \in \mathbb{N}} \subset S$.

2. Suppose that $x_n \rightharpoonup x \in S \cap \text{int} \text{dom} f$ and A2 is satisfied. Then $x_n \to x$ if and only if $\mathcal{G}(x_n)_{n \in \mathbb{N}} \neq \emptyset$. 

20
4. **D-Viable Operators of Class $\mathcal{B}$**

- Our task is to identify a large pool of $D$-monotone sequences generated by viable operators. An operator $T: \mathcal{X} \to 2^\mathcal{X}$ is $D$-viable if $\text{ran } T \subseteq \text{dom } T = \text{int } \text{dom } f$.

**The class $\mathcal{B}$.** For $x$ and $u$ in $\text{int } \text{dom } f$, set

\[
H(x, u) := \left\{ y \in \mathcal{X} \mid \langle y - u, \nabla f(x) - \nabla f(u) \rangle \leq 0 \right\}.
\]

Then define

\[
\mathcal{B} := \left\{ T: \mathcal{X} \to 2^\mathcal{X} \mid T \text{ is viable and } \forall (x, u) \in \text{gr } T, \quad \text{Fix } T \subset H(x, u) \right\}.
\]

- $\mathcal{B}$ is constructed to capture $D$-projections on hyper-planes.

When $\mathcal{X}$ is Hilbert, $f = \frac{1}{2} \| \cdot \|_2^2$, and single-valued operators are considered, $\mathcal{B}$ reverts to the operators introduced by Bauschke and Combettes (2001) which play a central role in the analysis of Fejér-monotone algorithms.
4b. $\mathcal{D}$-FIRM OPERATORS

**$\mathcal{D}$-firm operators.** $T : \mathcal{X} \rightarrow \mathcal{X}$ is firmly nonexpansive if for $x$ and $y$ in dom $T$ one has

\[(5) \quad \|Tx - Ty\| \leq \|\alpha(x - y) + (1 - \alpha)(Tx - Ty)\|,\]

\[\forall \alpha \in ]0, +\infty[.\]

- When $\mathcal{X}$ is smooth its normalized duality map $J$ is single-valued and (5) is equivalent to

\[\langle Tx - Ty, J(Tx - Ty) \rangle \leq \langle x - y, J(Tx - Ty) \rangle.\]

In Hilbert space, $J = \text{Id} = \nabla f$ for $f = w$, so equivalently

\[\langle Tx - Ty, \nabla f(Tx) - \nabla f(Ty) \rangle\]

\[\leq \langle Tx - Ty, \nabla f(x) - \nabla f(y) \rangle.\]

This suggests defining ...
... an operator $T: \mathcal{X} \to 2^\mathcal{X}$ with $\text{dom } T \cup \text{ran } T \subset \text{int dom } f$ to be $\mathcal{D}$-firm if

\begin{equation}
\langle u - v, \nabla f(u) - \nabla f(v) \rangle \leq \langle u - v, \nabla f(x) - \nabla f(y) \rangle
\quad \forall (x, u) \in \text{gr } T, \forall (y, v) \in \text{gr } T.
\end{equation}

**Proposition.** Let $T: \mathcal{X} \to 2^\mathcal{X}$ be $\mathcal{D}$-firm. Then $\text{Fix } T \subset H(x, u)$, $\forall (x, u) \in \text{gr } T$, and

1. $T \in \mathfrak{B}$ if $\text{int dom } f = \text{dom } T$;

2. $T$ is single-valued on $\text{dom } T$ whenever $f|_{\text{int dom } f}$ is strictly convex;

3. $\mathcal{D}(u, v) + \mathcal{D}(v, u) \\
\leq \mathcal{D}(u, y) + \mathcal{D}(v, x) - \mathcal{D}(u, x) - \mathcal{D}(v, y), \\
\forall (x, u) \in \text{gr } T, \forall (y, v) \in \text{gr } T.$
4c. $\mathcal{D}$-RESOLVENTS

- The classical (Hilbertian) \textit{resolvent} of $A: \mathcal{X} \to 2^{\mathcal{X}}$ is $(\text{Id} + A)^{-1}$. Also $T: \mathcal{X} \to \mathcal{X}$ is firmly non-expansive iff $T$ is the resolvent of an accretive $A: \mathcal{X} \to 2^{\mathcal{X}}$. So one defines the $\mathcal{D}$-\textit{resolvent} associated with any $A: \mathcal{X} \to 2^{\mathcal{X}^*}$ as

$$(7) \quad R_A := (\nabla f + A)^{-1} \circ \nabla f: \mathcal{X} \to 2^{\mathcal{X}}.$$

Indeed, $0 \in Ax \iff \nabla f(x) \in \nabla f(x) + \gamma A(x) = (\nabla f + \gamma A)(x) \iff x \in (\nabla f + \gamma A)^{-1} \circ \nabla f(x)$.

This is also consistent with previous attempts to define resolvents for monotone operators:

- Let $\mathcal{X}$ be smooth and set $f = w$. Then $\nabla f = J$ and $R_A = (J + A)^{-1} \circ J$.

- If $\mathcal{X}$ is Hilbertian and $f: x \mapsto \|\Pi x\|^2/2$, where $\Pi$ is the metric projector onto a closed vector subspace of $\mathcal{X}$, then $\nabla f = \Pi$ and $R_A = (\Pi + A)^{-1} \circ \Pi$. 

24
Proposition. $R_A$ satisfies the following:

1. $\text{dom } R_A \subset \text{int dom } f$.

2. $\text{ran } R_A \subset \text{int dom } f$.

3. $\text{Fix } R_A = (\text{int dom } f) \cap A^{-1}0$.

4. Suppose $A$ is monotone. Then

   (a) $R_A$ is $\mathcal{D}$-firm.

   (b) $R_A$ is single-valued on its domain if $f|_{\text{int dom } f}$ is strictly convex.

   (c) Suppose $\text{ran } \nabla f \subset \text{ran}(\nabla f + A)$. Then $R_A \in \mathcal{B}$. If, in addition, $f|_{\text{int dom } f}$ is strictly convex, then $\text{Fix } R_A$ is convex.
Theorem. (Resolvents in $\mathcal{B}$) Let $\gamma \in ]0, +\infty[$. Suppose $\mathcal{X}$ is reflexive, $A$ is maximal monotone with $(\text{int dom } f) \cap \text{dom } A = \text{dom } \partial f \cap \text{dom } A \neq \emptyset$, and that one of the following holds:

1. $\mathcal{X}$ is smooth and $f = \| \cdot \|_2 / 2$.

2. $(\nabla f + \gamma A)^{-1}$ is locally bounded on $\mathcal{X}^*$;

3. $\nabla f + \gamma A$ is weakly coercive;

4. $\text{dom } A \subset \text{int dom } f$ or $A$ is $3^*$-monotone, &:
   (a) $\text{ran } \nabla f + \gamma \text{ ran } A = \mathcal{X}^*$; or
   (b) $f$ is Legendre and cofinite; or
   (c) $\text{ran}(\nabla f + \gamma A)$ is closed and $0 \in \text{ran } A$; or
   (d) $\text{ran } \nabla f$ is open and $0 \in \text{ran } A$.

Then $R_{\gamma A} \in \mathcal{B}$. 
Corollary. Let $\gamma \in ]0, +\infty[. $ Suppose that $\mathcal{X}$ is reflexive, that $A$ is maximal monotone with $0 \in \text{ran } A$, and that one of the following conditions holds:

1. $\text{ran } \nabla f$ is open and $\text{dom } A \subset \text{int } \text{dom } f$;

2. $f$ is Legendre and $\text{dom } A \subset \text{int } \text{dom } f$;

3. $f$ is Legendre, $A$ is $3^*$-monotone*, and $\text{dom } A \cap \text{int } \text{dom } f \neq \emptyset$.

Then $R_{\gamma A} \in \mathcal{B}$.

*A $3^*$-monotone if it is monotone and $\forall (x, x^*) \in \text{dom } A \times \text{ran } A \sup \{ \langle x-y, y^*-x^* \rangle \mid (y, y^*) \in \text{gr } A \} < +\infty$. 

27
4d. \(D\)-PROX OPERATORS

- \(D\)-prox operators. The classical notion as introduced by Moreau (1962) in Hilbert space is the proximal operator

\[
\text{prox}^\varphi : y \mapsto \arg\min \varphi + \| \cdot - y \|^2 / 2,
\]

associated with \(\varphi : \mathcal{X} \to ]-\infty, +\infty]\). In \(\mathbb{R}^N\), the idea of defining proximal operators based on \(D\)-distance — rather than quadratic — penalizations was introduced by Censor.

- Let \(\varphi : \mathcal{X} \to ]-\infty, +\infty]\). The \(D\)-prox operator of index \(\gamma \in ]0, +\infty]\) associated with \(\varphi\) is the operator \(\text{prox}^\varphi_\gamma : \mathcal{X} \to 2^{\mathcal{X}}\) given by

\[
y \mapsto \left\{ x \in \text{dom} f \cap \text{dom} \varphi \mid \varphi(x) + D(x, y)/\gamma \right. \\
\quad \quad \quad \quad = \min \left( \varphi + \frac{1}{\gamma} D(\cdot, y) \right)(\mathcal{X}) < +\infty \left\}. \right.
\]
• The following special case of a more general result underscores the importance of the notion of Legendreness.

**Theorem.** Let \( \varphi : \mathcal{X} \to [\!-\infty, +\infty] \) be a lower semicontinuous convex function such that \((\text{int dom } f) \cap \text{dom } \varphi \neq \emptyset \) and let \( \gamma \in ]0, +\infty[. \) Suppose that \( \mathcal{X} \) is reflexive, that \( f \) is Legendre, and that \( \varphi \) is bounded below. Then

1. The mapping \( \text{prox}^\varphi_{\gamma} \) is single-valued on its domain and \( \text{prox}^\varphi_{\gamma} \in \mathcal{B}. \)

2. For every \( x \) and \( y \) in int dom \( f \),

\[
x = \text{prox}^\varphi_{\gamma} y \quad \Leftrightarrow \quad \forall z \in \text{dom } \varphi, \\
\langle z - x, \nabla f(y) - \nabla f(x) \rangle / \gamma + \varphi(x) \leq \varphi(z).
\]
4e. \( \mathcal{D} \)-PROJECTIONS

- The \( \mathcal{D} \)-projector onto a set \( C \subset \mathcal{X} \) is the mapping \( P_C : \mathcal{X} \to 2^\mathcal{X} \)

\[
y \mapsto \left\{ x \in C \cap \text{dom } f \mid D_f(x, y) = D_C(y) < +\infty \right\}.
\]

- The base concept goes back to Bregman. Clearly, for any \( \gamma \in ]0, +\infty[ \), \( P_C = \text{prox}_\gamma^{\mathcal{D}} \). Hence, the results above quickly yield results on \( \mathcal{D} \)-projections when specialized to \( \phi := \iota_C \).

**Corollary.** Suppose \( \mathcal{X} \) is reflexive, \( f \) is Legendre, and \( C \subset \mathcal{X} \) is closed and convex with \( C \cap \text{int } \text{dom } f \neq \emptyset \). Then

(a) \( C \) is \( \mathcal{D} \)-Chebyshev (i.e., \( P_C(x) \) is singleton on \( \text{dom } P_C = \text{int } \text{dom } f \)) and \( P_C \in \mathcal{B} \).

(b) For every \( x \) and \( y \) in \( \text{int } \text{dom } f \),

\[
(8) \quad x = P_C y \iff \left\{ \begin{array}{l}
x \in C \\
C \subset H(y, x).
\end{array} \right.
\]
Algorithm 1. Starting with $x_0 \in \text{int} \, \text{dom} \, f$, at every iteration $n \in \mathbb{N}$, first select $T_n \in \mathcal{B}$ and then select $x_{n+1} \in T_n \, x_n$.

Proposition. Let $(x_n)_{n \in \mathbb{N}}$ be an arbitrary orbit of Algorithm 1. Suppose that $S \cap \text{dom} \, f \neq \emptyset$ and

(9) \quad S \subset \bigcap_{n \in \mathbb{N}} \overline{\text{Fix}} \, T_n.

It follows that

1. if $f|_{\text{int} \, \text{dom} \, f}$ is strictly convex, then $(x_n)_{n \in \mathbb{N}}$ is $\mathcal{D}$-monotone with respect to $S$; and always

2. $\sum_{n \in \mathbb{N}} \mathcal{D}(x_{n+1}, x_n) < +\infty.$
**Theorem.** Let \((x_n)_{n \in \mathbb{N}}\) be an arbitrary bounded orbit of Algorithm 1. Suppose \(\mathcal{X}\) is reflexive, \(f|_{\text{int dom } f}\) is strictly convex, and (9) is satisfied. Suppose in addition, that Assumption 1 is satisfied and that

\[
\sum_{n \in \mathbb{N}} D(x_{n+1}, x_n) < +\infty \quad \Rightarrow \quad \mathcal{W}(x_n)_{n \in \mathbb{N}} \subset S.
\]

Then \((x_n)_{n \in \mathbb{N}}\) converges weakly to some point \(x \in S\).

Moreover, convergence is strong if \(x \in \text{int dom } f\), if Assumption 2 is satisfied, and if also

(10)

\[
\sum_{n \in \mathbb{N}} D(x_{n+1}, x_n) < +\infty \quad \Rightarrow \quad \mathcal{G}(x_n)_{n \in \mathbb{N}} \neq \emptyset.
\]
6a. EXAMPLES & APPLICATIONS

- By appropriately specializing one can recover and extend a variety of relaxed and unrelaxed parallel block-iterative $\mathcal{D}$-monotone algorithms, for sequential iterations, for best $\mathcal{D}$-approximation for subgradients, and diverse other iterations of the kind studied by many in attendance.

- Rather than specify them, I record that all our assumptions are met, for example, when

(i.) Let $\mathcal{X} = \mathbb{R}^N$ and let $f$ be a Bregman/Legendre function with closed domain; or

(ii.) Let $(\varphi_k)_{1 \leq k \leq N} : \mathbb{R} \to ]-\infty, +\infty]$ be a family of Legendre functions such that $(\text{dom } \varphi^*_k)_{k=1}^N$ are open. Let $\mathcal{X} = \mathbb{R}^N$ and let

$$f : (\xi_1, \cdots, \xi_N) \mapsto \sum_{k=1}^N \varphi_k(\xi_k); \text{ or}$$
(iii.) Let $\mathcal{X} = \ell_p(\mathbb{N})$ and take $f = \| \cdot \|^p/p$, for $p \in [2, +\infty[$.

- One of the key tools for (iii.) is:

**Proposition.** (*J* non-expansive) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $p \in [2, +\infty[$, and let $\mathcal{X} = L_p(\Omega, \mathcal{F}, \mu)$, with its canonical norm, and set $f = \| \cdot \|^2/2$. Then, for some $\kappa > 0$, the normalized duality mapping $\nabla f$ satisfies

$$\| J(u) - J(v) \| \leq \kappa \| u - v \|,$$

$\forall (u, v) \in S^2_{\mathcal{X}}$, and in consequence

$$\| J(x) - J(y) \| \leq (2\kappa + 1) \cdot \| x - y \|,$$

$\forall (u, v) \in \mathcal{X}^2$.

In addition, in $\ell_p(\mathbb{N})$ there is a sequentially weak-weak continuous duality map $(\nabla \| x \|_p^p/p)$.

- Additional examples can be generated in suitable product spaces such as $\ell_{p_1}(\mathbb{N}) \times \ell_{p_2}(\mathbb{N})$, equipped with the Euclidean product norm and with $\{p_1, p_2\} \subset [2, +\infty[$, or in certain spaces of power type 2.
6b. A MILDLY RELATED CHALLENGE

- Consider a network objective function $p_N$ given by

$$p_N(\vec{q}) = \sum_{\sigma \in S_N} \left( \prod_{i=1}^{N} \frac{q_{\sigma(i)}}{\sum_{j=i}^{N} q_{\sigma(j)}} \right) \left( \sum_{i=1}^{N} \frac{1}{\sum_{j=i}^{N} q_{\sigma(j)}} \right)$$

summed over all $N!$ permutations; so a typical term is

$$\left( \prod_{i=1}^{N} \frac{q_i}{\sum_{j=i}^{N} q_j} \right) \left( \sum_{i=1}^{N} \frac{1}{\sum_{j=i}^{n} q_j} \right).$$

- For $N = 3$ this is

$$q_1 q_2 q_3 \left( \frac{1}{q_1 + q_2 + q_3} \right) \left( \frac{1}{q_2 + q_3} \right) \left( \frac{1}{q_3} \right)$$

$$\times \left( \frac{1}{q_1 + q_2 + q_3} + \frac{1}{q_2 + q_3} + \frac{1}{q_3} \right).$$

- We wish to show $p_N$ is convex on the positive orthant. First we try to simplify the expression for $p_N$. 

35
• The *partial fraction decomposition* gives:

\[
p_1(x) = \frac{1}{x},
\]
\[
p_2(x_1, x_2) = \frac{1}{x_1} + \frac{1}{x_2} - \frac{1}{x_1 + x_2},
\]
\[
p_3(x_1, x_2, x_3) = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} - \frac{1}{x_1 + x_2} - \frac{1}{x_2 + x_3} - \frac{1}{x_1 + x_3} + \frac{1}{x_1 + x_2 + x_3}.
\]

So we predict the ‘same’ for \( N = 4 \) and

**CONJECTURE.** For each \( N \in \mathbb{N} \)

\[
p_N(x_1, \ldots, x_N) := \int_0^1 \left(1 - \prod_{i=1}^N (1 - t^{x_i})\right) \frac{dt}{t}
\]

is convex, indeed \( 1/p_N \) is concave.

• One may check for \( N \leq 5 \) by computing a large symbolic Hessian; but NOT for \( N = 6 \).
PROOF. A year later, consideration of joint expectations gave:

\[ p_N(x) = \int_{\mathbb{R}^n_+} e^{-(y_1+\cdots+y_n)} \max\left(\frac{y_1}{x_1}, \ldots, \frac{y_n}{x_n}\right) dy. \]

This represents \( p_N \) in a fashion that shows \( 1/p_N \) is concave. [See (BH), *SIAM Electronic Problems and Solutions*.]

QUESTION: What general tools can one develop to establish \( p_N \) is convex?
7. REFERENCES


and one good book?