# Quasi Bregman Monotonicity and Applications

Quang. V. NGUYEN
(under the supervision of Patrick L. COMBETTES)

Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie – Paris 6
75005 Paris, France

Sestri Levante, September 2014

# Fejér Monotonicity in Hilbert spaces

- Notation: given $C$ a nonempty subset and $(x_n)_{n\in\mathbb{N}}$ a sequence in $\mathcal{H}$.
  - $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$: set of weak accumulation points of $(x_n)_{n\in\mathbb{N}}$.
  - $\mathfrak{S}(x_n)_{n\in\mathbb{N}}$: set of strong accumulation points of $(x_n)_{n\in\mathbb{N}}$.
  - $d_C(x) = \inf\{\|y - x\| \mid y \in C\}$.

# Fejér Monotonicity in Hilbert spaces

- Notation: given $C$ a nonempty subset and $(x_n)_{n\in\mathbb{N}}$ a sequence in $\mathcal{H}$.
  - $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$: set of weak accumulation points of $(x_n)_{n\in\mathbb{N}}$.
  - $\mathfrak{S}(x_n)_{n\in\mathbb{N}}$: set of strong accumulation points of $(x_n)_{n\in\mathbb{N}}$.
  - $d_C(x) = \inf\{\|y-x\| \mid y \in C\}$.
- Fejér monotonicity:

$$(\forall x \in C)(\forall n \in \mathbb{N}) \quad \|x - x_{n+1}\|^2 \leqslant \|x - x_n\|^2.$$

# Fejér Monotonicity in Hilbert spaces

- Notation: given $C$ a nonempty subset and $(x_n)_{n\in\mathbb{N}}$ a sequence in $\mathcal{H}$.
  - $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$: set of weak accumulation points of $(x_n)_{n\in\mathbb{N}}$.
  - $\mathfrak{S}(x_n)_{n\in\mathbb{N}}$: set of strong accumulation points of $(x_n)_{n\in\mathbb{N}}$.
  - $d_C(x) = \inf\left\{\|y-x\| \mid y \in C\right\}$.
- Fejér monotonicity:

$$(\forall x \in C)(\forall n \in \mathbb{N}) \quad \|x - x_{n+1}\|^2 \leqslant \|x - x_n\|^2.$$

- Properties:
  - $(d_C(x_n))_{n\in\mathbb{N}}$ converges.
  - $x_n \rightharpoonup \overline{x} \in C \Leftrightarrow \mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset C$.
  - $x_n \to \overline{x} \in C \Leftrightarrow \underline{\lim}\, d_C(x_n) = 0$.

# Quasi Fejér Monotonicity in Hilbert spaces

- Quasi Fejér monotonicity:

  $$(\forall x \in C)(\forall n \in \mathbb{N}) \quad \|x - x_{n+1}\|^2 \leqslant (1 + \eta_n)\|x - x_n\|^2 + \varepsilon_n,$$

  where $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ and $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$.

# Quasi Fejér Monotonicity in Hilbert spaces

- Quasi Fejér monotonicity:

  $$(\forall x \in C)(\forall n \in \mathbb{N}) \quad \|x - x_{n+1}\|^2 \leqslant (1 + \eta_n)\|x - x_n\|^2 + \varepsilon_n,$$

  where $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ and $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$.
- Asymptotic properties: mostly preserved.

# Quasi Fejér Monotonicity in Hilbert spaces

- Quasi Fejér monotonicity:

$$(\forall x \in C)(\forall n \in \mathbb{N}) \quad \|x - x_{n+1}\|^2 \leqslant (1 + \eta_n)\|x - x_n\|^2 + \varepsilon_n,$$

where $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$ and $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$.
- Asymptotic properties: mostly preserved.
- Stochastic version (initiated by Ermol'ev 1968; P. L. Combettes and J. C. Pesquet 2014): Given a separable Hilbert space $\mathcal{H}$ together with its usual Borel $\sigma-$algebra, a probability space $(\Omega, \mathscr{F}, \mu)$, $(x_n)_{n\in\mathbb{N}}$ is a sequence of $\mathcal{H}-$valued random variables, $\mathscr{X} = (\mathcal{X}_n)_{n\in\mathbb{N}}$ is canonic filtration of $(x_n)_{n\in\mathbb{N}}$

$$(\forall z \in C)(\exists(\eta_n(z))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})(\exists(\varepsilon_n(z))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})(\forall n \in \mathbb{N})$$

$$\mathsf{E}(\|z - x_{n+1}\|^2 \,|\, \mathcal{X}_n) \leqslant (1 + \eta_n(z))\|z - x_n\|^2 + \varepsilon_n(z), \mu - \text{as},$$

where $\ell_+^1(\mathscr{X})$ is the set of all summable $[0, +\infty[-$valued stochastic process which are adapted to $\mathscr{X}$.

# Variable metric quasi Fejér monotonicity: definition (P. L. Combettes and B. C. Vũ 2013)

- These notions have seen extensively used to obtain convergence proofs of wide array of algorithms in a unified and simplified fashion.

# Variable metric quasi Fejér monotonicity: definition (P. L. Combettes and B. C. Vũ 2013)

- These notions have seen extensively used to obtain convergence proofs of wide array of algorithms in a unified and simplified fashion.
- A self-adjoint operator $U$ in $\mathcal{B}(H)$ with property

$$(\forall x \in \mathcal{H})\ \langle x \mid Ux\rangle \geqslant \alpha\|x\|^2, \text{where } \alpha \in \left]0, +\infty\right[$$

induces a scalar product and a norm

$$\langle x \mid y\rangle_U := \langle x \mid Uy\rangle \text{ and } \|\cdot\|_U := \sqrt{\langle\cdot \mid U\cdot\rangle}.$$

# Variable metric quasi Fejér monotonicity: analysis

- Variable metric Quasi-Fejér monotone sequence :

$$(\forall x \in C)(\forall n \in \mathbb{N}) \quad \|x - x_{n+1}\|^2_{U_{n+1}} \leqslant (1+\eta_n)\|x - x_n\|^2_{U_n} + \varepsilon_n,$$

where $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ and $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$.

# Variable metric quasi Fejér monotonicity: analysis

- Variable metric Quasi-Fejér monotone sequence :

$$(\forall x \in C)(\forall n \in \mathbb{N}) \quad \|x - x_{n+1}\|^2_{U_{n+1}} \leqslant (1+\eta_n)\|x - x_n\|^2_{U_n} + \varepsilon_n,$$

where $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ and $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$.

- Properties: set $d^U_C(x) = \inf\{\|y - x\|_U \mid y \in C\}$.
  1. $\left(d^{U_n}_C(x_n)\right)_{n\in\mathbb{N}}$ is quasi monotone.
  2. Weak convergence: $x_n \rightharpoonup \overline{x} \Leftrightarrow \mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset C$.
  3. Strong convergence: $x_n \to \overline{x} \Leftrightarrow \underline{\lim}\, d_C(x_n) = 0$.

# From Hilbert spaces to Banach spaces

- Notation:
  - $\mathcal{X}, \mathcal{Y}$: reflexive real Banach spaces.
  - $\langle \cdot, \cdot \rangle$: the canonic duality pairing, $\Delta$ is the duality mapping.
  - $\Gamma_0(\mathcal{X})$: the set of all proper lower semicontinuous convex functions defined on $\mathcal{X}$, $f^*$ is the Fenchel conjugate of $f$.

# From Hilbert spaces to Banach spaces

- Notation:
  - $\mathcal{X}, \mathcal{Y}$: reflexive real Banach spaces.
  - $\langle \cdot, \cdot \rangle$: the canonic duality pairing, $\Delta$ is the duality mapping.
  - $\Gamma_0(\mathcal{X})$: the set of all proper lower semicontinuous convex functions defined on $\mathcal{X}$, $f^*$ is the Fenchel conjugate of $f$.
- It is too difficult to manipulate the terms of the form $\|x - x_n\|^2$ in Banach spaces.

# From Hilbert spaces to Banach spaces

- Notation:
  - $\mathcal{X}, \mathcal{Y}$: reflexive real Banach spaces.
  - $\langle \cdot, \cdot \rangle$: the canonic duality pairing, $\Delta$ is the duality mapping.
  - $\Gamma_0(\mathcal{X})$: the set of all proper lower semicontinuous convex functions defined on $\mathcal{X}$, $f^*$ is the Fenchel conjugate of $f$.
- It is too difficult to manipulate the terms of the form $\|x - x_n\|^2$ in Banach spaces.
- The operator $P_C$ and its generalizations (proximal operators) becomes more complicated to manipulate, for example

$$p = P_C x \Leftrightarrow 0 \in N_C p + \Delta(p - x)$$

because the the gradient of the squared norm (the duality map) is not additive.

# Bregman monotonicity

- $f \in \Gamma_0(\mathcal{X})$ Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \varnothing$. The Bregman distance (L. M. Bregman 1967) associated with $f$ is

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

# Bregman monotonicity

- $f \in \Gamma_0(\mathcal{X})$ Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f \neq \varnothing$. The Bregman distance (L. M. Bregman 1967) associated with $f$ is

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

- $d_C^f(x) = \inf \left\{ D_f(y, x) \mid y \in C \right\}$.

# Bregman monotonicity

- $f \in \Gamma_0(\mathcal{X})$ Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f \neq \varnothing$. The Bregman distance (L. M. Bregman 1967) associated with $f$ is

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

- $d_C^f(x) = \inf \{ D_f(y, x) \mid y \in C \}$.
- Bregman monotonicity (H. H. Bauschke, J. M. Borwein, and P. L. Combettes 2003):
  - $(x_n)_{n\in\mathbb{N}}$ lies in $\operatorname{int} \operatorname{dom} f$.
  - $(\forall x \in C \cap \operatorname{int} \operatorname{dom} f)(\forall n \in \mathbb{N}) \quad D_f(x, x_{n+1}) \leqslant D_f(x, x_n)$.

# Bregman Monotonicity: convergence

- Conditions:
  1. $D_f(x, \cdot)$ is coercive for $x \in \operatorname{int}\operatorname{dom} f$.
  2. $\begin{cases} x' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}},\ x'' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \\ (x_n)_{n\in\mathbb{N}} \text{ is Bregman monotone w.r.t } C \end{cases} \Rightarrow \quad x' = x''.$
  3. For all bounded sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ in int dom $f$,

$$D_f(x_n, y_n) \to 0 \Rightarrow \quad x_n - y_n \to 0.$$

# Bregman Monotonicity: convergence

- Conditions:
  1. $D_f(x,\cdot)$ is coercive for $x \in \operatorname{int} \operatorname{dom} f$.
  2. $\begin{cases} x' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}},\ x'' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \\ (x_n)_{n\in\mathbb{N}} \text{ is Bregman monotone w.r.t } C \end{cases} \Rightarrow \quad x' = x''.$
  3. For all bounded sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ in int dom $f$,

$$D_f(x_n, y_n) \to 0 \Rightarrow \quad x_n - y_n \to 0.$$

- Properties:
  1. $\left(d_C^f(x_n)\right)_{n\in\mathbb{N}}$ converges.
  2. Suppose that Conditions (1) and (2) are satisfied. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $C \cap \overline{\operatorname{dom} f}$ if and only if $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset C$.
  3. Suppose that $x_n \rightharpoonup \overline{x} \in C \cap \operatorname{int} \operatorname{dom} f$ and Condition (3) is satisfied. Then $x_n \to \overline{x}$ if and only if $\mathfrak{S}(x_n)_{n\in\mathbb{N}} \neq \varnothing$.

# Quasi Bregman monotonicity: Notation and objectives

- Notation: Given $\alpha \in ]0,+\infty[$, set

$$\mathcal{F}_\alpha(f) = \{g \in \Gamma_0(\mathcal{X}) \mid D_g \geqslant \alpha D_f\}.$$

Note that $g \in \mathcal{F}_\alpha(f)$ if $g - \alpha f$ is convex, i.e., $g$ is more convex than $\alpha f$ in the terminology of J. J. MOREAU.

# Quasi Bregman monotonicity: Notation and objectives

- Notation: Given $\alpha \in \,]0,+\infty[$, set

$$\mathcal{F}_\alpha(f) = \{g \in \Gamma_0(\mathcal{X}) \mid D_g \geqslant \alpha D_f\}.$$

  Note that $g \in \mathcal{F}_\alpha(f)$ if $g - \alpha f$ is convex, i.e., $g$ is more convex than $\alpha f$ in the terminology of J. J. MOREAU.
- Objectives:
  - Give an unified framework for variable metric quasi Fejér monotonicity and Bregman monotonicity.
  - Apply to propose a forward-backward splitting in reflexive Banach spaces.

# Quasi Bregman monotonicity: Definitions

- Quasi Bregman monotonicity :
  - $(x_n)_{n\in\mathbb{N}}$ lies in int dom $f$.
  - $(\exists(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))(\forall x \in C \cap \text{int dom} f)(\exists(\varepsilon_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))(\forall n \in \mathbb{N})\ D_{f_{n+1}}(x, x_{n+1}) \leqslant (1+\eta_n)D_{f_n}(x, x_n) + \varepsilon_n$.

# Quasi Bregman monotonicity: Definitions

- Quasi Bregman monotonicity :
  - $(x_n)_{n\in\mathbb{N}}$ lies in int dom $f$.
  - $(\exists(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))(\forall x \in C \cap \text{int dom} f)(\exists(\varepsilon_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))(\forall n \in \mathbb{N})\ D_{f_{n+1}}(x, x_{n+1}) \leqslant (1+\eta_n)D_{f_n}(x, x_n) + \varepsilon_n$.
- Stationary quasi Bregman monotonicity :
  - $(x_n)_{n\in\mathbb{N}}$ lies in int dom $f$.
  - $(\exists(\varepsilon_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))(\exists(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))(\forall x \in C\cap\text{int dom} f)(\forall n \in \mathbb{N})\ D_{f_{n+1}}(x, x_{n+1}) \leqslant (1+\eta_n)D_{f_n}(x, x_n) + \varepsilon_n$.

# Quasi Bregman monotonicity: weak convergence

- Assumptions: $(f_n)_{n\in\mathbb{N}}$ is a sequence of functions in $\mathcal{F}_\alpha(f)$, and $(x_n)_{n\in\mathbb{N}}$ is quasi Bregman monotone with respect to $C$ relative to $(f_n)_{n\in\mathbb{N}}$.

# Quasi Bregman monotonicity: weak convergence

- Assumptions: $(f_n)_{n\in\mathbb{N}}$ is a sequence of functions in $\mathfrak{F}_\alpha(f)$, and $(x_n)_{n\in\mathbb{N}}$ is quasi Bregman monotone with respect to $C$ relative to $(f_n)_{n\in\mathbb{N}}$.
- $x_n \rightharpoonup \overline{x}$ if
  1. $D_f(x,\cdot)$ is coercive for every $x \in \operatorname{int}\operatorname{dom} f$.
  2. $(\exists g \in \Gamma_0(\mathcal{X}))(\exists \gamma \in\, ]0,+\infty[)\ \gamma g \in \bigcap_{n\in\mathbb{N}} \mathfrak{F}_1(f_n)$.
  3. $\begin{cases} x' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}, \quad x'' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \\ (\langle x'-x'', \nabla f_n(x_n)\rangle)_{n\in\mathbb{N}} \text{ converges} \end{cases} \Rightarrow \quad x' = x''.$
  4. $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset C \cap \operatorname{int}\operatorname{dom} f$.

# Quasi Bregman monotonicity: strong convergence

- Assumptions: $(f_n)_{n\in\mathbb{N}}$ is a sequence of functions in $\mathfrak{F}_\alpha(f)$, and $(x_n)_{n\in\mathbb{N}}$ is stationarily quasi Bregman monotone with respect to $C$ relative to $(f_n)_{n\in\mathbb{N}}$.

# Quasi Bregman monotonicity: strong convergence

- Assumptions: $(f_n)_{n\in\mathbb{N}}$ is a sequence of functions in $\mathcal{F}_\alpha(f)$, and $(x_n)_{n\in\mathbb{N}}$ is stationarily quasi Bregman monotone with respect to $C$ relative to $(f_n)_{n\in\mathbb{N}}$.
- $x_n \to \overline{x}$ if
  1. $C$ is a closed convex subset of int dom $f$.
  2. $f$ is uniformly convex on bounded subsets.
  3. $(\exists \beta \in\ ]0,+\infty[)\ \beta f \in \bigcap_{n\in\mathbb{N}} \mathcal{F}_1(f_n)$.
  4. $\underline{\lim}\, d_C^f(x_n) = 0$.

# Variable Bregman proximal point algorithm: Tools

- $f \in \Gamma_0(\mathcal{X})$ Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f \neq \varnothing$, $\varphi \in \Gamma_0(\mathcal{X})$, and $\gamma \in \,]0, +\infty[$.

$$\operatorname{prox}^f_{\gamma\varphi} : \mathcal{X} \to 2^{\mathcal{X}} : x \mapsto \underset{y \in \mathcal{X}}{\operatorname{Argmin}} \; \varphi(y) + \frac{1}{\gamma} D_f(y, x).$$

# Variable Bregman proximal point algorithm: Tools

- $f \in \Gamma_0(\mathcal{X})$ Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \varnothing$, $\varphi \in \Gamma_0(\mathcal{X})$, and $\gamma \in ]0, +\infty[$.

$$\operatorname{prox}^f_{\gamma\varphi} \colon \mathcal{X} \to 2^{\mathcal{X}} \colon x \mapsto \underset{y \in \mathcal{X}}{\operatorname{Argmin}}\ \varphi(y) + \frac{1}{\gamma} D_f(y, x).$$

- Suppose that $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom}\varphi)$ and $\operatorname{dom}\partial f \cap \operatorname{dom}\partial\varphi \subset \operatorname{int}\operatorname{dom} f$. Then
  1. $\operatorname{dom}\operatorname{prox}^f_{\gamma\varphi} \subset \operatorname{int}\operatorname{dom} f$; $\operatorname{ran}\operatorname{prox}^f_{\gamma\varphi} \subset \operatorname{int}\operatorname{dom} f$.
  2. $\operatorname{prox}^f_{\gamma\varphi} = (\nabla f + \gamma\partial\varphi)^{-1} \circ \nabla f$.
  3. $\operatorname{Fix}\operatorname{prox}^f_{\gamma\varphi} = \operatorname{int}\operatorname{dom} f \cap \operatorname{Argmin}\varphi$.
  4. Let $x \in \operatorname{Fix}\operatorname{prox}^f_{\gamma\varphi}$ and $(y, v) \in \operatorname{gra}\operatorname{prox}^f_{\gamma\varphi}$. Then $D_f(x, v) \leqslant D_f(x, y) - D_f(v, y)$.
  5. $\operatorname{prox}^f_{\gamma\varphi}$ is single-valued on its domain if $f|_{\operatorname{int}\operatorname{dom} f}$ is strictly convex.

# Variable Bregman proximal point algorithm: Tools

- $f \in \Gamma_0(\mathcal{X})$ Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \varnothing$, $\varphi \in \Gamma_0(\mathcal{X})$, and $\gamma \in \left]0,+\infty\right[$.

$$\operatorname{Prox}^f_{\gamma\varphi} \colon \mathcal{X}^* \to 2^{\mathcal{X}} \colon x^* \mapsto \underset{x\in\mathcal{X}}{\operatorname{Argmin}} \; \varphi(x) + \frac{1}{\gamma}\left(f(x) - \langle x, x^*\rangle\right).$$

# Variable Bregman proximal point algorithm: Tools

- $f \in \Gamma_0(\mathcal{X})$ Gâteaux differentiable on int dom $f \neq \varnothing$, $\varphi \in \Gamma_0(\mathcal{X})$, and $\gamma \in \left]0,+\infty\right[$.

$$\mathrm{Prox}^f_{\gamma\varphi} \colon \mathcal{X}^* \to 2^{\mathcal{X}} \colon x^* \mapsto \underset{x\in\mathcal{X}}{\mathrm{Argmin}}\ \ \varphi(x) + \frac{1}{\gamma}\left(f(x) - \langle x, x^*\rangle\right).$$

- Properties: suppose that $0 \in \mathrm{sri}(\mathrm{dom}\, f - \mathrm{dom}\, \varphi)$. Then
  1. $\mathrm{int}\,(\mathrm{dom}\, f^* + \gamma \mathrm{dom}\, \varphi^*) \subset \mathrm{dom}\, \mathrm{Prox}^f_{\gamma\varphi}$.
  2. $\mathrm{Prox}^f_{\gamma\varphi}$ is single-valued on its domain if $f|_{\mathrm{int\,dom}\, f}$ is strictly convex.
  3. $\mathrm{ran}\mathrm{Prox}^f_{\gamma\varphi} \subset \mathrm{int\ dom}\, f$ and $\mathrm{Prox}^f_{\gamma\varphi} = (\nabla f + \gamma\partial\varphi)^{-1}$ if $\mathrm{dom}\,\partial f \cap \mathrm{dom}\,\partial\varphi \subset \mathrm{int\,dom}\, f$.

# Variable Bregman proximal point algorithm: Algorithm

- $\varphi \in \Gamma_0(\mathcal{X})$ with $\operatorname{Argmin} \varphi \neq \varnothing$.
- $f \in \Gamma_0(\mathcal{X})$ strictly convex and Gâteaux differentiable on $\operatorname{int} \operatorname{dom} f \supset \operatorname{dom} \partial\varphi$.
- $\alpha \in \left]0, +\infty\right[$, $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, $(f_n)_{n\in\mathbb{N}}$ a sequence of functions in $\mathcal{F}_\alpha(f)$ such that $(\forall n \in \mathbb{N})$ $(1+\eta_n)f_n \in \mathcal{F}_1(f_{n+1})$.
- $(\gamma_n)_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}$ such that $0 < \inf_{n\in\mathbb{N}} \gamma_n \leqslant \sup_{n\in\mathbb{N}} \gamma_n < +\infty$,
- Iteration:

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}^{f_n}_{\gamma_n\varphi} x_n = \underset{x\in\mathcal{X}}{\operatorname{argmin}} \; \varphi(x) + \frac{1}{\gamma_n} D_{f_n}(x, x_n).$$

# Variable Bregman proximal point algorithm: analysis

The following hold for some $\overline{x} \in \text{Argmin}\varphi$.

- $(x_n)_{n\in\mathbb{N}}$ is stationarily quasi Bregman monotone w.r.t Argmin$\varphi$ relative to $(f_n)_{n\in\mathbb{N}}$ and is a minimizing sequence of $\varphi$.

# Variable Bregman proximal point algorithm: analysis

The following hold for some $\overline{x} \in \operatorname{Argmin}\varphi$.

- $(x_n)_{n\in\mathbb{N}}$ is stationarily quasi Bregman monotone w.r.t $\operatorname{Argmin}\varphi$ relative to $(f_n)_{n\in\mathbb{N}}$ and is a minimizing sequence of $\varphi$.
- $x_n \rightharpoonup \overline{x}$ if:
  1. $D_f(x,\cdot)$ is coercive for every $x \in \operatorname{int}\operatorname{dom} f$.
  2. $(\exists g \in \Gamma_0(\mathcal{X}))(\exists \beta \in \left]0,+\infty\right[)\ \beta g \in \bigcap_{n\in\mathbb{N}} \mathcal{F}_1(f_n)$.
  3. $\begin{cases} x' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}},\ x'' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \\ (\langle x'-x'', \nabla f_n(x_n)\rangle)_{n\in\mathbb{N}} \text{ converges} \end{cases} \quad\Rightarrow\quad x' = x''.$

# Variable Bregman proximal point algorithm: analysis

The following hold for some $\overline{x} \in \operatorname{Argmin}\varphi$.

- $(x_n)_{n\in\mathbb{N}}$ is stationarily quasi Bregman monotone w.r.t $\operatorname{Argmin}\varphi$ relative to $(f_n)_{n\in\mathbb{N}}$ and is a minimizing sequence of $\varphi$.
- $x_n \rightharpoonup \overline{x}$ if:
  1. $D_f(x,\cdot)$ is coercive for every $x \in \operatorname{int}\operatorname{dom} f$.
  2. $(\exists g \in \Gamma_0(\mathcal{X}))(\exists \beta \in\ ]0,+\infty[)\ \beta g \in \bigcap_{n\in\mathbb{N}} \mathcal{F}_1(f_n)$.
  3. $\begin{cases} x' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}},\ x'' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \\ (\langle x'-x'', \nabla f_n(x_n)\rangle)_{n\in\mathbb{N}} \text{ converges} \end{cases} \Rightarrow \quad x' = x''.$
- $x_n \to \overline{x}$ if:
  1. $f$ is uniformly convex on bounded subsets.
  2. $(\exists \beta \in\ ]0,+\infty[)\ \beta f \in \bigcap_{n\in\mathbb{N}} \mathcal{F}_1(f_n)$.
  3. $\underline{\lim}\, d^f_{\operatorname{Argmin}\varphi}(x_n) = 0$.

# Application: forward-backward splitting

- Problem: $\varphi \in \Gamma_0(\mathcal{X}), \psi \in \Gamma_0(\mathcal{Y})$ is differentiable, $L \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$. The problem is to

$$\underset{x \in \mathcal{X}}{\text{minimize}} \;\; \varphi(x) + \psi(Lx).$$

# Application: forward-backward splitting

- Problem: $\varphi \in \Gamma_0(\mathcal{X}), \psi \in \Gamma_0(\mathcal{Y})$ is differentiable, $L \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$. The problem is to

$$\underset{x \in \mathcal{X}}{\text{minimize}} \ \ \varphi(x) + \psi(Lx).$$

- Forward-Backward splitting in Hilbert spaces:
  - Algorithm :

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad x_{n+1} &= \text{prox}_{\gamma_n \varphi}(x_n - \gamma_n L^* \nabla \psi(Lx_n)) \\ &= \operatorname*{argmin}_{x \in \mathcal{H}} \ \varphi(x) + \langle x - x_n \mid L^* \nabla \psi(Lx_n) \rangle + \psi(Lx_n) \\ &\quad + \frac{1}{\gamma_n} \|x - x_n\|^2 \\ &= \text{prox}_{\gamma_n \Phi_{x_n}} x_n,\end{aligned}$$

  where $\Phi_y(x) = \langle x - y \mid L^* \nabla \psi(Ly) \rangle + \psi(Ly)$.
  - Algorithm requires that $\psi$ is differentiable on the whole space with a Lipschit gradient and the closed form of the proximal operators.

# Forward-backward splitting in Banach spces

- Assumptions:
  1. $f \in \Gamma_0(\mathcal{X})$ is strictly convex and Gâteaux differentiable.
  2. $L(\text{int dom} f) \subset \text{int dom}\,\psi$ and $\text{int dom} f \supset \text{dom}\,(\varphi + \psi \circ L)$.
  3. Either $f$ is cofinite or $-L^*(\text{ran}\nabla\psi) \subset \text{dom}\,\varphi^*$.
  4. $(\exists \beta \in\ ]0, +\infty[)\ \beta f \in \mathfrak{F}_1(\psi \circ L)$.

# Forward-backward splitting in Banach spces

- Assumptions:
  1. $f \in \Gamma_0(\mathcal{X})$ is strictly convex and Gâteaux differentiable.
  2. $L(\text{int dom}\, f) \subset \text{int dom}\, \psi$ and $\text{int dom}\, f \supset \text{dom}\,(\varphi + \psi \circ L)$.
  3. Either $f$ is cofinite or $-L^*(\text{ran}\nabla\psi) \subset \text{dom}\, \varphi^*$.
  4. $(\exists \beta \in \,]0, +\infty[)\ \beta f \in \mathfrak{F}_1(\psi \circ L)$.
- Algorithm:

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \text{Prox}^f_{\gamma_n\varphi}\left(\nabla f(x_n) - \gamma_n L^* \nabla\psi(Lx_n)\right).$$

where, $\varepsilon \in \,]0, \beta/(\beta+1)[$, $(\gamma_n)_{n\in\mathbb{N}} \in \mathbb{R}$, $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$,

$$(\forall n \in \mathbb{N}) \quad \varepsilon \leqslant \gamma_n \leqslant \beta(1-\varepsilon) \text{ and } (1+\eta_n)\gamma_n - \gamma_{n+1} \leqslant \beta\eta_n.$$

# Forward-backward splitting in Banach spces

$$
\begin{aligned}
x_{n+1} &= \underset{x\in\mathcal{X}}{\operatorname{argmin}}\ \varphi(x) + \langle x - x_n, L^*\nabla\psi(Lx_n)\rangle + \psi(Lx_n) + \frac{1}{\gamma_n}D_f(x, x_n) \\
&= \underset{x\in\mathcal{X}}{\operatorname{argmin}}\ (\varphi(x) + \psi(Lx)) - (\psi(Lx) - \langle x - x_n, L^*\nabla\psi(Lx_n)\rangle \\
&\qquad - \psi(Lx_n)) + \frac{1}{\gamma_n}D_f(x, x_n) \\
&= \underset{x\in\mathcal{X}}{\operatorname{argmin}}\ (\varphi(x) + \psi(Lx)) + \frac{1}{\gamma_n}(D_f(x, x_n) - \gamma_n D_\psi(Lx, Lx_n)) \\
&= \underset{x\in\mathcal{X}}{\operatorname{argmin}}\ (\varphi(x) + \psi(Lx)) + \frac{1}{\gamma_n}D_{f-\gamma_n\psi\circ L}(x, x_n) \\
&= \operatorname{prox}^{f_n}_{\gamma_n(\varphi+\psi\circ L)} x_n,
\end{aligned}
$$

where $(\forall n \in \mathbb{N})\ f_n = f - \gamma_n\psi \circ L$.

# Forward-backward: Convergence

The following hold for some $\overline{x} \in \text{Argmin}(\varphi + \psi \circ L)$:

- Weak convergence: suppose that
  1. $D_f(x, \cdot)$ is coercive for every $x \in \text{int dom}\, f$.
  2. $\nabla f$ and $\nabla \psi$ are weakly sequentially continuous.

  Then $x_n \rightharpoonup \overline{x}$.

# Forward-backward: Convergence

The following hold for some $\overline{x} \in \operatorname{Argmin}(\varphi + \psi \circ L)$:

- Weak convergence: suppose that
  1. $D_f(x, \cdot)$ is coercive for every $x \in \operatorname{int} \operatorname{dom} f$.
  2. $\nabla f$ and $\nabla \psi$ are weakly sequentially continuous.

  Then $x_n \rightharpoonup \overline{x}$.
- Strong convergence: suppose that
  1. $f$ is uniformly convex on bounded subsets.
  2. One of the following conditions holds:
     1. $\varphi$ is uniformly convex at $\overline{x}$.
     2. $\psi$ is uniformly convex at $L\overline{x}$ and $L$ is bounded below.
     3. $\underline{\lim}\, d^f_{\operatorname{Argmin}(\varphi+\psi\circ L)}(x_n) = 0$.

  Then $x_n \to \overline{x}$.

# Examples: notation

- Notation: $m$ and $K$ are strictly positive integers, $\mathbb{R}^m$ and $\mathbb{R}^K$ are usual Euclidean spaces, $(a_{ik})_{1\leqslant i\leqslant m, 1\leqslant k\leqslant K} \in \mathbb{R}_+^{m\times K}\setminus\{0\}$,
  $\alpha = \min\{a_{ik} \neq 0 \mid 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant K\}$,
  $\beta = \max\{a_{ik} \mid 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant K\}$.
- Some functions of interest:
  1. Burg entropy: $b \colon \xi \mapsto \begin{cases} -\ln \xi, & \text{if } \xi > 0; \\ +\infty, & \text{otherwise.} \end{cases}$
  2. Boltzmann-Shannon entropy:
     $$s \colon \xi \mapsto \begin{cases} \xi \ln \xi - \xi, & \text{if } \xi \geqslant 0; \\ +\infty, & \text{otherwise.} \end{cases}$$
- The Bregman distance associated with Burg entropy is the Itakura-Saito divergence and the Bregman distance associated with Boltzmann-Shannon entropy is the Kullback-Leibler divergence.

# Example 1

- Problem:

$$\underset{(\xi_1,\ldots,\xi_m)\in\,]0,+\infty[^m}{\text{minimize}} \quad \sum_{i=1}^{m}\varphi_i(\xi_i)+\sum_{k=1}^{K}\left(-\ln\frac{\sum_{i=1}^m a_{ik}\xi_i}{\varrho_k}+\frac{\sum_{i=1}^m a_{ik}\xi_i}{\varrho_k}-1\right),$$

which becomes

$$\underset{x\in\mathbb{R}^m}{\text{minimize}} \quad \varphi(x)+D_\psi(Lx,r),$$

where $\varphi\colon(\xi_i)_{1\leqslant i\leqslant m}\mapsto\sum_{i=1}^m\varphi_i(\xi_i)$, $L\colon(\xi_i)_{1\leqslant i\leqslant m}\mapsto(\sum_{i=1}^m a_{ik}\xi_i)_{1\leqslant k}$ $(\varrho_k)_{1\leqslant k\leqslant K}$, $\psi\colon(\rho_k)_{1\leqslant k\leqslant K}\mapsto\begin{cases}-\sum_{k=1}^K\ln\rho_k & \text{if } \min_{1\leqslant k\leqslant K}\rho_k>0,\\ +\infty & \text{otherwise}.\end{cases}$

# Example 1

- Problem:

$$\underset{(\xi_1,\ldots,\xi_m)\in\,]0,+\infty[^m}{\text{minimize}} \quad \sum_{i=1}^{m}\varphi_i(\xi_i)+\sum_{k=1}^{K}\Bigg(-\ln\frac{\sum_{i=1}^{m}a_{ik}\xi_i}{\varrho_k} + \frac{\sum_{i=1}^{m}a_{ik}\xi_i}{\varrho_k}-1\Bigg),$$

which becomes

$$\underset{x\in\mathbb{R}^m}{\text{minimize}} \quad \varphi(x)+D_\psi(Lx,r),$$

where $\varphi\colon(\xi_i)_{1\leqslant i\leqslant m}\mapsto\sum_{i=1}^{m}\varphi_i(\xi_i)$, $L\colon(\xi_i)_{1\leqslant i\leqslant m}\mapsto(\sum_{i=1}^{m}a_{ik}\xi_i)_{1\leqslant k}$
$(\varrho_k)_{1\leqslant k\leqslant K}$, $\psi\colon(\rho_k)_{1\leqslant k\leqslant K}\mapsto\begin{cases}-\sum_{k=1}^{K}\ln\rho_k & \text{if } \min_{1\leqslant k\leqslant K}\rho_k>0,\\ +\infty & \text{otherwise}.\end{cases}$

- Because $D_\psi(L\cdot,r)$ is not differentiable on $\mathbb{R}^m$, the classcical forward-backward can not be applied.

# Example 1

- Algorithm:

For $n = 0, 1, \ldots$

$\quad$ For $i = 1 \ldots, m$

$$\xi_{i,n+1} = \operatorname{Prox}^{b}_{\gamma_n \varphi_i}\left(\frac{-1}{\xi_{i,n}} - \gamma_n \sum_{k=1}^{K} a_{ik}\left(\frac{-1}{\sum_{j=1}^{m} a_{jk}\xi_{j,n}} + \frac{1}{\varrho_k}\right)\right).$$

Then $(\xi_{i,n})_{1 \leqslant i \leqslant m}$ converges to a solution.

# Example 1

- Algorithm:

For $n = 0, 1, \ldots$
  For $i = 1 \ldots, m$

$$\xi_{i,n+1} = \mathrm{Prox}^{b}_{\gamma_n \varphi_i}\left(\frac{-1}{\xi_{i,n}} - \gamma_n \sum_{k=1}^{K} a_{ik}\left(\frac{-1}{\sum_{j=1}^{m} a_{jk}\xi_{j,n}} + \frac{1}{\varrho_k}\right)\right).$$

Then $(\xi_{i,n})_{1\leqslant i\leqslant m}$ converges to a solution.

- In some cases, $\mathrm{Prox}^{b}_{\gamma_n\varphi_i}$ is easier to calculate than the classic proximal operator.

# Example 2

- Problem:

$$\underset{(\xi_1,\ldots,\xi_m)\in[0,+\infty[^m}{\text{minimize}} \sum_{i=1}^{m} \varphi_i(\xi_i) + \sum_{k=1}^{K}\left(\sum_{i=1}^{m} \alpha_{ik}\xi_i\right) \ln \frac{\sum_{i=1}^{m} \alpha_{ik}\xi_i}{\varrho_k} - \sum_{k=1}^{K}\sum_{i=1}^{m} \alpha_{ik}\xi_i + \sum_{k=1}^{K} \varrho_k,$$

which becomes

$$\underset{x\in\mathbb{R}^m}{\text{minimize}} \ \ \varphi(x) + D_\psi(Lx, r),$$

where $\varphi\colon (\xi_i)_{1\leqslant i\leqslant m} \mapsto \sum_{i=1}^{m} \varphi_i(\xi_i)$, $L\colon (\xi_i)_{1\leqslant i\leqslant m} \mapsto (\sum_{i=1}^{m} \alpha_{ik}\xi_i)_{1\leqslant k}$
$(\varrho_k)_{1\leqslant k\leqslant K}$, $\psi\colon (\rho_k)_{1\leqslant k\leqslant K} \mapsto \begin{cases} \sum_{k=1}^{K} \rho_k \ln \rho_k - \rho_k & \text{if } \min_{1\leqslant k\leqslant K} \rho_k \\ +\infty & \text{otherwise.} \end{cases}$

# Example 2

- Problem:

$$\underset{(\xi_1,\ldots,\xi_m)\in[0,+\infty[^m}{\text{minimize}} \sum_{i=1}^{m}\varphi_i(\xi_i)+\sum_{k=1}^{K}\left(\sum_{i=1}^{m}a_{ik}\xi_i\right)\ln\frac{\sum_{i=1}^{m}a_{ik}\xi_i}{\varrho_k} - \sum_{k=1}^{K}\sum_{i=1}^{m}a_{ik}\xi_i+\sum_{k=1}^{K}\varrho_k,$$

  which becomes

$$\underset{x\in\mathbb{R}^m}{\text{minimize}} \ \varphi(x)+D_\psi(Lx,r),$$

  where $\varphi\colon(\xi_i)_{1\leqslant i\leqslant m}\mapsto\sum_{i=1}^{m}\varphi_i(\xi_i)$, $L\colon(\xi_i)_{1\leqslant i\leqslant m}\mapsto(\sum_{i=1}^{m}a_{ik}\xi_i)_{1\leqslant k}$ $(\varrho_k)_{1\leqslant k\leqslant K}$, $\psi\colon(\rho_k)_{1\leqslant k\leqslant K}\mapsto\begin{cases}\sum_{k=1}^{K}\rho_k\ln\rho_k-\rho_k & \text{if } \min_{1\leqslant k\leqslant K}\rho_k\\ +\infty & \text{otherwise.}\end{cases}$

- Because $D_\psi(L\cdot,r)$ is not differentiable on $\mathbb{R}^m$,the classcical forward-backward can not be applied.

# Example 2

- Algorithm:

For $n = 0, 1, \ldots$

$\quad$ For $i = 1 \ldots, m$

$$\xi_{i,n+1} = \mathrm{prox}^{s}_{\gamma_n \varphi_i}\left(\xi_{i,n} \prod_{k=1}^{K} \left(\frac{\varrho_k}{\sum_{j=1}^{m} a_{jk}\xi_{j,n}}\right)^{\gamma_n a_{ik}}\right).$$

Then $(\xi_{i,n})_{1 \leqslant i \leqslant m}$ converges to a solution.

# Example 2

- Algorithm:

$$\begin{array}{l} \text{For } n = 0, 1, \dots \\ \left\lfloor \begin{array}{l} \text{For } i = 1 \dots, m \\ \left\lfloor \xi_{i,n+1} = \text{prox}^{s}_{\gamma_n \varphi_i} \left( \xi_{i,n} \prod_{k=1}^{K} \left( \frac{\varrho_k}{\sum_{j=1}^{m} a_{jk} \xi_{j,n}} \right)^{\gamma_n a_{ik}} \right). \right. \end{array} \right. \end{array}$$

  Then $(\xi_{i,n})_{1 \leqslant i \leqslant m}$ converges to a solution.
- In some cases, $\text{Prox}^{s}_{\gamma_n \varphi_i}$ is easier to calculate than the classic proximal operator.

# Some remarks

- Our algorithm can be applied for some problems set in non-hilbertian reflexive real Banach spaces, such as $\ell^p(\mathbb{N})$ for some $p \in ]1, +\infty[$.

# Some remarks

- Our algorithm can be applied for some problems set in non-hilbertian reflexive real Banach spaces, such as $\ell^p(\mathbb{N})$ for some $p \in ]1, +\infty[$.
- Stochastic version: Given a probability space $(\Omega, \mathscr{F}, \mu)$. Suppose that $D_f(x, \cdot)$ is $\mathscr{F}-$measurable. We can generalize the above Bregman monotonicity to a stochastic process.

# Some remarks

- Our algorithm can be applied for some problems set in non-hilbertian reflexive real Banach spaces, such as $\ell^p(\mathbb{N})$ for some $p \in ]1, +\infty[$.
- Stochastic version: Given a probability space $(\Omega, \mathscr{F}, \mu)$. Suppose that $D_f(x, \cdot)$ is $\mathscr{F}-$measurable. We can generalize the above Bregman monotonicity to a stochastic process.
- Open problems:
  - Forward-Backward splitting for sum of two monotone operators.
  - Recently, M. A. Alghamdi, A. Alotaibi, P. L. Combettes, and N. Shahzad 2014 proposed a primal-dual method of partial inverses for composite inclusions, which is, in fact, can be rewritten as a proximal algorithm. We don't know if there exists a suitable generalization of partial inverse, initiated by Spingarn, in Banach spaces.

# References


- H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.*, vol. 42, pp. 596–636, 2003.
- M. A. Alghamdi, A. Alotaibi, P. L. Combettes, and N. Shahzad, A primal-dual method of partial inverses for composite inclusions, *Optimization Letters*, 2014.
- L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programmings, *USSR Comp. Math. Math. Phys.*, vol 7, pp. 200–217, 1967.

# References


- P. L. Combettes and J. C. Pesquet, Stochastic quasi-Fejr block-coordinate fixed point iterations with random sweeping, *arXiv:1404.7536*, 2014.
- P. L. Combettes and B. C. Vũ, Variable metric quasi-Fejér monotonicity, *Nonlinear Anal.*, vol. 78, pp. 17–31, 2013.