

Horizontal and Vertical Block Decomposition of Monotone Operator Splitting Algorithms

Patrick L. Combettes

(joint work with J.-C. Pesquet)

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

Limoges, May 21, 2015, celebrating Terry's 80th birthday!

References

- PLC and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping, *SIAM J. Optim.*, 2015 (arxiv, April 2014)
- R. T. Rockafellar, “In a way, all you need is two,” personal communication, RTR’s kitchen, Seattle, May 1998

Framework

- Problem: Find a point in a closed convex subset F of a Hilbert space H
- Mathematical model:

$$\text{find } x \in F \subset \bigcap_{n \in \mathbb{N}} \text{Fix} T_n, \quad T_n: H \rightarrow H \quad \text{quasi-nonexpansive}$$

- Algorithmic model:

$$x_{n+1} = T_n x_n$$

- Asymptotic analysis: under suitable assumptions, $(x_n)_{n \in \mathbb{N}}$ converges (weakly/strongly/linearly) to a point in F
- Application areas: variational inequalities, game theory, optimization, statistics, partial differential equations, inverse problems, mechanics, signal and image processing, machine learning, computer vision, transport theory, optics,...

Basic convergence principle

- Algorithmic model: $x_{n+1} = T_n x_n$
- T_n is quasi-nonexpansive, i.e.,

$$(\forall x \in H)(\forall y \in \text{Fix } T_n) \quad \|T_n x - y\| \leq \|x - y\|,$$

with $F \subset \text{Fix } T_n$. Then:

- **Fejér monotonicity:**

$$(\forall n \in \mathbb{N})(\forall y \in F) \quad \|x_{n+1} - y\| \leq \|x_n - y\|$$

- Suppose that the set $\mathfrak{W}(x_n)_{n \in \mathbb{N}}$ of weak cluster points of $(x_n)_{n \in \mathbb{N}}$ is in F . Then $x_n \rightharpoonup x \in F$

Basic convergence principle

- Algorithmic model: $x_{n+1} = T_n x_n$
- T_n is quasi-nonexpansive, i.e.,

$$(\forall x \in H)(\forall y \in \text{Fix } T_n) \quad \|T_n x - y\| \leq \|x - y\|,$$

with $F \subset \text{Fix } T_n$. Then:

- **Fejér monotonicity:**

$$(\forall n \in \mathbb{N})(\forall y \in F) \quad \|x_{n+1} - y\| \leq \|x_n - y\|$$

- Suppose that the set $\mathfrak{W}(x_n)_{n \in \mathbb{N}}$ of weak cluster points of $(x_n)_{n \in \mathbb{N}}$ is in F . Then $x_n \rightharpoonup x \in F$
- Example 1: alternating projection method for finding a point in the intersection of two closed convex sets C_1 and C_2 . Set $T_{2n} = P_2$ and $T_{2n+1} = P_1$.
- Example 2: proximal point algorithm for finding a zero of a maximally monotone operator A . Set $T_n = J_{\gamma_n A}$.

Example: Special cases

- Fixed point methods: Krasnosel'skiĭ–Mann, string averaging, extrapolated barycentric methods, etc
- Projections methods for convex feasibility problems
- Projections methods for split feasibility problems
- Subgradient projections methods for systems of convex inequalities
- Splitting methods for monotone inclusions
- Convex optimization methods: projected gradient method, augmented Lagrangian method, ADMM, various proximal splitting methods, etc.
- Iterative methods for variational inequalities in mechanics, traffic theory, and finance
- etc.

Very large scale problems: Block-coordinate approach

- The basic iteration $\mathbf{x}_{n+1} = \mathbf{T}_n \mathbf{x}_n$ in the Hilbert space \mathbf{H} may be too involved (computations, memory) to be operational
- We assume that \mathbf{H} can be decomposed in m factors

$$\mathbf{H} = H_1 \times \cdots \times H_m$$

in which each \mathbf{T}_n has an explicit decomposition

$$\mathbf{T}_n: \mathbf{x} \mapsto (\mathbf{T}_{1,n}\mathbf{x}, \dots, \mathbf{T}_{m,n}\mathbf{x})$$

- The strategy is to update only arbitrarily chosen coordinates of $\mathbf{x}_{n+1} = (x_{1,n+1}, \dots, x_{m,n+1})$ up to some tolerance:

$$x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} (\mathbf{T}_{i,n}\mathbf{x}_n + \mathbf{a}_{i,n} - x_{i,n}),$$

where $\varepsilon_{i,n} \in \{0, 1\}$ (activation variable)

Very large scale problems: Block-coordinate approach

- The basic iteration $\mathbf{x}_{n+1} = \mathbf{T}_n \mathbf{x}_n$ in the Hilbert space \mathbf{H} may be too involved (computations, memory) to be operational
- We assume that \mathbf{H} can be decomposed in m factors

$$\mathbf{H} = H_1 \times \cdots \times H_m$$

in which each \mathbf{T}_n has an explicit decomposition

$$\mathbf{T}_n: \mathbf{x} \mapsto (\mathbf{T}_{1,n}\mathbf{x}, \dots, \mathbf{T}_{m,n}\mathbf{x})$$

- The strategy is to update only arbitrarily chosen coordinates of $\mathbf{x}_{n+1} = (x_{1,n+1}, \dots, x_{m,n+1})$ up to some tolerance:

$$x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}(\mathbf{T}_{i,n}\mathbf{x}_n + \mathbf{a}_{i,n} - x_{i,n}),$$

where $\varepsilon_{i,n} \in \{0, 1\}$ (activation variable)

- *Our goal is to extend available fixed points methods to this block-coordinate setting **while preserving their convergence properties***

A roadblock

- The nice properties of an operator $\mathbf{T}: \mathbf{H} \rightarrow \mathbf{H}$ are destroyed by coordinate sampling
- For instance, consider the nonexpansive (1-Lipschitz) operator

$$\mathbf{T}: (x_1, x_2) \mapsto (-x_2, x_1)$$

Then

$$\begin{cases} \mathcal{Q}_1: (x_1, x_2) \mapsto (x_1, x_1) \\ \mathcal{Q}_2: (x_1, x_2) \mapsto (-x_2, x_2) \\ \mathcal{Q}_1 \circ \mathcal{Q}_2 \\ \mathcal{Q}_2 \circ \mathcal{Q}_1 \end{cases}$$

are no longer nonexpansive

- Fejér monotonicity is destroyed

A roadblock

- The nice properties of an operator $\mathbf{T}: \mathbf{H} \rightarrow \mathbf{H}$ are destroyed by coordinate sampling
- For instance, consider the nonexpansive (1-Lipschitz) operator

$$\mathbf{T}: (x_1, x_2) \mapsto (-x_2, x_1)$$

Then

$$\left\{ \begin{array}{l} \mathcal{Q}_1: (x_1, x_2) \mapsto (x_1, x_1) \\ \mathcal{Q}_2: (x_1, x_2) \mapsto (-x_2, x_2) \\ \mathcal{Q}_1 \circ \mathcal{Q}_2 \\ \mathcal{Q}_2 \circ \mathcal{Q}_1 \end{array} \right.$$

are no longer nonexpansive

- Fejér monotonicity is destroyed
- .. and even for jointly convex functions with a unique minimizer, alternating minimizations fail, etc.

A roadblock

- The nice properties of an operator $\mathbf{T}: \mathbf{H} \rightarrow \mathbf{H}$ are destroyed by coordinate sampling
- For instance, consider the nonexpansive (1-Lipschitz) operator

$$\mathbf{T}: (x_1, x_2) \mapsto (-x_2, x_1)$$

Then

$$\left\{ \begin{array}{l} \mathcal{Q}_1: (x_1, x_2) \mapsto (x_1, x_1) \\ \mathcal{Q}_2: (x_1, x_2) \mapsto (-x_2, x_2) \\ \mathcal{Q}_1 \circ \mathcal{Q}_2 \\ \mathcal{Q}_2 \circ \mathcal{Q}_1 \end{array} \right.$$

are no longer nonexpansive

- Fejér monotonicity is destroyed
- .. and even for jointly convex functions with a unique minimizer, alternating minimizations fail, etc.
- \rightsquigarrow introduce **stochasticity** and renorming

Notation

- H : separable real Hilbert space; \rightharpoonup : weak convergence
- $\mathfrak{W}(x_n)_{n \in \mathbb{N}}$: set of weak cluster points of $(x_n)_{n \in \mathbb{N}} \in H^{\mathbb{N}}$
- $\Gamma_0(H)$: proper lower semicontinuous convex functions from H to $] -\infty, +\infty]$
- (Ω, \mathcal{F}, P) : underlying probability space
- Given a sequence $(x_n)_{n \in \mathbb{N}}$ of H -valued random variables,

$$\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}, \quad \text{where } (\forall n \in \mathbb{N}) \quad \mathcal{X}_n = \sigma(x_0, \dots, x_n)$$

- $l_+(\mathcal{X})$: set of sequences of $[0, +\infty[$ -valued random variables $(\xi_n)_{n \in \mathbb{N}}$ such that, for every $n \in \mathbb{N}$, ξ_n is \mathcal{X}_n -measurable
- $l_+^p(\mathcal{X}) = \left\{ (\xi_n)_{n \in \mathbb{N}} \in l_+(\mathcal{X}) \mid \sum_{n \in \mathbb{N}} \xi_n^p < +\infty \text{ P-a.s.} \right\}, p \in]0, +\infty[$
- $l_+^\infty(\mathcal{X}) = \left\{ (\xi_n)_{n \in \mathbb{N}} \in l_+(\mathcal{X}) \mid \sup_{n \in \mathbb{N}} \xi_n < +\infty \text{ P-a.s.} \right\}$

Stochastic quasi-Fejér sequences

- $\emptyset \neq F \subset H$, $\phi: [0, +\infty[\rightarrow [0, +\infty[$, $\phi(t) \uparrow +\infty$ as $t \rightarrow +\infty$
- **Deterministic definition:** A sequence $(x_n)_{n \in \mathbb{N}}$ in H is Fejér monotone with respect to F if for every $z \in F$,

$$(\forall n \in \mathbb{N}) \quad \phi(\|x_{n+1} - z\|) \leq \phi(\|x_n - z\|)$$

Stochastic quasi-Fejér sequences

- $\emptyset \neq F \subset H$, $\phi: [0, +\infty[\rightarrow [0, +\infty[$, $\phi(t) \uparrow +\infty$ as $t \rightarrow +\infty$
- **Stochastic definition 1:** A sequence $(x_n)_{n \in \mathbb{N}}$ of H -valued random variables is *stochastically Fejér monotone* with respect to F if, for every $z \in F$,

$$(\forall n \in \mathbb{N}) \mathbf{E}(\phi(\|x_{n+1} - z\|) \mid \mathcal{X}_n) \leq \phi(\|x_n - z\|)$$

Stochastic quasi-Fejér sequences

- $\emptyset \neq F \subset H$, $\phi: [0, +\infty[\rightarrow [0, +\infty[$, $\phi(t) \uparrow +\infty$ as $t \rightarrow +\infty$
- **Stochastic definition 2:** A sequence $(x_n)_{n \in \mathbb{N}}$ of H -valued random variables is *stochastically quasi-Fejér monotone* with respect to F if, for every $z \in F$, there exist $(\chi_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathcal{X})$, $(\vartheta_n(z))_{n \in \mathbb{N}} \in \ell_+(\mathcal{X})$, and $(\eta_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathcal{X})$ such that

$$(\forall n \in \mathbb{N}) \mathbb{E}(\phi(\|x_{n+1} - z\|) | \mathcal{X}_n) + \vartheta_n(z) \leq (1 + \chi_n(z))\phi(\|x_n - z\|) + \eta_n(z)$$

Stochastic quasi-Fejér sequences

- $\emptyset \neq F \subset H$, $\phi: [0, +\infty[\rightarrow [0, +\infty[$, $\phi(t) \uparrow +\infty$ as $t \rightarrow +\infty$
- **Stochastic definition 2:** A sequence $(x_n)_{n \in \mathbb{N}}$ of H -valued random variables is *stochastically quasi-Fejér monotone* with respect to F if, for every $z \in F$, there exist $(\chi_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathcal{X})$, $(\vartheta_n(z))_{n \in \mathbb{N}} \in \ell_+(\mathcal{X})$, and $(\eta_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathcal{X})$ such that

$$(\forall n \in \mathbb{N}) \mathbb{E}(\phi(\|x_{n+1} - z\|) | \mathcal{X}_n) + \vartheta_n(z) \leq (1 + \chi_n(z))\phi(\|x_n - z\|) + \eta_n(z)$$

Theorem

Suppose $(x_n)_{n \in \mathbb{N}}$ is stochastically quasi-Fejér monotone. Then

- $(\forall z \in F) \left[\sum_{n \in \mathbb{N}} \vartheta_n(z) < +\infty \text{ P-a.s.} \right]$
- $[\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset F \text{ P-a.s.}] \Leftrightarrow [(x_n)_{n \in \mathbb{N}} \text{ converges weakly P-a.s. to an } F\text{-valued random variable}]$

An abstract stochastic iterative scheme

Theorem

Let $\emptyset \neq F \subset H$. Suppose that $(x_n)_{n \in \mathbb{N}}$, $(t_n)_{n \in \mathbb{N}}$, and $(e_n)_{n \in \mathbb{N}}$ are sequences of H -valued random variables such that:

- $(\forall n \in \mathbb{N}) x_{n+1} = x_n + \lambda_n(t_n + e_n - x_n), \quad \lambda_n \in]0, 1]$
- $\sum_{n \in \mathbb{N}} \lambda_n \sqrt{E(\|e_n\|^2 | \mathcal{X}_n)} < +\infty$ P-a.s.
- For every $z \in F$, there exist $(\theta_n(z))_{n \in \mathbb{N}} \in \ell_+(\mathcal{X})$, $(\mu_n(z))_{n \in \mathbb{N}} \in \ell_+^\infty(\mathcal{X})$, and $(\nu_n(z))_{n \in \mathbb{N}} \in \ell_+^\infty(\mathcal{X})$ such that $(\lambda_n \mu_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathcal{X})$, $(\lambda_n \nu_n(z))_{n \in \mathbb{N}} \in \ell_+^{1/2}(\mathcal{X})$, and

$$(\forall n \in \mathbb{N}) \quad E(\|t_n - z\|^2 | \mathcal{X}_n) + \theta_n(z) \leq (1 + \mu_n(z))\|x_n - z\|^2 + \nu_n(z)$$

Then $(\forall z \in F) \left[\sum_{n \in \mathbb{N}} \lambda_n \theta_n(z) < +\infty \text{ P-a.s.} \right]$ and $[\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset F \text{ P-a.s.}] \Rightarrow (x_n)_{n \in \mathbb{N}}$ converges weakly P-a.s. to an F -valued random variable.

Single-layer algorithm

- $\mathbf{H} = H_1 \times \cdots \times H_m$, $(H_i)_{1 \leq i \leq m}$ separable real Hilbert spaces
- $\mathbf{T}_n: \mathbf{H} \rightarrow \mathbf{H}: \mathbf{x} \mapsto (\mathbf{T}_{i,n} \mathbf{x})_{1 \leq i \leq m}$ quasinonexpansive
- $\mathbf{F} = \bigcap_{n \in \mathbb{N}} \text{Fix} \mathbf{T}_n \neq \emptyset$
- \mathbf{x}_0 and the errors $(\mathbf{a}_n)_{n \in \mathbb{N}}$ are \mathbf{H} -valued random variables
- $(\varepsilon_n)_{n \in \mathbb{N}}$ identically distributed D -valued random variables with $D = \{0, 1\}^m \setminus \{\mathbf{0}\}$
- Algorithm:

for $n = 0, 1, \dots$

for $i = 1, \dots, m$	
	$x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \lambda_n (\mathbf{T}_{i,n}(x_{1,n}, \dots, x_{m,n}) + \mathbf{a}_{i,n} - x_{i,n})$

Single-layer algorithm

Theorem

Set $(\forall n \in \mathbb{N}) \mathbf{x}_n = \sigma(\mathbf{x}_0, \dots, \mathbf{x}_n)$ and $\mathcal{E}_n = \sigma(\varepsilon_n)$. Assume that

- $\inf_{n \in \mathbb{N}} \lambda_n > 0$ and $\sup_{n \in \mathbb{N}} \lambda_n < 1$.
- $\sum_{n \in \mathbb{N}} \sqrt{\mathbb{E}(\|\mathbf{a}_n\|^2 | \mathcal{X}_n)} < +\infty$.
- $\mathfrak{W}(\mathbf{x}_n)_{n \in \mathbb{N}} \subset \mathbf{F}$ P-a.s.
- For every $n \in \mathbb{N}$, \mathcal{E}_n and \mathcal{X}_n are independent.
- For every $i \in \{1, \dots, m\}$, $p_i = \mathbb{P}[\varepsilon_{i,0} = 1] > 0$.

Then $(\mathbf{x}_n)_{n \in \mathbb{N}}$ converges weakly P-a.s. to an \mathbf{F} -valued r.v.

Proof.

We achieve stochastic quasi-Fejér monotonicity w.r.t. the norm

$$\|\mathbf{x}\|^2 = \sum_{i=1}^m \|\mathbf{x}_i\|^2 / p_i \quad \square$$

Example: Krasnosel'skiĭ–Mann iteration

- $\mathbf{T}: \mathbf{H} \rightarrow \mathbf{H}: \mathbf{x} \mapsto (\mathbb{T}_i \mathbf{x})_{1 \leq i \leq m}$ nonexpansive operator
- $\mathbf{F} = \text{Fix} \mathbf{T} \neq \emptyset$
- \mathbf{x}_0 and the errors $(\mathbf{a}_n)_{n \in \mathbb{N}}$ are \mathbf{H} -valued random variables
- $(\varepsilon_n)_{n \in \mathbb{N}}$ identically distributed D -valued random variables with $D = \{0, 1\}^m \setminus \{\mathbf{0}\}$
- Algorithm:

$$\begin{array}{l} \text{for } n = 0, 1, \dots \\ \quad \left[\begin{array}{l} \text{for } i = 1, \dots, m \\ \quad \left[X_{i,n+1} = X_{i,n} + \varepsilon_{i,n} \lambda_n (\mathbb{T}_i (X_{1,n}, \dots, X_{m,n}) + \mathbf{a}_{i,n} - X_{i,n}) \end{array} \right. \end{array} \right.$$

Example: Krasnosel'skiĭ–Mann iteration

Theorem

Set $(\forall n \in \mathbb{N}) \mathbf{x}_n = \sigma(\mathbf{x}_0, \dots, \mathbf{x}_n)$ and $\mathcal{E}_n = \sigma(\varepsilon_n)$. Assume that

- $\inf_{n \in \mathbb{N}} \lambda_n > 0$ and $\sup_{n \in \mathbb{N}} \lambda_n < 1$
- $\sum_{n \in \mathbb{N}} \sqrt{\mathbb{E}(\|\mathbf{a}_n\|^2 | \mathcal{X}_n)} < +\infty$
- For every $n \in \mathbb{N}$, \mathcal{E}_n and \mathcal{X}_n are independent
- For every $i \in \{1, \dots, m\}$, $\mathbb{P}[\varepsilon_{i,0} = 1] > 0$

Then $(\mathbf{x}_n)_{n \in \mathbb{N}}$ converges weakly P-a.s. to an \mathbf{F} -valued r.v.

Proof.

Apply the single-layer theorem with $\mathbf{T}_n = \mathbf{T}$. □

Example: Block-coordinate, primal-dual splitting of coupled composite monotone inclusions

- Let \mathbf{F} be the set of solutions to the problem

find $x_1 \in H_1, \dots, x_m \in H_m$ such that

$$(\forall i \in \{1, \dots, m\}) \quad 0 \in A_i x_i + \sum_{k=1}^p L_{ki}^* B_k \left(\sum_{j=1}^m L_{kj} x_j \right),$$

where each $A_i: H_i \rightarrow 2^{H_i}$, $B_k: G_k \rightarrow 2^{G_k}$ are maximally monotone, $L_{ki}: H_i \rightarrow G_k$ linear & bounded

- Let \mathbf{F}^* be the set of solutions to the dual problem

find $v_1 \in G_1, \dots, v_p \in G_p$ such that

$$(\forall k \in \{1, \dots, p\}) \quad 0 \in - \sum_{i=1}^m L_{ki} A_i^{-1} \left(- \sum_{l=1}^p L_{li}^* v_l \right) + B_k^{-1} v_k$$

Example: Block-coordinate, primal-dual splitting of coupled composite monotone inclusions

Let \mathcal{Q}_j ($1 \leq j \leq m+p$) be the j th component of the projector P_V onto the subspace

$$\mathbf{V} = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbf{H} \times \mathbf{G} \mid (\forall k \in \{1, \dots, p\}) y_k = \sum_{i=1}^m L_{ki} x_i \right\}$$

Algorithm: $\gamma \in]0, +\infty[$ and for $n = 0, 1, \dots$

$$\mu_n \in]0, 2[$$

for $i = 1, \dots, m$

$$\begin{cases} Z_{i,n+1} = Z_{i,n} + \varepsilon_{i,n} (\mathcal{Q}_i(x_{1,n}, \dots, x_{m,n}, y_{1,n}, \dots, y_{p,n}) + c_{i,n} - Z_{i,n}) \\ x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \mu_n (J_{\gamma A_i}(2Z_{i,n+1} - x_{i,n}) + a_{i,n} - Z_{i,n+1}) \end{cases}$$

for $k = 1, \dots, p$

$$\begin{cases} w_{k,n+1} = w_{k,n} + \varepsilon_{m+k,n} (\mathcal{Q}_{m+k}(x_{1,n}, \dots, x_{m,n}, y_{1,n}, \dots, y_{p,n}) + d_{k,n} - w_{k,n}) \\ y_{k,n+1} = y_{k,n} + \varepsilon_{m+k,n} \mu_n (J_{\gamma B_k}(2w_{k,n+1} - y_{k,n}) + b_{k,n} - w_{k,n+1}). \end{cases}$$

Example: Block-coordinate, primal-dual splitting of coupled composite monotone inclusions

- Under the same conditions as before:
 - $(\mathbf{z}_n)_{n \in \mathbb{N}}$ converges weakly P-a.s. to an \mathbf{F} -valued random variable
 - $(\gamma^{-1}(\mathbf{w}_n - \mathbf{y}_n))_{n \in \mathbb{N}}$ converges weakly P-a.s. to an \mathbf{F}^* -valued random variable
- Proof: This relies on the single-layer theorem and a nonstandard implementation of the Douglas-Rachford algorithm in the product space $\mathbf{H} \times \mathbf{G} = \mathbf{H}_1 \times \cdots \times \mathbf{H}_m \times \mathbf{G}_1 \times \cdots \times \mathbf{G}_p$:

solve $(0, 0) \in \mathbf{Ax} \times \mathbf{By} + N_{\mathbf{V}}(\mathbf{x}, \mathbf{y})$, where

$$\begin{cases} \mathbf{A}: \mathbf{H} \rightarrow 2^{\mathbf{H}}: \mathbf{x} \mapsto \bigtimes_{i=1}^m \mathbf{A}_i \mathbf{x}_i \\ \mathbf{B}: \mathbf{G} \rightarrow 2^{\mathbf{G}}: \mathbf{x} \mapsto \bigtimes_{k=1}^p \mathbf{B}_k \mathbf{y}_k \\ \mathbf{V} = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbf{H} \times \mathbf{G} \mid (\forall k \in \{1, \dots, p\}) \mathbf{y}_k = \sum_{i=1}^m \mathbf{L}_{ki} \mathbf{x}_i \right\} \end{cases}$$

Double-layer random block-coordinate algorithms

- $\mathbf{T}_n: \mathbf{H} \rightarrow \mathbf{H}: \mathbf{x} \mapsto (\mathbb{T}_{i,n} \mathbf{x})_{1 \leq i \leq m}$ is α_n -averaged ($\text{Id} + \alpha_n^{-1}(\mathbf{T}_n - \text{Id})$ is nonexpansive)
- $\mathbf{R}_n: \mathbf{H} \rightarrow \mathbf{H}$ is β_n -averaged
- $\mathbf{F} = \bigcap_{n \in \mathbb{N}} \text{Fix} \mathbf{T}_n \circ \mathbf{R}_n \neq \emptyset$
- $\mathbf{x}_0, (\mathbf{a}_n)_{n \in \mathbb{N}}$, and $(\mathbf{b}_n)_{n \in \mathbb{N}}$ are \mathbf{H} -valued random variables
- $(\varepsilon_n)_{n \in \mathbb{N}}$ identically distributed D -valued random variables with $D = \{0, 1\}^m \setminus \{\mathbf{0}\}$
- Algorithm:

$$\begin{array}{l} \text{for } n = 0, 1, \dots \\ \left[\begin{array}{l} \mathbf{y}_n = \mathbf{R}_n \mathbf{x}_n + \mathbf{b}_n \\ \text{for } i = 1, \dots, m \\ \quad \left[\begin{array}{l} x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} (\mathbb{T}_{i,n} \mathbf{y}_n + \mathbf{a}_{i,n} - x_{i,n}) \end{array} \right. \end{array} \right. \end{array}$$

Double-layer random block-coordinate algorithms

Theorem

Set $(\forall n \in \mathbb{N}) \mathcal{X}_n = \sigma(\mathbf{x}_0, \dots, \mathbf{x}_n)$ and $\mathcal{E}_n = \sigma(\varepsilon_n)$. Assume that

- $\sum_{n \in \mathbb{N}} \sqrt{\mathbb{E}(\|\mathbf{a}_n\|^2 | \mathcal{X}_n)} < +\infty, \sum_{n \in \mathbb{N}} \sqrt{\mathbb{E}(\|\mathbf{b}_n\|^2 | \mathcal{X}_n)} < +\infty$
- $\mathfrak{W}(\mathbf{x}_n)_{n \in \mathbb{N}} \subset \mathbf{F}$ P-a.s.
- For every $n \in \mathbb{N}$, \mathcal{E}_n and \mathcal{X}_n are independent
- For every $i \in \{1, \dots, m\}$, $p_i = \mathbb{P}[\varepsilon_{i,0} = 1] > 0$

Then $(\mathbf{x}_n)_{n \in \mathbb{N}}$ converges weakly P-a.s. to an \mathbf{F} -valued r.v.

Proof.

We achieve stochastic quasi-Fejér monotonicity w.r.t. the norm

$$\|\mathbf{x}\|^2 = \sum_{i=1}^m \|\mathbf{x}_i\|^2 / p_i \quad \square$$

Block-coordinate forward-backward splitting

- Let \mathbf{F} be the set of solutions to the problem

find $x_1 \in H_1, \dots, x_m \in H_m$ such that

$$(\forall i \in \{1, \dots, m\}) \quad 0 \in A_i x_i + B_i(x_1, \dots, x_m)$$

where $A_j: H_j \rightarrow 2^{H_j}$ is maximally monotone and, for some $\vartheta > 0$,

$$(\forall \mathbf{x} \in \mathbf{H})(\forall \mathbf{y} \in \mathbf{H}) \quad \sum_{i=1}^m \langle x_i - y_i \mid B_i \mathbf{x} - B_i \mathbf{y} \rangle \geq \vartheta \sum_{i=1}^m \|B_i \mathbf{x} - B_i \mathbf{y}\|^2$$

- Algorithm:

for $n = 0, 1, \dots$

$$\left[\begin{array}{l} \varepsilon \leq \gamma_n \leq (2 - \varepsilon)\vartheta \\ \text{for } i = 1, \dots, m \\ \left[\begin{array}{l} x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \left(J_{\gamma_n A_i} \left(x_{i,n} - \gamma_n (B_i(x_{1,n}, \dots, x_{m,n}) + c_{i,n}) \right) \right. \\ \left. + a_{i,n} - x_{i,n} \right) \end{array} \right. \end{array} \right.$$

Block-coordinate forward-backward splitting

- Under the same conditions as before almost sure weak convergence of $(\mathbf{x}_n)_{n \in \mathbb{N}}$ to a point in \mathbf{F} is achieved.
- Proof: In double-layer theorem, set

- $\mathbf{A}: \mathbf{H} \rightarrow 2^{\mathbf{H}}: \mathbf{x} \mapsto \prod_{i=1}^m \mathbf{A}_i \mathbf{x}_i$
- $\mathbf{B}: \mathbf{H} \rightarrow \mathbf{H}: \mathbf{x} \mapsto (\mathbf{B}_i \mathbf{x})_{1 \leq i \leq m}$
- $\mathbf{T}_n = \mathbf{J}_{\gamma_n \mathbf{A}}$
- $\mathbf{R}_n = \text{Id} - \gamma_n \mathbf{B}$,
- $\mathbf{F} = \text{zer}(\mathbf{A} + \mathbf{B})$
- $\mathbf{b}_n = -\gamma_n \mathbf{c}_n$
- $\alpha_n = 1/2$
- $\beta_n = \gamma_n / (2^\vartheta)$

Block-coordinate forward-backward splitting: convex minimization

- Let \mathbf{F} be the set of solutions to the problem

$$\underset{x_1 \in H_1, \dots, x_m \in H_m}{\text{minimize}} \quad \sum_{i=1}^m f_i(x_i) + \sum_{k=1}^p g_k \left(\sum_{i=1}^m L_{ki} x_i \right)$$

where $f_i \in \Gamma_0(H_i)$, $g_k: G_k \rightarrow \mathbb{R}$ differentiable, convex, ∇g_k τ_k -Lipschitz, $L_{ki}: H_i \rightarrow G_k$ linear & bounded

- Let

$$\vartheta = \left(\sum_{k=1}^p \tau_k \left\| \sum_{i=1}^m L_{ki} L_{ki}^* \right\| \right)^{-1}$$

- Algorithm:

for $n = 0, 1, \dots$

 for $i = 1, \dots, m$

$$\left[\begin{array}{l} r_{i,n} = \varepsilon_{i,n} (x_{i,n} - \gamma_n (\sum_{k=1}^p L_{ki}^* \nabla g_k (\sum_{j=1}^m L_{kj} x_{j,n}) + c_{i,n})) \\ x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \lambda_n (\text{prox}_{\gamma_n f_i} r_{i,n} + a_{i,n} - x_{i,n}). \end{array} \right.$$

Summary

