## Splitting Techniques in the Face of Huge Problem Sizes: Block-Coordinate and Block-Iterative Approaches

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

Edinburgh, May 7, 2015

### Framework

A wide range of problems in applied nonlinear analysis can be reduced to finding a point in a closed convex subset F of a Hilbert space H. The solution set F is often constructed from prior information and the observation of data.

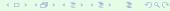
Mathematical model:

find 
$$x \in F \subset \bigcap_{n \in \mathbb{N}} FixT_n$$
,  $T_n : H \to H$  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

 $\blacksquare$  T<sub>n</sub> is quasi-nonexpansive, i.e.,

$$(\forall x \in H)(\forall y \in FixT_n) \quad ||T_nx - y|| \le ||x - y||,$$

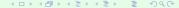
with  $F \subset Fix T_n$ . Then:

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

$$x_n \rightarrow x \in F$$



## Basic convergence principle

 $\blacksquare$  T<sub>n</sub> is quasi-nonexpansive, i.e.,

$$(\forall x \in H)(\forall y \in FixT_n) \quad ||T_nx - y|| \le ||x - y||,$$

with  $F \subset Fix T_n$ . 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

$$x_n \rightarrow x \in F$$

■ Elementary example: 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: Special cases

- Fixed point methods: Krasnosel'skii-Mann, string averaging, extrapolated barycentric methods, Martinet's cyclic firmly nonexpansive iteration, 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: forward-backward, Douglas-Rachford, forward-backward-forward, Spingarn, etc.
- 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 I: 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 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

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

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:

$$X_{i,n+1} = X_{i,n} + \varepsilon_{i,n} (T_{i,n} \mathbf{x}_n + \mathbf{c}_{i,n} - X_{i,n}),$$

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



# Very large scale problems I: 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 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

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

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:

$$X_{i,n+1} = X_{i,n} + \varepsilon_{i,n} (T_{i,n} \mathbf{X}_n + O_{i,n} - X_{i,n}),$$

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  $\mathbf{T} \colon \mathbf{H} \to \mathbf{H}$  are destroyed by coordinate sampling
- For instance, consider the nonexpansive (1-Lipschitz) operator

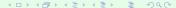
**T**: 
$$(x_1, x_2) \mapsto (-x_2, x_1)$$

Then

$$\begin{cases} Q_1 \colon (x_1, x_2) \mapsto (x_1, x_1) \\ Q_2 \colon (x_1, x_2) \mapsto (-x_2, x_2) \\ Q_1 \circ Q_2 \\ Q_2 \circ Q_1 \end{cases}$$

are no longer nonexpansive

Fejér monotonicity is destroyed



### A roadblock

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

$$\mathbf{T}\colon (\mathsf{X}_1,\mathsf{X}_2)\mapsto (-\mathsf{X}_2,\mathsf{X}_1)$$

Then

$$\begin{cases} Q_1 \colon (x_1, x_2) \mapsto (x_1, x_1) \\ Q_2 \colon (x_1, x_2) \mapsto (-x_2, x_2) \\ Q_1 \circ Q_2 \\ Q_2 \circ Q_1 \end{cases}$$

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  $\mathbf{T} \colon \mathbf{H} \to \mathbf{H}$  are destroyed by coordinate sampling
- For instance, consider the nonexpansive (1-Lipschitz) operator

**T**: 
$$(x_1, x_2) \mapsto (-x_2, x_1)$$

Then

$$\begin{cases} Q_1 \colon (x_1, x_2) \mapsto (x_1, x_1) \\ Q_2 \colon (x_1, x_2) \mapsto (-x_2, x_2) \\ Q_1 \circ Q_2 \\ Q_2 \circ Q_1 \end{cases}$$

are no longer nonexpansive

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



### **Notation**

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

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

- $\ell_+(\mathscr{X})$ : set of sequences of  $[0, +\infty[$ -valued random variables  $(\xi_n)_{n\in\mathbb{N}}$  such that, for every  $n\in\mathbb{N}$ ,  $\xi_n$  is  $\mathcal{X}_n$ -measurable

- **Deterministic definition**: A sequence  $(x_n)_{n \in \mathbb{N}}$  in H is Fejér monotone with respect to F if for every  $z \in F$ ,

$$(\forall n \in \mathbb{N}) \quad \phi(\|\mathbf{x}_{n+1} - \mathbf{z}\|) \leqslant \phi(\|\mathbf{x}_n - \mathbf{z}\|)$$

- 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  $z \in F$ ,

$$(\forall n \in \mathbb{N}) \ \mathsf{E}(\phi(\|x_{n+1} - \mathsf{z}\| | \ \mathfrak{X}_n) \leqslant \phi(\|x_n - \mathsf{z}\|)$$

- 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  $(\chi_n(z))_{n\in\mathbb{N}} \in \ell^1_+(\mathscr{X})$ ,  $(\vartheta_n(z))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$ , and  $(\eta_n(z))_{n\in\mathbb{N}} \in \ell^1_+(\mathscr{X})$  such that

$$(\forall n \in \mathbb{N}) \ \mathsf{E}(\phi(\|\mathsf{X}_{n+1} - \mathsf{z}\|) \,|\, \mathcal{X}_n) + \vartheta_n(\mathsf{z}) \leqslant (1 + \chi_n(\mathsf{z}))\phi(\|\mathsf{X}_n - \mathsf{z}\|) + \eta_n(\mathsf{z})$$

- 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  $(\chi_n(z))_{n\in\mathbb{N}} \in \ell^1_+(\mathscr{X})$ ,  $(\vartheta_n(z))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$ , and  $(\eta_n(z))_{n\in\mathbb{N}} \in \ell^1_+(\mathscr{X})$  such that

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

#### **Theorem**

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

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



### 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:

- $(\forall n \in \mathbb{N}) \ x_{n+1} = x_n + \lambda_n (t_n + e_n x_n), \quad \lambda_n \in ]0, 1]$
- 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

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

Then  $(\forall z \in F)$  [  $\sum_{n \in \mathbb{N}} \lambda_n \theta_n(z) < +\infty$  P-a.s. ] and [ $\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset F$  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
- $\mathbf{T}_n$ :  $\mathbf{H} \to \mathbf{H}$ :  $\mathbf{x} \mapsto (\mathbf{T}_{i,n} \mathbf{x})_{1 \leq i \leq m}$  quasinonexpansive
- $\mathbf{F} = \bigcap_{n \in \mathbb{N}} \operatorname{Fix} \mathbf{T}_n \neq \emptyset$
- **x**<sub>0</sub> and the errors  $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$  are **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:



## Single-layer algorithm

#### **Theorem**

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

- $\inf_{n\in\mathbb{N}} \lambda_n > 0$  and  $\sup_{n\in\mathbb{N}} \lambda_n < 1$ .
- $\blacksquare \mathfrak{W}(x_n)_{n\in\mathbb{N}}\subset \mathbf{F} \text{ P-}a.s.$
- For every  $n \in \mathbb{N}$ ,  $\mathcal{E}_n$  and  $\mathfrak{X}_n$  are independent.
- For every  $i \in \{1, ..., m\}$ ,  $p_i = P[\varepsilon_{i,0} = 1] > 0$ .

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

#### Proof.

We achieve stochastic quasi-Fejér monotonicity w.r.t. the norm  $|||\mathbf{x}|||^2 = \sum_{i=1}^{m} ||\mathbf{x}_i||^2/p_i$ 

### Example: Krasnosel'skii-Mann iteration

- **T**:  $\mathbf{H} \to \mathbf{H}$ :  $\mathbf{x} \mapsto (\mathsf{T}_i \mathbf{x})_{1 \leqslant i \leqslant m}$  nonexpansive operator
- $\blacksquare$  **F** = Fix **T**  $\neq$  Ø
- **x**<sub>0</sub> and the errors  $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$  are **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 
$$n = 0, 1, ...$$
  
for  $i = 1, ..., m$   

$$\begin{bmatrix} X_{i,n+1} = X_{i,n} + \varepsilon_{i,n} \lambda_n (T_i(X_{1,n}, ..., X_{m,n}) + C_{i,n} - X_{i,n}) \end{bmatrix}$$



## Example: Krasnosel'skiĭ-Mann iteration

#### **Theorem**

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

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

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

#### Proof.

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



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

Let F be the set of solutions to the problem

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

$$(\forall i \in \{1,\ldots,m\}) \quad 0 \in A_i X_i + \sum_{k=1}^{\rho} L_{ki}^* B_k \left(\sum_{j=1}^m L_{kj} X_j\right),$$

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

■ Let **F**\* be the set of solutions to the dual problem

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

$$(\forall k \in \{1, ..., p\}) \quad 0 \in -\sum_{i=1}^{m} L_{ki} A_{i}^{-1} \left(-\sum_{l=1}^{p} L_{li}^{*} v_{l}\right) + B_{k}^{-1} v_{k}$$

**◆□▶◆@▶◆≧▶◆≧▶ ● 釣९♡** 

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

Let  $Q_j$  (1  $\leq$   $j \leq$  m+p) be the jth component of the projector  $P_V$  onto the subspace

$$\mathbf{V} = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbf{H} \times \mathbf{G} \,\middle|\, (\forall k \in \{1, \dots, p\}) \, \mathbf{y}_k = \sum_{i=1}^m \mathsf{L}_{ki} \mathsf{x}_i \right\}$$

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

$$\mu_{n} \in ]0, 2[$$
for  $i = 1, ..., m$ 

$$z_{i,n+1} = z_{i,n} + \varepsilon_{i,n} (Q_{i}(x_{1,n}, ..., x_{m,n}, y_{1,n}, ..., y_{p,n}) + C_{i,n} - z_{i,n})$$

$$x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \mu_{n} (J_{\gamma A_{i}}(2z_{i,n+1} - x_{i,n}) + C_{i,n} - z_{i,n+1})$$

ロト (部) (注) (注) 注 りく()

# 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 **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  $\mathbf{H} \times \mathbf{G} = H_1 \times \cdots \times H_m \times G_1 \times \cdots \times G_p$ :

solve 
$$(0,0) \in \mathbf{A}\mathbf{x} \times \mathbf{B}\mathbf{y} + \mathcal{N}_{\mathbf{V}}(\mathbf{x},\mathbf{y})$$
, where

$$\begin{cases} \mathbf{A} \colon \mathbf{H} \to 2^{\mathbf{H}} \colon \mathbf{x} \mapsto X_{i=1}^{m} A_{i} X_{i} \\ \mathbf{B} \colon \mathbf{G} \to 2^{\mathbf{G}} \colon \mathbf{x} \mapsto X_{k=1}^{p} B_{k} y_{k} \\ \mathbf{V} = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbf{H} \times \mathbf{G} \middle| (\forall k \in \{1, \dots, p\}) \ y_{k} = \sum_{i=1}^{m} L_{ki} X_{i} \right\} \end{cases}$$

## Example: Nonsmooth, block-coordinate, primal-dual multivariate minimization

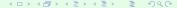
Let F be the set of solutions to the problem

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

where  $f_i \in \Gamma_0(H_i)$ ,  $g_k \in \Gamma_0(G_k)$ ,  $L_{ki}: H_i \to G_k$  linear&bounded

Let F\* be the set of solutions to the dual problem

$$\underset{\mathsf{V}_1 \in \mathsf{G}_1, \dots, \mathsf{V}_p \in \mathsf{G}_p}{\mathsf{minimize}} \ \sum_{i=1}^m \mathsf{f}_i^* \bigg( - \sum_{k=1}^p \mathsf{L}_{ki}^* \mathsf{V}_k \bigg) + \sum_{k=1}^p \mathsf{g}_k^* (\mathsf{V}_k)$$



# Example: Nonsmooth, block-coordinate, primal-dual multivariate minimization

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

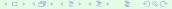
 $\mu_{n} \in ]0,2[$ for i = 1,...,m  $z_{i,n+1} = z_{i,n} + \varepsilon_{i,n} (Q_{i}(x_{1,n},...,x_{m,n},y_{1,n},...,y_{p,n}) + C_{i,n} - z_{i,n})$   $x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \mu_{n} (\text{prox}_{\gamma f_{i}}(2z_{i,n+1} - x_{i,n}) + \alpha_{i,n} - z_{i,n+1})$ 

Under the same conditions as before:

- $(z_n)_{n\in\mathbb{N}}$  converges weakly P-a.s. to an **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

# Double-layer random block-coordinate algorithms

- $\mathbf{T}_n$ :  $\mathbf{H} \to \mathbf{H}$ :  $\mathbf{x} \mapsto (\mathbf{T}_{i,n} \mathbf{x})_{1 \leqslant i \leqslant m}$  is  $\alpha_n$ -averaged ( $\mathrm{Id} + \alpha_n^{-1} (\mathbf{T}_n \mathrm{Id})$  is nonexpansive)
- **R**<sub>n</sub>:  $\mathbf{H} \to \mathbf{H}$  is  $\beta_n$ -averaged
- $\mathbf{F} = \bigcap_{n \in \mathbb{N}} \operatorname{Fix} \mathbf{T}_n \circ \mathbf{R}_n \neq \emptyset$
- **x**<sub>0</sub>,  $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$ , and  $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$  are **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:



# Double-layer random block-coordinate algorithms

#### **Theorem**

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

- $\mathbf{w}(x_n)_{n\in\mathbb{N}}\subset\mathbf{F}$  P-a.s.
- For every  $n \in \mathbb{N}$ ,  $\mathcal{E}_n$  and  $\mathfrak{X}_n$  are independent
- For every  $i \in \{1, ..., m\}$ ,  $p_i = P[\varepsilon_{i,0} = 1] > 0$

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

#### Proof.

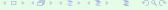
We achieve stochastic quasi-Fejér monotonicity w.r.t. the norm  $|||\mathbf{x}|||^2 = \sum_{i=1}^m ||\mathbf{x}_i||^2/p_i$ 



## Block-coordinate forward-backward splitting

The forward-backward splitting algorithm is important because:

- It models many problems of interest and tolerates errors:
  - PLC and Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Model. Simul., vol. 4, 2005.
- Applied to the dual problem of a strongly monotone/convex composite problem, it provides a primal-dual algorithm:
  - PLC, Dung, and Vū, Dualization of signal recovery problems, Set-Valued Var. Anal., vol. 18, 2010.
- Applied in a renormed product space, it covers/extends various methods (e.g., Chambolle-Pock):
  - Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, Adv. Comput. Math., vol. 38, 2013.
- It can be implemented with variable metrics:
  - PLC and Vū, Variable metric forward-backward splitting with applications to monotone inclusions in duality, Optimization, vol. 63, 2014.
- In minimization problems it provides Fejér-monotonicity, convergent sequences, and monotone minimizing sequences.



## Block-coordinate forward-backward splitting

Let 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  $A_i \colon H_i \to 2^{H_i}$  is maximally monotone and, for some  $\vartheta > 0$ ,

$$(\forall \mathbf{x} \in \mathbf{H})(\forall \mathbf{y} \in \mathbf{H}) \quad \sum_{i=1}^{m} \langle \mathbf{x}_i - \mathbf{y}_i \mid \mathbf{B}_i \mathbf{x} - \mathbf{B}_i \mathbf{y} \rangle \geqslant \vartheta \sum_{i=1}^{m} \|\mathbf{B}_i \mathbf{x} - \mathbf{B}_i \mathbf{y}\|^2$$

Algorithm:

$$\begin{split} &\text{for } n = 0, 1, \dots \\ & \quad \varepsilon \leqslant \gamma_n \leqslant (2 - \varepsilon) \vartheta \\ & \text{for } i = 1, \dots, m \\ & \quad \left\lfloor \begin{array}{l} x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \big( J_{\gamma_n A_i} \big( x_{i,n} - \gamma_n \big( B_i(x_{1,n}, \dots, x_{m,n} \big) + C_{i,n} \big) \big) \\ & \quad + C_{i,n} - x_{i,n} \big) \\ \end{split}$$

## 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 **F** is achieved.
- Proof: In double-layer theorem, set

**A**: 
$$\mathbf{H} \to 2^{\mathbf{H}}$$
:  $\mathbf{x} \mapsto X_{i=1}^m A_i X_i$ 

$$\blacksquare \mathbf{B} \colon \mathbf{H} \to \mathbf{H} \colon \mathbf{x} \mapsto (\mathsf{B}_i \mathbf{x})_{1 \leqslant i \leqslant m}$$

$$\blacksquare$$
  $\mathbf{I}_n = \mathbf{J}_{\gamma_n \mathbf{A}}$ 

$$\blacksquare$$
  $\mathbf{R}_n = \mathrm{Id} - \gamma_n \mathbf{B}$ ,

$$\blacksquare$$
  $\mathbf{F} = \operatorname{zer}(\mathbf{A} + \mathbf{B})$ 

$$lacksquare$$
  $oldsymbol{b}_n = -\gamma_n oldsymbol{c}_n$ 

$$\alpha_n = 1/2$$

$$\beta_n = \gamma_n/(2\vartheta)$$



## Block-coordinate forward-backward splitting: convex minimization

Let F be the set of solutions to the problem

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

where  $f_i \in \Gamma_0(H_i)$ ,  $g_k : G_k \to \mathbb{R}$  differentiable, convex,  $\nabla g_k \to \mathbb{R}$ 

Let

$$\vartheta = \left(\sum_{k=1}^{p} \tau_{k} \left\| \sum_{i=1}^{m} \mathsf{L}_{ki} \mathsf{L}_{ki}^{*} \right\| \right)^{-1}$$

Algorithm:

for 
$$n = 0, 1, ...$$
  
for  $i = 1, ..., m$   

$$\begin{bmatrix} r_{i,n} = \varepsilon_{i,n}(X_{i,n} - \gamma_n(\sum_{k=1}^{p} L_{kl}^* \nabla g_k(\sum_{j=1}^{m} L_{kl} X_{j,n}) + c_{i,n})) \\ X_{i,n+1} = X_{i,n} + \varepsilon_{i,n} \lambda_n(\operatorname{prox}_{\gamma_n f_i} r_{i,n} + \alpha_{i,n} - X_{i,n}). \end{bmatrix}$$



### References

- PLC and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping, SIAM J. Optim., 2015 (arxiv, April 2014)
- H. H. Bauschke and PLC, Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York, 2011. (2nd ed., Fall 2015)