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:

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

- 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

- T_n is quasi-nonexpansive, i.e.,

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

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

  - Fejér monotonicity:

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

  - 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

- T_n is quasi-nonexpansive, i.e.,

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

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

  - Fejér monotonicity:

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

  - 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

$$x_n \rightharpoonup x \in F$$

- 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'skiĭ–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 inequalites 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 $\mathbf{H}$ can be decomposed in m factors

$$\mathbf{H} = \mathrm{H}_1 \times \cdots \times \mathrm{H}_m$$

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

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

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

$$\mathrm{x}_{i,n+1} = \mathrm{x}_{i,n} + \varepsilon_{i,n}\big(\mathrm{T}_{i,n}\mathbf{x}_n + a_{i,n} - \mathrm{x}_{i,n}\big),$$

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 $\mathbf{H}$ can be decomposed in m factors

$$\mathbf{H} = \mathsf{H}_1 \times \cdots \times \mathsf{H}_m$$

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

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

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

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

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

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

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

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
$$\mathbf{T}\colon (x_1, x_2) \mapsto (-x_2, x_1)$$
Then
$$\begin{cases} Q_1 : (x_1, x_2) \mapsto (x_1, x_1) \\ Q_2 : (x_1, x_2) \mapsto (-x_2, x_2) \\ Q_1 \circ Q_2 \\ Q_2 \circ Q_1 \end{cases}$$
are no longer nonexpansive
- Fejér monotonicity is destroyed
- .. and even for jointly convex functions with a unique minimizer, alternating minimizations fail, etc.
- $\rightsquigarrow$ 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 $(x_n)_{n\in\mathbb{N}} \in \mathsf{H}^{\mathbb{N}}$
- $\Gamma_0(\mathsf{H})$: proper lower semicontinuous convex functions from H to $]-\infty,+\infty]$
- $(\Omega,\mathcal{F},\mathsf{P})$: underlying probability space
- Given a sequence $(x_n)_{n\in\mathbb{N}}$ of H-valued random variables,

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

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

Stochastic quasi-Fejér sequences

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

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

Stochastic quasi-Fejér sequences

- $\varnothing \neq \mathsf{F} \subset \mathsf{H}$, $\phi\colon [0,+\infty[\, \to [0,+\infty[$, $\phi(t) \uparrow +\infty$ as $t \to +\infty$
- **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 $\mathsf{z} \in \mathsf{F}$,

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

Stochastic quasi-Fejér sequences

- $\varnothing \neq \mathsf{F} \subset \mathsf{H}$, $\phi\colon [0,+\infty[\, \to [0,+\infty[$, $\phi(t) \uparrow +\infty$ as $t \to +\infty$
- **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 $\mathsf{z} \in \mathsf{F}$, there exist $(\chi_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$, $(\vartheta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$, and $(\eta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$ such that

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

Stochastic quasi-Fejér sequences

- $\varnothing \neq \mathsf{F} \subset \mathsf{H}$, $\phi\colon [0,+\infty[\to [0,+\infty[$, $\phi(t) \uparrow +\infty$ as $t \to +\infty$
- **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 $\mathsf{z} \in \mathsf{F}$, there exist $(\chi_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$, $(\vartheta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$, and $(\eta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$ such that

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

Theorem

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

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

An abstract stochastic iterative scheme

Theorem

Let $\varnothing \neq \mathsf{F} \subset \mathsf{H}$. *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:*

- $(\forall n \in \mathbb{N})\ x_{n+1} = x_n + \lambda_n(t_n + e_n - x_n), \quad \lambda_n \in \left]0,1\right]$
- $\sum_{n\in\mathbb{N}} \lambda_n \sqrt{\mathsf{E}(\|e_n\|^2 \,|\, \mathfrak{X}_n)} < +\infty$ P-*a.s.*
- *For every* $\mathsf{z} \in \mathsf{F}$, *there exist* $(\theta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$, $(\mu_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^\infty(\mathscr{X})$, *and* $(\nu_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^\infty(\mathscr{X})$ *such that* $(\lambda_n\mu_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$, $(\lambda_n\nu_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^{1/2}(\mathscr{X})$, *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 $(\forall \mathsf{z} \in \mathsf{F})$ $\left[\ \sum_{n\in\mathbb{N}} \lambda_n\theta_n(\mathsf{z}) < +\infty\ \mathsf{P}\text{-}a.s.\ \right]$ *and* $[\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{F}\ \ \mathsf{P}\text{-}a.s.] \Rightarrow (x_n)_{n\in\mathbb{N}}$ *converges weakly* P-*a.s. to an* F*-valued random variable.*

Single-layer algorithm

- $\mathbf{H} = \mathsf{H}_1 \times \cdots \times \mathsf{H}_m$, $(\mathsf{H}_i)_{1 \leq i \leq m}$ separable real Hilbert spaces
- $\mathbf{T}_n \colon \mathbf{H} \to \mathbf{H} \colon \mathbf{x} \mapsto (\mathsf{T}_{i,n}\, \mathbf{x})_{1 \leqslant i \leqslant m}$ quasinonexpansive
- $\mathbf{F} = \bigcap_{n\in\mathbb{N}} \operatorname{Fix} \mathbf{T}_n \neq \varnothing$
- $\boldsymbol{x}_0$ and the errors $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$ are $\mathbf{H}$-valued random variables
- $(\varepsilon_n)_{n\in\mathbb{N}}$ identically distributed D-valued random variables with $\mathsf{D} = \{0,1\}^m \smallsetminus \{\mathbf{0}\}$
- Algorithm:

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } i = 1, \ldots, m \\
\left\lfloor\; x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \lambda_n \big(\mathsf{T}_{i,n}\,(x_{1,n}, \ldots, x_{m,n}) + a_{i,n} - x_{i,n} \big) \right.
\end{array}
\right.
\end{array}
$$

Single-layer algorithm

Theorem

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

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

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

Proof.

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

Example: Krasnosel'skiĭ–Mann iteration

- $\mathbf{T}\colon \mathbf{H} \to \mathbf{H}\colon \mathbf{x} \mapsto (\mathrm{T}_i\,\mathbf{x})_{1\leqslant i\leqslant m}$ nonexpansive operator
- $\mathbf{F} = \operatorname{Fix}\mathbf{T} \neq \varnothing$
- $\boldsymbol{x}_0$ 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 $\mathrm{D} = \{0,1\}^m \smallsetminus \{\mathbf{0}\}$
- Algorithm:

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
\text{for } i = 1, \ldots, m \\
\left\lfloor \; x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\lambda_n\big(\mathrm{T}_i\,(x_{1,n}, \ldots, x_{m,n}) + a_{i,n} - x_{i,n}\big) \right.
\end{array} \right.
\end{array}
$$

Example: Krasnosel'skiĭ–Mann iteration

Theorem

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

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

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

Proof.

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

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

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

$$\text{find } \mathsf{x}_1 \in \mathsf{H}_1, \ldots, \mathsf{x}_m \in \mathsf{H}_m \text{ such that}$$

$$(\forall i \in \{1, \ldots, m\}) \quad 0 \in \mathsf{A}_i \mathsf{x}_i + \sum_{k=1}^{p} \mathsf{L}_{ki}^* \mathsf{B}_k \Bigg(\sum_{j=1}^{m} \mathsf{L}_{kj} \mathsf{x}_j \Bigg),$$

where each $\mathsf{A}_i \colon \mathsf{H}_i \to 2^{\mathsf{H}_i}$, $\mathsf{B}_k \colon \mathsf{G}_k \to 2^{\mathsf{G}_k}$ are maximally monotone, $\mathsf{L}_{ki} \colon \mathsf{H}_i \to \mathsf{G}_k$ linear&bounded

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

$$\text{find } \mathsf{v}_1 \in \mathsf{G}_1, \ldots, \mathsf{v}_p \in \mathsf{G}_p \text{ such that}$$

$$(\forall k \in \{1, \ldots, p\}) \quad 0 \in -\sum_{i=1}^{m} \mathsf{L}_{ki} \mathsf{A}_i^{-1} \Bigg(-\sum_{l=1}^{p} \mathsf{L}_{li}^* \mathsf{v}_l \Bigg) + \mathsf{B}_k^{-1} \mathsf{v}_k$$

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

Let Q_j ($1 \leqslant j \leqslant m+p$) be the jth component of the projector $P_{\mathbf{V}}$ onto the subspace

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

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

$$\begin{array}{l}
\mu_n \in \,]0,2[\\
\text{for } i = 1,\ldots,m\\
\left\lfloor \begin{array}{l}
z_{i,n+1} = z_{i,n} + \varepsilon_{i,n}\big(Q_i(x_{1,n},\ldots,x_{m,n},y_{1,n},\ldots,y_{p,n}) + c_{i,n} - z_{i,n}\big)\\
x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\,\mu_n\big(J_{\gamma \mathsf{A}_i}(2z_{i,n+1} - x_{i,n}) + a_{i,n} - z_{i,n+1}\big)
\end{array}\right.\\
\text{for } k = 1,\ldots,p\\
\left\lfloor \begin{array}{l}
w_{k,n+1} = w_{k,n} + \varepsilon_{m+k,n}\big(Q_{m+k}(x_{1,n},\ldots,x_{m,n},y_{1,n},\ldots,y_{p,n}) + d_{k,n} - w_{k,n}\big)\\
y_{k,n+1} = y_{k,n} + \varepsilon_{m+k,n}\,\mu_n\big(J_{\gamma \mathsf{B}_k}(2w_{k,n+1} - y_{k,n}) + b_{k,n} - w_{k,n+1}\big).
\end{array}\right.
\end{array}$$

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

- Under the same conditions as before:
  - $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ converges weakly P-a.s. to an $\mathbf{F}$-valued random variable
  - $(\gamma^{-1}(\boldsymbol{w}_n - \boldsymbol{y}_n))_{n\in\mathbb{N}}$ 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 $\mathbf{H} \times \mathbf{G} = \mathsf{H}_1 \times \cdots \times \mathsf{H}_m \times \mathsf{G}_1 \times \cdots \times \mathsf{G}_p$:

$$\text{solve } (0,0) \in \mathbf{A}\mathbf{x} \times \mathbf{B}\mathbf{y} + N_{\mathbf{V}}(\mathbf{x},\mathbf{y}), \quad \text{where}$$

$$\begin{cases} \mathbf{A}\colon \mathbf{H} \to 2^{\mathbf{H}}\colon \mathbf{x} \mapsto \bigtimes_{i=1}^{m} \mathsf{A}_i \mathsf{x}_i \\ \mathbf{B}\colon \mathbf{G} \to 2^{\mathbf{G}}\colon \mathbf{x} \mapsto \bigtimes_{k=1}^{p} \mathsf{B}_k \mathsf{y}_k \\ \mathbf{V} = \left\{ (\mathbf{x},\mathbf{y}) \in \mathbf{H} \times \mathbf{G} \;\middle|\; (\forall k \in \{1,\ldots,p\})\ \mathsf{y}_k = \sum_{i=1}^{m} \mathsf{L}_{ki} \mathsf{x}_i \right\} \end{cases}$$

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

- Let **F** be the set of solutions to the problem

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

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

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

$$\underset{v_1\in G_1,\ldots,v_p\in G_p}{\text{minimize}} \quad \sum_{i=1}^m f_i^*\Bigg(-\sum_{k=1}^p L_{ki}^* v_k\Bigg) + \sum_{k=1}^p g_k^*(v_k)$$

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

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

$$
\begin{array}{l}
\mu_n \in \,]0,2[\\
\text{for } i = 1,\ldots,m\\
\left\lfloor
\begin{array}{l}
z_{i,n+1} = z_{i,n} + \varepsilon_{i,n}\big(Q_i(x_{1,n},\ldots,x_{m,n},y_{1,n},\ldots,y_{p,n}) + c_{i,n} - z_{i,n}\big)\\
x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\,\mu_n\big(\operatorname{prox}_{\gamma f_i}(2z_{i,n+1} - x_{i,n}) + a_{i,n} - z_{i,n+1}\big)
\end{array}
\right.\\[2mm]
\text{for } k = 1,\ldots,p\\
\left\lfloor
\begin{array}{l}
w_{k,n+1} = w_{k,n} + \varepsilon_{m+k,n}\big(Q_{m+k}(x_{1,n},\ldots,x_{m,n},y_{1,n},\ldots,y_{p,n}) + d_{k,n} - w_{k,n}\big)\\
y_{k,n+1} = y_{k,n} + \varepsilon_{m+k,n}\,\mu_n\big(\operatorname{prox}_{\gamma g_k}(2w_{k,n+1} - y_{k,n}) + b_{k,n} - w_{k,n+1}\big).
\end{array}
\right.
\end{array}
$$

Under the same conditions as before:

- $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ converges weakly P-a.s. to an **F**-valued random variable
- $(\gamma^{-1}(\boldsymbol{w}_n - \boldsymbol{y}_n))_{n\in\mathbb{N}}$ converges weakly P-a.s. to an $\mathbf{F}^*$-valued random variable

Double-layer random block-coordinate algorithms

- $\mathbf{T}_n\colon \mathbf{H} \to \mathbf{H}\colon \mathbf{x} \mapsto (\mathsf{T}_{i,n}\mathbf{x})_{1\leqslant i\leqslant m}$ is α_n-averaged $(\mathrm{Id} + \alpha_n^{-1}(\mathbf{T}_n - \mathrm{Id})$ is nonexpansive)
- $\mathbf{R}_n\colon \mathbf{H} \to \mathbf{H}$ is β_n-averaged
- $\mathbf{F} = \bigcap_{n\in\mathbb{N}} \operatorname{Fix} \mathbf{T}_n \circ \mathbf{R}_n \neq \varnothing$
- $\boldsymbol{x}_0$, $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$, and $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$ are $\mathbf{H}$-valued random variables
- $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ identically distributed D-valued random variables with $\mathrm{D} = \{0,1\}^m \smallsetminus \{\mathbf{0}\}$
- Algorithm:

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{y}_n = \mathbf{R}_n \boldsymbol{x}_n + \boldsymbol{b}_n \\
\text{for } i = 1, \ldots, m \\
\left\lfloor\; x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\big(\mathsf{T}_{i,n}\boldsymbol{y}_n + a_{i,n} - x_{i,n}\big)\right.
\end{array}
\right.
\end{array}
$$

Double-layer random block-coordinate algorithms

Theorem

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

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

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

Proof.

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

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 Vũ, 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 Vũ, 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 $\mathbf{F}$ be the set of solutions to the problem

$$\text{find } \mathsf{x}_1 \in \mathsf{H}_1, \ldots, \mathsf{x}_m \in \mathsf{H}_m \text{ such that } (\forall i \in \{1, \ldots, m\}) \quad 0 \in \mathsf{A}_i \mathsf{x}_i + B_i(\mathsf{x}_1, \ldots, \mathsf{x}_m)$$

where $\mathsf{A}_i \colon \mathsf{H}_i \to 2^{\mathsf{H}_i}$ is maximally monotone and, for some $\vartheta > 0$,

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

- Algorithm:

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
\varepsilon \leqslant \gamma_n \leqslant (2 - \varepsilon)\vartheta \\
\text{for } i = 1, \ldots, m \\
\left\lfloor \begin{array}{l}
x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \Big(\mathsf{J}_{\gamma_n \mathsf{A}_i} \big(x_{i,n} - \gamma_n (\mathsf{B}_i(x_{1,n}, \ldots, x_{m,n}) + c_{i,n}) \big) \\
\qquad\qquad\qquad\qquad + a_{i,n} - x_{i,n} \Big)
\end{array} \right.
\end{array} \right.
\end{array}$$

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
  - $\mathbf{A}\colon \mathbf{H}\to 2^{\mathbf{H}}\colon \mathbf{x}\mapsto \bigtimes_{i=1}^m A_i x_i$
  - $\mathbf{B}\colon \mathbf{H}\to \mathbf{H}\colon \mathbf{x}\mapsto (B_i\mathbf{x})_{1\leqslant i\leqslant m}$
  - $\mathbf{T}_n = \mathbf{J}_{\gamma_n\mathbf{A}}$
  - $\mathbf{R}_n = \mathrm{Id} - \gamma_n\mathbf{B}$,
  - $\mathbf{F} = \mathrm{zer}(\mathbf{A}+\mathbf{B})$
  - $\boldsymbol{b}_n = -\gamma_n\boldsymbol{c}_n$
  - $\alpha_n = 1/2$
  - $\beta_n = \gamma_n/(2\vartheta)$

Block-coordinate forward-backward splitting: convex minimization

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

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

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

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

- Algorithm:

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
\text{for } i = 1, \ldots, m \\
\left\lfloor \begin{array}{l}
r_{i,n} = \varepsilon_{i,n}\big(x_{i,n} - \gamma_n\big(\sum_{k=1}^{p} L_{ki}^{*}\nabla g_k\big(\sum_{j=1}^{m} L_{kj}x_{j,n}\big) + c_{i,n}\big)\big) \\
x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\lambda_n\big(\operatorname{prox}_{\gamma_n f_i} r_{i,n} + a_{i,n} - x_{i,n}\big).
\end{array}\right.
\end{array}\right.
\end{array}$$

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)