Randomly activated proximal methods for nonsmooth convex minimization

Patrick L. Combettes & Javier I. Madariaga

Department of Mathematics
North Carolina State University
Raleigh, NC 27695, USA

TAGMaC Fall 2024
November 02, 2024

NC STATE UNIVERSITY

Background and motivation

- Throughout, $\mathsf{H}, \mathsf{G}_1, \ldots, \mathsf{G}_p$ are separable real Hilbert spaces.

Background and motivation

- Throughout, $\mathsf{H}, \mathsf{G}_1, \ldots, \mathsf{G}_p$ are separable real Hilbert spaces.
- $\Gamma_0(\mathsf{H})$ denotes the class of lower semicontinuous convex functions $\mathsf{f}\colon \mathsf{H} \to \left]-\infty, +\infty\right]$ such that

$$\operatorname{dom} \mathsf{f} = \left\{\mathsf{x} \in \mathsf{H} \mid \mathsf{f}(\mathsf{x}) < +\infty\right\} \neq \varnothing.$$

Background and motivation

- Throughout, $\mathsf{H}, \mathsf{G}_1, \ldots, \mathsf{G}_p$ are separable real Hilbert spaces.
- $\Gamma_0(\mathsf{H})$ denotes the class of lower semicontinuous convex functions $f\colon \mathsf{H} \to \left]-\infty, +\infty\right]$ such that

$$\operatorname{dom} f = \left\{x \in \mathsf{H} \mid f(x) < +\infty\right\} \neq \varnothing.$$

- Let $f \in \Gamma_0(\mathsf{H})$. The subdifferential of f is the operator

$$\partial f\colon \mathsf{H} \to 2^{\mathsf{H}}\colon x \mapsto \left\{x^* \in \mathsf{H} \mid (\forall z \in \mathsf{H})\ \langle z - x \mid x^* \rangle + f(x) \leqslant f(z)\right\}$$

and the proximity operator of f is

$$\operatorname{prox}_f\colon \mathsf{H} \to \mathsf{H}\colon x \mapsto \underset{z \in \mathsf{H}}{\operatorname{argmin}} \left(f(z) + \frac{1}{2}\|x - z\|^2\right).$$

Background and motivation

- Let C be a nonempty closed convex subset of H. Then ι_C denotes the indicator function of C and $\mathrm{proj}_C = \mathrm{prox}_{\iota_C}$ the projection operator onto C.

Background and motivation

- Let C be a nonempty closed convex subset of H. Then ι_{C} denotes the indicator function of C and $\mathrm{proj}_{\mathsf{C}} = \mathrm{prox}_{\iota_{\mathsf{C}}}$ the projection operator onto C.
- The underlying probability space is $(\Omega, \mathcal{F}, \mathsf{P})$. $\mathcal{B}_{\mathsf{H}}$ denotes the Borel σ-algebra of H. An H-valued random variable is a measurable mapping $x\colon (\Omega, \mathcal{F}) \to (\mathsf{H}, \mathcal{B}_{\mathsf{H}})$.

Background and motivation

- Let C be a nonempty closed convex subset of H. Then ι_{C} denotes the indicator function of C and $\mathrm{proj}_{\mathsf{C}} = \mathrm{prox}_{\iota_{\mathsf{C}}}$ the projection operator onto C.
- The underlying probability space is $(\Omega, \mathcal{F}, \mathsf{P})$. $\mathcal{B}_{\mathsf{H}}$ denotes the Borel σ-algebra of H. An H-valued random variable is a measurable mapping $x\colon (\Omega, \mathcal{F}) \to (\mathsf{H}, \mathcal{B}_{\mathsf{H}})$.
- We use sans-serif letters to denote deterministic variables and italicized serif letters to denote random variables.

Background and motivation

Problem 1

H is a separable real Hilbert space and $f \in \Gamma_0(H)$. For every $k \in \{1,\ldots,p\}$, G_k is a separable real Hilbert space, $g_k \in \Gamma_0(G_k)$, and $0 \neq L_k \colon H \to G_k$ is linear and bounded. It is assumed that

$$\operatorname{zer}\left(\partial f + \sum_{k=1}^{p} L_k^* \circ \partial g_k \circ L_k\right) \neq \varnothing.$$

The task is to

$$\underset{x \in H}{\operatorname{minimize}} \;\; f(x) + \sum_{k=1}^{p} g_k(L_k x)$$

and the set of solutions is denoted by Z.

Background and motivation

Problem 1

$\mathcal{H}$ is a separable real Hilbert space and $f \in \Gamma_0(\mathcal{H})$. For every $k \in \{1, \ldots, p\}$, $\mathcal{G}_k$ is a separable real Hilbert space, $g_k \in \Gamma_0(\mathcal{G}_k)$, and $0 \neq L_k \colon \mathcal{H} \to \mathcal{G}_k$ is linear and bounded. It is assumed that

$$\operatorname{zer}\left(\partial f + \sum_{k=1}^{p} L_k^* \circ \partial g_k \circ L_k\right) \neq \varnothing.$$

The task is to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{k=1}^{p} g_k(L_k x)$$

and the set of solutions is denoted by Z.

It covers a wide range of minimization models in data analysis and it is an essential tool in signal processing, statistics, inverse problems, and machine learning.

Background and motivation

We aim at designing **proximal algorithms** which are **stochastic** in the sense that they activate a randomly selected block of functions at each iteration

Background and motivation

We aim at designing **proximal algorithms** which are **stochastic** in the sense that they activate a randomly selected block of functions at each iteration

Furthermore, they should satisfy:

R1: They **guarantee the convergence** of the sequence of iterates to a solution to Problem 1 without any additional assumptions.

Background and motivation

We aim at designing **proximal algorithms** which are **stochastic** in the sense that they activate a randomly selected block of functions at each iteration

Furthermore, they should satisfy:

R1: They **guarantee the convergence** of the sequence of iterates to a solution to Problem 1 without any additional assumptions.

R2: At each iteration, more than one randomly selected function $(f, g_1, \ldots, g_p)$ can be activated.

Background and motivation

We aim at designing **proximal algorithms** which are **stochastic** in the sense that they activate a randomly selected block of functions at each iteration

Furthermore, they should satisfy:

- **R1:** They **guarantee the convergence** of the sequence of iterates to a solution to Problem 1 without any additional assumptions.
- **R2:** At each iteration, more than one randomly selected function $(f, g_1, \ldots, g_p)$ can be activated.
- **R3:** Knowledge of bounds on the norms of the linear operators $(L_k)_{1 \leqslant k \leqslant p}$ is not required.

Background and motivation

There does not seem to exist methods that satisfy simultaneously **R1–R3**.

Background and motivation

There does not seem to exist methods that satisfy simultaneously **R1–R3**.

Our main contribution is to proposed three algorithmic frameworks that comply with **R1–R3**.

Background and motivation

There does not seem to exist methods that satisfy simultaneously **R1–R3**.

Our main contribution is to proposed three algorithmic frameworks that comply with **R1–R3**.

The strategy consists in embedding Problem 1 into a **multivariate problem**.

General Framework

Problem 2

- Let $(X_i)_{1 \leqslant i \leqslant m}$ and $(Y_j)_{1 \leqslant j \leqslant r}$ be families of separable real Hilbert spaces with direct Hilbert sums $\mathbf{X} = X_1 \oplus \cdots \oplus X_m$ and $\mathbf{Y} = Y_1 \oplus \cdots \oplus Y_r$.

Problem 2

- Let $(X_i)_{1\leqslant i\leqslant m}$ and $(Y_j)_{1\leqslant j\leqslant r}$ be families of separable real Hilbert spaces with direct Hilbert sums $\mathbf{X} = X_1 \oplus \cdots \oplus X_m$ and $\mathbf{Y} = Y_1 \oplus \cdots \oplus Y_r$.
- For every $i \in \{1,\ldots,m\}$, let $f_i \in \Gamma_0(X_i)$ and, for every $j \in \{1,\ldots,r\}$, let $h_j \in \Gamma_0(Y_j)$, and let $M_{ji}\colon X_i \to Y_j$ be linear and bounded.

Problem 2

- Let $(\mathsf{X}_i)_{1\leqslant i\leqslant m}$ and $(\mathsf{Y}_j)_{1\leqslant j\leqslant r}$ be families of separable real Hilbert spaces with direct Hilbert sums $\boldsymbol{\mathsf{X}} = \mathsf{X}_1 \oplus \cdots \oplus \mathsf{X}_m$ and $\boldsymbol{\mathsf{Y}} = \mathsf{Y}_1 \oplus \cdots \oplus \mathsf{Y}_r$.
- For every $i \in \{1,\ldots,m\}$, let $f_i \in \Gamma_0(\mathsf{X}_i)$ and, for every $j \in \{1,\ldots,r\}$, let $h_j \in \Gamma_0(\mathsf{Y}_j)$, and let $M_{ji}\colon \mathsf{X}_i \to \mathsf{Y}_j$ be linear and bounded.

The task is to

$$\underset{\boldsymbol{x}\in\boldsymbol{\mathsf{X}}}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x_i) + \sum_{j=1}^{r} h_j\Bigg(\sum_{i=1}^{m} M_{ji}x_i\Bigg).$$

The set of solutions to Problem 2 is denoted by $\mathbf{Z}$.

General Framework

The set of solutions to Problem 2 is denoted by **Z**. Further, the projection operator onto the subspace

$$\mathbf{V} = \left\{ (\mathbf{x}, \mathbf{y}) \in \mathbf{X} \oplus \mathbf{Y} \,\middle|\, (\forall j \in \{1, \ldots, r\})\ y_j = \sum_{i=1}^{m} M_{ji} x_i \right\}$$

is decomposed as $\mathrm{proj}_{\mathbf{V}} \colon (\mathbf{x}, \mathbf{y}) \mapsto (Q_l(\mathbf{x}, \mathbf{y}))_{1 \leqslant l \leqslant m+r}$, where for every $i \in \{1, \ldots, m\}$, $Q_i \colon \mathbf{X} \oplus \mathbf{Y} \to X_i$ and, for every $j \in \{1, \ldots, r\}$, $Q_{m+j} \colon \mathbf{X} \oplus \mathbf{Y} \to Y_j$.

General Framework

Theorem 3...

Douglas–Rachford method yields the following result.

General Framework

Theorem 3...

Douglas–Rachford method yields the following result.

- Let $\gamma \in]0, +\infty[$,

General Framework

Theorem 3...

Douglas–Rachford method yields the following result.

- Let $\gamma \in \left]0, +\infty\right[$,
- let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, 2\right[$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$ and $\sup_{n\in\mathbb{N}} \lambda_n < 2$,

General Framework

Theorem 3...

Douglas–Rachford method yields the following result.

- Let $\gamma \in \,]0, +\infty[$,
- let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $]0, 2[$ such that $\inf_{n \in \mathbb{N}} \lambda_n > 0$ and $\sup_{n \in \mathbb{N}} \lambda_n < 2$,
- let $\boldsymbol{x}_0$ and $\boldsymbol{z}_0$ be X-valued random variables,

General Framework

Theorem 3...

Douglas–Rachford method yields the following result.

- Let $\gamma \in]0, +\infty[$,
- let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 2[$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$ and $\sup_{n\in\mathbb{N}} \lambda_n < 2$,
- let $\boldsymbol{x}_0$ and $\boldsymbol{z}_0$ be X-valued random variables,
- let $\boldsymbol{y}_0$ and $\boldsymbol{w}_0$ be Y-valued random variables,

General Framework

Theorem 3...

Douglas–Rachford method yields the following result.

- Let $\gamma \in]0, +\infty[$,
- let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 2[$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$ and $\sup_{n\in\mathbb{N}} \lambda_n < 2$,
- let $\boldsymbol{x}_0$ and $\boldsymbol{z}_0$ be **X**-valued random variables,
- let $\boldsymbol{y}_0$ and $\boldsymbol{w}_0$ be **Y**-valued random variables,
- set $\mathrm{D} = \{0, 1\}^{m+r} \smallsetminus \{\mathbf{0}\}$, and let $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ be identically distributed D-valued random variables. (**)

General Framework

...Theorem 3

Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } i = 1, \ldots, m \\
\left\lfloor
\begin{array}{l}
x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\big(Q_i(\boldsymbol{z}_n, \boldsymbol{w}_n) - x_{i,n}\big) \\
z_{i,n+1} = z_{i,n} + \varepsilon_{i,n}\lambda_n\big(\mathrm{prox}_{\gamma f_i}(2x_{i,n+1} - z_{i,n}) - x_{i,n+1}\big)
\end{array}
\right. \\
\text{for } j = 1, \ldots, r \\
\left\lfloor
\begin{array}{l}
y_{j,n+1} = y_{j,n} + \varepsilon_{m+j,n}\big(Q_{m+j}(\boldsymbol{z}_n, \boldsymbol{w}_n) - y_{j,n}\big) \\
w_{j,n+1} = w_{j,n} + \varepsilon_{m+j,n}\lambda_n\big(\mathrm{prox}_{\gamma h_j}(2y_{j,n+1} - w_{j,n}) - y_{j,n+1}\big).
\end{array}
\right.
\end{array}
\right.
\end{array}
$$

General Framework

...Theorem 3

Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } i = 1, \ldots, m \\
\left\lfloor
\begin{array}{l}
x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\big(Q_i(\boldsymbol{z}_n, \boldsymbol{w}_n) - x_{i,n}\big) \\
z_{i,n+1} = z_{i,n} + \varepsilon_{i,n}\lambda_n\big(\mathrm{prox}_{\gamma f_i}(2x_{i,n+1} - z_{i,n}) - x_{i,n+1}\big)
\end{array}
\right. \\
\text{for } j = 1, \ldots, r \\
\left\lfloor
\begin{array}{l}
y_{j,n+1} = y_{j,n} + \varepsilon_{m+j,n}\big(Q_{m+j}(\boldsymbol{z}_n, \boldsymbol{w}_n) - y_{j,n}\big) \\
w_{j,n+1} = w_{j,n} + \varepsilon_{m+j,n}\lambda_n\big(\mathrm{prox}_{\gamma h_j}(2y_{j,n+1} - w_{j,n}) - y_{j,n+1}\big).
\end{array}
\right.
\end{array}
\right.
\end{array}
$$

Then $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly P-a.s. to a **Z**-valued random variable.

Framework 1

How can we use this general framework to solve Problem 1?

Framework 1

How can we use this general framework to solve Problem 1?

$$\underset{\mathbf{x}\in\mathbf{X}}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x_i) + \sum_{j=1}^{r} h_j\left(\sum_{i=1}^{m} M_{ji} x_i\right). \tag{GF}$$

$$\underset{x\in H}{\text{minimize}} \quad f(x) + \sum_{k=1}^{p} g_k(L_k x). \tag{P1}$$

Framework 1

How can we use this general framework to solve Problem 1?

$$\underset{\mathbf{x}\in\mathbf{X}}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x_i) + \sum_{j=1}^{r} h_j\left(\sum_{i=1}^{m} M_{ji}x_i\right). \tag{GF}$$

$$\underset{x\in H}{\text{minimize}} \quad f(x) + \sum_{k=1}^{p} g_k(L_k x). \tag{P1}$$

The same for $m = 1$, $r = p$, $X_1 = H$, $f_1 = f$, and $(\forall k \in \{1, \ldots, p\})$ $Y_k = G_k$, $M_{k,1} = L_k$, and $h_k = g_k$.

Framework 1

Proposition 4

Set $D = \{0, 1\}^{1+p} \setminus \{\mathbf{0}\}$ and consider Theorem 3 with

Framework 1

Proposition 4

Set $\mathsf{D} = \{0,1\}^{1+\mathsf{p}} \smallsetminus \{\mathbf{0}\}$ and consider Theorem 3 with

$$
\begin{array}{l}
\text{for } \mathsf{n} = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_{\mathsf{n}} = \left(\mathsf{Id} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_{\mathsf{k}}^{*} \circ \mathsf{L}_{\mathsf{k}}\right)^{-1} \left(z_{\mathsf{n}} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_{\mathsf{k}}^{*} w_{\mathsf{k},\mathsf{n}}\right) \\
\mathsf{Q}_1(\boldsymbol{z}_{\mathsf{n}}, \boldsymbol{w}_{\mathsf{n}}) = q_{\mathsf{n}} \\
\\
\\
\end{array}
\right.
\end{array}
$$

Framework 1

Proposition 4

Set $\mathsf{D} = \{0,1\}^{1+\mathsf{p}} \smallsetminus \{\mathbf{0}\}$ and consider Theorem 3 with

$$
\begin{array}{l}
\text{for } \mathsf{n} = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_{\mathsf{n}} = \left(\mathsf{Id} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_{\mathsf{k}}^{*} \circ \mathsf{L}_{\mathsf{k}}\right)^{-1}\left(z_{\mathsf{n}} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_{\mathsf{k}}^{*} w_{\mathsf{k},\mathsf{n}}\right) \\
\mathsf{Q}_1(\boldsymbol{z}_{\mathsf{n}}, \boldsymbol{w}_{\mathsf{n}}) = q_{\mathsf{n}} \\
\text{for } \mathsf{k} = 1, \ldots, \mathsf{p} \\
\left\lfloor \; \mathsf{Q}_{1+\mathsf{k}}(\boldsymbol{z}_{\mathsf{n}}, \boldsymbol{w}_{\mathsf{n}}) = \mathsf{L}_{\mathsf{k}} q_{\mathsf{n}}. \right.
\end{array}
\right.
\end{array}
$$

Then $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ converges weakly P-a.s. to a Z-valued random variable.

Framework 2

Can we use (GF) to solve the original problem differently?

Framework 2

Can we use (GF) to solve the original problem differently? Set $\mathbf{G} = G_1 \oplus \cdots \oplus G_p$ and

$$\mathbf{W} = \left\{ \mathbf{x} \in H \oplus \mathbf{G} \;\middle|\; (\forall k \in \{1, \ldots, p\})\ x_{k+1} = L_k x_1 \right\}.$$

Framework 2

Can we use (GF) to solve the original problem differently? Set $\mathbf{G} = \mathsf{G}_1 \oplus \cdots \oplus \mathsf{G}_p$ and

$$\mathbf{W} = \left\{ \mathbf{x} \in \mathsf{H} \oplus \mathbf{G} \;\middle|\; (\forall k \in \{1, \ldots, p\}) \; x_{k+1} = L_k x_1 \right\}.$$

Problem 5

Set $f_1 = f$ and, for every $i \in \{2, \ldots, p+1\}$, $f_i = g_{i-1}$. Denote by $\mathbf{x} = (x_1, \ldots, x_{p+1})$ a generic element in $\mathsf{H} \oplus \mathbf{G}$. The task is to

$$\underset{\mathbf{x} \in \mathsf{H} \oplus \mathbf{G}}{\text{minimize}} \; \sum_{i=1}^{p+1} f_i(x_i)$$

Framework 2

Can we use (GF) to solve the original problem differently? Set $\mathbf{G} = G_1 \oplus \cdots \oplus G_p$ and

$$\mathbf{W} = \left\{ \mathbf{x} \in H \oplus \mathbf{G} \;\middle|\; (\forall k \in \{1, \ldots, p\})\ x_{k+1} = L_k x_1 \right\}.$$

Problem 5

Set $f_1 = f$ and, for every $i \in \{2, \ldots, p+1\}$, $f_i = g_{i-1}$. Denote by $\mathbf{x} = (x_1, \ldots, x_{p+1})$ a generic element in $H \oplus \mathbf{G}$. The task is to

$$\underset{\mathbf{x} \in H \oplus \mathbf{G}}{\text{minimize}} \quad \sum_{i=1}^{p+1} f_i(x_i) + \iota_{\mathbf{W}}(\mathbf{x}).$$

Framework 2

Proposition 6

Set $D = \{0,1\}^{p+2} \setminus \{\mathbf{0}\}$ and consider Theorem 3 with

Framework 2

Proposition 6

Set $\mathsf{D} = \{0,1\}^{\mathsf{p}+2} \smallsetminus \{\mathbf{0}\}$ and consider Theorem 3 with

for $\mathsf{n} = 0, 1, \ldots$
  for $\mathsf{i} = 1, \ldots, \mathsf{p} + 1$
    $\mathsf{Q}_{\mathsf{i}}(\boldsymbol{z}_{\mathsf{n}}, \boldsymbol{w}_{\mathsf{n}}) = \frac{1}{2}(z_{\mathsf{i},\mathsf{n}} + w_{\mathsf{i},\mathsf{n}})$

Framework 2

Proposition 6

Set $D = \{0,1\}^{p+2} \smallsetminus \{\mathbf{0}\}$ and consider Theorem 3 with

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } i = 1, \ldots, p+1 \\
\left\lfloor\; Q_i(\boldsymbol{z}_n, \boldsymbol{w}_n) = \frac{1}{2}(z_{i,n} + w_{i,n}) \right. \\
Q_{p+2}(\boldsymbol{z}_n, \boldsymbol{w}_n) = \frac{1}{2}(\boldsymbol{z}_n + \boldsymbol{w}_n) \\
q_n = \left(\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k\right)^{-1}\left(2y_{1,n+1} - w_{1,n} + \sum_{k=1}^{p} L_k^*(2y_{k+1,n+1} - w_{k+1,n})\right) \\
\mathrm{prox}_{\iota_W}(2y_{k,n+1} - w_{k,n}) = (q_n, L_1 q_n, \ldots, L_p q_n)
\end{array}
\right.
\end{array}
$$

Framework 2

Proposition 6

Set $\mathsf{D} = \{0,1\}^{\mathsf{p}+2} \smallsetminus \{\mathbf{0}\}$ and consider Theorem 3 with

$$
\begin{array}{l}
\text{for } \mathsf{n} = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } \mathsf{i} = 1, \ldots, \mathsf{p}+1 \\
\left\lfloor \; \mathsf{Q}_\mathsf{i}(\boldsymbol{z}_\mathsf{n}, \boldsymbol{w}_\mathsf{n}) = \frac{1}{2}(z_{\mathsf{i},\mathsf{n}} + w_{\mathsf{i},\mathsf{n}}) \right. \\
\mathsf{Q}_{\mathsf{p}+2}(\boldsymbol{z}_\mathsf{n}, \boldsymbol{w}_\mathsf{n}) = \frac{1}{2}(\boldsymbol{z}_\mathsf{n} + \boldsymbol{w}_\mathsf{n}) \\
q_\mathsf{n} = (\mathrm{Id} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_\mathsf{k}^* \circ \mathsf{L}_\mathsf{k})^{-1}\big(2y_{1,\mathsf{n}+1} - w_{1,\mathsf{n}} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_\mathsf{k}^*(2y_{\mathsf{k}+1,\mathsf{n}+1} - w_{\mathsf{k}+1,\mathsf{n}})\big) \\
\mathrm{prox}_{\iota_\mathsf{W}}(2y_{\mathsf{k},\mathsf{n}+1} - w_{\mathsf{k},\mathsf{n}}) = (q_\mathsf{n}, \mathsf{L}_1 q_\mathsf{n}, \ldots, \mathsf{L}_\mathsf{p} q_\mathsf{n})
\end{array}
\right.
\end{array}
$$

Then $(x_{1,\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ converges weakly P-a.s. to a Z-valued random variable.

Framework 3

We extend the same idea...

Problem 7

Consider the problem

$$\underset{\mathbf{x}\in H\oplus \mathbf{G}}{\text{minimize}} \sum_{i=1}^{p+1} f_i(x_i)$$

Framework 3

We extend the same idea...

Problem 7

Consider the problem

$$\underset{\mathbf{x}\in\mathcal{H}\oplus\boldsymbol{\mathcal{G}}}{\text{minimize}} \sum_{i=1}^{p+1} f_i(x_i) + \sum_{k=1}^{p} \iota_{\{0\}}\left(\sum_{i=1}^{p+1} C_{ki}x_i\right),$$

where

$$C_{ki} = \begin{cases} L_k, & \text{if } i = 1; \\ -\mathrm{Id}, & \text{if } i = k+1; \\ 0, & \text{otherwise.} \end{cases}$$

Framework 3

Proposition 8

Set $\mathsf{D} = \{0,1\}^{2\mathsf{p}+1} \smallsetminus \{\mathbf{0}\}$ and consider Theorem 3 with

Framework 3

Proposition 8

Set $\mathsf{D} = \{0,1\}^{2\mathsf{p}+1} \smallsetminus \{\mathbf{0}\}$ and consider Theorem 3 with

$$
\begin{array}{l}
\text{for } \mathsf{n} = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_{\mathsf{n}} = \left(2\mathsf{Id} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_{\mathsf{k}}^{*} \circ \mathsf{L}_{\mathsf{k}}\right)^{-1}\Big(2\mathsf{z}_{1,\mathsf{n}} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_{\mathsf{k}}^{*}(\mathsf{z}_{\mathsf{k}+1,\mathsf{n}} + \mathsf{w}_{\mathsf{k},\mathsf{n}})\Big) \\
\mathsf{Q}_{1}(\boldsymbol{z}_{\mathsf{n}}, \boldsymbol{w}_{\mathsf{n}}) = q_{\mathsf{n}} \\
\text{for } \mathsf{i} = 1, \ldots, \mathsf{p} \\
\left\lfloor \mathsf{Q}_{1+\mathsf{i}}(\boldsymbol{z}_{\mathsf{n}}, \boldsymbol{w}_{\mathsf{n}}) = \frac{1}{2}(\mathsf{L}_{\mathsf{i}} q_{\mathsf{n}} + z_{\mathsf{i}+1,\mathsf{n}} - w_{\mathsf{i},\mathsf{n}}) \right. \\
\\
\\
\text{for } \mathsf{k} = 1, \ldots, \mathsf{p} \\
\left\lfloor
\begin{array}{l}
\mathsf{Q}_{1+\mathsf{p}+\mathsf{k}}(\boldsymbol{z}_{\mathsf{n}}, \boldsymbol{w}_{\mathsf{n}}) = \frac{1}{2}(\mathsf{L}_{\mathsf{k}} q_{\mathsf{n}} - z_{\mathsf{k}+1,\mathsf{n}} - w_{\mathsf{k},\mathsf{n}}) \\
\mathsf{prox}_{\iota_{\{0\}}}(2y_{\mathsf{k},\mathsf{n}+1} - w_{\mathsf{k},\mathsf{n}}) = 0.
\end{array}
\right.
\end{array}
\right.
\end{array}
$$

Framework 3

Proposition 8

Set $D = \{0,1\}^{2p+1} \smallsetminus \{\mathbf{0}\}$ and consider Theorem 3 with

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_n = \left(2\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k\right)^{-1}\left(2z_{1,n} + \sum_{k=1}^{p} L_k^*(z_{k+1,n} + w_{k,n})\right) \\
Q_1(\boldsymbol{z}_n, \boldsymbol{w}_n) = q_n \\
\quad \text{for } i = 1, \ldots, p \\
\quad \left\lfloor
\begin{array}{l}
Q_{1+i}(\boldsymbol{z}_n, \boldsymbol{w}_n) = \frac{1}{2}(L_i q_n + z_{i+1,n} - w_{i,n})
\end{array}
\right. \\
\\
\\
\quad \text{for } k = 1, \ldots, p \\
\quad \left\lfloor
\begin{array}{l}
Q_{1+p+k}(\boldsymbol{z}_n, \boldsymbol{w}_n) = \frac{1}{2}(L_k q_n - z_{k+1,n} - w_{k,n}) \\
\mathrm{prox}_{\iota_{\{0\}}}(2y_{k,n+1} - w_{k,n}) = 0.
\end{array}
\right.
\end{array}
\right.
\end{array}
$$

Then $(x_{1,n})_{n\in\mathbb{N}}$ converges weakly P-a.s. to a Z-valued random variable.

Numerical experiment

Consider the **overlapping group lasso regression** described as follows:

Consider the **overlapping group lasso regression** described as follows: $H = \mathbb{R}^N$ and, for every $k \in \{1, \dots, q\}$, $\varnothing \neq I_k \subset \{1, \dots, N\}$ and

$$L_k \colon \mathbb{R}^N \to \mathbb{R}^{\operatorname{card} I_k} \colon x = (\xi_j)_{1 \leqslant j \leqslant N} \mapsto (\xi_j)_{j \in I_k}.$$

Further, $\bigcup_{k=1}^q I_k = \{1, \dots, N\}$. The goal is to

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \quad \frac{\alpha}{2}\|Ax - b\|^2 + \frac{1}{q}\sum_{k=1}^q \|L_k x\|,$$

where $A \in \mathbb{R}^{M \times N}$, $b \in \mathbb{R}^M$, and $\alpha \in \left]0, +\infty\right[$.

Numerical experiment

- In the experiment $M = 1000$, $N = 3610$, $q = 40$, and $\lambda = 5/q^2$.

Numerical experiment

- In the experiment $\mathsf{M} = 1000$, $\mathsf{N} = 3610$, $\mathsf{q} = 40$, and $\lambda = 5/\mathsf{q}^2$.
- The entries of A are i.i.d. samples from a $\mathcal{N}(0,1)$ distribution, and the entries of b are i.i.d. samples from a $\mathcal{N}(100,100)$ distribution.

Numerical experiment

- In the experiment $\mathsf{M} = 1000$, $\mathsf{N} = 3610$, $\mathsf{q} = 40$, and $\lambda = 5/\mathsf{q}^2$.
- The entries of A are i.i.d. samples from a $\mathcal{N}(0,1)$ distribution, and the entries of b are i.i.d. samples from a $\mathcal{N}(100,100)$ distribution.
- We split the function $\|Ax - b\|^2$ into 30 blocks of 40 entries each, where the entries are selected in order without overlap.

Numerical experiment

- In the experiment $M = 1000$, $N = 3610$, $q = 40$, and $\lambda = 5/q^2$.
- The entries of A are i.i.d. samples from a $\mathcal{N}(0,1)$ distribution, and the entries of b are i.i.d. samples from a $\mathcal{N}(100,100)$ distribution.
- We split the function $\|Ax - b\|^2$ into 30 blocks of 40 entries each, where the entries are selected in order without overlap.
- Finally,

$$(\forall k \in \{1,\ldots,p\}) \quad I_k = \{90k - 89, \ldots, 90k + 10\}.$$

Numerical experiment

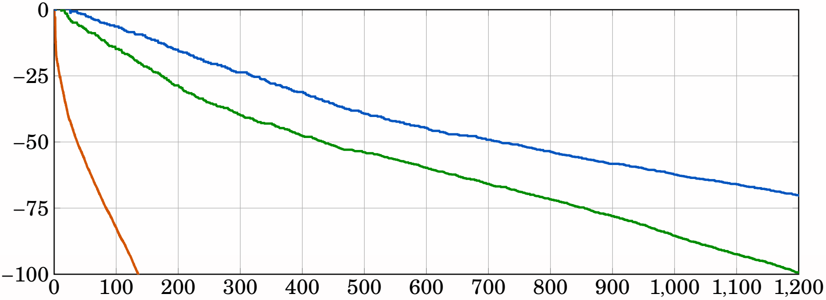

Figure 1: Normalized error $20\log(\|x_{1,n} - x_\infty\|/\|x_{1,0} - x_\infty\|)$ (dB) versus execution time (s). **Green**: Framework 1. **Orange**: Framework 2. **Blue**: Framework 3.

Discussion

The frameworks differ in terms of storage requirements, use of proximity operators, and use of linear operators.

Discussion

The frameworks differ in terms of storage requirements, use of proximity operators, and use of linear operators.

- Framework 1: It stores $2p + 3$ vectors and, for each of the $p + 1$ random activation indices, there is one proximity evaluation.

Discussion

The frameworks differ in terms of storage requirements, use of proximity operators, and use of linear operators.

- Framework 1: It stores $2p + 3$ vectors and, for each of the $p + 1$ random activation indices, there is one proximity evaluation.
- Framework 2: It stores $4p + 5$ vectors. The first $p + 1$ indices involve a proximity operator. The linear operators are used only if index $p + 2$ is activated.

Discussion

The frameworks differ in terms of storage requirements, use of proximity operators, and use of linear operators.

- Framework 1: It stores $2p + 3$ vectors and, for each of the $p + 1$ random activation indices, there is one proximity evaluation.
- Framework 2: It stores $4p + 5$ vectors. The first $p + 1$ indices involve a proximity operator. The linear operators are used only if index $p + 2$ is activated.
- Framework 3: It stores $4p + 3$ vectors. Moreover, out of the $2p+1$ random activation indices, those in $\{1, \ldots, p+1\}$ involve a proximity operator, while those in $\{p+2, \ldots, p+r+1\}$ do not require it.

Discussion

Although Framework 1 is the most efficient in terms of storage, it may not always be the fastest, especially when proximity operators are computationally expensive or when the linear operators are costly, which is the case in the experiment.

Randomly activated proximal methods for nonsmooth convex minimization

Patrick L. Combettes & Javier I. Madariaga

Department of Mathematics
North Carolina State University
Raleigh, NC 27695, USA

TAGMaC Fall 2024
November 02, 2024

NC STATE UNIVERSITY