Coordinate-Update Algorithms can Efficiently Detect Infeasible Optimization Problems

Jinhee Paeng

Sep 21, 2023

Seoul National University

Table of Contents

Table of Contents

Fixed Point Iteration

Definition (Fixed Point Iteration)

For a given operator $\mathbb{T}$, we call an iterative method

$$x^{k+1} = \mathbb{T}x^k, \quad k = 0, 1, 2, \cdots$$

as *Fixed Point Iteration*, or *FPI*.

Theorem (Banach-Fixed Point Theorem)

When (X, d) is a non-empty complete metric space and an operator $\mathbb{T} : X \to X$ is L-Lipschitz with $L < 1$, then

- *there exists an unique fixed point of $\mathbb{T}$, $x^* = \mathbb{T}x^*$,*
- *FPI with $\mathbb{T}$ converges to x^*, $\lim_{k\to\infty} x^k = x^*$.*

Fixed Point Iteration

Remark. However, in most of *FPI*-based convex optimization solvers, the operator is only guaranteed to be a 1-Lipschitz.

Definition (θ-averaged operator)

An operator $\mathbb{T}$ is θ-averaged if it can be described as

$$\mathbb{T} = (1-\theta)\mathbb{I} + \theta\mathbb{C}, \quad \theta \in (0,1], \quad \mathbb{C} \text{ is 1-Lipschitz.}$$

The set of fixed points $\operatorname{Fix}\mathbb{T}$ coincides with $\operatorname{Fix}\mathbb{C}$.

Theorem (Averaged Fixed Point Theorem)

When $\mathcal{H}$ is a non-empty Real Hilbert space and an operator $\mathbb{T}:\mathcal{H}\to\mathcal{H}$ is θ-averaged with $\theta \in (0,1)$ and has a nonempty fixed point, $\operatorname{Fix}\mathbb{T} \neq \emptyset$, then

$$\lim_{k\to\infty} x^k = x^*, \quad x^* \in \operatorname{Fix}\mathbb{T}.$$

Fixed Point Iteration - Inconsistent case

Remark. $\operatorname{Fix}\mathbb{T} = \emptyset$ is possible when $\mathbb{T}$ is 1-Lipschitz or θ-averaged mapping. For example, consider a translation mapping.

Theorem (Pazy, 1971)

When $\mathcal{H}$ is a non-empty Real Hilbert space and an operator $\mathbb{T} : \mathcal{H} \to \mathcal{H}$ is 1-Lipschitz mapping, then

$$\lim_{k\to\infty} \frac{x^k}{k} = -\mathbf{v},$$

where $\mathbf{v}$ is a minimal norm vector of closed convex set $\overline{\operatorname{range}(\mathbb{I} - \mathbb{T})}$.

- We call $\mathbf{v}$ as an *infimal displacement vector* of $\mathbb{T}$.
- We call x^k/k as a *normalized iterate* of *FPI* by $\mathbb{T}$.

Remark. If $\operatorname{Fix}\mathbb{T} \neq \emptyset$, then $\mathbf{v} = 0$.

Table of Contents

RC-FPI

Definition (Classical RC-FPI)

Partition $\mathbb{R}^n$ into a m block coordinates and write a vector in $\mathbb{R}^n$ as

$$x = (x_1, x_2, \cdots, x_m).$$

Consider an operator $\mathbb{T}$ on $\mathbb{R}^n$ and define a *randomized operator* $\mathbb{T}_i$ as:

$$\mathbb{T}x = ((\mathbb{T}x)_1, (\mathbb{T}x)_2, \cdots, (\mathbb{T}x)_m), \quad \mathbb{T}_i x = (x_1, x_2, \cdots, (\mathbb{T}x)_i, \cdots, x_m).$$

A randomized update by $\mathbb{T}_{i^k}$ with IID i^k's is called *RC-FPI*, short for Randomized Coordinate Fixed Point Iteration:

$$x^{k+1} = \mathbb{T}_{i^k} x^k, \quad i^0, i^1, \cdots \overset{iid}{\sim} \text{Uniform}(m).$$

Convergence of RC-FPI

Theorem (Convergence of RC-FPI)

When an operator $\mathbb{T} : \mathbb{R}^n \to \mathbb{R}^n$ *is* θ*-averaged with* $\theta \in (0, 1)$ *and has a nonempty fixed point,* $\operatorname{Fix} \mathbb{T} \neq \emptyset$*, then with probability* 1*,*

$$\lim_{k \to \infty} x^k = x^*, \quad x^* \in \operatorname{Fix} \mathbb{T}.$$

Remark. Many optimization solvers use (RC-FPI) due to its faster convergence speed. Faster speed is not guaranteed but it is not slower if the operator is *coordinate friendly*.

Main Goal of the Paper

Question. What is the asymptotic behavior of RC-FPI when $\operatorname{Fix}\mathbb{T} = \emptyset$?

Answers. Here's the results this work has found for the first time.

- Convergence result analogous of Pazy's work, both in L^2 and *a.s.*:

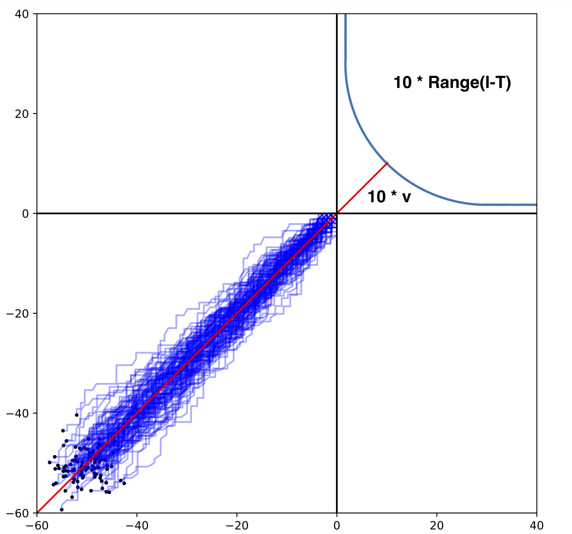
$$\lim_{k\to\infty} \frac{x^k}{k} = -\alpha \mathbf{v}.$$

- Upper bound of the variance analogous to the CLT:

$$\limsup_{k\to\infty} k \operatorname{Var}\left(\frac{x^k}{k}\right) \leq \left(\alpha - \alpha^2\right) \left\|\mathbf{v}\right\|^2.$$

Here, α is a probability of an update for each coordinate.

Main Goal of the Paper



Example of (RC-FPI), 100 iterations, ran 100 times.

RC-FPI setting: Underlying space

Let's extend RC-FPI to more general setting.

Underlying space:

- The underlying space is a real Hilbert space $\mathcal{H}$, consisted of m real Hilbert spaces.

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \dots \mathcal{H}_m.$$

- An element $u \in \mathcal{H}$ can be decomposed into m blocks as

$$u = (u_1, u_2, \dots, u_m), \quad u_i \in \mathcal{H}_i,$$

and u_i is called the ith block coordinates of u.

RC-FPI setting: Underlying space

- The Hilbert space $\mathcal{H}$ has its induced norm and inner product as

$$\|x\|^2 = \sum_{i=1}^{m} \|x_i\|_i^2, \qquad \langle x, y \rangle = \sum_{i=1}^{m} \langle x_i, y_i \rangle_i,$$

  for all $x, y \in \mathcal{H}$, where $\| \cdot \|_i$ and $\langle \cdot, \cdot \rangle_i$ are from $\mathcal{H}_i$.
- Consider a linear, bounded, self-adjoint and positive definite operator $M : \mathcal{H} \to \mathcal{H}$. The M-norm and M-inner product of $\mathcal{H}$ are defined as

$$\|x\|_M = \sqrt{\langle x, Mx \rangle}, \qquad \langle x, y \rangle_M = \langle x, My \rangle,$$

- The M-variance of a random variable X with the domain $\mathcal{H}$ as

$$\mathrm{Var}_M[X] = \mathbb{E}[\|X\|_M^2] - \|\mathbb{E}[X]\|_M^2.$$

RC-FPI setting: Randomized operator

Remark. When an operator $\mathbb{C} : \mathcal{H} \to \mathcal{H}$ is 1-Lipschitz of a real Hilbert space $\mathcal{H}$, then $\mathbb{S} = \mathbb{I} - \mathbb{C}$ is (1/2)-cocoersive operator:

$$\langle x - y, \mathbb{S}x - \mathbb{S}y \rangle \geq \frac{1}{2} \left\| \mathbb{S}x - \mathbb{S}y \right\|^2, \quad \forall x, y \in \mathcal{H}.$$

Consider a θ-averaged $\mathbb{T} = (1-\theta)\mathbb{I} + \theta\mathbb{C}$, then we can rewrite $\mathbb{T}$ as

$$\mathbb{T} = \mathbb{I} - \theta\mathbb{S}, \quad \mathbb{S} \text{ is (1/2)-cocoersive.}$$

For (1/2)-cocoercive operator $\mathbb{S} = \theta^{-1}(\mathbb{I} - \mathbb{T})$, define $\mathbb{S}_i : \mathcal{H} \to \mathcal{H}$ as:

$$\mathbb{S} = \sum_{i=1}^{m} \mathbb{S}_i, \quad \mathbb{S}_i : \mathcal{H} \to 0 \times 0 \times \cdots \times \mathcal{H}_i \times \cdots \times 0.$$

RC-FPI setting: Randomized operator

Randomized operator: Consider a θ-averaged operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$.

- For a *selection vector* $\mathcal{I} \in [0,1]^m$, define $\mathbb{S}_{\mathcal{I}}, \mathbb{T}_{\mathcal{I}}\colon \mathcal{H} \to \mathcal{H}$ as:

$$\mathbb{S}_{\mathcal{I}} = \sum_{i=1}^{m} \mathcal{I}_i \mathbb{S}_i, \qquad \mathbb{T}_{\mathcal{I}} = \mathbb{I} - \theta \mathbb{S}_{\mathcal{I}}.$$

Randomized vector: Similarly, for $u \in \mathcal{H}$, define $u_{\mathcal{I}}$ as:

$$u_{\mathcal{I}} = \sum_{i=1}^{m} \mathcal{I}_i u_i, \quad u = \sum_{i=1}^{m} u_i, \quad u_i \in 0 \times 0 \times \cdots \times \mathcal{H}_i \times \cdots \times 0.$$

RC-FPI setting

Coefficients:

- α: expected amount of update in each coordinate

Definition (Uniform expected step-size condition)

Uniform expected step-size condition is satisfied when $\mathcal{I}$ is randomly sampled from a distribution on $[0,1]^m$ that satisfies:

$$\mathbb{E}_{\mathcal{I}}\left[\mathcal{I}\right] = \alpha \mathbf{1}.$$

- β: depends on the distribution of $\mathcal{I}$ and the choice of the norm.

$$\mathbb{E}_{\mathcal{I}}\left[\|u_{\mathcal{I}}\|_M^2\right] \le \beta \|u\|_M^2, \qquad \forall\, u \in \mathcal{H}.$$

RC-FPI setting

Lemma

Let's consider the norm of $\mathcal{H}$ as $\|\cdot\|$-norm, i.e. M is the identity map. If $\mathcal{I}$ satisfies the uniform expected step-size condition, then we can say:

$$\beta \leq \alpha.$$

Thus, we can choose α as a value of β.

Remark. β always satisfies $\beta \geq \alpha^2$ regardless of the choice of M-norm. However, smaller β is preferred.

Remark. We develop the theory with the general M-norm so that the theory can be extended to non-orthogonal basis.

RC-FPI setting

Definition (RC-FPI)

The randomized coordinate fixed-point iteration (RC-FPI) is:

$$x^{k+1} = \mathbb{T}_{\mathcal{I}^k} x^k, \qquad k = 0, 1, 2, \dots, \tag{RC-FPI}$$

where $\mathcal{I}^0, \mathcal{I}^1, \dots$ is sampled IID and $x^0 \in \mathcal{H}$ is a starting point.

Remark. With uniform expected step-size condition, define $\bar{\mathbb{T}} \colon \mathcal{H} \to \mathcal{H}$:

$$\bar{\mathbb{T}} x = \mathbb{E}_{\mathcal{I}} \left[\mathbb{T}_{\mathcal{I}} x \right], \qquad \forall x \in \mathcal{H}.$$

Equivalently, $\bar{\mathbb{T}} = \mathbb{I} - \alpha\theta\mathbb{S}$. We can define a FPI by $\bar{\mathbb{T}}$:

$$z^{k+1} = \bar{\mathbb{T}} z^k, \qquad k = 0, 1, 2, \dots. \tag{FPI by $\bar{\mathbb{T}}$}$$

Table of Contents

Convergence of Normalized iterates

Goal. The normalized iterate of (RC-FPI) converges to $-\alpha\mathbf{v}$:

$$\frac{x^k}{k} \xrightarrow{L^2} -\alpha\mathbf{v}, \quad \frac{x^k}{k} \xrightarrow{\text{a.s.}} -\alpha\mathbf{v}.$$

both in L^2 and almost surely.

Inequality on expectation of randomized operator

Here's an inequality we will use repeatedly throughout the proofs.

Lemma

Consider the situation:

- $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ *be* θ*-averaged respect to* $\|\cdot\|_M$ *with* $\theta \in (0, 1]$.
- $\mathcal{I}$ *satisfies the uniform expected step-size condition.*

Then for any $x, z \in \mathcal{H}$,

$$\underset{\mathcal{I}}{\mathbb{E}}\left[\left\|\mathbb{T}_{\mathcal{I}} x - \bar{\mathbb{T}} z\right\|_M^2\right] \leq \|x - z\|_M^2 + \theta^2\left(\beta - \alpha^2\right)\|\mathbb{S}x\|_M^2 - \alpha\theta\left(1 - \alpha\theta\right)\|\mathbb{S}x - \mathbb{S}z\|_M^2 .$$

Notation. We will refer this inequality as *One-step inequality*.

Inequality on expectation of randomized operator

Proof.

First, substitute $\mathbb{T}_{\mathcal{I}} = \mathbb{I} - \theta \mathbb{S}_{\mathcal{I}}$ and $\bar{\mathbb{T}} = \mathbb{I} - \alpha\theta\mathbb{S}$ at $\mathbb{E}_{\mathcal{I}}\left[\left\|\mathbb{T}_{\mathcal{I}} x - \bar{\mathbb{T}} z\right\|_M^2\right]$. Then, use $(1/2)$-cocoercive property of the operator $\mathbb{S}$.

$$
\begin{aligned}
&\mathbb{E}\left[\left\|\mathbb{T}_{\mathcal{I}} x - \bar{\mathbb{T}} z\right\|_M^2\right] \\
&\quad = \|x - z\|_M^2 + \theta^2 \mathbb{E}\left[\|\mathbb{S}_{\mathcal{I}} x - \alpha \mathbb{S} z\|_M^2\right] - 2\alpha\theta \langle x - z, \mathbb{S}x - \mathbb{S}z \rangle_M \\
&\quad \leq \|x - z\|_M^2 + \theta^2 \mathbb{E}\left[\|\mathbb{S}_{\mathcal{I}} x - \alpha \mathbb{S} z\|_M^2\right] - \alpha\theta \|\mathbb{S}x - \mathbb{S}z\|_M^2 .
\end{aligned}
$$

Finally, apply the following inequality to conclude the proof.

$$
\begin{aligned}
\mathbb{E}\left[\|\mathbb{S}_{\mathcal{I}} x - \alpha \mathbb{S} z\|_M^2\right] &= \mathbb{E}\left[\|(\mathbb{S}_{\mathcal{I}} x - \alpha \mathbb{S} x) + \alpha (\mathbb{S}x - \mathbb{S}z)\|_M^2\right] \\
&\leq \left(\beta - \alpha^2\right) \|\mathbb{S}x\|_M^2 + \alpha^2 \|\mathbb{S}x - \mathbb{S}z\|_M^2 .
\end{aligned}
$$

□

Table of Contents

L^2 convergence

Theorem (L^2 convergence of normalized iterate)

Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be θ-averaged with respect to $\|\cdot\|$-norm with $\theta \in (0,1]$. Assume $\mathcal{I}^0, \mathcal{I}^1, \ldots$ is sampled IID from a distribution satisfying the uniform expected step-size condition.
Let $x^0, x^1, x^2, \ldots$ be the iterates of (RC-FPI)*. Then*

$$\frac{x^k}{k} \xrightarrow{L^2} -\alpha \mathbf{v}$$

as $k \to \infty$, where $\mathbf{v}$ is the infimal displacement vector of $\mathbb{T}$.

Remark. We will prove for the M-norm with the assumption:

$$\alpha \geq \theta\beta.$$

Proof of L^2 convergence

As mentioned, we will alternatively prove the following result.

Lemma

Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be θ-averaged with respect to $\|\cdot\|_M$ with $\theta \in (0,1]$. Assume $\mathcal{I}^0, \mathcal{I}^1, \ldots$ is sampled IID from a distribution satisfying the uniform expected step-size condition and β satisfies that $\beta \leq \alpha/\theta$. Let $x^0, x^1, x^2, \ldots$ be the iterates of (RC-FPI) *and let $z^0, z^1, z^2, \ldots$ be the iterates of* (FPI by $\bar{\mathbb{T}}$). *Then,*

$$\mathbb{E}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right] \leq \frac{1}{k^2}\left\|x^0 - z^0\right\|_M^2 + \frac{1}{k}A,$$

where $\mathbf{v}$ is the infimal displacement vector of $\mathbb{T}$ and A is a constant

$$A = (1-\alpha\theta)\left[2\sqrt{\alpha\theta}\left\|\mathbb{S}x^0\right\|_M\left\|\mathbb{S}z^0\right\|_M - \frac{\alpha}{\theta}\left\|\mathbf{v}\right\|_M^2\right].$$

Proof of L^2 convergence

Abstract Proof.

First take a full expectation on *One-step inequality* with z^k, x^k.
With the next two lemmas, we can prove that A is an uniform upper bound of full expectation on the last two terms of *One-step inequality*:

$$\theta\left(\theta\beta-\alpha\right)\left\|\mathbb{S}x^{k-1}\right\|_M^2+\alpha\theta\left(1-\alpha\theta\right)\left[2\left\langle \mathbb{S}x^{k-1},\mathbb{S}z^{k-1}\right\rangle_M-\left\|\mathbb{S}z^{k-1}\right\|_M^2\right].$$

Thus, we get

$$\mathbb{E}\left[\left\|x^k-z^k\right\|_M^2\right]\leq\mathbb{E}\left[\left\|x^{k-1}-z^{k-1}\right\|_M^2\right]+A.$$

Apply above inequality repeatedly, divide by k^2 to conclude the proof. □

Remark. To prove the main theorem of L^2 convergence, apply Pazy's theorem on z^k to conclude the proof.

Proof of L^2 convergence

First, let's bound $\left\|\mathbb{S}z^k\right\|_M$ independent from k.

Lemma

$\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ is a θ-averaged with $\theta \in (0,1]$ and $\mathbb{S} = \theta^{-1}(\mathbb{I} - \mathbb{T})$.
Let $z^0, z^1, z^2, \cdots$ to be the iterates of (RC-FPI) *with starting point z^0.*
Then,

$$\left\|\mathbb{S}z^k\right\|_M \leq \left\|\mathbb{S}z^{k-1}\right\|_M \leq \cdots \leq \left\|\mathbb{S}z^0\right\|_M .$$

Proof.

With $\mathbb{T}z - z = -\theta\mathbb{S}z$ and $\mathbb{S}$ being $(1/2)$-cocoercive operator,

$$\theta \left\langle \mathbb{S}\mathbb{T}z - \mathbb{S}z, -\mathbb{S}z - \mathbb{S}\mathbb{T}z \right\rangle_M \geq (1-\theta)\left\|\mathbb{S}\mathbb{T}z - \mathbb{S}z\right\|_M^2 \geq 0,$$

Thus $\left\|\mathbb{S}\mathbb{T}z\right\|_M \leq \left\|\mathbb{S}z\right\|_M$ holds for any $z \in \mathcal{H}$, concluding the proof. $\square$

Proof of L^2 convergence

Next, let's bound $\left\|\mathbb{E}\left[\mathbb{S}x^k\right]\right\|_M$ independent from k.

Lemma

$\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ is a θ-averaged with $\theta \in (0,1]$ and $\mathbb{S} = \theta^{-1}(\mathbb{I} - \mathbb{T})$. Let $x^0, x^1, x^2, \cdots$ to be the iterates of (RC-FPI) with starting point x^0. Then,

$$\left\|\mathbb{E}\left[\mathbb{S}\mathbb{T}_{\mathcal{I}^k} \dots \mathbb{T}_{\mathcal{I}^0} x^0\right]\right\|_M \le \beta^{1/2}\alpha^{-1}\left\|\mathbb{S}x^0\right\|_M$$

holds if $\mathcal{I}^0, , \mathcal{I}^1, \dots, \mathcal{I}^k$ follow IID distribution with the uniform expected step-size condition with $\alpha \in (0,1]$ and β satisfies $\beta \le \alpha/\theta$.

Proof of L^2 convergence

Proof.

With the same distribution of $\mathcal{I}$, it is possible to show

$$\mathbb{E}_{\mathcal{I},X,Y}\left[\|\mathbb{T}_{\mathcal{I}}X - \mathbb{T}_{\mathcal{I}}Y\|_M^2\right] \leq \mathbb{E}_{X,Y}\left[\|X - Y\|_M^2\right].$$

for any random variable X, Y on $\mathcal{H}$. ($\because \beta \leq \alpha/\theta$ and $(1/2)$-cocoersivity.)
Apply above inequality repeatedly and use Jensen's inequality:

$$\|\mathbb{E}\left[\mathbb{T}_{\mathcal{I}^k}\ldots\mathbb{T}_{\mathcal{I}^1}X - \mathbb{T}_{\mathcal{I}^k}\ldots\mathbb{T}_{\mathcal{I}^1}Y\right]\|_M^2 \leq \mathbb{E}\left[\|X - Y\|_M^2\right].$$

Now set up X, Y as $X = \mathbb{T}_{\mathcal{I}^0}x^0$, $Y = x^0$. Shift the index using IID,

$$\left\|\alpha\mathbb{E}\left[\theta\mathbb{S}x^k\right]\right\|_M = \left\|\mathbb{E}\left[(\mathbb{T}_{\mathcal{I}^k} - \mathbb{I})\mathbb{T}_{\mathcal{I}^{k-1}}\ldots\mathbb{T}_{\mathcal{I}^0}x^0\right]\right\|_M \leq \beta^{1/2}\left\|\theta\mathbb{S}x^0\right\|_M.$$

□

Table of Contents

Almost sure convergence

Theorem (Almost sure convergence of normalized iterate)

Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ *be* θ*-averaged with respect to* $\|\cdot\|$*-norm with* $\theta \in (0,1)$*. Assume* $\mathcal{I}^0, \mathcal{I}^1, \dots$ *is sampled IID from a distribution satisfying the uniform expected step-size condition.*
Let $x^0, x^1, x^2, \dots$ *be the iterates of* (RC-FPI)*. Then*

$$\frac{x^k}{k} \overset{\text{a.s.}}{\to} -\alpha \mathbf{v}$$

as $k \to \infty$*.*

Remark. We will prove for the M-norm with the assumption:

$$\alpha > \theta\beta.$$

Proof of almost sure convergence

As mentioned, we will alternatively prove the following result.

Lemma

Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be θ-averaged with respect to $\|\cdot\|_M$ with $\theta \in (0, 1]$. Assume $\mathcal{I}^0, \mathcal{I}^1, \ldots$ is sampled IID from a distribution satisfying the uniform expected step-size condition and β satisfies that $\beta < \alpha/\theta$. Let $x^0, x^1, x^2, \ldots$ be the iterates of (RC-FPI)*. Then, x^k/k is strongly convergent to $-\alpha\mathbf{v}$ in probability 1, i.e.*

$$\frac{x^k}{k} \overset{a.s.}{\to} -\alpha\mathbf{v}$$

as $k \to \infty$, where $\mathbf{v}$ is the infimal displacement vector of $\mathbb{T}$.

Almost supermartingale convergence theorem

Lemma (Robbins-Siegmund quasi-martingale theorem)

$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ is a sequence of sub-σ-algebras of $\mathcal{F}$ where $(\Omega, \mathcal{F}, P)$ is a probability space. When $X_k, b_k, \tau_k, \zeta_k$ are non-negative $\mathcal{F}_k$-random variables such that

$$\mathbb{E}\left[X_{k+1} \mid \mathcal{F}_k\right] \leq (1 + b_k)\, X_k + \tau_k - \zeta_k,$$

$\lim_{k\to\infty} X_k$ exists and is finite and $\sum_{k=1}^{\infty} \zeta_k < \infty$ almost surely if $\sum_{k=1}^{\infty} b_k < \infty, \sum_{k=1}^{\infty} \tau_k < \infty$.

Proof of almost sure convergence

Abstract Proof.

The idea is to apply *almost supermartingale theorem*.
To do this, we first need to make an upper bound of the last two terms in *One-step inequality* without taking full expectation.

$$
\begin{aligned}
&-\alpha\theta\left(1-\alpha\theta\right)\left\|\mathbb{S}x-\mathbb{S}z\right\|_M^2+\theta^2\left(\beta-\alpha^2\right)\left\|\mathbb{S}x\right\|_M^2 \\
&\quad=\underbrace{-\theta\left(\alpha-\beta\theta\right)\left\|\mathbb{S}x-\frac{\alpha-\alpha^2\theta}{\alpha-\beta\theta}\mathbb{S}z\right\|_M^2}_{\leq 0}+\underbrace{\frac{\alpha\theta^2(1-\alpha\theta)(\beta-\alpha^2)}{\alpha-\beta\theta}}_{=:B\geq 0}\left\|\mathbb{S}z\right\|_M^2 \\
&\quad\leq B\left\|\mathbb{S}z\right\|_M^2.
\end{aligned}
$$

Now consider a sequence $z^0, z^1, z^2, \cdots$ of (FPI by $\bar{\mathbb{T}}$).

Proof of almost sure convergence

Abstract Proof continued.

From $\left\|\mathbb{S}z^k\right\|_M \leq \left\|\mathbb{S}z^0\right\|_M$,

$$\mathbb{E}_{\mathcal{I}^k}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2 \mid \mathcal{F}_{k-1}\right] \leq \left\|\frac{x^{k-1}}{k-1} - \frac{z^{k-1}}{k-1}\right\|_M^2 + \frac{B}{k^2}\left\|\mathbb{S}z^0\right\|_M^2.$$

Since $\sum \frac{B}{k^2}\left\|\mathbb{S}z^0\right\|_M^2 < \infty$, apply the R-S quasi-martingale theorem.
Then, $\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2$ converges almost surely to some random variable.
With Fatou's lemma and L^2 convergence,

$$\mathbb{E}\left[\lim_{k\to\infty}\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right] \leq \lim_{k\to\infty}\mathbb{E}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right] = 0.$$

Thus, with probability 1, $\lim_{k\to\infty}\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2 = 0.$ □

Table of Contents

Variance of Normalized iterates

Goal. Here's the main result we will prove in this section.

Theorem

Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ *be* θ*-averaged with respect to* $\|\cdot\|_M$ *with* $\theta \in (0,1]$. *Assume* $\mathcal{I}^0, \mathcal{I}^1, \ldots$ *is sampled IID from a distribution satisfying the uniform expected step-size condition and* β *satisfies that* $\beta < \alpha/\theta$. *Let* $x^0, x^1, x^2, \ldots$ *be the iterates of* (RC-FPI).

(a) *If* $\mathbf{v} \in \operatorname{range}(\mathbb{I} - \mathbb{T})$, *then*

$$\limsup_{k\to\infty} k\mathbb{E}\left[\left\|\frac{x^k}{k} + \alpha\mathbf{v}\right\|_M^2\right] \le (\beta - \alpha^2)\,\|\mathbf{v}\|_M^2.$$

(b) *In general, regardless of whether* $\mathbf{v}$ *is in* $\operatorname{range}(\mathbb{I} - \mathbb{T})$ *or not,*

$$\limsup_{k\to\infty} k\operatorname{Var}_M\left(\frac{x^k}{k}\right) \le (\beta - \alpha^2)\,\|\mathbf{v}\|_M^2.$$

Variance of Normalized iterates

Example. Before moving on to the proofs, here's check some examples.

Ex 1. Equality holds: Consider the translation operator $\mathbb{T}(x) = x - \mathbf{v}$. Choose the norm as $\|\cdot\|$ and consider the distribution of $\mathcal{I}$ as an uniform distribution on standard basis.
When $x^0, x^1, x^2, \ldots$ are the iterates of (RC-FPI) with $\mathbb{T}$, then

$$k \mathrm{Var}\left(\frac{x^k}{k}\right) = \alpha\left(1-\alpha\right)\|\mathbf{v}\|^2$$

for $k = 1, 2, \ldots$, and the variance bound holds with equality.

Remark. In this scenario, each step is independent to each other.
The result is identical to the result of *Central Limit Theorem*.

Variance of Normalized iterates

Ex 2. Strict inequality Define $\mathbb{T}\colon \mathbb{R}^2 \to \mathbb{R}^2$ as $\mathbb{T} = \mathbb{I} - \mathrm{Proj}_{x+y=1}$, or

$$\mathbb{T}\colon (x, y) \mapsto \left(x - \frac{1+x-y}{2}, y - \frac{1+y-x}{2} \right),$$

which is 1/2-averaged and has the infimal displacement vector $(1/2, 1/2)$.

When $(x^0, y^0), (x^1, y^1), (x^2, y^2), \dots$ are the iterates of (RC-FPI) with $\mathbb{T}$,

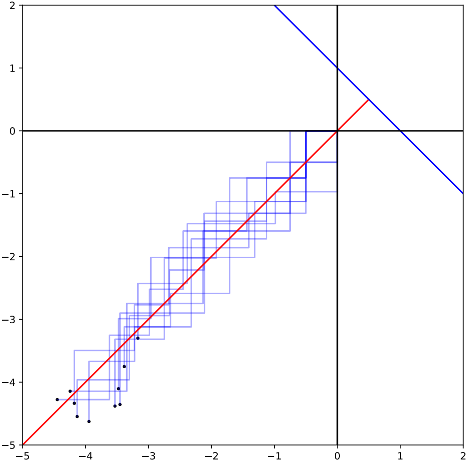
$$\limsup_{k\to\infty} k\mathrm{Var}_M \left(\frac{\left(x^k, y^k\right)}{k} \right) = \frac{1}{24}.$$

On the other hand, the right hand side of the inequality is

$$\alpha\left(1-\alpha\right)\left\|\mathbf{v}\right\|^2 = \frac{1}{2}\left(1 - \frac{1}{2}\right)\left\|\mathbf{v}\right\|^2 = \frac{1}{8}.$$

Remark. This is also an example that $\lim_{k\to\infty} \mathbb{E}\left[\left\|\theta \mathbb{S} x^k\right\|_M^2\right] \neq \|\mathbf{v}\|_M^2$.

Variance of Normalized iterates



Simulation of prior example, 15 iterations, ran 10 times.

Idea of proof

Idea of proof.

First generate a sequence $z^0, z^1, z^2, \cdots$ as (FPI by $\bar{\mathbb{T}}$) with $z^0 = z$. From the *One-step inequality*, with full expectation,

$$\mathbb{E}\left[k\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right] \leq \frac{1}{k}\left\|x^0 - z^0\right\|_M^2 + \mathbb{E}\left[\frac{1}{k}\sum_{j=0}^{k-1} U^j\right]$$

where $U^0, U^1, U^2, \ldots$ is a sequence of random variables:

$$U^k = -\alpha\left(\theta^{-1} - \alpha\right)\left\|\theta \mathbb{S} x^k - \theta \mathbb{S} z^k\right\|_M^2 + \theta^2\left(\beta - \alpha^2\right)\left\|\mathbb{S} x^k\right\|_M^2.$$

The key of this proof is to bound U^k asymptotically as $k \to \infty$,

$$U^k \lesssim (\beta - \alpha^2)\left\|\mathbf{v}\right\|_M^2.$$

Then due to Cesàro mean we may conclude the proof. □

Proof of Variance of Normalized iterates

Abstract Proof.

1. Choose z with $\|\mathbf{v}\|_M \le \left\|\theta \mathbb{S} z^k\right\|_M \le \|\theta \mathbb{S} z\|_M \le \|\mathbf{v}\|_M + \epsilon$.
   From convexity of range set, this gives $\mathbf{v} \simeq \theta \mathbb{S} z^k \simeq \theta \mathbb{S} z$.
2. $\mathbb{S} x^k - \mathbb{S} z$ and $\mathbb{S} z$ are nearly orthogonal for after sufficient iteration:
   $$\left\|\theta \mathbb{S} x^k\right\|_M^2 \simeq \left\|\theta \mathbb{S} x^k - \theta \mathbb{S} z\right\|_M^2 + \|\theta \mathbb{S} z\|_M^2 .$$
   Nearly orthogonal property will be handled in the next lemma.
3. Then we use this to build an inequality of a form:
   $$U^k \lesssim \underbrace{-\theta^{-1}(\alpha - \beta\theta)\left\|\theta \mathbb{S} x^k - \mathbf{v}\right\|_M^2}_{\le 0} + \underbrace{\left(\beta - \alpha^2\right)\|\mathbf{v}\|_M^2}_{\text{RHS of the theorem}},$$
   as $k \to \infty$.

Proof of Variance of Normalized iterates

Abstract Proof.

To be precise, with probability 1, when $x^0, x^1, x^2, \cdots$ are generated, for $\delta \in (0, \pi/2)$ and $z \in \mathcal{H}$, there exists $N_{\delta,z}$ s.t. for all $k > N_{\delta,z}$,

$$U^k \leq \theta^2 \left(\beta - \alpha^2\right) \|\mathbb{S}z\|_M^2 + \frac{\theta}{\alpha - \beta\theta} \tilde{\tau}_{\delta,z}^2,$$

where $\tilde{\tau}_{\delta,z}$ depends only on δ, z and $\tilde{\tau}_{\delta,z} \to 0$ as $\delta \to 0$ and $\theta\mathbb{S}z \to \mathbf{v}$.

1. Since $N_{\delta,z}$ also depends on $\{x^k\}$, apply Cesàro mean:

$$\limsup_{k\to\infty} \left[\frac{1}{k} \sum_{j=0}^{k-1} U^j\right] \leq \theta^2 \left(\beta - \alpha^2\right) \|\mathbb{S}z\|_M^2 + \frac{\theta}{\alpha - \beta\theta} \tilde{\tau}_{\delta,z}^2.$$

2. Apply expectation, then use Reverse Fatou's lemma (since $U^j \leq B \|\mathbb{S}z\|_M^2$ holds) to bound the variance term.
3. Finally, take the limit $\delta \to 0$ and $\theta\mathbb{S}z \to \mathbf{v}$ to conclude the proof.

□

Nearly orthogonal

Lemma (Nearly orthogonal)

Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be θ-averaged with respect to $\|\cdot\|_M$ with $\theta \in (0,1]$. Consider a sequence $y^0, y^1, y^2, \ldots$ in $\mathcal{H}$ such that its normalized iterate converges strongly to $-\gamma\mathbf{v}$ for some $\gamma > 0$:

$$\lim_{k\to\infty} \frac{y^k}{k} = -\gamma\mathbf{v}.$$

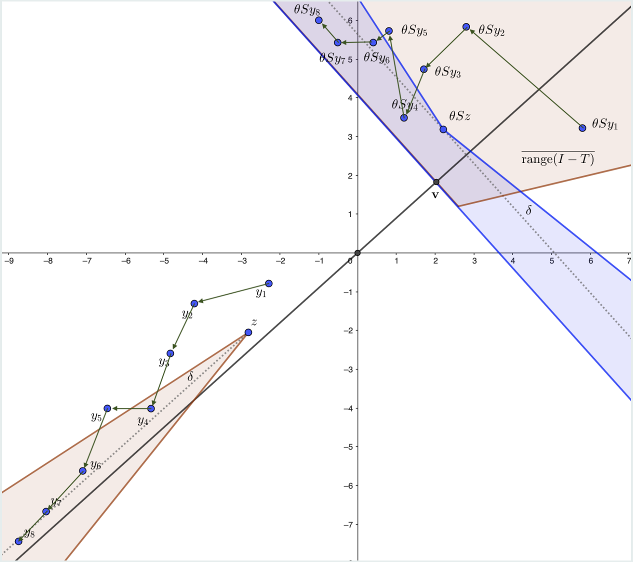
Then, for any $\delta \in (0, \pi/2)$ and $z \in \mathcal{H}$, $\exists N_{\delta,z} \in \mathbb{N}$ s.t., $\forall k > N_{\delta,z}$,

$$\left\langle \mathbf{v}, \mathbb{S}y^k - \mathbb{S}z \right\rangle_M \leq \|\mathbf{v}\|_M \left\|\mathbb{S}y^k - \mathbb{S}z\right\|_M \sin\delta.$$

The proof is quite technical. We replace the proof with a figure describing the dynamics, while the full proof is available in the paper.

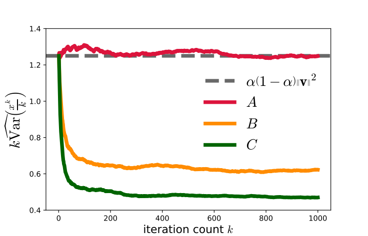
Nearly orthogonal

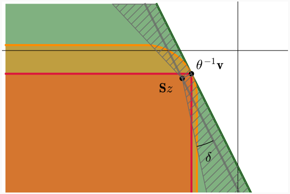
Idea of proof.



Relation between variance and range set

Remark. Throughout observations, we could check that there was a weak relation between the shape of the range set $\operatorname{range}(\mathbb{I} - \mathbb{T})$ near $\mathbf{v}$ and the variance value.





We interpret such phenomenon happen when the following inequality can be non-tight even after sufficient iterations.

$$-\theta^{-1}(\alpha - \beta\theta)\left\|\theta \mathbb{S} x^k - \mathbf{v}\right\|_M^2 \le 0$$

Table of Contents

Infeasibility Detection

Goal. Infeasibility detection method for (RC-FPI) by hypothesis testing.

Theorem

Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ *be* θ*-averaged with respect to* $\|\cdot\|_M$ *with* $\theta \in (0, 1]$. *Consider a null hypothesis of* $\|\mathbf{v}\|_M \leq \delta$ *with small* δ *satisfying* $\alpha\delta < \epsilon$. *Assume* $\mathcal{I}^0, \mathcal{I}^1, \ldots$ *is sampled IID from a distribution satisfying the uniform expected step-size condition and* β *satisfies that* $\beta < \alpha/\theta$. *Let* $x^0, x^1, x^2, \ldots$ *be the iterates of* (RC-FPI). *Then*

$$\mathbb{P}\left(\left\|\frac{x^k}{k}\right\|_M \geq \varepsilon\right) \lesssim \frac{\left(\beta - \alpha^2\right)\delta^2}{k(\varepsilon - \alpha\delta)^2}$$

as $k \to \infty$, *where* $\mathbf{v}$ *is the infimal displacement vector of* $\mathbb{T}$.

Infeasibility Detection

Therefore, for any statistical significance level $p \in (0, 1)$, the test

$$\left\| \frac{x^k}{k} \right\|_M \geq \varepsilon$$

with

$$k \gtrsim \frac{\left(\beta - \alpha^2\right) \delta^2}{p \left(\varepsilon - \alpha\delta\right)^2}$$

can reject the null hypothesis and conclude that $\|\mathbf{v}\|_M > \delta$, which implies that the problem is inconsistent.

Proof of infeasibility detection

Proof.

Here, we provide a simpler case where $\mathbf{v} \in \operatorname{range}(\mathbb{I} - \mathbb{T})$.

By the triangle inequality, Markov inequality, and the variance theorem we just proved, under the null hypothesis,

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{x^k}{k}\right\|_M \geq \varepsilon\right) &\leq \mathbb{P}\left(\left\|\frac{x^k}{k} + \alpha\mathbf{v}\right\|_M \geq \varepsilon - \alpha\delta\right) \\
&\leq \frac{1}{(\varepsilon - \alpha\delta)^2}\mathbb{E}\left[\left\|\frac{x^k}{k} + \alpha\mathbf{v}\right\|_M^2\right] \\
&\lesssim \frac{\left(\beta - \alpha^2\right)\delta^2}{k(\varepsilon - \alpha\delta)^2},
\end{aligned}$$

as $k \to \infty$. The full proof use the inequality about the variance. □

Table of Contents

Review

Review.

In general $\|\cdot\|_M$-norm,

- if $\alpha \geq \theta\beta$,

$$\frac{x^k}{k} \stackrel{L^2}{\rightarrow} -\alpha \mathbf{v}.$$

- If $\alpha > \theta\beta$,

$$\frac{x^k}{k} \stackrel{\text{a.s.}}{\rightarrow} -\alpha \mathbf{v},$$

and $\limsup_{k\to\infty} k\mathrm{Var}_M\left(\frac{x^k}{k}\right) \leq (\beta - \alpha^2)\|\mathbf{v}\|_M^2$.

When we use $\|\cdot\|$-norm, we can choose $\beta = \alpha$. Thus the conditions $\alpha \geq \theta\beta$ and $\alpha > \theta\beta$ are replaced with $\theta \in (0, 1]$ and $\theta \in (0, 1)$.

Remark. However, optimization methods such as (PG-EXTRA) uses averaged operator that's 1-Lipschitz in $\|\cdot\|_M$-norm with $M \neq \mathbb{I}$.

Extension to Nonorthogonal basis

Now let's consider an update with nonorthogonal basis.

- The underlying space $\mathcal{H}$ with extra $\mathcal{H}_0$ block, making $\mathcal{H}$ as

$$\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \dots \mathcal{H}_m.$$

- Consider two subspaces U_1 and U_2 of $\mathcal{H}$ as

$$U_1 = \mathcal{H}_0 \times 0 \times 0 \times \cdots \times 0, \quad U_2 = 0 \times \mathcal{H}_1 \times \mathcal{H}_2 \times \cdots \times \mathcal{H}_m.$$

- We assume that with M-inner product of $\mathcal{H}$,
  - Block components in U_2 are orthogonal.
  - U_1 and U_2 need not be orthogonal.

Question. When are the conditions $\alpha \geq \theta\beta$ and $\alpha > \theta\beta$ satisfied?

Friedrichs angle

Definition (Friedrichs angle)

The cosine of the *Friedrichs angle* c_F between two subspaces U_1 and U_2 is defined as a smallest value among $c \leq 1$ such that satisfies:

$$|\langle u_1, u_2 \rangle_M| \leq c \, \|u_1\|_M \, \|u_2\|_M \quad \forall \, u_1 \in U_1, \; u_2 \in U_2.$$

Remark. When $M = \mathbb{I}$, $c_F = 0$.

Extension to Nonorthogonal basis

Lemma

Suppose the subspaces U_1, U_2 of $\mathcal{H}$ with $U_1 \cap U_2 = \{0\}$ satisfy:

$$|\langle u_1, u_2 \rangle_M| \leq c_F \left\| u_1 \right\|_M \left\| u_2 \right\|_M, \quad c_F \leq \sqrt{\frac{1-\theta}{1-\alpha\theta}}$$

for any $u_1 \in U_1, u_2 \in U_2$.
Then, there exists $\beta \geq 0$ such that $\beta\theta \leq \alpha$ and

$$\mathbb{E}_{\mathcal{I}}\left[u_{\mathcal{I}}\right] = \alpha u, \quad \mathbb{E}_{\mathcal{I}}\left[\left\| u_{\mathcal{I}} \right\|_M^2\right] \leq \beta \left\| u \right\|_M^2.$$

If $c_F < \sqrt{\frac{1-\theta}{1-\alpha\theta}}$, then there exists β with $\beta\theta < \alpha$.

Remark. Setting β as $\beta = \alpha^2 + \frac{\alpha-\alpha^2}{1-c_F^2}$ proves the lemma.

Extension to Nonorthogonal basis

Thus, in this case of (RC-FPI),

Extension to nonorthogonal basis.

In general $\|\cdot\|_M$-norm,

- if $c_F \leq \sqrt{\frac{1-\theta}{1-\alpha\theta}}$,

$$\frac{x^k}{k} \overset{L^2}{\to} -\alpha \mathbf{v}.$$

- If $c_F < \sqrt{\frac{1-\theta}{1-\alpha\theta}}$,

$$\frac{x^k}{k} \overset{\text{a.s.}}{\to} -\alpha \mathbf{v},$$

and $\limsup_{k\to\infty} k \mathrm{Var}_M \left(\frac{x^k}{k}\right) \leq (\beta - \alpha^2) \|\mathbf{v}\|_M^2$.

Remark. β value is guaranteed to be at most $\beta = \alpha^2 + \frac{\alpha-\alpha^2}{1-c_F^2}$.

Application in PG-EXTRA

Definition (PG-EXTRA)

Consider the convex optimization problem

$$\underset{x\in\mathbb{R}^d}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x),$$

where $f_i\colon \mathbb{R}^d \to \mathbb{R}$ is closed, convex, and proper function.
The decentralized algorithm *PG-EXTRA* is a iterative solver:

$$\begin{aligned} x_i^{k+1} &= \mathrm{Prox}_{\tau f_i}\left(\textstyle\sum_{j=1}^m W_{ij} x_i^k - w_i^k\right) \\ w_i^{k+1} &= w_i^k + \frac{1}{2}\left(x_i^k - \textstyle\sum_{j=1}^m W_{ij} x_j^k\right) \end{aligned} \qquad \text{(PG-EXTRA)}$$

Application in PG-EXTRA

Remark. When using (PG-EXTRA), it is computed with m agents.

W is a matrix representing the connection between agents, satisfying:

- W is symmetric m by m matrix.
- $W_{ij} = 0$ if $i \neq j$ and agents i and j are not connected.
- $N(I - W) = \text{span}(\mathbf{1})$ and $1 = \lambda_1 > \max\{|\lambda_2|, \cdots, |\lambda_m|\}$.

Each agent i handles x_i and w_i value. At each iteration, only communication between connected agents are required.

Application in PG-EXTRA

A randomized coordinate-update version of (PG-EXTRA) performs:

- randomly chooses i among $1, 2, \ldots, m$ to update x_i^k,
- every $w_1, w_2, \ldots, w_m$ gets updated at each iterations.

RC-PG-EXTRA.

while Not converged **do**

  Sample: $\mathcal{I}$

  for i such that $\mathcal{I}_i \neq 0$ **do**

    $\Delta x_i = \mathrm{Prox}_{\tau f_i}([Wx]_i - w_i) - x_i$

    Update: $x_i = x_i + \mathcal{I}_i \Delta x_i$

    for $j \in N_i \cup \{i\}$ **do**

      Send: Δx_i From ith agent to jth agent.

      $[Wx]_j = [Wx]_j + W_{ij} \Delta x_i$

    end for

  end for

end while

Application in PG-EXTRA

Corollary

Assume $\mathcal{I}^0, \mathcal{I}^1, \ldots$ *is sampled IID from a distribution satisfying the uniform expected step-size condition. Perform (RC-PG-EXTRA).*
If the minimum eigenvalue of the mixing matrix $W \in \mathbb{R}^m$ *satisfies:*

$$\lambda_{min}(W) > -\frac{\alpha}{2-\alpha},$$

then the results of previous theorems holds:

$$\frac{x^k}{k} \xrightarrow{\text{a.s.}} -\alpha \mathbf{v}, \quad \limsup_{k\to\infty} k \mathrm{Var}_M \left(\frac{x^k}{k}\right) \leq (\beta - \alpha^2) \|\mathbf{v}\|_M^2.$$

Application in PG-EXTRA

Additionally, here is the infimal displacement vector of (PG-EXTRA).

Lemma

The infimal displacement vector $\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$ *of* (PG-EXTRA) *is*

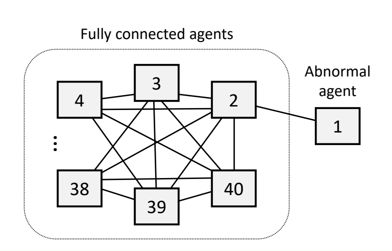
$$\mathbf{v}_i = \begin{bmatrix} \frac{\tau}{m} \sum_{j=1}^m g_j \\ -\frac{1}{2} \left(y_i - \sum_{j=1}^m W_{ij} y_j \right) \end{bmatrix}$$

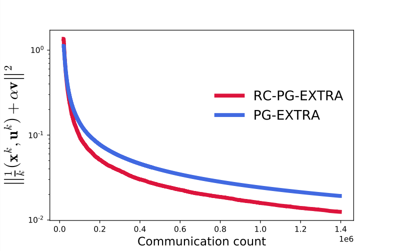
for $i = 1, \ldots, m$, *where* $(y_1, y_2, \ldots, y_m)$ *and* $(g_1, g_2, \ldots, g_m)$ *are*

$$\underset{\substack{y_1, y_2, \ldots y_m \in \mathbb{R}^d \\ g_j \in \partial f_j(y_j), 1 \le j \le m}}{\operatorname{argmin}} \left\| \frac{\tau}{m} \sum_{j=1}^m g_j \right\|^2 + \frac{1}{2} \sum_{i,j=1}^m W_{ij} \left\| y_i - y_j \right\|^2.$$

Application in PG-EXTRA

Remark. We observed faster convergence to the infimal displacement vector when randomized, in the inconsistent case of (PG-EXTRA). Here, the notion of *faster* regards on the communication count, not iterations.





(Left) Network used in our experiment, consisting of $m = 40$ agents, with agents $2, \ldots, 40$ densely connected.
(Right) Graph of $\left\|\left(\mathbf{x}^k, \mathbf{u}^k\right)/k + \alpha\mathbf{v}\right\|^2$ against the communication count for (PG-EXTRA) and (RC-PG-EXTRA).

Reference i

J. Paeng, J. Park, and E. K. Ryu, *"Coordinate-Update Algorithms can Efficiently Detect Infeasible Optimization Problems"*, Preprint (2023) , Submitted to JMAA, arXiv:2305.12211

A. Pazy, *"Asymptotic behavior of contractions in Hilbert space"*, Israel Journal of Mathematics 9 (1971) 235–240.

H. Robbins, D. Siegmund, *"A convergence theorem for non negative almost supermartingales and some applications"*, Optimizing Methods in Statistics (1971) 233–257.