Coordinate-Update Algorithms can Efficiently Detect Infeasible Optimization Problems

Jinhee Paeng[a], Jisun Park[a], Ernest K. Ryu[a,*]

[a]*Department of Mathematical Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, 08826, Seoul, Korea*

Abstract

Coordinate update/descent algorithms are widely used in large-scale optimization due to their low per-iteration cost and scalability, but their behavior on infeasible or misspecified problems has not been much studied compared to the algorithms that use full updates. For coordinate-update methods to be as widely adopted to the extent so that they can be used as engines of general-purpose solvers, it is necessary to also understand their behavior under pathological problem instances. In this work, we show that the normalized iterates of randomized coordinate-update fixed-point iterations (RC-FPI) converge to the infimal displacement vector and use this result to design an efficient infeasibility detection method. We then extend the analysis to the setup where the coordinates are defined by non-orthonormal basis using the Friedrichs angle and then apply the machinery to decentralized optimization problems.

Keywords: convex optimization, monotone operator theory, fixed-point iterations

1. Introduction

Coordinate update/descent algorithms are widely used in large-scale optimization due to their low per-iteration cost and scalability. These algorithms update only a single block of coordinates of an optimization variable per iteration in contrast to full or stochastic gradient algorithms, which update all variables every iteration. The convergence of coordinate update algorithms has been analyzed extensively, and they have been shown to achieve strong practical and theoretical performance in many large-scale machine learning and optimization problems [1] for non-pathological problem instances.

However, the behavior of coordinate update algorithms on infeasible or misspecified problems has not been analyzed, which sharply contrasts with algorithms that use full (deterministic) updates. The recent interest in building general-purpose optimization solvers with first-order algorithms has led to much work analyzing the behavior of full-update

*Corresponding author.

Email addresses: jhpaeng306@snu.ac.kr (Jinhee Paeng), colleenp0515@snu.ac.kr (Jisun Park), ernestryu@snu.ac.kr (Ernest K. Ryu)

first-order algorithms on pathological problem instances so that the solvers can robustly detect such instances. For coordinate-update methods to be as widely adopted, to the extent that they can be used as engines of general-purpose solvers, it is necessary to also understand their behavior under pathological problem instances.

1.1. Summary of results, contribution, and organization

In this work, we analyze the behavior of randomized coordinate-update fixed-point iterations (RC-FPI) applied to inconsistent problem instances. Analogous to the classical results of the full-update fixed-point iterations, we show that the normalized iterate x^k/k generated by RC-FPI converges toward the infimal displacement vector, which serves as a certificate of infeasibility, in the sense of both L^2 and almost sure convergence. We then bound the asymptotic bias and variance of the estimator, thereby establishing an asymptotic convergence rate. Finally, we extend the analysis to the setup where the coordinates are defined by non-orthonormal basis using the Friedrichs angle and then apply the machinery to decentralized optimization problems.

Section 3 defines the randomized-coordinate update. The *uniform expected step-size condition* is defined, which is a condition ensuring that each coordinate is updated equally in expectation. α value is defined as the expected scale of the update on each coordinate.

Section 4 presents the convergence of normalized iterate. It starts with Section 4.1, presenting properties about expectation values in (RC-FPI). The upper bound condition for the expectation value of squared norm is presented with β value. We show that when $M = \mathbb{I}$, namely orthonormal basis, such condition is satisfied with $\beta = \alpha$. Additionally, a non-expansive behavior in expectation of squared norm is shown, which in result bounds the expected difference between iterates from (RC-FPI) and (FPI with $\bar{\mathbb{T}}$).

Section 4.2 and Section 4.4 present first two key achievements of this paper, the convergence of normalized iterate when the basis is orthonormal. Section 4.2 handles the L^2 convergence, and Section 4.4 handles almost sure convergence when the operator is averaged, i.e. $\theta < 1$. The normalized iterate x^k/k converges to the $-\alpha\mathbf{v}$, where $\mathbf{v}$ is the infimal displacement vector.

Section 5 presents the third main achievement, the asymptotic upper bound of the variance of normalized iterate. We show that as the iteration count $k \to \infty$, the limit supremum of the variance is bounded in $\mathcal{O}(1/k)$. We further find the example with the equality to show the bound is strict. Then, we present an experiment about a relation between range set and the variance.

Section 6 presents the infeasibity detection method for (RC-FPI). Results from preceeding sections are used to construct such method. The method focuses on rejecting the null hypothesis $\|\mathbf{v}\|_M \leq \delta$, by checking $\left\|x^k/k\right\|_M \geq \epsilon$, after certain iteration count. We provide the required iteration count to use such method.

Section 7 presents an extension of our results to the non-orthogonal basis. We show that a certain condition on Friedrichs angle need to be satisfied to obtain the same results. Then we apply this result on a decentralized optimization, (PG-EXTRA). Furthermore, in the experiment of (PG-EXTRA) on infeasible problem, the convergence of the normalized iterate was found to be faster in (RC-FPI) than (FPI).

The paper is organized as follows. Section 2 sets up notations and reviews known results and notions. Section 3 provides clear definition of randomized-coordinate update setting. Section 4 presents the L^2 and almost sure convergence of the normalized iterate. Section 5 provides the asymptotic upper bound for the normalized iterate. Section 6 then uses these results to build the infeasibility detection in (RC-FPI). Section 7 extends our result to the non-orthogonal basis, allowing application to the optimization methods such as (PG-EXTRA). Section 8 concludes the paper.

1.2. Prior work

1.2.1. FPI of inconsistent case.

Behavior of the inconsistent fixed-point iteration has been first characterized by Browder and Petryshyn [2], who showed that the iterates are not bounded. Later, Pazy [3] showed that the iterates actually diverge in a sense that $\lim_{k\to\infty} \frac{\mathbb{T}^k x^0}{k} = -\mathbf{v}$, and this work also led to the similar results in more general Banach space settings [4, 5, 6, 7] or geodesic spaces [8, 9]. If the operator is more than just non-expansive, then the difference of iterates is also convergent to $\mathbf{v}$; see Bruck Jr [10], Bailion et al. [11], Reich and Shafrir [12].

There are also in-depth analyses on the characteristics of infimal displacement vector, regarding its direction [13, 14, 15] and the composition and convex combinations of non-expansive operators [16, 17].

1.2.2. Infeasibility detection and numerical solvers.

Fixed-point iteration covers a broad range of optimization algorithms, including Douglas-Rachford splitting (DRS) [18] or alternating direction method of multipliers (ADMM) [19, 20], which are commonly used as a first-order methods for solving general convex optimization problems. The infimal displacement vector of DRS and ADMM operator have been recently studied [21, 22, 23, 24, 25, 26], and it was proven to have meaning in terms of primal and dual problems as well [27, 28]. Related to such behaviors, ADMM-based infeasibility detecting algorithms have been suggested [29, 30, 31], which led to the first-order numerical solvers like OSQP [32] and COSMO [33]. Apart from above, SCS [34, 35] uses homogeneous self-dual embedding [36, 37].

1.2.3. Randomized coordinate update and RC-FPI.

The coordinate descent is a method which updates one coordinate or blocks at each iteration [38, 39, 40, 41, 42, 43, 44]. Such methods are also popular in proximal setup [45, 46, 47], prox-linear [48, 49, 1, 50, 51, 52, 53, 54, 55], distributed (or asynchronous) setup [56, 57], and even in discrete optimizations [58, 59, 60]. There are in-depth complexity analysis and accelerated variants of coordinate descent method as well [61, 62, 63, 64, 65, 66, 67, 68, 69]. Furthermore, there are attempts to hybrid coordinate update with full update in primal-dual algorithms [70, 71].

Randomized coordinate-update for fixed-point iteration has been first proposed by Verkama [72]. General framework for randomized block-coordinate fixed-point iteration was suggested by Combettes and Pesquet [73, 74], followed by similar line of works including block-coordinate update fixed-point iteration in asynchronous parallel setup [57], forward-backward splitting [75, 76], Douglas-Rachford splitting [77], and so on. It also led to the refined analysis in cyclic fixed-point iterations [78, 79], and the iteration complexity of coordinate update fixed-point iterations and their variants [80, 55, 81, 82, 83].

1.2.4. Friedrichs angle and splitting methods.

Friedrichs angle [84, 85, 86] measures an angle between a number of subspaces, and is often used to characterize the convergence rate of projection methods. [87, 88, 89, 90, 91, 92, 93, 94, 95, 96]. This kind of approach has been extended to cover splitting methods such as DRS and ADMM as well [97, 98, 99, 100, 101, 102, 103].

2. Preliminaries and notations

In this section, we set up notations and review known results. First, let's clarify the underlying space. Throughout this paper, a Hilbert space refers to a real Hilbert space. The underlying space is a real Hilbert space $\mathcal{H}$, which is 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.

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 the norm and inner product of $\mathcal{H}_i$ and x_i, y_i are ith block coordinates of x, y, respectively.

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

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

which can also be a pair of norm and inner product of the space $\mathcal{H}$. $\|\cdot\|$ and $\langle \cdot, \cdot\rangle$ are simply the instances of M-norm and M-inner product with M as an identity map. For the remark, the map M can be expressed as a symmetric positive definite matrix if $\mathcal{H} = \mathbb{R}^n$. In this case, M-inner product and M-norm are

$$\|x\|_M = \sqrt{x^T M x}, \qquad \langle x, y\rangle_M = x^T M y.$$

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

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

We develop the theory of Sections 4 and 5 with the general M-norm so that the theory is applicable to the applications of Section 7.

2.1. Operators

Denote $\mathbb{I}\colon \mathcal{H} \to \mathcal{H}$ as the identity operator. For an operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$, let $\operatorname{range} \mathbb{T}$ be a range of $\mathbb{T}$. If $x_\star \in \mathcal{H}$ is a point such that $x_\star = \mathbb{T}x_\star$, it is called a fixed point of $\mathbb{T}$.

We say an operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ is non-expansive with respect to $\|\cdot\|_M$ if

$$\|\mathbb{T}x - \mathbb{T}y\|_M \leq \|x - y\|_M\,, \qquad \forall x, y \in \mathcal{H},$$

and is θ-averaged for $\theta \in (0, 1)$ if $\mathbb{T} = (1-\theta)\mathbb{I} + \theta\mathbb{C}$ for some non-expansive operator $\mathbb{C}$. For notational convenience, we will refer to non-expansive operators as θ-averaged operators with $\theta = 1$, even though, strictly speaking, $\theta = 1$ means the operator is not averaged. An operator $\mathbb{S}\colon \mathcal{H} \to \mathcal{H}$ is $(1/2)$-cocoercive with respect to $\|\cdot\|_M$ if

$$\langle \mathbb{S}x - \mathbb{S}y,\, x - y\rangle_M \geq \frac{1}{2}\,\|\mathbb{S}x - \mathbb{S}y\|_M^2\,, \qquad \forall x, y \in \mathcal{H}.$$

$\mathbb{T}$ is non-expansive if and only if $\mathbb{S} = \mathbb{I} - \mathbb{T}$ is $(1/2)$-cocoercive. Also, $\mathbb{T}$ is θ-averaged for some $\theta \in (0, 1)$ if and only if $\mathbb{S} = \theta^{-1}(\mathbb{I} - \mathbb{T})$ is $(1/2)$-cocoercive.

2.2. Inconsistent operators and infimal displacement vector

We say an operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ is consistent if it has a fixed point, and inconsistent if it does not have a fixed point. $\mathbb{T}$ is consistent if and only if $0 \in \operatorname{range}(\mathbb{I} - \mathbb{T})$. When $\mathbb{T}$ is non-expansive, the closure $\overline{\operatorname{range}(\mathbb{I} - \mathbb{T})}$ is a nonempty closed convex set, so it has a unique minimum-norm element, which we denote by $\mathbf{v}$ [3].

We call $\mathbf{v}$ the *infimal displacement vector* of $\mathbb{T}$ [104, 21]. Alternatively, $\mathbf{v}$ is the projection of 0 onto $\overline{\operatorname{range}(\mathbb{I} - \mathbb{T})}$. Equivalently, $\mathbf{v} \in \overline{\operatorname{range}(\mathbb{I} - \mathbb{T})}$ is the infimal displacement vector of $\mathbb{T}$ if and only if

$$\langle y - \mathbf{v},\, \mathbf{v}\rangle_M \geq 0, \qquad \forall y \in \operatorname{range}(\mathbb{I} - \mathbb{T}). \tag{1}$$

For a convex optimization problem, let $\mathbb{T}$ be an operator corresponding to an iterative first-order method, such as the Douglas-Rachford splitting (DRS) operator [18], and let $\mathbf{v}$ be its infimal displacement vector. Loosely speaking, if the optimization problem is feasible and the problem is well-behaved, then $\mathbf{v} = 0$. (However, it is possible for "weakly infeasible" problems to have $\mathbf{v} = 0$, so $\mathbf{v} = 0$ does not guarantee feasibility.) On the other hand, $\mathbf{v} \neq 0$ implies that the problem or its dual problem is infeasible, so $\mathbf{v} \neq 0$ serves as a certificate of infeasibility [30, 29].

2.3. Fixed point iteration and normalized iterate

The fixed-point iteration (FPI) with respect to an operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ is defined as

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

where $x^0 \in \mathcal{H}$ is a starting point.

Let $x^0, x^1, x^2, \ldots$ be the iterates of (FPI). We call x^k/k the kth *normalized iterate* of (FPI) for $k = 1, 2, \ldots$. The seminal work of Pazy [3] characterizes the dynamics of normalized iterates of (FPI).

Theorem 2.1 (Pazy [3])**.** *Let* $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ *be non-expansive. Let* $x^0, x^1, x^2, \dots$ *be the iterates of* (FPI)*. Then, the normalized iterate* x^k/k *converges strongly,*

$$\frac{x^k}{k} \to -\mathbf{v}$$

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

Since the underlying space is a Hilbert space, we clarify that the convergence in the space $\mathcal{H}$ throughout this paper refers to the strong convergence of Hilbert space.

3. Randomized-coordinate update setup

In this section, we focus on a variant of (FPI) with randomized coordinate updates. Consider a θ-averaged operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ with its corresponding (1/2)-cocoercive operator $\mathbb{S} = \theta^{-1}(\mathbb{I} - \mathbb{T})$ with $\theta \in (0, 1]$. To clarify, we will refer to non-expansive operators as θ-averaged operators with $\theta = 1$. Define $\mathbb{S}_i\colon \mathcal{H} \to \mathcal{H}$ for $i = 1, 2, \dots, m$ as $\mathbb{S}_i x = (0, \dots, 0, (\mathbb{S}x)_i, 0, \dots, 0)$, where $(\mathbb{S}x)_i \in \mathcal{H}_i$.

We call $\mathcal{I} = (\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_m) \in [0, 1]^m \subset \mathbb{R}^m$ a *selection vector* and use it as follows. Define $\mathbb{S}_{\mathcal{I}}\colon \mathcal{H} \to \mathcal{H}$ and $\mathbb{T}_{\mathcal{I}}\colon \mathcal{H} \to \mathcal{H}$ as

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

We can think of $\mathbb{S}_{\mathcal{I}}$ as the selection of blocks based on $\mathcal{I}$ and $\mathbb{T}_{\mathcal{I}}$ as the update based on the selected blocks. Throughout this paper, we assume that $\mathcal{I}$ is randomly sampled from a distribution on $[0, 1]^m$ that satisfies the *uniform expected step-size condition*

$$\mathbb{E}_{\mathcal{I}}[\mathcal{I}] = \alpha \mathbf{1} \tag{2}$$

for some $\alpha \in (0, 1]$, where $\mathbf{1} \in \mathbb{R}^m$ is the vector with all entries equal to 1. (Note, $\mathcal{I} \in [0, 1]^m$ already implies $\alpha \in [0, 1]$ so we are additionally assuming that $\alpha > 0$.) The randomized coordinate fixed-point iteration (RC-FPI) is defined as

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

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

(RC-FPI) is a randomized variant of (FPI). The uniform expected step-size condition (2) allows us to view one step of (RC-FPI) to be corresponding to a step of (FPI) with $\bar{\mathbb{T}}\colon \mathcal{H} \to \mathcal{H}$ defined as

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

Equivalently, $\bar{\mathbb{T}} = \mathbb{I} - \alpha\theta\mathbb{S}$.

4. Convergence of normalized iterates

In this section, we show that (RC-FPI) exhibits behavior similar to Theorem 2.1. Let $x^0, x^1, x^2, \ldots$ be the iterates of (RC-FPI). For each $k \in \mathbb{N}$, we likewise call x^k/k the kth *normalized iterate* of (RC-FPI) for $k = 1, 2, \ldots$. Then, the normalized iterate converges to $-\alpha \mathbf{v}$ both in L^2 and almost surely.

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

To clarify, $\xrightarrow{L^2}$ and $\xrightarrow{\text{a.s.}}$ respectively denote L^2 and almost sure convergence of random variables. The almost sure convergence means that x^k/k being strongly convergent to $-\alpha\mathbf{v}$ happens with the probability 1.

4.1. Properties of expectation on RC-FPI

We first characterize certain aspects of (RC-FPI) before establishing our main results. For any $u \in \mathcal{H}$ and selection vector $\mathcal{I}$, define

$$u_{\mathcal{I}} = \sum_{i=1}^{m} \underbrace{\mathcal{I}_i}_{\in \mathbb{R}} \underbrace{(0, \ldots, 0, u_i, 0, \ldots, 0)}_{\in \mathcal{H}},$$

where $u_i \in \mathcal{H}_i$ for $i = 1, \ldots, m$. If $\mathcal{I}$ satisfies the uniform expected step-size condition (2) with $\alpha \in (0, 1]$, then clearly $\mathbb{E}_{\mathcal{I}}[u_{\mathcal{I}}] = \alpha u$. Let $\beta > 0$ be a coefficient such that

$$\mathbb{E}_{\mathcal{I}}\left[\|u_{\mathcal{I}}\|_M^2\right] \leq \beta \|u\|_M^2, \qquad \forall\, u \in \mathcal{H}. \tag{3}$$

Lemma 4.1. *Consider a Hilbert space $\mathcal{H}$ with its norm $\|\cdot\|$. If $\mathcal{I}$ satisfies the uniform expected step-size condition* (2) *with $\alpha \in (0, 1]$, then $\beta = \alpha$ satisfies* (3)*.*

Proof. Since $\mathcal{I}_i \in [0, 1]$,

$$\begin{aligned} \mathbb{E}_{\mathcal{I}}\left[\left\|\sum_{i=1}^{m} \mathcal{I}_i(0, \ldots, 0, u_i, 0, \ldots, 0)\right\|^2\right] &= \mathbb{E}_{\mathcal{I}}\left[\sum_{i=1}^{m} \|\mathcal{I}_i u_i\|_i^2\right] \\ &\leq \mathbb{E}_{\mathcal{I}}\left[\sum_{i=1}^{m} \mathcal{I}_i \|u_i\|_i^2\right] = \sum_{i=1}^{m} \alpha \|u_i\|^2 = \alpha \|u\|^2. \end{aligned}$$

Thus, choose α as β to satisfy the inequality. □

Next, we present a lemma exhibiting a non-expansiveness.

Lemma 4.2. *Let $\mathbf{T}\colon \mathcal{H} \to \mathcal{H}$ be θ-averaged with respect to $\|\cdot\|_M$ with $\theta \in (0, 1]$. Let $\mathcal{I}$ be a random selection vector with distribution satisfying the uniform expected step-size condition* (2) *with $\alpha \in (0, 1]$. Assume* (3) *holds with some β that $\beta \leq \alpha/\theta$. Let X and Y be random variables on $\mathcal{H}$ that are independent of $\mathcal{I}$. (However, X and Y need not be independent.) Then,*

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

Proof. Substitute $\mathbb{T}_{\mathcal{I}} = \mathbb{I} - \theta \mathbb{S}_{\mathcal{I}}$ at $\mathbb{E}_{\mathcal{I},X,Y}\left[\|\mathbb{T}_{\mathcal{I}}X - \mathbb{T}_{\mathcal{I}}Y\|_M^2\right]$ and apply (3) with u as $\mathbb{S}X - \mathbb{S}Y$ to get

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathcal{I},X,Y}\left[\|\mathbb{T}_{\mathcal{I}}X - \mathbb{T}_{\mathcal{I}}Y\|_M^2\right] = \mathop{\mathbb{E}}_{\mathcal{I},X,Y}\left[\|X - Y - \theta\left(\mathbb{S}_{\mathcal{I}}X - \mathbb{S}_{\mathcal{I}}Y\right)\|_M^2\right] \\
&\quad = \mathop{\mathbb{E}}_{X,Y}\left[\|X - Y\|_M^2\right] + \theta^2 \mathop{\mathbb{E}}_{\mathcal{I},X,Y}\left[\|\mathbb{S}_{\mathcal{I}}X - \mathbb{S}_{\mathcal{I}}Y\|_M^2\right] - 2\theta \mathop{\mathbb{E}}_{\mathcal{I},X,Y}\left[\langle X - Y, \mathbb{S}_{\mathcal{I}}X - \mathbb{S}_{\mathcal{I}}Y\rangle_M\right] \\
&\quad \leq \mathop{\mathbb{E}}_{X,Y}\left[\|X - Y\|_M^2\right] + \beta\theta^2 \mathop{\mathbb{E}}_{X,Y}\left[\|(\mathbb{S}X - \mathbb{S}Y)\|_M^2\right] - 2\alpha\theta \mathop{\mathbb{E}}_{X,Y}\left[\langle X - Y, \mathbb{S}X - \mathbb{S}Y\rangle_M\right].
\end{aligned}$$

Since $\beta\theta \leq \alpha$ and $\mathbb{S}$ is (1/2)-cocoercive,

$$\beta\theta^2 \mathop{\mathbb{E}}_{X,Y}\left[\|\mathbb{S}X - \mathbb{S}Y\|_M^2\right] \leq \alpha\theta \mathop{\mathbb{E}}_{X,Y}\left[\|\mathbb{S}X - \mathbb{S}Y\|_M^2\right] \leq 2\alpha\theta \mathop{\mathbb{E}}_{X,Y}\left[\langle X - Y, \mathbb{S}X - \mathbb{S}Y\rangle_M\right].$$

Thus, we can reach the conclusion

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

□

Lemma 4.3. *Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be θ-averaged respect to $\|\cdot\|_M$ with $\theta \in (0,1]$. Let $\mathcal{I}$ be a random selection vector with distribution satisfying the uniform expected step-size condition* (2) *with $\alpha \in (0,1]$. Assume* (3) *holds with some β. For any $x, z \in \mathcal{H}$,*

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

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 the expectation $\mathbb{E}_{\mathcal{I}}\left[\|\mathbb{T}_{\mathcal{I}}x - \bar{\mathbb{T}}z\|_M^2\right]$.

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

Then, use (1/2)-cocoercive property of the operator $\mathbb{S}$.

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

Finally, apply an inequality

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

we get the desired inequality

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

□

4.2. L^2 convergence of normalized iterate

Theorem 4.4. *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* (2) *with $\alpha \in (0,1]$. Let $x^0, x^1, x^2, \ldots$ be the iterates of* (RC-FPI)*. Then*

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

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

Before presenting the full proof, here is the key outline for the proof. Define another sequence $z^0, z^1, z^2, \ldots$ with

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

Apply Lemma 4.3 on the iterates of (RC-FPI) starting from x^0 and the iterates of (FPI with $\bar{\mathbb{T}}$) starting from $z^0 = x^0$. In Section 4.3, we obtain a bound on the last two terms of Lemma 4.3 that is independent of k. More specifically, for all $k = 1, 2, \ldots$,

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

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

$$\mathbb{E}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right] \le \frac{A}{k}$$

and appealing to Theorem 2.1, we conclude with the L^2 convergence. We defer the detailed proof to Section 4.3.

4.3. Proof of Theorem 4.4

In the proof, the norm $\|\cdot\|$ and the inner product $\langle\cdot,\cdot\rangle$ are not used until the final part. Lemmas prior to the main proof of Theorem 4.4 uses the general M-norm and M-inner product.

We start the proof of Theorem 4.4 by presenting two lemmas to upper bound the terms $\left\|\mathbb{S}z^k\right\|_M$ and $\left\|\mathbb{E}\left[\mathbb{S}x^k\right]\right\|_M$.

Lemma 4.5. *$\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ is a θ-averaged with $\theta \in (0,1]$ and choose any starting point $z^0 \in \mathcal{H}$ for* (FPI with $\bar{\mathbb{T}}$)*. When $\mathbb{S} = \theta^{-1}(\mathbb{I} - \mathbb{T})$,*

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

Proof of Lemma 4.5. All we need to prove is,

$$\left\|\mathbb{S}\mathbb{T}z\right\|_M \le \left\|\mathbb{S}z\right\|_M, \qquad \forall z \in \mathcal{H}.$$

From $\mathbb{S}$ being (1/2)-cocoercive operator,

$$2\left\langle \mathbb{S}\mathbb{T}z - \mathbb{S}z, \mathbb{T}z - z\right\rangle_M \geq \|\mathbb{S}\mathbb{T}z - \mathbb{S}z\|_M^2 .$$

With $\mathbb{T}z - z = -\theta\mathbb{S}z$, we get

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

which is equivalent to

$$\|\mathbb{S}\mathbb{T}z\|_M^2 \leq \|\mathbb{S}z\|_M^2 .$$

□

Lemma 4.6. $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ *is a* θ*-averaged with* $\theta \in (0,1]$ *and choose any starting point* $x^0 \in \mathcal{H}$ *for* (RC-FPI)*. When* $\mathbb{S} = \theta^{-1}(\mathbb{I} - \mathbb{T})$,

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

holds if $\mathcal{I}^0,,\mathcal{I}^1,\dots,\mathcal{I}^k$ *follow IID distribution with the condition* (2) *with* $\alpha \in (0,1]$ *and* (3) *holds with some* β *that* $\beta \leq \alpha/\theta$.

Proof of Lemma 4.6. Apply Lemma 4.2 repeatedly, we get

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

for arbitrary random variable X, Y. From Jensen's inequality,

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

Now set up X, Y as

$$X = \mathbb{T}_{\mathcal{I}^0}x^0, \quad Y = x^0.$$

Then as a result, we have an inequality

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

Since $\mathcal{I}_0, \mathcal{I}_1, \dots, \mathcal{I}_n$ are independent and identically distributed, the following equivalence holds.

$$\mathbb{E}\left[\mathbb{T}_{\mathcal{I}^k}\dots\mathbb{T}_{\mathcal{I}^1}x^0\right] = \mathbb{E}\left[\mathbb{T}_{\mathcal{I}^{k-1}}\dots\mathbb{T}_{\mathcal{I}^0}x^0\right].$$

This equality gives the conclusion

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

□

Lemma 4.7. *Let* $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ *be* θ*-averaged with respect to* $\|\cdot\|_M$ *with* $\theta \in (0,1]$*, and let* $x^0, x^1, x^2, \dots$ *be the iterates of* (RC-FPI) *and let* $z^0, z^1, z^2, \dots$ *be the iterates of* (FPI with $\bar{\mathbb{T}}$)*. Assume that the distribution of* $\mathcal{I}$ *satisfies the uniform expected step-size condition* (2) *with* $\alpha \in (0,1]$ *and* (3) *holds with some* β *that* $\beta \le \alpha/\theta$*. Then*

$$\mathbb{E}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right] \le \frac{1}{k^2}\left\|x^0 - z^0\right\|_M^2 + \frac{1}{k}(1-\alpha\theta)\left[2\sqrt{\alpha\theta}\left\|\mathbb{S}x^0\right\|_M\left\|\mathbb{S}z^0\right\|_M - \frac{\alpha}{\theta}\|\mathbf{v}\|_M^2\right],$$

where $\mathbf{v}$ *is the infimal displacement vector of* $\mathbb{T}$*.*

Proof. The key step of proving Lemma 4.7 is to bound the last two terms in Lemma 4.3. To achieve this, rewrite the terms as

$$\begin{aligned}
&-\alpha\theta(1-\alpha\theta)\|\mathbb{S}x - \mathbb{S}z\|_M^2 + \theta^2\left(\beta - \alpha^2\right)\|\mathbb{S}x\|_M^2 \\
&\quad = -\theta(\alpha - \beta\theta)\|\mathbb{S}x\|_M^2 + 2\alpha\theta(1-\alpha\theta)\langle \mathbb{S}x, \mathbb{S}z\rangle_M - \alpha\theta(1-\alpha\theta)\|\mathbb{S}z\|_M^2 \\
&\quad \le -\theta^{-1}\alpha(1-\alpha\theta)\|\mathbf{v}\|_M^2 + 2\alpha\theta(1-\alpha\theta)\langle \mathbb{S}x, \mathbb{S}z\rangle_M,
\end{aligned}$$

where the last inequality is from $\mathbf{v}$ being infimal displacement vector, i.e. $\|\mathbf{v}\|_M \le \|\theta\mathbb{S}x\|_M, \|\theta\mathbb{S}z\|_M$.

Now use Lemma 4.3 with x as x^k and z as z^k. Take a full expectation among $\mathcal{I}^0, \mathcal{I}^1, \dots, \mathcal{I}^{k-1}$, then we get

$$\begin{aligned}
&\mathbb{E}\left[\left\|x^k - z^k\right\|_M^2\right] \\
&\quad \le \mathbb{E}\left[\left\|x^{k-1} - z^{k-1}\right\|_M^2\right] - \theta^{-1}\alpha(1-\alpha\theta)\|\mathbf{v}\|_M^2 + 2\alpha\theta(1-\alpha\theta)\mathbb{E}\left[\left\langle \mathbb{S}x^{k-1}, \mathbb{S}z^{k-1}\right\rangle_M\right] \\
&\quad \le \mathbb{E}\left[\left\|x^{k-1} - z^{k-1}\right\|_M^2\right] - \theta^{-1}\alpha(1-\alpha\theta)\|\mathbf{v}\|_M^2 + 2\alpha\theta(1-\alpha\theta)\left\|\mathbb{E}\left[\mathbb{S}x^{k-1}\right]\right\|_M\left\|\mathbb{S}z^{k-1}\right\|_M \\
&\quad \le \mathbb{E}\left[\left\|x^{k-1} - z^{k-1}\right\|_M^2\right] - \theta^{-1}\alpha(1-\alpha\theta)\|\mathbf{v}\|_M^2 + 2\beta^{1/2}\theta(1-\alpha\theta)\left\|\mathbb{S}x^0\right\|_M\left\|\mathbb{S}z^0\right\|_M \\
&\quad \le \mathbb{E}\left[\left\|x^{k-1} - z^{k-1}\right\|_M^2\right] - \theta^{-1}\alpha(1-\alpha\theta)\|\mathbf{v}\|_M^2 + 2\sqrt{\alpha\theta}(1-\alpha\theta)\left\|\mathbb{S}x^0\right\|_M\left\|\mathbb{S}z^0\right\|_M,
\end{aligned}$$

where the third inequality uses Lemma 4.5 and Lemma 4.6. Note that the term

$$-\theta^{-1}\alpha(1-\alpha\theta)\|\mathbf{v}\|_M^2 + 2\sqrt{\alpha\theta}(1-\alpha\theta)\left\|\mathbb{S}x^0\right\|_M\left\|\mathbb{S}z^0\right\|_M$$

is independent from the random process and iterations. Thus, above inequality can be applied through iterations, resulting in

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

Divide both sides by k^2 to obtain the desired result

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right] \\
&\quad \le \frac{1}{k}\left(2\sqrt{\alpha\theta}(1-\alpha\theta)\left\|\mathbb{S}x^0\right\|_M\left\|\mathbb{S}z^0\right\|_M - \frac{\alpha}{\theta}(1-\alpha\theta)\|\mathbf{v}\|_M^2\right) + \frac{1}{k^2}\left\|x^0 - z^0\right\|_M^2.
\end{aligned}$$

$\square$

Proof of and Theorem 4.4. Since the $M = \mathbb{I}$, we have $\beta = \alpha \le \alpha/\theta$ due to Lemma 4.1. Thus, we may apply Lemma 4.7 with $z^0 = x^0$.

$$\mathbb{E}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|^2\right] \le \frac{1}{k}\left(2\sqrt{\alpha\theta}\,(1-\alpha\theta)\left\|\mathbb{S}x^0\right\|^2 - \frac{\alpha}{\theta}\,(1-\alpha\theta)\left\|\mathbf{v}\right\|^2\right).$$

When the limit $k \to \infty$ is taken,

$$\lim_{k\to\infty} \mathbb{E}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|^2\right] = 0, \quad \lim_{k\to\infty}\left\|\frac{z^k}{k} + \alpha\mathbf{v}\right\| = 0,$$

where the second equation is from Theorem 2.1. These two limits provide L^2 convergence of normalized iterate, namely

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

as $k \to \infty$. $\square$

4.4. Almost sure convergence of normalized iterate

Theorem 4.8. *Under the conditions of Theorem 4.4 with $\theta \in (0,1)$, x^k/k is strongly convergent to $-\alpha\mathbf{v}$ in probability* 1*. In other words,*

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

as $k \to \infty$.

While the full proof is presented in the next subsection, here is the outline of the proof of the theorem. Let $z^0, z^1, z^2, \dots$ be the iterates of (FPI with $\bar{\mathbb{T}}$). Assume $\beta < \alpha/\theta$. From Lemma 4.3 and further analysis in Section 4.5, we obtain

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

for $k = 2, 3, \dots$, where $B = \alpha\theta^2(1-\alpha\theta)(\beta-\alpha^2)/(\alpha-\beta\theta) \ge 0$ and $\mathcal{F}_k$ is a filtration consisting of information up to the kth iterate.

We then apply the Robbins–Siegmund quasi-martingale theorem [105], restated as Lemma 4.9, to conclude that the random sequence $\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2$ converges almost surely to some random variable. Then, by Fatou's lemma and the L^2 convergence of Theorem 4.4, we have

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

Thus, as $k \to \infty$,

$$\left\| \frac{x^k}{k} - \frac{z^k}{k} \right\|_M^2 \overset{\text{a.s.}}{\to} 0,$$

and appealing to Theorem 2.1, we conclude the almost sure convergence. Finally, in the case of $\|\cdot\|$-norm, the assumption $\beta < \alpha/\theta$ is satisfied by Lemma 4.1. We defer the detailed proof to Section 4.5.

4.5. Proof of Theorem 4.8

First, let's recall the Robbins-Siegmund quasi-martingale theorem [105].

Lemma 4.9 (Robbins and Siegmund [105]). *$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ 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*

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

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

Now, we present a proof of Theorem 4.8 with the norm of $\|\cdot\|_M$.

Lemma 4.10. *Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be θ-averaged with respect to $\|\cdot\|_M$ with $\theta \in (0, 1]$, and let $x^0, x^1, x^2, \dots$ be the iterates of* (RC-FPI) *and let $z^0, z^1, z^2, \dots$ be the iterates of* (FPI with $\bar{\mathbb{T}}$). *Assume that the distribution of $\mathcal{I}$ satisfies the uniform expected step-size condition* (2) *with $\alpha \in (0, 1]$ and* (3) *holds with some β that $\beta < \alpha/\theta$. 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 $\mathbb{T}$.

Proof. To use the Robbins-Siegmund quasi-martingale theorem Lemma 4.9, we cannot take full expectation to bound the extra terms in Lemma 4.3. Here, we provide alternate way to bound the last two terms in Lemma 4.3.

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

Note that this inequality only holds when $\beta < \alpha/\theta$. Also, $\beta - \alpha^2 \ge 0$ comes from $\|\mathbb{E}\left[u_{\mathcal{I}}\right]\|_M^2 \le \mathbb{E}\left[\|u_{\mathcal{I}}\|_M^2\right]$.

From Lemma 4.3,

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

where $x^0, x^1, x^2, \dots$ is a random sequence generated by (RC-FPI) with $\mathbb{T}$, $z^0, z^1, z^2, \dots$ is a sequence generated by (FPI with $\bar{\mathbb{T}}$) and starting point $z^0 = x^0$, and $\mathcal{F}_k$ is a filtration consisting of information up to nth iteration.

With $\left\|\mathbb{S}z^{k-1}\right\|_M \le \left\|\mathbb{S}z^0\right\|_M$ from the Lemma 4.5,

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

Now we may apply the Robbins-Siegmund quasi-martingale theorem, Lemma 4.9, and conclude that the random sequence $\left\|\frac{x^k}{k}-\frac{z^k}{k}\right\|_M^2$ converges almost surely to some random variable since $\sum_{n=1}^{\infty} n^{-2}B\left\|\mathbb{S}z^0\right\|_M^2 < \infty$. Then,

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

holds, where the inequality comes from the Fatou's lemma and the equality comes from L^2 convergence by taking $k \to \infty$ in Lemma 4.7. Thus,

$$\lim_{k\to\infty}\left\|\frac{x^k}{k}-\frac{z^k}{k}\right\|_M^2 = 0, \qquad a.s.$$

happens, which, with the strong convergent z^k/k to $-\alpha\mathbf{v}$ by Theorem 2.1, gives the almost sure convergence of Lemma 4.10. □

Proof of Theorem 4.8. In the case of $M = \mathbb{I}$, with $\theta \in (0,1)$, we have $\beta = \alpha < \alpha/\theta$ from Lemma 4.1. Thus, from Lemma 4.10, we can conclude

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

as $k \to \infty$. □

4.6. Infeasibility detection.

Since $\mathbf{v} \neq 0$ implies the problem is inconsistent and $x^k/k \to -\alpha\mathbf{v}$, we propose

$$\frac{1}{k}\|x^k\| > \varepsilon \tag{4}$$

as a test of inconsistency with sufficiently large $k \in \mathbb{N}$ and sufficiently small $\varepsilon > 0$. The remaining question of how to choose the iteration count k and threshold ε will be considered in Section 6. (This test is not able to detect inconsistency in the pathological case where the problem is inconsistent despite $\mathbf{v} = 0$.)

5. Bias and variance of normalized iterates

Previously in Section 4, we showed that the normalized iterate x^k/k of (RC-FPI) converges to the scaled infimal displacement vector $-\alpha\mathbf{v}$. However, to use x^k/k as an estimator of $-\alpha\mathbf{v}$ and to use $x^k/k \neq 0$ as a test for inconsistency, we need to characterize the error $\|x^k/k+\alpha\mathbf{v}\|^2$. In this section, we provide an asymptotic upper bound of the bias and variance of x^k/k as an estimator of $-\alpha\mathbf{v}$.

Theorem 5.1. *Let $\mathbb{T}\colon \mathcal{H}\to\mathcal{H}$ be θ-averaged with respect to $\|\cdot\|_M$ with $\theta\in(0,1]$. Let $\mathbf{v}$ be the infimal displacement vector of $\mathbb{T}$. Assume $\mathcal{I}^0,\mathcal{I}^1,\dots$ is sampled IID from a distribution satisfying the uniform expected step-size condition* (2) *with $\alpha\in(0,1]$, and assume* (3) *holds with some $\beta>0$ such that $\beta<\alpha/\theta$. Let $x^0,x^1,x^2,\dots$ be the iterates of* (RC-FPI).

(a) If $\mathbf{v}\in\operatorname{range}(\mathbb{I}-\mathbb{T})$, then as $k\to\infty$,

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

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

$$\operatorname{Var}_M\left(\frac{x^k}{k}\right) \lesssim \frac{(\beta-\alpha^2)\|\mathbf{v}\|_M^2}{k}$$

as $k\to\infty$.

To clarify, the precise meaning of the first asymptotic statement of (a) is

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

The precise meaning of the asymptotic statement of (b) is defined similarly.

Here, we outline the proof for the special case, $\mathbf{v}\in\operatorname{range}(\mathbb{I}-\mathbb{T})$, while deferring the full proof to Section 5.1. Let $z^0,z^1,z^2,\dots$ be the iterates of (FPI with $\bar{\mathbb{T}}$) with z^0 satisfying $\theta\mathbb{S}z^0=\mathbf{v}$. Then, $\theta\mathbb{S}z^k=\mathbf{v}$ for all $k\in\mathbb{N}$. Apply Lemma 4.3 on x^k and z^k and take full expectation to get

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

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

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

The key idea is to bound the U^k terms. This can be done by showing that $\mathbf{v}$ and $\theta\mathbb{S}x^k-\mathbf{v}$ are almost orthogonal. Such property is exhibited in the next two lemmas.

Lemma 5.2. *Suppose θ-averaged operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ has an infimal displacement vector $\mathbf{v}$. Consider a closed cone C_δ in $\mathcal{H}$ with $\delta \in (0, \pi/2)$, which is a set of vectors whose angle between them and $\mathbf{v}$ being less than $\pi/2 - \delta$.*

$$C_\delta = \left\{x : \langle \mathbf{v}, x \rangle_M \geq \sin\delta \left\| \mathbf{v} \right\|_M \left\| x \right\|_M \right\}.$$

When the points $y, z \in \mathcal{H}$ satisfy that $\mathbb{S}y \in \mathbb{S}z + C_\delta$ and $\mathbb{S}y \neq \mathbb{S}z$, then the following inequality holds.

$$\langle -\mathbf{v}, y - z \rangle_M \leq \cos\delta \left\| \mathbf{v} \right\|_M \left\| y - z \right\|_M.$$

Proof. Since the inequality holds if $\mathbf{v} = 0$, let's assume that $\mathbf{v} \neq 0$.

Define x as $y - z$, u as $\mathbb{S}y - \mathbb{S}z$. Since $u \in C_\delta$, there exist $\phi \in [\delta, \pi/2]$ such that

$$\langle \mathbf{v}, u \rangle_M = \sin\phi \left\| \mathbf{v} \right\|_M \left\| u \right\|_M.$$

Due to cocoersivity of $\mathbb{S}$, x and u must satisfy

$$\langle x, u \rangle_M = \langle y - z, \mathbb{S}y - \mathbb{S}z \rangle_M \geq 0.$$

Since $\mathbf{v}$ is nonzero, decompose x and u as

$$x = \langle x, \mathbf{v} \rangle_M \frac{\mathbf{v}}{\left\| \mathbf{v} \right\|_M^2} + \mathbf{v}_x^\perp, \quad u = \langle u, \mathbf{v} \rangle_M \frac{\mathbf{v}}{\left\| \mathbf{v} \right\|_M^2} + \mathbf{v}_u^\perp.$$

Both $\mathbf{v}_x^\perp$ and $\mathbf{v}_u^\perp$ are orthogonal to $\mathbf{v}$ and

$$\left\| \mathbf{v}_u^\perp \right\|_M = \cos\phi \left\| u \right\|_M, \quad \left\| \mathbf{v}_x^\perp \right\|_M^2 = \left\| x \right\|_M^2 - \left(\frac{\langle x, \mathbf{v} \rangle_M}{\left\| \mathbf{v} \right\|_M} \right)^2.$$

Compute $\langle x, u \rangle$ using the decomposition above,

$$\begin{aligned}
\langle x, u \rangle_M &= \frac{1}{\left\| \mathbf{v} \right\|_M^2} \langle x, \mathbf{v} \rangle_M \langle u, \mathbf{v} \rangle_M + \left\langle \mathbf{v}_x^\perp, u \right\rangle_M \\
&= \frac{1}{\left\| \mathbf{v} \right\|_M} \langle x, \mathbf{v} \rangle_M \sin\phi \left\| u \right\|_M + \left\langle \mathbf{v}_x^\perp, \mathbf{v}_u^\perp \right\rangle_M \\
&\leq \frac{1}{\left\| \mathbf{v} \right\|_M} \langle x, \mathbf{v} \rangle_M \sin\phi \left\| u \right\|_M + \cos\phi \left\| u \right\|_M \sqrt{\left\| x \right\|_M^2 - \left(\frac{\langle x, \mathbf{v} \rangle}{\left\| \mathbf{v} \right\|_M} \right)^2}.
\end{aligned}$$

If $\langle x, -\mathbf{v} \rangle_M \leq 0$, then $\langle -\mathbf{v}, y - z \rangle_M \leq 0 \leq \cos\delta \left\| \mathbf{v} \right\|_M \left\| y - z \right\|_M$ and the conclusion holds. Thus, consider only the case where $\langle x, -\mathbf{v} \rangle_M > 0$. In such case, $\langle x, u \rangle_M \geq 0$ with $\left\| u \right\|_M > 0$ gives

$$0 < \frac{1}{\left\| \mathbf{v} \right\|_M} \langle x, -\mathbf{v} \rangle_M \sin\phi \leq \cos\phi \sqrt{\left\| x \right\|_M^2 - \left(\frac{\langle x, -\mathbf{v} \rangle_M}{\left\| \mathbf{v} \right\|_M} \right)^2},$$

which leads to a conclusion by squaring each sides :

$$\langle x, -\mathbf{v} \rangle_M \leq \cos\phi \left\| \mathbf{v} \right\|_M \left\| x \right\|_M \leq \cos\delta \left\| \mathbf{v} \right\|_M \left\| x \right\|_M.$$

□

Lemma 5.3. *Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be a θ-averaged operator with respect to $\|\cdot\|_M$. Let $\mathbf{v}$ be the infimal displacement vector of $\mathbb{T}$. Let $\mathbb{S} = \theta^{-1}(\mathbb{I} - \mathbb{T})$. Consider a sequence $y^0, y^1, y^2, \ldots$ in $\mathcal{H}$ such that its normalized iterate converges strongly to $-\gamma\mathbf{v}$,*

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

for some $\gamma > 0$. Then, for any $\delta \in (0, \pi/2)$ and $z \in \mathcal{H}$, there exists $N_{\delta,z} \in \mathbb{N}$ such that, for all $k > N_{\delta,z}$,

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

Proof. Choose a point z in $\mathcal{H}$. To prove by contradiction, suppose that for any l, there exists $k_l > l$ such that

$$\mathbb{S}y^{k_l} \in \mathbb{S}z + C_\delta, \quad \mathbb{S}y^{k_l} \ne \mathbb{S}z.$$

The subsequence $y^{k_1}, y^{k_2}, y^{k_3}, \ldots$ satisfies the inequality below for all l, due to Lemma 5.2.

$$\left\langle -\mathbf{v}, y^{k_l} - z \right\rangle_M \le \cos\delta \|\mathbf{v}\|_M \left\|y^{k_l} - z\right\|_M.$$

Divide each side by k_l and take a limit as $l \to \infty$. Since $\lim_{l\to\infty} y^{k_l}/k_l = -\gamma\mathbf{v}$ strongly,

$$\gamma \|\mathbf{v}\|_M^2 = \langle -\mathbf{v}, -\gamma\mathbf{v}\rangle_M \le \cos\delta \|\mathbf{v}\|_M \|-\gamma\mathbf{v}\|_M < \gamma \|\mathbf{v}\|_M^2,$$

which yields a contradiction.

Thus, when z is given, for any $\delta \in (0, \pi/2)$, there exist a $N_{\delta,z}$ such that for all $k > N_{\delta,z}$, it is either $\mathbb{S}y^k = \mathbb{S}z$ or $\mathbb{S}y^k \notin \mathbb{S}z + C_\delta$. As a conclusion, for all $k > N_{\delta,z}$,

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

□

Returning to the proof outline of Theorem 5.1, by the Inequality 1 and Lemma 5.3,

$$0 \le \left\langle \mathbf{v}, \theta\mathbb{S}x^k - \mathbf{v} \right\rangle_M \le \|\mathbf{v}\|_M \left\|\theta\mathbb{S}x^k - \mathbf{v}\right\|_M \sin\delta \approx 0$$

for small δ. Therefore, for k large enough,

$$\left\|\theta\mathbb{S}x^k\right\|_M^2 \lesssim \|\mathbf{v}\|_M^2 + \left\|\theta\mathbb{S}x^k - \mathbf{v}\right\|_M^2.$$

Since $\theta\mathbb{S}x^k \to \mathbf{v}$ as $k \to \infty$, we have

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

Finally, we conclude

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

Note that for (b), $k\mathrm{Var}_M\left(x^k/k\right)$ can be bounded by

$$k\mathrm{Var}_M\left(\frac{x^k}{k}\right) \le \mathbb{E}\left[k \left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right],$$

and so is (a) with an extra $\mathcal{O}\left(1/\sqrt{k}\right)$ term.

5.1. Proof of Theorem 5.1

Proof. When $x^0, x^1, x^2, \dots$ is a random sequence generated by RC-FPI of $\mathbb{T}$ and $z^0, z^1, z^2, \dots$ is a sequence generated by FPI of $\bar{\mathbb{T}}$ with $z^0 = z$, from Lemma 4.3, it is already known that for all k,

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

For convenience, define

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

Consequently, by taking a full expectation,

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

The key of this proof is to bound the term $\left\|\mathbb{S}x^k\right\|_M^2$ in U^k using Lemma 5.3, since x^k/k is strongly convergent, i.e., $\lim_{k\to\infty} x^k/k = -\alpha\mathbf{v}$ almost surely. Suppose the case where $\lim_{k\to\infty} x^k/k = -\alpha\mathbf{v}$ holds, which actually does hold almost surely. For such x^k, for an arbitrary $\delta \in (0, \pi/2)$, there exists a $N_{\delta,z}$ such that for all $k > N_{\delta,z}$,

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

From the inequality above, for all sufficiently large $k > N_{\delta,z}$,

$$\begin{aligned}
\left\|\mathbb{S}x^k\right\|_M^2 &= \left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 + 2\left\langle \mathbb{S}x^k - \mathbb{S}z, \mathbb{S}z\right\rangle_M + \left\|\mathbb{S}z\right\|_M^2 \\
&= \left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 + 2\left\langle \mathbb{S}x^k - \mathbb{S}z, \mathbb{S}z - \frac{1}{\theta}\mathbf{v}\right\rangle_M + 2\left\langle \mathbb{S}x^k - \mathbb{S}z, \frac{1}{\theta}\mathbf{v}\right\rangle_M + \left\|\mathbb{S}z\right\|_M^2 \\
&\le \left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 + 2\left\{\left\|\mathbb{S}z - \frac{1}{\theta}\mathbf{v}\right\|_M + \sin\delta\left\|\frac{1}{\theta}\mathbf{v}\right\|_M\right\}\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M + \left\|\mathbb{S}z\right\|_M^2.
\end{aligned}$$

By the inequality above, the term U^k can be bounded as :

$$\begin{aligned}
U^k &\le -\left(\alpha\theta - \alpha^2\theta^2\right)\left\|\mathbb{S}x^k - \mathbb{S}z^k\right\|_M^2 \\
&\quad + \theta^2\left(\beta-\alpha^2\right)\left[\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 + 2\left\{\left\|\mathbb{S}z - \frac{1}{\theta}\mathbf{v}\right\|_M + \sin\delta\left\|\frac{1}{\theta}\mathbf{v}\right\|_M\right\}\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M + \left\|\mathbb{S}z\right\|_M^2\right] \\
&= \theta^2\left(\beta-\alpha^2\right)\left\|\mathbb{S}z\right\|_M^2 - \theta\left(\alpha - \beta\theta\right)\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 \\
&\quad + 2\theta^2\left(\beta-\alpha^2\right)\left\{\left\|\mathbb{S}z - \frac{1}{\theta}\mathbf{v}\right\|_M + \sin\delta\left\|\frac{1}{\theta}\mathbf{v}\right\|_M\right\}\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M \\
&\quad + \left(\alpha\theta - \alpha^2\theta^2\right)\left\{\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 - \left\|\mathbb{S}x^k - \mathbb{S}z^k\right\|_M^2\right\}.
\end{aligned}$$

Here, the term $\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 - \left\|\mathbb{S}x^k - \mathbb{S}z^k\right\|_M^2$ is bounded above by

$$\begin{aligned}
&\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 - \left\|\mathbb{S}x^k - \mathbb{S}z^k\right\|_M^2 \\
&\quad \leq \left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 - \left(\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 + \left\|\mathbb{S}z - \mathbb{S}z^k\right\|_M^2 + 2\left\langle \mathbb{S}x^k - \mathbb{S}z, \mathbb{S}z - \mathbb{S}z^k\right\rangle_M\right) \\
&\quad \leq -\left\|\mathbb{S}z - \mathbb{S}z^k\right\|_M^2 + 2\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M \left\|\mathbb{S}z - \mathbb{S}z^k\right\|_M,
\end{aligned}$$

from the triangular inequality.

Thus, we can bound U^k as

$$\begin{aligned}
U^k &\leq \theta^2\left(\beta - \alpha^2\right)\|\mathbb{S}z\|_M^2 - \theta\left(\alpha - \beta\theta\right)\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 \\
&\quad + 2\theta^2\left(\beta - \alpha^2\right)\left\{\left\|\mathbb{S}z - \frac{1}{\theta}\mathbf{v}\right\|_M + \sin\delta\left\|\frac{1}{\theta}\mathbf{v}\right\|_M\right\}\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M \\
&\quad + \left(\alpha\theta - \alpha^2\theta^2\right)\left\{-\left\|\mathbb{S}z - \mathbb{S}z^k\right\|_M^2 + 2\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M\left\|\mathbb{S}z - \mathbb{S}z^k\right\|_M\right\} \\
&= \theta^2\left(\beta - \alpha^2\right)\|\mathbb{S}z\|_M^2 - \left(\alpha\theta - \alpha^2\theta^2\right)\left\|\mathbb{S}z - \mathbb{S}z^k\right\|_M^2 \\
&\quad - \theta\left(\alpha - \beta\theta\right)\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 + 2\theta\tau_{\delta,z,k}\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M \\
&\leq \theta^2\left(\beta - \alpha^2\right)\|\mathbb{S}z\|_M^2 \\
&\quad - \theta\left(\alpha - \beta\theta\right)\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M^2 + 2\theta\tau_{\delta,z,k}\left\|\mathbb{S}x^k - \mathbb{S}z\right\|_M,
\end{aligned}$$

where $\tau_{\delta,z,k}$ is defined as

$$\tau_{\delta,z,k} = \left(\beta - \alpha^2\right)\left(\|\theta\mathbb{S}z - \mathbf{v}\|_M + \sin\delta\|\mathbf{v}\|_M\right) + \left(\alpha - \alpha^2\theta\right)\left\|\mathbb{S}z - \mathbb{S}z^k\right\|_M.$$

To make an upper bound of $\tau_{\delta,z,k}$ regardless of k, an upper bound of $\left\|\mathbb{S}z - \mathbb{S}z^k\right\|_M$ independent from k is required. Since $\mathbf{v}$ is the infimal displacement vector,

$$\left\langle \mathbb{S}z^k - \frac{1}{\theta}\mathbf{v}, \mathbf{v}\right\rangle_M \geq 0, \quad \|\mathbb{S}z\|_M^2 - \left\|\frac{1}{\theta}\mathbf{v}\right\|_M^2 \geq 0,$$

hold. With $\left\|\mathbb{S}z^k\right\|_M \leq \|\mathbb{S}z\|_M$ from Lemma 4.5, such uniform upper bound can be built as

$$\begin{aligned}
\left\|\mathbb{S}z - \mathbb{S}z^k\right\|_M &\leq \left\|\mathbb{S}z - \frac{1}{\theta}\mathbf{v}\right\|_M + \left\|\mathbb{S}z^k - \frac{1}{\theta}\mathbf{v}\right\|_M \\
&= \left\|\mathbb{S}z - \frac{1}{\theta}\mathbf{v}\right\|_M + \sqrt{\|\mathbb{S}z^k\|_M^2 - 2\left\langle \mathbb{S}z^k, \frac{1}{\theta}\mathbf{v}\right\rangle_M + \left\|\frac{1}{\theta}\mathbf{v}\right\|_M^2} \\
&= \left\|\mathbb{S}z - \frac{1}{\theta}\mathbf{v}\right\|_M + \sqrt{\|\mathbb{S}z^k\|_M^2 - 2\left\langle \mathbb{S}z^k - \frac{1}{\theta}\mathbf{v}, \frac{1}{\theta}\mathbf{v}\right\rangle_M - \left\|\frac{1}{\theta}\mathbf{v}\right\|_M^2} \\
&\leq \left\|\mathbb{S}z - \frac{1}{\theta}\mathbf{v}\right\|_M + \sqrt{\|\mathbb{S}z\|_M^2 - \left\|\frac{1}{\theta}\mathbf{v}\right\|_M^2}.
\end{aligned}$$

Now define $\tilde{\tau}_{\delta,z}$ as

$$\tilde{\tau}_{\delta,z} = \left(\beta - \alpha^2\right)\left(\|\theta \mathbb{S} z - \mathbf{v}\|_M + \sin\delta \|\mathbf{v}\|_M\right) + \left(\alpha - \alpha^2\theta\right)\left\{\left\|\mathbb{S} z - \frac{1}{\theta}\mathbf{v}\right\|_M + \sqrt{\|\mathbb{S} z\|_M^2 - \left\|\frac{1}{\theta}\mathbf{v}\right\|_M^2}\right\},$$

then we have $\tau_{\delta,z,k} \leq \tilde{\tau}_{\delta,z}$ for any $k \in \mathbb{N}$.

Thus, we can bound U^k as

$$\begin{aligned} U^k &\leq \theta^2\left(\beta - \alpha^2\right)\|\mathbb{S} z\|_M^2 - \theta\left(\alpha - \beta\theta\right)\left\|\mathbb{S} x^k - \mathbb{S} z\right\|_M^2 + 2\theta\tau_{\delta,z,k}\left\|\mathbb{S} x^k - \mathbb{S} z\right\|_M \\ &\leq \theta^2\left(\beta - \alpha^2\right)\|\mathbb{S} z\|_M^2 - \theta\left(\alpha - \beta\theta\right)\left\|\mathbb{S} x^k - \mathbb{S} z\right\|_M^2 + 2\theta\tilde{\tau}_{\delta,z}\left\|\mathbb{S} x^k - \mathbb{S} z\right\|_M. \end{aligned}$$

From the fact that $-at^2 + 2bt \leq b^2/a$ for any $a, b > 0$ and $t \in \mathbb{R}$, U^k has an upper bound which is completely independent from x^k and k.

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

However, this upper bound holds only at $k > N_{\delta,z}$. Since $N_{\delta,z}$ depends on the choice of the sequence $x^0, x^1, x^2, \ldots$, such upper bound only works when the sequence $x^0, x^1, x^2, \ldots$ is fixed. To avoid this problem, take a limit supremum of U^k over k,

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

Furthermore, due to Cesàro mean,

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

Now, this inequality always holds with only one condition on the choice of sequence $x^0, x^1, x^2 \ldots$,

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

regardless of z, δ. Since $\lim_{k\to\infty} x^k/k = -\alpha\mathbf{v}$ holds almost surely,

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

holds almost surely for any z, δ.

By Fatou's lemma, we also have

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

Thus, $\limsup_{k\to\infty} \mathbb{E}\left[k\left\|\frac{x^k}{k}-\frac{z^k}{k}\right\|_M^2\right]$ has a upper bound of

$$\begin{aligned}\limsup_{k\to\infty} \mathbb{E}\left[k\left\|\frac{x^k}{k}-\frac{z^k}{k}\right\|_M^2\right] &\leq \limsup_{k\to\infty} \mathbb{E}\left[\frac{1}{k}\left\|x^0-z\right\|_M^2+\frac{1}{k}\sum_{j=0}^{k-1}U^j\right] \\ &\leq \theta^2\left(\beta-\alpha^2\right)\left\|\mathbb{S}z\right\|_M^2+\frac{\theta}{\alpha-\beta\theta}\tilde{\tau}_{\delta,z}^2.\end{aligned} \tag{5}$$

Proof of statement (b). First, let's prove the statement (b) of Theorem 5.1. Since

$$\limsup_{k\to\infty} k\mathrm{Var}_M\left(\frac{x^k}{k}\right) \leq \limsup_{k\to\infty} \mathbb{E}\left[k\left\|\frac{x^k}{k}-\frac{z^k}{k}\right\|_M^2\right],$$

from (5) we have

$$\limsup_{k\to\infty} k\mathrm{Var}_M\left(\frac{x^k}{k}\right) \leq \theta^2\left(\beta-\alpha^2\right)\left\|\mathbb{S}z\right\|_M^2+\frac{\theta}{\alpha-\beta\theta}\tilde{\tau}_{\delta,z}^2.$$

At start, we chose z and δ arbitrarily. Since $\mathbf{v}$ is the infimal displacement vector, there exists a sequence of z's that allows us to take a limit $\left\|\mathbb{S}z-\theta^{-1}\mathbf{v}\right\|_M\to 0$. When we take a limit $\left\|\mathbb{S}z-\theta^{-1}\mathbf{v}\right\|_M\to 0$ and $\delta\to 0$,

$$\lim_{\left\|\mathbb{S}z-\frac{1}{\theta}\mathbf{v}\right\|_M\to 0,\delta\to 0} \tilde{\tau}_{\delta,z}=0.$$

Since $\limsup_{k\to\infty} k\mathrm{Var}_M\left(x^k/k\right)$ is independent from δ and z, by $\left\|\mathbb{S}z-\theta^{-1}\mathbf{v}\right\|_M\to 0$ and $\delta\to 0$ we have

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

Proof of statement (a). Next, to prove the statement (a) of Theorem 5.1, let's start again from inequality (5),

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

Expand the term $\left\|\frac{x^k}{k}+\alpha\mathbf{v}\right\|_M^2$ as

$$\mathbb{E}\left[\left\|\frac{x^k}{k}+\alpha\mathbf{v}\right\|_M^2\right]=\mathbb{E}\left[\left\|\frac{x^k}{k}-\frac{z^k}{k}\right\|_M^2\right]+2\left\langle\mathbb{E}\left[\frac{x^k}{k}-\frac{z^k}{k}\right],\frac{z^k}{k}+\alpha\mathbf{v}\right\rangle_M+\left\|\frac{z^k}{k}+\alpha\mathbf{v}\right\|_M^2.$$

From Lemma 4.7 we have

$$\begin{aligned}\left\|\mathbb{E}\left[\frac{x^k}{k}-\frac{z^k}{k}\right]\right\|_M^2 &\leq \mathbb{E}\left[\left\|\frac{x^k}{k}-\frac{z^k}{k}\right\|_M^2\right] \\ &\leq \frac{1}{k}(1-\alpha\theta)\left[2\sqrt{\alpha\theta}\left\|\mathbb{S}x^0\right\|_M\left\|\mathbb{S}z^0\right\|_M-\frac{\alpha}{\theta}\left\|\mathbf{v}\right\|_M^2\right]+\frac{1}{k^2}\left\|x^0-z^0\right\|_M^2.\end{aligned}$$

When $x_\star$ is a point such that $x_\star - \mathbb{T}x_\star = v$, set $z^0 = x_\star$. Then,

$$z^k = -k\alpha\mathbf{v} + x_\star,$$

since $\left\|\theta\mathbb{S}z^k\right\|_M \le \|\theta\mathbb{S}z\|_M = \|\mathbf{v}\|_M$ makes $\theta\mathbb{S}z^k = \mathbf{v}$ for all $k \in \mathbb{N}$. Thus,

$$\mathbb{E}\left[\left\|\frac{x^k}{k} + \alpha\mathbf{v}\right\|_M^2\right]^2 \le \mathbb{E}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right]$$
$$+ 2\frac{1}{\sqrt{k}}\sqrt{(1-\alpha\theta)\left[2\sqrt{\alpha\theta}\,\|\mathbb{S}x^0\|_M\,\|\mathbb{S}x_\star\|_M - \frac{\alpha}{\theta}\|\mathbf{v}\|_M^2\right] + \frac{1}{k}\|x^0 - z^0\|_M^2}\left\|\frac{x_\star}{k}\right\|_M + \left\|\frac{x_\star}{k}\right\|_M^2.$$

Note that the last two terms are $\mathcal{O}\left(k^{-3/2}\right)$. By taking lim sup as $k \to \infty$,

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

Thus,

$$\limsup_{k\to\infty}\mathbb{E}\left[k\left\|\frac{x^k}{k} + \alpha\mathbf{v}\right\|_M^2\right] \le \theta^2\left(\beta - \alpha^2\right)\|\mathbb{S}x_\star\|_M^2 + \frac{\theta}{\alpha - \beta\theta}\tilde{\tau}_{\delta,x_\star}^2.$$

Since $\theta\mathbb{S}x_\star = \mathbf{v}$, by $\delta \to 0$, we have $\tilde{\tau}_{\delta,x_\star} \to 0$ and

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

□

5.2. Tightness of variance bounds

In this section, we provide examples for which the variance bound of Theorem 5.1 holds with equality and with a strict inequality. We then discuss how the geometry of range $(\mathbb{I} - \mathbb{T})$ influences the tightness of the inequality. Throughout this section, we consider the setting where the norm and inner product is $\|\cdot\|$-norm and $\langle\cdot,\cdot\rangle$, with $\mathcal{H} = \mathbb{R}^m$, $\mathcal{H}_i = \mathbb{R}$, and $\mathcal{I}$ follows uniform distribution on the set of standard unit vectors of $\mathcal{H}$. In this case, the smallest β we can choose is $\alpha = 1/m$.

5.2.1. Example: Theorem 5.1(b) holds with equality.

Consider the translation operator $\mathbb{T}(x) = x - \mathbf{v}$. When $x^0, x^1, x^2, \dots$ are the iterates of (RC-FPI) with $\mathbb{T}$, then

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

for $k = 1, 2, \dots$, and the variance bound of Theorem 5.1 holds with equality.

5.2.2. Example: Theorem 5.1(b) holds with strict inequality.

Define $\mathbb{T}\colon \mathbb{R}^2 \to \mathbb{R}^2$ as

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

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

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

$$\limsup_{k\to\infty}\ k\mathrm{Var}_M\left(\frac{(x^k, y^k)}{k}\right) = \frac{1}{24}. \tag{6}$$

On the other hand, the right hand side of the inequality in Theorem 5.1 (b) is

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

Proof of Equation (6). An $(1/2)-$averaged operator $\mathbb{T}$ in $\mathbb{R}^2$ is defined as,

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

with the infimal displacement vector $\mathbf{v}$ of range $\mathbb{I} - \mathbb{T}$ as $(1/2, 1/2)$. The RC-FPI by $\mathbb{T}$ with the distribution as a uniform distribution on $\{(1,0),(0,1)\}$. The random coordinate operators are respectively,

$$\mathbb{T}_{(1,0)} : (x,y) \mapsto \left(x - \frac{1+x-y}{2}, y\right),$$
$$\mathbb{T}_{(0,1)} : (x,y) \mapsto \left(x, y - \frac{1+y-x}{2}\right).$$

When we set the initial point (x^0, y^0) as the origin, from the relations

$$\mathbb{E}\left[x^{k+1}\right] = \mathbb{E}\left[x^k\right] - \frac{1}{4} - \frac{1}{4}\mathbb{E}\left[x^k - y^k\right],$$
$$\mathbb{E}\left[y^{k+1}\right] = \mathbb{E}\left[y^k\right] - \frac{1}{4} + \frac{1}{4}\mathbb{E}\left[x^k - y^k\right],$$
$$\mathbb{E}\left[x^{k+1} - y^{k+1}\right] = \frac{1}{2}\mathbb{E}\left[x^k - y^k\right],$$

each expectations have a value of $\mathbb{E}\left[x^k\right] = \mathbb{E}\left[y^k\right] = -k/4$.

Next, an expectation $\mathbb{E}\left[\left\|x^k - y^k\right\|^2\right]$ has a recurrence relation of

$$\begin{aligned}\mathbb{E}\left[\left\|x^{k+1} - y^{k+1}\right\|^2\right] &= \mathbb{E}_{(x^k,y^k)}\mathbb{E}\left[\left\|x^{k+1} - y^{k+1}\right\|^2 \mid (x^k, y^k)\right] \\ &= \mathbb{E}_{(x^k,y^k)}\left[\frac{1}{4}\left(x^k - y^k\right)^2 + \frac{1}{4}\right],\end{aligned}$$

which obtains $\mathbb{E}\left[\left\|x^k - y^k\right\|^2\right] = \left(1 - 4^{-k}\right)/3$ as a solution.

Finally, an expectation $\mathbb{E}\left[\left\|x^k\right\|^2 + \left\|y^k\right\|^2\right]$ has a relation

$$
\begin{aligned}
\mathbb{E}\left[\left\|x^{k+1}\right\|^2 + \left\|y^{k+1}\right\|^2\right] &= \mathbb{E}_{\left(x^k, y^k\right)}\left[\frac{1}{2}\left(\left\|\mathbb{T}_{(1,0)}\left(x^k, y^k\right)\right\|^2 + \left\|\mathbb{T}_{(0,1)}\left(x^k, y^k\right)\right\|^2\right)\right] \\
&= \mathbb{E}\left[\left\|x^k\right\|^2 + \left\|y^k\right\|^2\right] + \frac{1}{4} - \frac{1}{2}\mathbb{E}\left[x^k + y^k\right] - \frac{1}{4}\mathbb{E}\left[\left\|x^k - y^k\right\|^2\right],
\end{aligned}
$$

which can be applied inductively, and as a result,

$$
\mathbb{E}\left[\left\|x^k\right\|^2 + \left\|y^k\right\|^2\right] = \frac{1}{8}k^2 + \frac{1}{24}k + \frac{1}{9}\left(1 - 4^{-k}\right).
$$

From above computations, a variance of $\left(x^k, y^k\right)$ can be estimated explicitly as

$$
\mathrm{Var}_M\left(x^k, y^k\right) = \frac{1}{24}k + \frac{1}{9}\left(1 - 4^{-k}\right).
$$

Thus, $\limsup_{k\to\infty} k\mathrm{Var}_M\left((x^k, y^k)/k\right)$ is

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

□

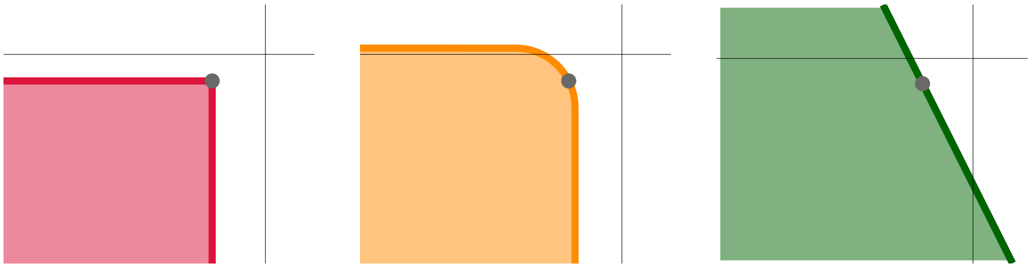
5.3. Relationship between the variance and the range set.

Consider the three convex sets A, B, and C in Figure 1 as a subset of $\mathcal{H} = \mathbb{R}^2$. The explicit definitions are

$$
\begin{aligned}
A &= \{(x, y) \mid x \le -10,\ y \le -5\} \\
B &= \left\{(x, y) \mid \operatorname{dist}((x, y), 3A) \le 2\sqrt{5^2 + 10^2}\right\} \\
C &= \{(x, y) \mid -2x - y \ge 25\},
\end{aligned}
$$

where $\operatorname{dist}((x, y), 3A)$ denotes the (Euclidean) distance of (x, y) to the set $3A = \{(3x, 3y) \mid (x, y) \in A\}$. The minimum norm elements in each set are all identically equal to $(-10, -5)$.

Let $\mathbb{T} = \mathbb{I} - \theta\mathrm{Proj}$, where Proj denotes the projections onto A, B, and C. Then $\mathbb{T}$ is θ-averaged and $\operatorname{range}\left(\theta^{-1}(\mathbb{I} - \mathbb{T})\right)$ is equal to A, B, and C, respectively. These sets are designed for $\mathbb{T}$ to have the same infimal displacement vector. Figure 2 (left), shows that the normalized iterates of the three instances have different asymptotic variances despite identical $\mathbf{v}$. In the experiment, θ was set as 0.2, and as a consequence, $\mathbf{v} = (-2, -1)$ is the infimal displacement vector for each experiments. (RC-FPI) is performed with $x^0 = (0, 0)$, $m = 2$ and $\mathcal{H}_1 = \mathcal{H}_2 = \mathbb{R}$.



(a) A (b) B (c) C

Figure 1: Visualization A, B, and C as defined in Section 5.2. The grey dot is $\theta^{-1}\mathbf{v}$, where $\mathbf{v}$ is the infimal displacement vector of $\mathbb{T} = \mathbb{I} - \theta\text{Proj}$.

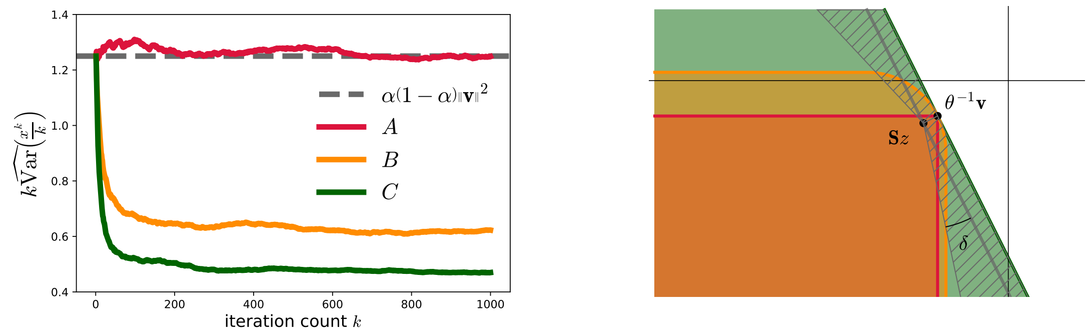


Figure 2: (Left) Graph of $k\widehat{\text{Var}}\left(x^k/k\right)$ by k, where $\widehat{\text{Var}}\left(x^k/k\right)$ is the variance estimate with 10,000 samples. (Right) Visualization of A, B, and C as red, yellow, green regions and $D_{z,\delta}$ as the hatched area, where the sets are as defined in Section 5.2. We conjecture that the broader intersection with $D_{z,\delta}$ leads to smaller asymptotic variance.

We conjecture that the asymptotic variance is intimately related to the geometry of the set $\mathrm{range}\,(\theta^{-1}(\mathbb{I}-\mathbb{T}))$. For $z \in \mathbb{R}^n$ and $\delta > 0$, let

$$D_{z,\delta} = \left\{ u \in \mathbb{R}^2 \middle| \langle \mathbf{v}, u - \mathbb{S}z \rangle \leq \|\mathbf{v}\| \, \|u - \mathbb{S}z\| \sin\delta \right\}.$$

Lemma 5.3 states that eventually, $\mathbb{S}x^k \in D_{z,\delta}$ for sufficiently large k. Since $\mathbb{S}x^k \in \mathrm{range}\,(\theta^{-1}(\mathbb{I}-\mathbb{T}))$ for all k, the shaded region in the Figure 2 (right) depicting $D_{z,\delta} \cap \mathrm{range}\,(\theta^{-1}(\mathbb{I}-\mathbb{T}))$ actually shows the region where $\mathbb{S}x^k$ lies for large k. In the proof of Theorem 5.1, loosely speaking, we establish the upper bound using

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

Therefore, the variance can be strictly smaller than the upper bound when $\left\| \mathbb{S}x^k - \theta^{-1}\mathbf{v} \right\|_M^2$ is large, which can happen when the area of intersection $D_{z,\delta} \cap \mathrm{range}\,(\theta^{-1}(\mathbb{I}-\mathbb{T}))$ is large near $\theta^{-1}\mathbf{v}$. This can be observed in Figure 2, which shows that the range set having large intersection with $D_{z,\delta}$ have smaller asymptotic variance.

6. Infeasibility detection

In this section, we present the infeasibility detection method for (RC-FPI) using the hypothesis testing.

Theorem 6.1. *Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be θ-averaged with respect to $\|\cdot\|_M$ with $\theta \in (0,1]$. Let $\mathbf{v}$ be the infimal displacement vector of $\mathbb{T}$. Assume $\mathcal{I}^0, \mathcal{I}^1, \dots$ is sampled IID from a distribution satisfying the uniform expected step-size condition* (2) *with $\alpha \in (0,1]$, and assume* (3) *holds with some $\beta > 0$ such that $\beta < \alpha/\theta$. Let $x^0, x^1, x^2, \dots$ be the iterates of* (RC-FPI)*. Then under the null hypothesis $\|\mathbf{v}\|_M \leq \delta$ with nonzero δ, for ϵ that satisfies $\alpha\delta < \epsilon$,*

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

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

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

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

with

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

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

For the proof of the Theorem 6.1, we begin with the simpler case where $\mathbf{v} \in \text{range}\,(\mathbb{I}-\mathbb{T})$. Let $\|\mathbf{v}\|_M \le \delta$ be the null hypothesis with δ satisfying $\alpha\delta < \epsilon$. By the triangle inequality, Markov inequality, and Theorem 5.1, under the null hypothesis,

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

as $k\to\infty$.

When $\mathbf{v} \notin \text{range}\,(\mathbb{I}-\mathbb{T})$, we can still obtain the same (asymptotic) statistical significance with the same test and the same iteration count $k \gtrsim \frac{(\beta-\alpha^2)\delta^2}{p(\varepsilon-\alpha\delta)^2}$. Below, we present the full proof of this general case.

Proof. First by the triangle inequality and Markov inequality,

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{x^k}{k}\right\|_M \ge \varepsilon\right) &\le \mathbb{P}\left(\left\|\frac{x^k}{k}-\mathbb{E}\left[\frac{x^k}{k}\right]\right\|_M \ge \varepsilon-\left\|\mathbb{E}\left[\frac{x^k}{k}\right]\right\|_M\right) \\
&\le \left(\varepsilon-\left\|\mathbb{E}\left[\frac{x^k}{k}\right]\right\|_M\right)^{-2}\text{Var}_M\left(\frac{x^k}{k}\right).
\end{aligned}$$

Next, let's bound the term $\left\|\mathbb{E}\left[x^k/k\right]\right\|_M$. Due to triangle inequality, Jensen's inequality and Lemma 4.7 with $z^0 = x^0$, we have

$$\left\|\mathbb{E}\left[\frac{x^k}{k}\right]\right\|_M \le \left\|\mathbb{E}\left[\frac{x^k}{k}\right]-\frac{z^k}{k}\right\|_M + \left\|\frac{z^k}{k}\right\|_M \le \mathcal{O}\left(\frac{1}{\sqrt{k}}\right)+\left\|\frac{z^k}{k}\right\|_M.$$

By [106, Theorem 3], for any $\omega>0$, there exist ω-dependent constant C_ω such that

$$\left\|\frac{z^k}{k}\right\|_M \le \alpha\|\mathbf{v}\|_M + \frac{1}{k}C_\omega + \omega.$$

Thus,

$$\left\|\mathbb{E}\left[\frac{x^k}{k}\right]\right\|_M \le \alpha\|\mathbf{v}\|_M + \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) + \frac{1}{k}C_\omega + \omega.$$

Substitute this inequality at $\left\|\mathbb{E}\left[x^k/k\right]\right\|_M$ and obtain

$$\begin{aligned}
k\mathbb{P}\left(\left\|\frac{x^k}{k}\right\|_M \ge \varepsilon\right) &\le \left(\varepsilon-\left\|\mathbb{E}\left[\frac{x^k}{k}\right]\right\|_M\right)^{-2} k\text{Var}_M\left(\frac{x^k}{k}\right) \\
&\le \left(\varepsilon-\alpha\|\mathbf{v}\|_M-\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)-\frac{1}{k}C_\omega-\omega\right)^{-2} k\text{Var}_M\left(\frac{x^k}{k}\right),
\end{aligned}$$

when choice of ω is sufficiently small and that of k is sufficiently large to keep

$$\varepsilon - \alpha \|\mathbf{v}\|_M - \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) - \frac{1}{k}C_\omega + \omega > 0.$$

Take a limit supremum by $k \to \infty$. Then by Theorem 5.1,

$$\begin{aligned}
&\limsup_{k\to\infty} k\mathbb{P}\left(\left\|\frac{x^k}{k}\right\|_M \geq \epsilon\right) \\
&\leq \limsup_{k\to\infty}\left(\varepsilon - \alpha \|\mathbf{v}\|_M - \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) - \frac{1}{k}C_\omega - \omega\right)^{-2} k\mathrm{Var}_M\left(\frac{x^k}{k}\right) \\
&\leq \frac{(\beta-\alpha^2)\|\mathbf{v}\|_M^2}{(\varepsilon - \alpha\|\mathbf{v}\|_M - \omega)^2}
\end{aligned}$$

holds for all sufficiently small $\omega > 0$. Thus, by $\omega \to 0$,

$$\limsup_{k\to\infty} k\mathbb{P}\left(\left\|\frac{x^k}{k}\right\|_M \geq \epsilon\right) \leq \frac{(\beta-\alpha^2)\|\mathbf{v}\|_M^2}{(\varepsilon - \alpha\|\mathbf{v}\|_M)^2}$$

Now consider a null hypothesis of $\|\mathbf{v}\|_M \leq \delta$. Under the null hypothesis,

$$\limsup_{k\to\infty} k\mathbb{P}\left(\left\|\frac{x^k}{k}\right\|_M \geq \epsilon\right) \leq \frac{(\beta-\alpha^2)\delta^2}{(\varepsilon-\alpha\delta)^2},$$

or in other word,

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

as $k \to \infty$.

□

7. Extension to non-orthogonal basis and applications to decentralized optimization

Operator splitting methods such as ADMM/DRS [20, 19, 18] or PDHG [107] are fixed-point iterations with operators that are non-expansive with respect to M-norms where $M \neq \mathbb{I}$, and in such cases, the coordinates form a non-orthogonal basis. Our analyses of Sections 4 and 5 were mostly general, accommodating any M-norm, with the sole exception of Lemma 4.1, which only applies to the case where $M = \mathbb{I}$. In this section, we use the notion of the Friedrichs angle to extend our analysis to general M-norms. We then apply our framework to decentralized optimization and present a numerical experiment.

7.1. Convergence condition in non-orthogonal basis

Let's modify the underlying space $\mathcal{H}$ with extra $\mathcal{H}_0$ block, making $\mathcal{H}$ as

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

where each $\mathcal{H}_i$ is a Hilbert space. Consider two subspaces U_1 and U_2 of $\mathcal{H}$ as

$$U_1 = \{(x_0, 0, 0, \dots, 0) \,|\, x_0 \in \mathcal{H}_0\}, \quad U_2 = \{(0, x_1, x_2, \dots, x_m) \,|\, x_i \in \mathcal{H}_i,\ 1 \le i \le m\},$$

so that $U_1 \cap U_2 = \{0\}$. We further assume that with M-inner product of $\mathcal{H}$, block components of U_2 are orthogonal to each other :

$$\langle (0, 0, \dots, x_i, \dots, 0), (0, 0, \dots, x_j, \dots, 0) \rangle_M = 0, \quad x_i \in \mathcal{H}_i,\ x_j \in \mathcal{H}_j, \quad 1 \le i < j \le m.$$

Note that every vector in $\mathcal{H}$ can be uniquely expressed as a linear combination of vectors in U_1 and U_2. Given $\mathbb{T} \colon \mathcal{H} \to \mathcal{H}$ and $\mathbb{S} = \theta^{-1}(\mathbb{I} - \mathbb{T})$, define $\mathbb{G}$ and $\mathbb{H}$ as

$$\mathbb{S}x = \mathbb{G}x + \mathbb{H}x, \quad \mathbb{G}x \in U_1, \quad \mathbb{H}x \in U_2$$

for all $x \in \mathcal{H}$. We decompose U_2 into m block coordinates, which is also the set of orthogonal subspaces. (To clarify, the m blocks of U_2 are orthogonal with respect to the M-norm.) With a selection vector $\mathcal{I} \in [0, 1]^m$, define a randomized coordinate operator as

$$\begin{aligned} \mathbb{S}_{\mathcal{I}} &= \alpha \mathbb{G} + \sum_{i=1}^{m} \mathcal{I}_i \mathbb{H}_i \\ \mathbb{T}_{\mathcal{I}} &= \mathbb{I} - \theta \mathbb{S}_{\mathcal{I}}, \end{aligned} \tag{7}$$

where $\mathbb{H}_i$ is defined similarly to how $\mathbb{S}_i$ was defined in Section 3.

The cosine of the Friedrichs angle c_F between U_1 and U_2 [84] is defined as a smallest value among $c \le 1$ such that satisfies

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

The RC-FPI by (7) converges, almost surely and in L^2, if the cosine of the Friedrichs angle is sufficiently small.

Theorem 7.1. *Let $\mathbb{T} \colon \mathcal{H} \to \mathcal{H}$ be θ-averaged with respect to $\|\cdot\|_M$ with $\theta \in (0, 1]$. Let $\mathbf{v}$ be the infimal displacement vector of $\mathbb{T}$. Assume $\mathcal{I}^0, \mathcal{I}^1, \dots$ is sampled IID from a distribution satisfying the uniform expected step-size condition* (2) *with $\alpha \in (0, 1]$. Let $x^0, x^1, x^2, \dots$ be the iterates of* (RC-FPI) $x^{k+1} = \mathbb{T}_{\mathcal{I}^k} x^k$*, where $\mathbb{T}_{\mathcal{I}^k}$ is as defined in* (7)*. Let c_F be the cosine of the Friedrichs angle between U_1, U_2.*

(a) If $c_F \le \sqrt{\frac{1-\theta}{1-\alpha\theta}}$, then $x^k/k \stackrel{L^2}{\to} -\alpha \mathbf{v}$ as $k \to \infty$.

(b) If $c_F < \sqrt{\frac{1-\theta}{1-\alpha\theta}}$, then $x^k/k \stackrel{a.s.}{\to} -\alpha \mathbf{v}$ as $k \to \infty$. (x^k/k converges strongly to $-\alpha \mathbf{v}$ in probability 1.*) Furthermore, the results of Theorem 5.1 hold.*

7.2. Proof of Theorem 7.1

Lemma 7.2. *Suppose the subspaces U_1, U_2 of $\mathcal{H}$ with $U_1 \cap U_2 = \{0\}$ satisfy the condition*

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

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

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

where $u_{\mathcal{I}}$ and u are defined as

$$u_{\mathcal{I}} = \alpha g + \sum_{i=1}^{m} \mathcal{I}_i h_i, \quad u = g + h, \quad g \in U_1, \quad h \in U_2.$$

Proof. First equation comes from,

$$\mathbb{E}_{\mathcal{I}}[u_{\mathcal{I}}] = \alpha g + \mathbb{E}_{\mathcal{I}}\left[\sum_{i=1}^{m} \mathcal{I}_i h_i\right] = \alpha g + \alpha h = \alpha u.$$

The expectation in the second equation is

$$\begin{aligned}
\mathbb{E}_{\mathcal{I}}\left[\|u_{\mathcal{I}}\|_M^2\right] &= \alpha^2 \|g\|_M^2 + 2\mathbb{E}\left[\left\langle \sum_{i=1}^{m} \mathcal{I}_i h_i, \alpha g \right\rangle_M\right] + \mathbb{E}\left[\left\|\sum_{i=1}^{m} \mathcal{I}_i h_i\right\|_M^2\right] \\
&= \alpha^2 \|g\|_M^2 + 2\langle \alpha g, \alpha h \rangle_M + \sum_{i=1}^{m} \mathbb{E}\left[\mathcal{I}_i^2\right] \|h_i\|_M^2 \\
&\leq \alpha^2 \|g\|_M^2 + 2\langle \alpha g, \alpha h \rangle_M + \sum_{i=1}^{m} \mathbb{E}[\mathcal{I}_i] \|h_i\|_M^2 \\
&= \alpha^2 \|g\|_M^2 + 2\alpha^2 \langle g, h \rangle_M + \alpha \|h\|_M^2 \\
&= \alpha^2 \|u\|_M^2 + (\alpha - \alpha^2) \|h\|_M^2.
\end{aligned}$$

Note that

$$\|u\|_M^2 = \|h\|_M^2 + 2\langle h, g \rangle_M + \|g\|_M^2 \geq \|h\|_M^2 - 2c_F \|h\|_M \|g\|_M + \|g\|_M^2 \geq \left(1 - c_F^2\right) \|h\|_M^2.$$

Thus, set β as

$$\beta = \alpha^2 + \frac{\alpha - \alpha^2}{1 - c_F^2}.$$

Then,

$$\mathbb{E}_{\mathcal{I}}\left[\|u_{\mathcal{I}}\|_M^2\right] \leq \alpha^2 \|u\|_M^2 + (\alpha - \alpha^2) \|h\|_M^2 \leq \left(\alpha^2 + \frac{\alpha - \alpha^2}{1 - c_F^2}\right) \|u\|_M^2 = \beta \|u\|_M^2,$$

with

$$\theta\beta \le \theta\left(\alpha + \frac{1-\alpha\theta}{\theta}\right)\alpha = \alpha.$$

□

Additionally, if $c_F < \sqrt{\frac{1-\theta}{1-\alpha\theta}}$ in Lemma 7.2, we have $\theta\beta < \alpha$.

Proof of Theorem 7.1. With Lemma 7.2, we know that β is dependent on the value of the cosine of Friedrichs angle c_F as :

$$\mathbb{E}_{\mathcal{I}}\left[u_{\mathcal{I}}\right] = \alpha u, \quad \mathbb{E}_{\mathcal{I}}\left[\|u_{\mathcal{I}}\|_M^2\right] \le \beta \|u\|_M^2, \quad \beta = \alpha^2 + \frac{\alpha - \alpha^2}{1 - c_F^2}.$$

Hence, when $c_F \le \sqrt{\frac{1-\theta}{1-\alpha\theta}}$, we have $\beta \le \alpha/\theta$, and when $c_F < \sqrt{\frac{1-\theta}{1-\alpha\theta}}$, we have $\beta < \alpha/\theta$.

Proof of statement (a). Since $c_F \le \sqrt{\frac{1-\theta}{1-\alpha\theta}}$, we have $\beta \le \alpha/\theta$. Therefore, we may use the result of Lemma 4.7 with $z^0 = x^0$.

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

When the limit $k \to \infty$ is taken,

$$\lim_{k\to\infty}\mathbb{E}\left[\left\|\frac{x^k}{k} - \frac{z^k}{k}\right\|_M^2\right] = 0, \quad \lim_{k\to\infty}\left\|\frac{z^k}{k} + \alpha\mathbf{v}\right\|_M = 0,$$

where the second equation is from Theorem 2.1. These two limits provide L^2 convergence of normalized iterate, namely

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

as $k \to \infty$.

Proof of statement (b). Since $c_F < \sqrt{\frac{1-\theta}{1-\alpha\theta}}$, we have $\beta < \alpha/\theta$. Thus, from Lemma 4.10, we can conclude the strong convergence in probability 1,

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

as $k \to \infty$. Furthermore, since $\beta < \alpha/\theta$, we now satisfy every conditions of Theorem 5.1. Thus, identical results of Theorem 5.1 are obtained in this case. □

7.3. Application of Theorem 7.1 in PG-EXTRA

Consider the convex optimization problem

$$\underset{x\in\mathbb{R}^d}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(x), \tag{8}$$

where $f_i\colon \mathbb{R}^d \to \mathbb{R}$ is closed, convex, and proper function for $i = 1, \ldots, m$. Consider the decentralized algorithm PG-EXTRA [108]

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

for $i = 1, 2, \ldots, m$. In decentralized optimization, we use network of agents to compute the algorithm. If a pair of agents could communicate, we say that they are connected. For each agents $i = 1, 2 \ldots, m$, N_i is a set of agents connected to agent i. A matrix W is called a mixing matrix, and it is a symmetric m by m matrix with $W_{ij} = 0$ if $i \neq j$ and $j \notin N_i$.

A randomized coordinate-update version of PG-EXTRA randomly chooses i among $1, 2, \ldots, m$ to update x_i^k, while every $w_1, w_2, \ldots, w_m$ gets updated at each iterations.

Algorithm 1 RC-PG-EXTRA

```
for i ∈ {1, 2, ..., m} do
    Initialize: w_i = 0, x_i = 0, [Wx]_i = 0
end for
for j ∈ {1, 2, ..., m} do
    Update: w_j = w_j + (α/2)(x_i − [Wx]_i)
end for
while Not converged do
    Sample: I
    for i such that I_i ≠ 0 do
        Δx_i = Prox_{τf_i}([Wx]_i − w_i) − x_i
        Update: x_i = x_i + I_i Δx_i
        for j ∈ N_i ∪ {i} do
            Send: Δx_i From ith agent to jth agent.
            [Wx]_j = [Wx]_j + W_ij Δx_i
        end for
    end for
end while
```

Note that Δx_i is the only quantity communicated across agents. $|N_i|$ communications happen each iteration, while values $x_i, w_i, [Wx]_i$ are stored in ith agent.

(PG-EXTRA) is a fixed-point iteration with an averaged operator with respect to M-norm where $M \neq \mathbb{I}$. Under the conditions of Corollary 7.3, the condition regarding the Friedrichs angle of Theorem 7.1 holds and Algorithm 1 converges.

Corollary 7.3. *Suppose $\mathcal{I}^0, \mathcal{I}^1, \dots$ is sampled IID from a distribution satisfying the uniform expected step-size condition* (2) *with $\alpha \in (0, 1]$. Consider Algorithm 1 with $\mathcal{I} = \mathcal{I}^0, \mathcal{I}^1, \dots$. If the minimum eigenvalue of the symmetric mixing matrix $W \in \mathbb{R}^m$ satisfies*

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

the normalized iterate of Algorithm 1 converges to $-\alpha\mathbf{v}$, where $\mathbf{v}$ is the infimal displacement vector of (PG-EXTRA)*, both in L^2 and almost surely.*

Proof. In the proofs, we use a stack notation for convenience. With stack notation, $\mathbf{x} \in \mathbb{R}^{m\times d}$ refers

$$\mathbf{x} = \begin{bmatrix} — & x_1^\intercal & — \\ — & x_2^\intercal & — \\ & \vdots & \\ — & x_m^\intercal & — \end{bmatrix}, \quad [W\mathbf{x}]_i = \sum_{j=1}^m W_{ij}x_j.$$

(PG-EXTRA) originates from Condat-Vũ [109, 110] with $\mathbf{w}^k = \tau U\mathbf{u}^k$, where Condat-Vũ is a method defined as

$$\begin{aligned} \mathbf{x}^{k+1} &= \mathrm{Prox}_{\tau f}(W\mathbf{x}^k - \tau U\mathbf{u}^k) \\ \mathbf{u}^{k+1} &= \mathbf{u}^k + \frac{1}{\tau}U\mathbf{x}^k, \end{aligned}$$

which is a fixed-point iteration with an operator that's (1/2)-averaged in M-norm. Thus, θ value in (PG-EXTRA) is $\theta = 1/2$.

The matrix M in (PG-EXTRA) is

$$M = \begin{bmatrix} \frac{1}{\tau}I & U \\ U & \tau I \end{bmatrix},$$

where U is a positive semidefinite matrix such that $U^2 = \frac{1}{2}(I - W)$. Note that the inner product in this case is

$$\left\langle \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix}, \begin{bmatrix} \mathbf{y} \\ \mathbf{v} \end{bmatrix} \right\rangle_M = \mathrm{tr}\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix}^T M \begin{bmatrix} \mathbf{y} \\ \mathbf{v} \end{bmatrix} \right).$$

Due to the given inner product, two subspaces $V_1 = (\mathbb{R}^m \times \{0\}^m)^d$ and $V_2 = (\{0\}^m \times \mathbb{R}^m)^d$ are no longer orthogonal to each other. On the other hand, m subspaces of V_1,

$$\left(\{0\}^{i-1} \times \mathbb{R} \times \{0\}^{m-i} \times \{0\}^m\right)^d, \quad i = 1, 2, \dots, m,$$

are orthogonal to each other. Inner product between V_1 and V_2 is constrained as

$$\left| \left\langle \begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{0} \\ \mathbf{u} \end{bmatrix} \right\rangle_M \right| = \left|\mathbf{x}^T U\mathbf{u}\right| \le \lambda_{\max}^U \left\| \begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix} \right\| \left\| \begin{bmatrix} \mathbf{0} \\ \mathbf{u} \end{bmatrix} \right\| = \lambda_{\max}^U \left\| \begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix} \right\|_M \left\| \begin{bmatrix} \mathbf{0} \\ \mathbf{u} \end{bmatrix} \right\|_M.$$

Since $\lambda^W_{\min} > -\alpha/(2-\alpha)$,

$$\lambda^U_{\max} = \sqrt{\frac{1-\lambda^W_{\min}}{2}} < \sqrt{\frac{1}{2-\alpha}} = \sqrt{\frac{1-\frac{1}{2}}{1-\frac{\alpha}{2}}},$$

and we may apply Theorem 7.1 with

$$\mathbf{u} \in V_2 = U_1, \quad \mathbf{x} \in V_1 = U_2, \quad \mathcal{H}_0 = \mathbb{R}^{m\times d}, \quad \mathcal{H}_1 = \mathcal{H}_2 = \cdots = \mathcal{H}_m = d$$

and block coordinate update with each orthogonal blocks as

$$\left(\{0\}^{i-1} \times \mathbb{R} \times \{0\}^{m-i} \times \{0\}^m\right)^d, \quad i = 1, 2, \dots, m,$$

to conclude Corollary 7.3.

□

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

Lemma 7.4. *The infimal displacement vector* $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_m)$ *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 $i = 1, \dots, m$, *where* $(y_1, y_2, \dots, y_m)$ *and* $(g_1, g_2, \dots, g_m)$ *are*

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

Proof. Recall that the PG-EXTRA originated from Condat-Vũ, an FPI with

$$\mathbb{T}\begin{bmatrix}\mathbf{x}\\ \mathbf{u}\end{bmatrix} = \begin{bmatrix}\mathrm{Prox}_{\tau f}(W\mathbf{x} - \tau U\mathbf{u}) \\ \mathbf{u} + \frac{1}{\tau}U\mathbf{x}\end{bmatrix}$$

which is a non-expansive mapping in M-norm, where

$$M = \begin{bmatrix}\frac{1}{\tau}I & U \\ U & \tau I\end{bmatrix}.$$

Finding the infimal displacement vector of $\mathbb{T}$ is equivalent to

$$\underset{\mathbf{x},\mathbf{u}}{\operatorname{argmin}} \left\| \begin{bmatrix}\Delta\mathbf{x}\\ \Delta\mathbf{u}\end{bmatrix} \right\|_M^2, \quad \begin{bmatrix}\Delta\mathbf{x}\\ \Delta\mathbf{u}\end{bmatrix} = (\mathbb{I} - \mathbb{T})\begin{bmatrix}\mathbf{x}\\ \mathbf{u}\end{bmatrix} = \begin{bmatrix}\mathbf{x} - \mathrm{Prox}_{\tau f}(W\mathbf{x} - \tau U\mathbf{u}) \\ -\frac{1}{\tau}U\mathbf{x}\end{bmatrix}.$$

From $\Delta \mathbf{u} = -\frac{1}{\tau} U \mathbf{x}$,

$$\begin{aligned}
\left\| \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{u} \end{bmatrix} \right\|_M^2 &= \frac{1}{\tau} \|\Delta \mathbf{x}\|^2 + \tau \|\Delta \mathbf{u}\|^2 + 2\mathrm{tr}\left(\Delta \mathbf{x}^T U \Delta \mathbf{u}\right) \\
&= \frac{1}{\tau} \|\Delta \mathbf{x}\|^2 + \frac{1}{\tau} \|U \mathbf{x}\|^2 - \frac{2}{\tau}\mathrm{tr}\left(\Delta \mathbf{x}^T U^2 \mathbf{x}\right) \\
&= \frac{1}{\tau} \left[\|\Delta \mathbf{x}\|^2 + \frac{1}{2}\mathrm{tr}\left(\mathbf{x}^T (I - W) \mathbf{x}\right) - \mathrm{tr}\left(\Delta \mathbf{x}^T (I - W) \mathbf{x}\right) \right].
\end{aligned}$$

When $\Delta \mathbf{x} = \mathbf{x_C} + \mathbf{x}_\perp$, where $\mathbf{x_C} = \mathbf{1} \tilde{x}^T$ for some $\tilde{x} \in \mathbb{R}^d$ and $\mathbf{1}^T \mathbf{x}_\perp = \mathbf{0}$, we have

$$\begin{aligned}
&\Delta \mathbf{x} = \mathbf{x} - \mathrm{Prox}_{\tau f}(W\mathbf{x} - \tau U \mathbf{u}) \\
&\Leftrightarrow \mathbf{x_C} = (\mathbf{x} - \mathbf{x}_\perp) - \mathrm{Prox}_{\tau f}(W\left(\mathbf{x} - \mathbf{x}_\perp\right) - (\tau U \mathbf{u} - W \mathbf{x}_\perp)).
\end{aligned}$$

Since

$$\left\{ \tau U \mathbf{u} : \mathbf{u} \in \mathbb{R}^{m \times d} \right\} = \left\{ \mathbf{w} \in \mathbb{R}^{m \times d} : \mathbf{1}^T \mathbf{w} = \mathbf{0} \right\}, \quad \mathbf{1}^T W \mathbf{x}_\perp = \mathbf{1}^T \mathbf{x}_\perp = \mathbf{0},$$

we have $\tau U \mathbf{u} - W \mathbf{x}_\perp = \tau U \tilde{\mathbf{u}}$ for some $\tilde{\mathbf{u}}$. Thus,

$$\begin{bmatrix} \Delta\left(\mathbf{x} - \mathbf{x}_\perp\right) \\ \Delta \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{x_C} \\ -\frac{1}{\tau} U \left(\mathbf{x} - \mathbf{x}_\perp\right) \end{bmatrix},$$

and its M-norm is

$$\left\| \begin{bmatrix} \Delta\left(\mathbf{x} - \mathbf{x}_\perp\right) \\ \Delta \tilde{\mathbf{u}} \end{bmatrix} \right\|_M^2 = \frac{1}{\tau} \left[\|\mathbf{x_C}\|^2 + \frac{1}{2}\mathrm{tr}\left(\left(\mathbf{x} - \mathbf{x}_\perp\right)^T (I - W) \left(\mathbf{x} - \mathbf{x}_\perp\right) \right) \right].$$

Due to the inequality $\|\mathbf{x}_\perp\|^2 \geq \mathrm{tr}\left(\mathbf{x}_\perp{}^T (I - W) \mathbf{x}_\perp\right)$ with equality only when $\mathbf{x}_\perp = \mathbf{0}$,

$$\begin{aligned}
\left\| \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{u} \end{bmatrix} \right\|_M^2 &= \frac{1}{\tau} \left[\|\Delta \mathbf{x}\|^2 + \frac{1}{2}\mathrm{tr}\left(\mathbf{x}^T (I - W) \mathbf{x}\right) - \mathrm{tr}\left(\Delta \mathbf{x}^T (I - W) \mathbf{x}\right) \right] \\
&= \frac{1}{\tau} \left[\|\mathbf{x_C}\|^2 + \|\mathbf{x}_\perp\|^2 + \frac{1}{2}\mathrm{tr}\left(\mathbf{x}^T (I - W) \mathbf{x}\right) - \mathrm{tr}\left(\mathbf{x}_\perp{}^T (I - W) \mathbf{x}\right) \right] \\
&\geq \left\| \begin{bmatrix} \Delta\left(\mathbf{x} - \mathbf{x}_\perp\right) \\ \Delta \tilde{\mathbf{u}} \end{bmatrix} \right\|_M^2,
\end{aligned}$$

with equality only when $\mathbf{x}_\perp = \mathbf{0}$. Thus, the infimal displacement vector $\tilde{\mathbf{v}}$ of Condat-Vũ follows a form of

$$\tilde{\mathbf{v}} = \begin{bmatrix} \mathbf{v}_x \\ \mathbf{v}_u \end{bmatrix}, \quad \mathbf{v}_x = \mathbf{1} \tilde{x}^T,$$

for some $\tilde{x} \in \mathbb{R}^d$. Now we may consider only the case where $\Delta \mathbf{x} = \mathbf{1} x^T$. In this case,

$$\left\| \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{u} \end{bmatrix} \right\|_M^2 = \frac{1}{\tau} \left[\|\Delta \mathbf{x}\|^2 + \frac{1}{2}\mathrm{tr}\left(\mathbf{x}^T (I - W) \mathbf{x}\right) \right],$$

and the relation

$$\mathbf{1}x^T = \mathbf{x} - \text{Prox}_{\tau f}(W\mathbf{x} - \tau U\mathbf{u})$$

must hold. This relation is equivalent to

$$0 \in \tau \partial f(\mathbf{x} - \mathbf{1}x^T) + \mathbf{x} - \mathbf{1}x^T - W\mathbf{x} + \tau U\mathbf{u}.$$

By taking direction of $\mathbf{1}$ to consideration,

$$0 \in \tau \mathbf{1}^T \partial f(\mathbf{x} - \mathbf{1}x^T) + \mathbf{1}^T\mathbf{x} - mx^T - \mathbf{1}^T\mathbf{x}.$$

When we set the new variable $\mathbf{y} = \mathbf{x} - \mathbf{1}x^T$, x^T is expressed as

$$x^T \in \tau \frac{1}{m} \mathbf{1}^T \partial f(\mathbf{y}),$$

which makes for some $\mathbf{g} \in \partial f(\mathbf{y})$,

$$\begin{aligned}
\left\| \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{u} \end{bmatrix} \right\|_M^2 &= \frac{1}{\tau} \left[\|\Delta\mathbf{x}\|^2 + \frac{1}{2}\text{tr}\left(\mathbf{y}^T(I - W)\mathbf{y}\right) \right] \\
&= \frac{1}{\tau} \left[\tau^2 \frac{1}{m} \left\|\mathbf{1}^T\mathbf{g}\right\|^2 + \frac{1}{2}\text{tr}\left(\mathbf{y}^T(I - W)\mathbf{y}\right) \right].
\end{aligned}$$

Thus, the infimal displacement vector of Condat-Vũ is

$$\tilde{\mathbf{v}} = \begin{bmatrix} \tau \frac{1}{m} \mathbf{1}\mathbf{1}^T\mathbf{g} \\ -\frac{1}{\tau} U\mathbf{y} \end{bmatrix},$$

where $\mathbf{y}$ and $\mathbf{g}$ are

$$\underset{\substack{\mathbf{y} \in \mathbb{R}^{m \times d} \\ \mathbf{g} \in \partial f(\mathbf{y})}}{\text{argmin}} \left[\tau^2 \frac{1}{m} \left\|\mathbf{1}^T\mathbf{g}\right\|^2 + \frac{1}{2}\text{tr}\left(\mathbf{y}^T(I - W)\mathbf{y}\right) \right].$$

As a conclusion, the infimal displacement vector 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

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

□

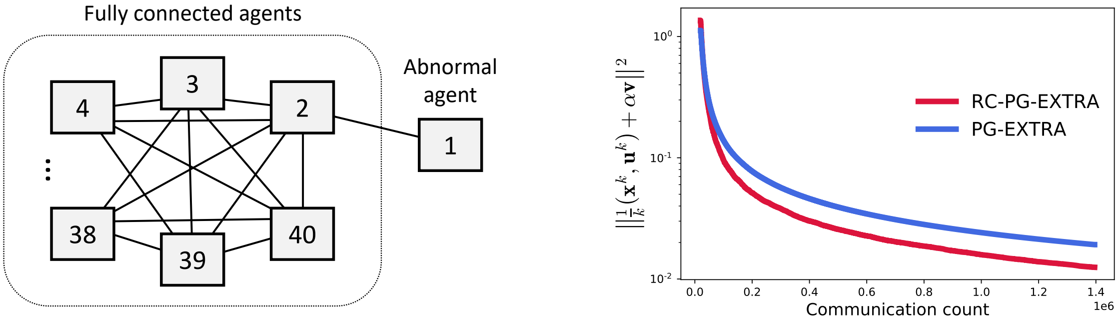


Figure 3: (Left) Network used in our experiment, consisting of $m = 40$ agents, with agents $2, \dots, 40$ densely connected. (Right) Graph of $\left\| \left(\mathbf{x}^k, \mathbf{u}^k \right) / k + \alpha \mathbf{v} \right\|^2$ against the communication count for (PG-EXTRA) and (RC-PG-EXTRA), Algorithm 1.

7.4. Experiment of the infeasible case in PG-EXTRA

We perform an experiment on an instance of (8) using Algorithm 1. Figure 3 shows that (RC-PG-EXTRA), Algorithm 1, converges to the infimal displacement vector faster in terms of communication count.

Specifically, define $f_i \colon \mathbb{R}^2 \to \mathbb{R}$ for $i = 1, \cdots, m$ as

$$f_i(x) = \begin{cases} 0 & \text{if } x \in C_i \\ \infty & \text{otherwise.} \end{cases}$$

with $C_1 = \{(x, y) \mid x \le -10\}$ and $C_2 = C_3 = \cdots = C_m = \{(x, y) \mid x > 0,\, xy \le -1\}$. The network is depicted in Figure 3. We use Metropolis constant edge weight matrix [111, 112] for our mixing matrix W. Metropolis mixing matrix is a symmetric matrix of the form

$$W_{ij} = \begin{cases} \frac{1}{\max(|N_i|,|N_j|)+\epsilon} & \text{if } j \in N_i \\ 1 - \sum_{l \in N_i} W_{il} & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

with $\epsilon > 0$. We choose $\epsilon = 0.05$ in our experiment.

In this setting, the infimal displacement vector has the analytical form :

$$\mathbf{v}_i = \frac{b_i}{2(m - 1 + \epsilon)} \begin{bmatrix} \mathbf{0} \\ u_1 - u_2 \end{bmatrix}, \quad b_i = \begin{cases} 1 & \text{if } i = 1 \\ -1 & \text{if } i = 2 \\ 0 & \text{if } i > 2, \end{cases}$$

where $u_1, u_2 \in \mathbb{R}^d$ is a vector defined as

$$(u_1, u_2) = \operatorname*{argmin}_{u_1 \in \overline{C_1}, u_2 \in \overline{C_2}} \|u_1 - u_2\|.$$

Calculation of the infimal displacement vector. From Lemma 7.4, the infimal displacement vector of (PG-EXTRA) is

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

where $(y_1, y_2, \ldots, y_m)$ and $(g_1, g_2, \ldots, g_m)$ are

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

Note that the subgradient of the indicator function is the normal cone operator

$$\partial \delta_C(x) = \mathbb{N}_C = \begin{cases} \emptyset & \text{if } z \notin C \\ \{y \mid \langle y, z - x \rangle \le 0, \forall z \in C\} & \text{if } z \in C. \end{cases}$$

Thus, with the choice $g_j = 0$, the problem of $(y_1, y_2, \ldots, y_m)$ is equivalent to

$$(y_1, y_2, \ldots, y_m) = \underset{y_1 \in C_1, y_2, \ldots y_m \in C_2}{\operatorname{argmin}} \frac{1}{2} \sum_{i,j=1}^m W_{ij} \left\| y_i - y_j \right\|^2.$$

Since

$$\sum_{i,j=1}^m W_{ij} \left\| y_i - y_j \right\|^2$$
$$= \frac{1}{m-1+\epsilon} \left\| y_1 - y_2 \right\|^2 + \sum_{j>2}^m \frac{1}{m-1+\epsilon} \left\| y_2 - y_j \right\|^2 + \sum_{i,j>2, i \ne j}^m \frac{1}{m-2+\epsilon} \left\| y_i - y_j \right\|^2,$$

$(y_1, y_2, \ldots, y_m)$ take value of $y_2 = y_3 = \cdots = y_m$ with

$$(y_1, y_2) = \underset{y_1 \in C_1, y_2 \in C_2}{\operatorname{argmin}} \left\| y_1 - y_2 \right\|^2.$$

Now chose $g_j = 0$ for each j,

$$\mathbf{v}_i = \begin{bmatrix} \mathbf{0} \\ -\frac{1}{2}\left(y_i - \sum_{j=1}^m W_{ij} y_j\right) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ -\frac{1}{2}\sum_{j \ne i}^m W_{ij}(y_i - y_j) \end{bmatrix}.$$

With $y_2 = y_3 = \cdots = y_m$,

$$\mathbf{v}_1 = \begin{bmatrix} \mathbf{0} \\ \frac{1}{2(m-1+\epsilon)}(y_1 - y_2) \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} \mathbf{0} \\ \frac{1}{2(m-1+\epsilon)}(y_2 - y_1) \end{bmatrix}, \quad \mathbf{v}_i = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad i > 2.$$

Thus, the infimal displacement vector is

$$\mathbf{v}_i = \frac{b_i}{2(m-1+\epsilon)} \begin{bmatrix} \mathbf{0} \\ u \end{bmatrix}, \quad b_i = \begin{cases} 1 & \text{if } i = 1 \\ -1 & \text{if } i = 2 \\ 0 & \text{if } i > 2, \end{cases}$$

with

$$u = \underset{u \in \{u_1 - u_2 | u_1 \in C_1, u_2 \in C_2\}}{\operatorname{argmin}} \|u\|.$$

□

The distribution of $\mathcal{I}$ used for the experiment is

$$P(\mathcal{I}) = \begin{cases} 0.3 & \text{if } \mathcal{I} = \frac{0.7}{0.3 \times (m-1)} e_1 \\ \frac{0.7}{m-1} & \text{if } \mathcal{I} = e_i \text{ for some } i \geq 2 \\ 0 & \text{otherwise,} \end{cases}$$

where $e_i \in \mathbb{R}^m$ is the ith standard unit vector.

8. Conclusion

This work analyzes the asymptotic behavior of the (RC-FPI) and establishes convergence of the normalized iterates to the infimal displacement vector, and this allows us to use the normalized iterates to test for infeasibility. We also extend our analyses to the setup with non-orthogonal basis, thereby making our results applicable to the decentralized optimization algorithm (PG-EXTRA).

One possible direction of future work would be to use variance reduction techniques in the style of, say, SVRG [113] or [114] to improve the convergence rate. Such techniques allow stochastic-gradient-type methods to exhibit a rate faster than $\mathcal{O}(1/k)$, and may be applicable in to the coordinate-update setup accelerate the infeasibility detection.

Acknowledgments

We thank Kibeom Myoung for providing careful reviews and valuable feedback. The authors have been funded by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIP) [grant number 2020R1F1A1A01072877], the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIP) [grant number 2017R1A5A1015626], and the Samsung Science and Technology Foundation [Project Number SSTF-BA2101-02].

References

[1] Y. Nesterov, Efficiency of coordinate descent methods on huge-scale optimization problems, SIAM Journal on Optimization 22 (2012) 341–362.

[2] F. E. Browder, W. Petryshyn, The solution by iteration of nonlinear functional equations in Banach spaces, Bulletin of the American Mathematical Society 72 (1966) 571–575.

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

[4] S. Reich, Asymptotic behavior of contractions in Banach spaces, Journal of Mathematical Analysis and Applications 44 (1973) 57–70.

[5] S. Reich, On the asymptotic behavior of nonlinear semigroups and the range of accretive operators. I, Journal of Mathematical Analysis and Applications 79 (1981) 113–126.

[6] S. Reich, On the asymptotic behavior of nonlinear semigroups and the range of accretive operators. II, Journal of Mathematical Analysis and Applications 87 (1982) 134–146.

[7] A. T. Plant, S. Reich, The asymptotics of nonexpansive iterations, Journal of Functional Analysis 54 (1983) 308–319.

[8] D. Ariza-Ruiz, L. Leuştean, G. López-Acedo, Firmly nonexpansive mappings in classes of geodesic spaces, Transactions of the American Mathematical Society 366 (2014) 4299–4322.

[9] A. Nicolae, Asymptotic behavior of averaged and firmly nonexpansive mappings in geodesic spaces, Nonlinear Analysis: Theory, Methods & Applications 87 (2013) 102–115.

[10] R. E. Bruck Jr, On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space, Journal of Mathematical Analysis and Applications 61 (1977) 159–164.

[11] J. Bailion, R. E. Bruck, S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, Houston Journal of Mathematics 4 (1978) 1–9.

[12] S. Reich, I. Shafrir, The asymptotic behavior of firmly nonexpansive mappings, Proceedings of the American Mathematical Society (1987) 246–250.

[13] H. H. Bauschke, G. R. Douglas, W. M. Moursi, On a result of Pazy concerning the asymptotic behaviour of nonexpansive mappings, Journal of Fixed Point Theory and Applications 18 (2016) 297–307.

[14] E. K. Ryu, Cosmic divergence, weak cosmic convergence, and fixed points at infinity, Journal of Fixed Point Theory and Applications 20 (2018) 1–16.

[15] A. W. Gutiérrez, A. Karlsson, Comments on the cosmic convergence of nonexpansive maps, Journal of Fixed Point Theory and Applications 23 (2021) 1–10.

[16] H. H. Bauschke, W. M. Moursi, The magnitude of the minimal displacement vector for compositions and convex combinations of firmly nonexpansive mappings, Optimization Letters 12 (2018) 1465–1474.

[17] H. H. Bauschke, W. M. Moursi, On the minimal displacement vector of compositions and convex combinations of nonexpansive mappings, Foundations of Computational Mathematics 20 (2020) 1653–1666.

[18] P.-L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM Journal on Numerical Analysis 16 (1979) 964–979.

[19] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, Computers & Mathematics with Applications 2 (1976) 17–40.

[20] R. Glowinski, A. Marroco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires, Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique 9 (1975) 41–76.

[21] H. H. Bauschke, W. L. Hare, W. M. Moursi, Generalized solutions for the sum of two maximally monotone operators, SIAM Journal on Control and Optimization 52 (2014) 1034–1047.

[22] H. H. Bauschke, M. N. Dao, W. M. Moursi, The Douglas–Rachford algorithm in the affine-convex case, Operations Research Letters 44 (2016) 379–382.

[23] H. H. Bauschke, W. M. Moursi, The Douglas–Rachford algorithm for two (not necessarily intersecting) affine subspaces, SIAM Journal on Optimization 26 (2016) 968–985.

[24] H. H. Bauschke, W. M. Moursi, On the behavior of the Douglas–Rachford algorithm for minimizing a convex function subject to a linear constraint, SIAM Journal on Optimization 30 (2020) 2559–2576.

[25] H. H. Bauschke, W. M. Moursi, On the Douglas-Rachford algorithm for solving possibly inconsistent optimization problems, arXiv preprint arXiv:2106.11547 (2021).

[26] W. M. Moursi, M. Saurette, On the Douglas-Rachford and Peaceman-Rachford algorithms in the presence of uniform monotonicity and the absence of minimizers, arXiv preprint arXiv:2201.06661 (2022).

[27] G. Banjac, J. Lygeros, On the asymptotic behavior of the Douglas–Rachford and proximal-point algorithms for convex optimization, Optimization Letters 15 (2021) 2719–2732.

[28] G. Banjac, On the minimal displacement vector of the Douglas–Rachford operator, Operations Research Letters 49 (2021) 197–200.

[29] G. Banjac, P. Goulart, B. Stellato, S. Boyd, Infeasibility detection in the alternating direction method of multipliers for convex optimization, Journal of Optimization Theory and Applications 183 (2019) 490–519.

[30] Y. Liu, E. K. Ryu, W. Yin, A new use of Douglas–Rachford splitting for identifying infeasible, unbounded, and pathological conic programs, Mathematical Programming 177 (2019) 225–253.

[31] E. K. Ryu, Y. Liu, W. Yin, Douglas–Rachford splitting and ADMM for pathological convex optimization, Computational Optimization and Applications 74 (2019) 747–778.

[32] B. Stellato, G. Banjac, P. Goulart, A. Bemporad, S. Boyd, OSQP: an operator splitting solver for quadratic programs, Mathematical Programming Computation 12 (2020) 637–672.

[33] M. Garstka, M. Cannon, P. Goulart, COSMO: A conic operator splitting method for convex conic problems, Journal of Optimization Theory and Applications 190 (2021) 779–810.

[34] B. O'Donoghue, E. Chu, N. Parikh, S. Boyd, SCS: Splitting conic solver, version 3.2.2, `https://github.com/cvxgrp/scs`, 2022.

[35] P. Sopasakis, K. Menounou, P. Patrinos, Superscs: fast and accurate large-scale conic optimization, European Control Conference (2019).

[36] B. O'Donoghue, E. Chu, N. Parikh, S. Boyd, Conic optimization via operator splitting and homogeneous self-dual embedding, Journal of Optimization Theory and Applications 169 (2016) 1042–1068.

[37] B. O'Donoghue, Operator splitting for a homogeneous embedding of the linear complementarity problem, SIAM Journal on Optimization 31 (2021) 1999–2023.

[38] C. Hildreth, A quadratic programming procedure, Naval Research Logistics Quarterly 4 (1957) 79–85.

[39] D. d'Esopo, A convex programming procedure, Naval Research Logistics Quarterly 6 (1959) 33–42.

[40] J. Warga, Minimizing certain convex functions, Journal of the Society for Industrial and Applied Mathematics 11 (1963) 588–593.

[41] A. L. Garkavi, Method of cyclic descent in the problem of best approximation, Mathematical notes of the Academy of Sciences of the USSR 27 (1980) 270–274.

[42] Z.-Q. Luo, P. Tseng, On the convergence of the coordinate descent method for convex differentiable minimization, Journal of Optimization Theory and Applications 72 (1992) 7–35.

[43] L. Grippo, M. Sciandrone, On the convergence of the block nonlinear Gauss–Seidel method under convex constraints, Operations Research Letters 26 (2000) 127–136.

[44] P. Tseng, Convergence of a block coordinate descent method for nondifferentiable minimization, Journal of Optimization Theory and Applications 109 (2001) 475–494.

[45] A. Auslender, Asymptotic properties of the Fenchel dual functional and applications to decomposition problems, Journal of Optimization Theory and Applications 73 (1992) 427–449.

[46] M. Razaviyayn, M. Hong, Z.-Q. Luo, A unified convergence analysis of block successive minimization methods for nonsmooth optimization, SIAM Journal on Optimization 23 (2013) 1126–1153.

[47] Y. Xu, W. Yin, A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion, SIAM Journal on Imaging Sciences 6 (2013) 1758–1789.

[48] P. Tseng, S. Yun, A coordinate gradient descent method for nonsmooth separable minimization, Mathematical Programming 117 (2009) 387–423.

[49] S. Yun, K.-C. Toh, A coordinate gradient descent method for ℓ_1-regularized convex minimization, Computational Optimization and Applications 48 (2011) 273–307.
[50] A. Beck, L. Tetruashvili, On the convergence of block coordinate descent type methods, SIAM Journal on Optimization 23 (2013) 2037–2060.
[51] J. V. Shi, Y. Xu, R. G. Baraniuk, Sparse bilinear logistic regression, UCLA CAM Reports 14-12 (2014).
[52] Y. Xu, Alternating proximal gradient method for sparse nonnegative Tucker decomposition, Mathematical Programming Computation 7 (2015) 39–70.
[53] N. Zhou, Y. Xu, H. Cheng, J. Fang, W. Pedrycz, Global and local structure preserving sparse subspace learning: An iterative approach to unsupervised feature selection, Pattern Recognition 53 (2016) 87–101.
[54] Y. Xu, W. Yin, A globally convergent algorithm for nonconvex optimization based on block coordinate update, Journal of Scientific Computing 72 (2017) 700–734.
[55] M. Hong, X. Wang, M. Razaviyayn, Z.-Q. Luo, Iteration complexity analysis of block coordinate descent methods, Mathematical Programming 163 (2017) 85–114.
[56] O. Fercoq, Z. Qu, P. Richtárik, M. Takáč, Fast distributed coordinate descent for non-strongly convex losses, International Workshop on Machine Learning for Signal Processing (2014).
[57] Z. Peng, Y. Xu, M. Yan, W. Yin, ARock: an algorithmic framework for asynchronous parallel coordinate updates, SIAM Journal on Scientific Computing 38 (2016) A2851–A2879.
[58] D. Z. Farsa, S. Rahnamayan, Discrete coordinate descent (DCD), International Conference on Systems, Man, and Cybernetics (2020).
[59] H. Hazimeh, R. Mazumder, Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms, Operations Research 68 (2020) 1517–1537.
[60] S. Jäger, A. Schöbel, The blockwise coordinate descent method for integer programs, Mathematical Methods of Operations Research 91 (2020) 357–381.
[61] Q. Lin, Z. Lu, L. Xiao, An accelerated proximal coordinate gradient method, Neural Information Processing Systems (2014).
[62] O. Fercoq, P. Richtárik, Accelerated, parallel, and proximal coordinate descent, SIAM Journal on Optimization 25 (2015) 1997–2023.
[63] Z. Allen-Zhu, Z. Qu, P. Richtárik, Y. Yuan, Even faster accelerated coordinate descent using non-uniform sampling, International Conference on Machine Learning (2016) 1110–1119.
[64] Z. Qu, P. Richtárik, Coordinate descent with arbitrary sampling I: Algorithms and complexity, Optimization Methods and Software 31 (2016) 829–857.
[65] Y. Nesterov, S. U. Stich, Efficiency of the accelerated coordinate descent method on structured optimization problems, SIAM Journal on Optimization 27 (2017) 110–123.
[66] F. Hanzely, P. Richtárik, Accelerated coordinate descent with arbitrary sampling and best rates for minibatches, International Conference on Artificial Intelligence and Statistics (2019).
[67] Q. Bertrand, M. Massias, Anderson acceleration of coordinate descent, International Conference on Artificial Intelligence and Statistics (2021).
[68] A. Alacaoglu, V. Cevher, S. J. Wright, On the complexity of a practical primal-dual coordinate method, arXiv preprint arXiv:2201.07684 (2022).
[69] J. Nutini, I. Laradji, M. Schmidt, Let's make block coordinate descent go fast: Faster greedy rules, message-passing, active-set complexity, and superlinear convergence, Journal of Machine Learning Research (JMLR) 23 (2022) 1–74.
[70] A. Chambolle, M. J. Ehrhardt, P. Richtárik, C.-B. Schönlieb, Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications, SIAM Journal on Optimization 28 (2018) 2783–2808.
[71] O. Fercoq, P. Bianchi, A coordinate-descent primal-dual algorithm with large step size and possibly nonseparable functions, SIAM Journal on Optimization 29 (2019) 100–134.
[72] M. Verkama, Random relaxation of fixed-point iteration, SIAM Journal on Scientific Computing 17 (1996) 906–912.

[73] P. L. Combettes, J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping, SIAM Journal on Optimization 25 (2015) 1221–1248.
[74] P. L. Combettes, J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping II: Mean-square and linear convergence, Mathematical Programming 174 (2019) 433–451.
[75] E. Chouzenoux, J.-C. Pesquet, A. Repetti, A block coordinate variable metric forward–backward algorithm, Journal of Global Optimization 66 (2016) 457–485.
[76] S. Salzo, S. Villa, Parallel random block-coordinate forward–backward algorithm: a unified convergence analysis, Mathematical Programming 193 (2022) 225–269.
[77] L. M. Briceño-Arias, G. Chierchia, E. Chouzenoux, J.-C. Pesquet, A random block-coordinate Douglas–Rachford splitting method with low computational complexity for binary logistic regression, Computational Optimization and Applications 72 (2019) 707–726.
[78] Y. T. Chow, T. Wu, W. Yin, Cyclic coordinate-update algorithms for fixed-point problems: Analysis and applications, SIAM Journal on Scientific Computing 39 (2017) A1280–A1300.
[79] B. Peng, H.-K. Xu, A cyclic coordinate-update fixed point algorithm, Carpathian Journal of Mathematics 35 (2019) 365–370.
[80] Z. Lu, L. Xiao, On the complexity analysis of randomized block-coordinate descent methods, Mathematical Programming 152 (2015) 615–642.
[81] P. L. Combettes, J.-C. Pesquet, Linear convergence of stochastic block-coordinate fixed point algorithms, European Signal Processing Conference (2018) 742–746.
[82] R. Sun, Y. Ye, Worst-case complexity of cyclic coordinate descent: $O(n^2)$ gap with randomized version, Mathematical Programming 185 (2021) 487–520.
[83] Q. Tran-Dinh, Accelerated randomized block-coordinate algorithms for co-coercive equations and applications, arXiv preprint arXiv:2301.03113 (2023).
[84] K. O. Friedrichs, On certain inequalities and characteristic value problems for analytic functions and for functions of two variables, Transactions of the American Mathematical Society 41 (1937) 321–364.
[85] F. Deutsch, Rate of convergence of the method of alternating projections, in: Parametric Optimization and Approximation, Springer, 1985, pp. 96–107.
[86] F. Deutsch, The angle between subspaces of a Hilbert space, in: Approximation Theory, Wavelets and Applications, Springer, 1995, pp. 107–130.
[87] N. Aronszajn, Theory of reproducing kernels, Transactions of the American Mathematical Society 68 (1950) 337–404.
[88] S. Kayalar, H. L. Weinert, Error bounds for the method of alternating projections, Mathematics of Control, Signals and Systems 1 (1988) 43–59.
[89] H. H. Bauschke, J. M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Review 38 (1996) 367–426.
[90] C. Badea, S. Grivaux, V. Müller, A generalization of the Friedrichs angle and the method of alternating projections, Comptes Rendus Mathématique 348 (2010) 53–56.
[91] R. Nishihara, S. Jegelka, M. I. Jordan, On the convergence rate of decomposable submodular function minimization, Neural Information Processing Systems (2014).
[92] M. Fält, P. Giselsson, Optimal convergence rates for generalized alternating projections, in: 2017 IEEE 56th Annual Conference on Decision and Control (CDC), 2017, pp. 2268–2274. doi:10.1109/CDC.2017.8263980.
[93] S. Reich, R. Zalas, The optimal error bound for the method of simultaneous projections, Journal of Approximation Theory 223 (2017) 96–107.
[94] F. J. Aragón Artacho, R. Campoy, A new projection method for finding the closest point in the intersection of convex sets, Computational Optimization and Applications 69 (2018) 99–132.
[95] F. J. Aragón Artacho, R. Campoy, Optimal rates of linear convergence of the averaged alternating modified reflections method for two subspaces, Numerical Algorithms 82 (2019) 397–421.
[96] M. Fält, P. Giselsson, Generalized alternating projections on manifolds and convex sets, arXiv preprint arXiv:2101.07286 (2021).

[97] H. H. Bauschke, J. B. Cruz, T. T. Nghia, H. M. Phan, X. Wang, The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle, Journal of Approximation Theory 185 (2014) 63–79.

[98] A. U. Raghunathan, S. Di Cairano, ADMM for convex quadratic programs: Q-linear convergence and infeasibility detection, arXiv preprint arXiv:1411.7288 (2014).

[99] H. H. Bauschke, J. Bello Cruz, T. T. Nghia, H. M. Pha, X. Wang, Optimal rates of linear convergence of relaxed alternating projections and generalized Douglas-Rachford methods for two subspaces, Numerical Algorithms 73 (2016) 33–76.

[100] J. Liang, J. Fadili, G. Peyré, Local convergence properties of Douglas–Rachford and alternating direction method of multipliers, Journal of Optimization Theory and Applications 172 (2017) 874–913.

[101] G. Banjac, P. J. Goulart, Tight global linear convergence rate bounds for operator splitting methods, IEEE Transactions on Automatic Control 63 (2018) 4126–4139.

[102] R. Behling, J. Y. Bello Cruz, L.-R. Santos, Circumcentering the Douglas–Rachford method, Numerical Algorithms 78 (2018) 759–776.

[103] J. Liang, J. Fadili, G. Peyré, Local linear convergence analysis of primal–dual splitting methods, Optimization 67 (2018) 821–853.

[104] H. H. Bauschke, J. M. Borwein, A. S. Lewis, The method of cyclic projections for closed convex sets in Hilbert space, Contemporary Mathematics 204 (1997) 1–38.

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

[106] D. Applegate, M. Díaz, H. Lu, M. Lubin, Infeasibility detection with primal-dual hybrid gradient for large-scale linear programming, arXiv preprint arXiv:2102.04592 (2021).

[107] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, Journal of Mathematical Imaging and Vision 40 (2011) 120–145.

[108] W. Shi, Q. Ling, G. Wu, W. Yin, A proximal gradient algorithm for decentralized composite optimization, IEEE Transactions on Signal Processing 63 (2015) 6013–6023.

[109] L. Condat, A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, Journal of Optimization Theory and Applications 158 (2013) 460–479.

[110] B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, Advances in Computational Mathematics 38 (2013) 667–681.

[111] S. Boyd, P. Diaconis, L. Xiao, Fastest mixing Markov chain on a graph, SIAM Review 46 (2004) 667–689.

[112] L. Xiao, S. Boyd, S.-J. Kim, Distributed average consensus with least-mean-square deviation, Journal of Parallel and Distributed Computing 67 (2007) 33–46.

[113] R. Johnson, T. Zhang, Accelerating stochastic gradient descent using predictive variance reduction, Neural Information Processing Systems (2013).

[114] M. Schmidt, N. L. Roux, F. Bach, Minimizing finite sums with the stochastic average gradient, Mathematical Programming 162 (2017) 83–112.