Accelerated Infeasibility Detection of Constrained Optimization and Fixed-Point Iterations

Jisun Park [1] **Ernest K. Ryu** [1 2]

Abstract

As first-order optimization methods become the method of choice for solving large-scale optimization problems, optimization solvers based on first-order algorithms are being built. Such general-purpose solvers must robustly detect infeasible or misspecified problem instances, but the computational complexity of first-order methods for doing so has yet to be formally studied. In this work, we characterize the optimal accelerated rate of infeasibility detection. We show that the standard fixed-point iteration achieves a $\mathcal{O}(1/k^2)$ and $\mathcal{O}(1/k)$ rates, respectively, on the normalized iterates and the fixed-point residual converging to the infimal displacement vector, while the accelerated fixed-point iteration achieves $\mathcal{O}(1/k^2)$ and $\tilde{\mathcal{O}}(1/k^2)$ rates. We then provide a matching complexity lower bound to establish that $\Theta(1/k^2)$ is indeed the optimal accelerated rate.

1. Introduction

First-order optimization methods have become the method of choice for solving the large-scale optimization problems of the modern era. As first-order methods scale more favorably than classical interior-point methods (O'Donoghue et al., 2016; Stellato et al., 2020; Garstka et al., 2021), new optimization solvers based on first-order algorithms are being built with the goal of replacing classical solvers based on interior-point methods or simplex methods in large-scale applications.

However, these new first-order solvers are far less equipped to robustly detect infeasible or misspecified problem instances. A general-purpose solver must robustly detect infeasible problem instances arising from user misspecification or from applications such as embedded application, mixed-integer optimization with branch-and-bound technique, or combinatorial optimization (Naik & Bemporad, 2017; De Loera et al., 2012). Classical solvers based on interior point methods or simplex methods, in their first phase, determines whether the problem is feasible or infeasible by finding a feasible point. The behavior of such classical solvers under pathologies is well understood through extensive theoretical research and through the decades-long deployment of open-source and commercial solvers. The analysis of first-order algorithms such as Douglas-Rachford splitting (DRS) and ADMM applied to pathological problem instances has started to gain attention. However, the computational complexity of determining the infeasibility of a given problem instance has yet to be formally studied.

In this work, we characterize the optimal accelerated rate of infeasibility detection by analyzing the convergence rates of fixed-point iterations towards the infimal displacement vector, which serves as a certificate of infeasibility. We show that the standard fixed-point iteration achieves a $\mathcal{O}(1/k^2)$ and $\mathcal{O}(1/k)$ rates, respectively, on the normalized iterates and the fixed-point residual, while the accelerated fixed-point iteration achieves $\mathcal{O}(1/k^2)$ and $\tilde{\mathcal{O}}(1/k^2)$ rates. We then provide a matching complexity lower bound to establish that $\Theta(1/k^2)$ is indeed the optimal accelerated rate.

1.1. Preliminaries and notations

We use the standard notations in Ryu & Yin (2022).

Sets and operators. Let $\mathcal{H}$ be a real Hilbert space. For a set $C \subseteq \mathcal{H}$, we denote by $\operatorname{conv} C$ a convex hull of C, $\overline{C}$ a closure of C, and $\overline{\operatorname{conv}}\, C$ a closure of a convex hull of C. If C is a nonempty closed convex set, for any $x \in \mathcal{H}$, there exists a unique vector $z \in C$ such that $z \in \operatorname{argmin}_{y\in C} \|x-y\|^2$, which is denoted as $\Pi_C(x)$ and called a projection of x onto C. We also let δ_C refer to the indicator function of set C. We denote by $\mathbb{S}^n$ the set of all $n \times n$ symmetric matrices and by $\mathbb{S}^n_+$ the set of all $n \times n$ symmetric positive semidefinite matrices. We say $X \succeq Y$ if $X - Y \in \mathbb{S}^n_+$ for $X, Y \in \mathbb{S}^n$.

Let $\mathbb{T} \colon \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-valued operator. $\operatorname{dom} \mathbb{T} = \{x \mid \mathbb{T}x \neq \emptyset\}$ is called the domain of $\mathbb{T}$, and $\mathcal{R}(\mathbb{T}) = \{y \mid \exists x \in$

[1]Department of Mathematical Sciences, Seoul National University [2]Interdisciplinary Program in Artificial Intelligence, Seoul National University. Correspondence to: Ernest K. Ryu <ernestryu@snu.ac.kr>.

Proceedings of the 40^th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

$\mathcal{H}$ *s.t.* $y \in \mathbb{T}x\}$ is called the range of $\mathbb{T}$. For a single-valued operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$, it is called nonexpansive if $\|\mathbb{T}x - \mathbb{T}y\| \le \|x-y\|$ for all $x, y \in \operatorname{dom}\mathbb{T}$ and γ-contractive with $0 < \gamma < 1$ if $\|\mathbb{T}x - \mathbb{T}y\| \le \gamma\|x-y\|$ holds for all $x, y \in \operatorname{dom}\mathbb{T}$, and θ-averaged if there exists nonexpansive operator $\mathbb{S}$ and identity operator $\mathbb{I}$ such that $\mathbb{T} = \theta\mathbb{S} + (1-\theta)\mathbb{I}$. $\mathbb{T}$ is called maximal nonexpansive (contractive) if $\operatorname{dom}\mathbb{T} = \mathcal{H}$. If $x_\star \in \mathcal{H}$ is a point such that $x_\star = \mathbb{T}x_\star$, we call $x_\star$ a fixed point of $\mathbb{T}$. $\operatorname{Fix}\mathbb{T} \subseteq \mathcal{H}$ denotes a set of fixed points of $\mathbb{T}$.

Fixed-point iteration. Classical Banach fixed-point theorem illustrates that if $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ is a contraction, then $\operatorname{Fix}\mathbb{T}$ is nonempty and the *Picard iteration* (Picard)

$$x^{k+1} = \mathbb{T}x^k, \qquad k = 0, 1, \dots \tag{Picard}$$

starting from $x^0 \in \mathcal{H}$ converges to some $x_\star \in \operatorname{Fix}\mathbb{T}$. When $\mathbb{T}$ is nonexpansive but not necessarily contractive, (Picard) may not converge to the fixed point of $\mathbb{T}$. In such cases, to guarantee the convergence, one may use *Krasnosel'skiĭ-Mann iteration* (Krasnosel'skiĭ, 1955; Mann, 1953) or *Halpern iteration* (Halpern, 1967), whose forms are described in Section 3.

Constrained optimization and fixed-point iterations. Consider a constrained optimization problem

$$\begin{array}{ll} \underset{x\in\mathcal{H}}{\text{minimize}} & f(x) \\ \text{subject to} & x \in C, \end{array}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex and $C \subseteq \mathbb{R}^n$ is a nonempty closed convex set. Problems of this type can be solved with various first-order methods including projected gradient method, proximal gradient method, alternating direction method of multipliers (ADMM), and primal-dual hybrid gradient (PDHG). These methods can be understood and analyzed as nonexpansive fixed-point iterations (Ryu & Yin, 2022). Therefore, the analysis of fixed-point iteration broadly applies to this broad class of first-order methods.

Inconsistent operators. We say $\mathbb{T}$ is consistent if $\operatorname{Fix}\mathbb{T} \neq \emptyset$ and inconsistent if $\operatorname{Fix}\mathbb{T} = \emptyset$. $\mathbb{T}$ is inconsistent if and only if $0 \notin \mathcal{R}(\mathbb{I} - \mathbb{T})$. From the well-known fact that $\overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$ is closed and convex (Pazy, 1971, Lemma 4), $v = \Pi_{\overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}}(0)$ is well-defined, and v is called infimal displacement vector. If $\mathbb{T}$ is consistent, then $v = 0$. If $v \neq 0$, then $\mathbb{T}$ is inconsistent.

Any nonexpansive operator $\mathbb{T}$ is of exactly one of these three cases: (i) $\operatorname{Fix}\mathbb{T} \neq \emptyset$, (ii) $v \neq 0$, and (iii) $\operatorname{Fix}\mathbb{T} = \emptyset$ with $v = 0$. In convex optimization, (i) corresponds to the case where primal and dual solution exist and primal-dual gap being 0, and (ii) corresponds to the case where either the primal problem or the dual problem is infeasible. (iii) corresponds to pathological weakly feasible and weakly infeasible cases (Banjac et al., 2019; Liu et al., 2019; Ryu et al., 2019). Our main focus will be on cases (i) and (ii).

1.2. Prior work

Inconsistent fixed-point iteration. Browder & Petryshyn (1966) first proved that the iterates of Picard iteration is bounded if and only if $\mathbb{T}$ is consistent, and later followed by work of Pazy (1971) showing the convergence of $-x^k/k$ to the infimal displacement vector. This result has been extended to Banach space setup by Reich (1973). If $\mathbb{T}$ is more than just a nonexpansive operator, then difference of iterates $\mathbb{T}^k x^0 - \mathbb{T}^{k+1}x^0$ also converges to the infimal displacement vector; see Bailion et al. (1978), Reich & Shafrir (1987), and Bruck Jr (1977) for averaged, firmly-nonexpansive, and strongly nonexpansive operators in Banach spaces. For more on general settings, see Reich (1981; 1982); Plant & Reich (1983); Ariza-Ruiz et al. (2014); Nicolae (2013). Despite its numerous appearance, it was not until in late 1990s where the term 'minimal displacement vector' was coined (Bauschke et al., 1997). It was later called 'infimal displacement vector' (Bauschke et al., 2014), and its properties have been analyzed with depth as well (Bauschke et al., 2016; Ryu, 2018; Bauschke & Moursi, 2018; 2020b).

First-order numerical solvers. The interior point method (Nesterov & Nemirovskii, 1994) has been successful in solving convex optimization problems, and a number of numerical solvers based on this exists (Nesterov & Nemirovskii, 1994; Sturm, 1999; Gurobi Optimization, LLC, 2023; ApS, 2019; Mattingley & Boyd, 2012). Recently, first-order method solving conic optimization programs has gained huge interest, due to its scalability to very large and high-dimensional problems. ADMM-based solvers such as SCS (O'Donoghue et al., 2016; Sopasakis et al., 2019), OSQP (Stellato et al., 2020), and COSMO (Garstka et al., 2021), and also include PDHG-based solver PDLP (Chambolle & Pock, 2011; Applegate et al., 2021a) are first-order numerical solvers.

Constrained optimization and infeasibility. For convex feasibility problem, primary choice of methods are cyclic projection (Von Neumann, 1951), Dykstra's algorithm (Dykstra, 1983), AAR method (Bauschke et al., 2004), and so on. These methods have been analyzed extensively (Boyle & Dykstra, 1986; Bauschke & Borwein, 1994; Bauschke et al., 1997; Artacho et al., 2014; Borwein & Tam, 2015; Aragón Artacho et al., 2016). For general constrained convex optimization problem, Douglas-Rachford splitting (DRS) (Lions & Mercier, 1979) and alternating direction method of multipliers (ADMM) (Glowinski & Marroco, 1975; Gabay & Mercier, 1976) are popular choices of algorithm, and their behavior on infeasible primal or dual problems has been recently analyzed (Eckstein & Bertsekas, 1992; Bauschke et al., 2014; Raghunathan & Di Cairano, 2014; Banjac et al., 2019; Liu et al., 2019; Bauschke & Moursi, 2020a; Banjac, 2021; Banjac & Lygeros, 2021; Bauschke & Moursi, 2021; O'Donoghue, 2021; Moursi & Saurette, 2022). Recently, PDHG (Chambolle & Pock, 2011) has been used as a first-order algorithm solving possibly inconsistent LP and QP (Applegate et al., 2021b).

Accelerated fixed-point iterations. Picard iteration converges when the operator $\mathbb{T}$ is contractive, but does not converge with nonexpansivity alone. If $\mathbb{T}$ is averaged, fixed-point residual of Picard iteration converges in $\mathcal{O}(1/k)$ rate (Davis, 2015). But rather than adding conditions on operators, interpolation or extrapolation schemes (Krasnosel'skiĭ, 1955; Mann, 1953; Anderson, 1965; Ishikawa, 1976; Xu, 2004; Maingé, 2008; Dong et al., 2018; Shehu, 2018; Themelis & Patrinos, 2019; Reich et al., 2021; Walker & Ni, 2011; Zhang et al., 2020; Shehu & Gibali, 2020; Shehu et al., 2020; Scieur et al., 2020; Barré et al., 2022a) may result in faster convergence rate, which is the case for Halpern iteration (Halpern, 1967), which exhibits $\mathcal{O}(1/k^2)$ rate (Sabach & Shtern, 2017; Lieder, 2021).

For the inconsistent fixed-point iteration, the rate of convergence to infimal displacement vector is measured. Unlike the convergence itself (Bailion et al., 1978), the $\mathcal{O}(1/k)$ rate of convergence was not known until late 2010s (Liu et al., 2019). Another sequence converging to infimal displacement vector is normalized iterates, and it is proven to converge in $\mathcal{O}(1/k^2)$ rate (Applegate et al., 2021b).

Complexity lower bound. Using the information-based complexity framework (Nemirovski, 1992), lower bounds to the iteration complexity has been thoroughly studied for first-order convex optimization methods (Nesterov, 2004; Drori, 2017; Carmon et al., 2020; Drori & Shamir, 2020; Carmon et al., 2021; Dragomir et al., 2022; Drori & Taylor, 2022; Yoon & Ryu, 2021; Park & Ryu, 2022). For the fixed-point iterations, Diakonikolas (2020) first proved $\Omega(1/k^2)$-lower bound, and Park & Ryu (2022) later closed the constant gap by showing that Halpern iteration of Lieder (2021) has exactly matching $\Theta(1/k^2)$-complexity to the lower bound of Park & Ryu (2022). However, these works are restricted to the consistent fixed-point iterations.

Performance estimation problem (PEP). From the seminal work of Drori & Teboulle (2014), performance estimation problem (PEP) has been widely used to obtain the worst-case complexity of algorithms, including first-order methods (Kim & Fessler, 2017; Taylor et al., 2017; De Klerk et al., 2017; Kim & Fessler, 2018; Taylor et al., 2018b; Barré et al., 2020; De Klerk et al., 2020; Kim & Fessler, 2021; Abbaszadehpeivasti et al., 2022a;c; Barré et al., 2022b; Kamri et al., 2022; Rotaru et al., 2022; Gupta et al., 2023), operator splitting methods (Ryu et al., 2020), minimax algorithms (Abbaszadehpeivasti et al., 2021; Gorbunov et al., 2022; Zamani et al., 2022), proximal point methods (Gu & Yang, 2020; Kim, 2021; Gu & Yang, 2022; 2023), decentralized methods (Colla & Hendrickx, 2021; 2022a;b), coordinate descent methods (Abbaszadehpeivasti et al., 2022b), and even the continuous-time models (Moucer et al., 2022). PEP also finds the optimal method with optimal worst-case complexity (Drori & Teboulle, 2016; Kim & Fessler, 2016; Drori & Taylor, 2020; Taylor & Drori, 2022; Kim, 2021; Park & Ryu, 2022), and is even used to construct the Lyapunov function for the proof of convergence (Taylor et al., 2018a) and complexity lower bound (Dragomir et al., 2022). All these works assume the existence of the solution or optimal value.

1.3. Contribution

We summarize the contribution of this work as follows. First, we prove upper bounds on the rates of convergence of certain sequences to the infimal displacement vector, which can serve as a certificate of infeasibility. In particular, we establish a $\mathcal{O}(1/k^2)$-rate for the normalized iterates and $\tilde{\mathcal{O}}(1/k^2)$-rate for the fixed-point residual of the Halpern iteration. Second, we extend the performance estimation problem (PEP) methodology to inconsistent fixed-point iterations based on a new interpolability result and demonstrate how we used this methodology to discover the upper bounds. Third, we prove a matching $\Omega(1/k^2)$-complexity lower bound and thereby establish that the $\mathcal{O}(1/k^2)$ upper bound is the optimal accelerated rate. Finally, we complement our theoretical results with a numerical experiment on a decentralized semidefinite program (SDP).

2. Measure of optimality

Consider a nonexpansive operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$. Then $\operatorname{Fix}\mathbb{T} = \emptyset$ if and only if $0 \notin \overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}$. In such case, since $\overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}$ is a closed convex set, it has a unique minimum element $v = \operatorname{argmin}_{y \in \overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}} \|y\|^2$. Roughly, v represents the distance from $\mathcal{R}(\mathbb{I}-\mathbb{T})$ to containing 0, or $\mathbb{T}$ being consistent. As long as v remains nonzero, $\mathbb{T}$ will never have a fixed point. For an operator $\mathbb{T}$ which we do not have full access to, if we are able to obtain v approximately from only a sufficient number of first-order oracle calls, then this will save resources including time and computational power.

Given a nonexpansive operator $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$, we measure the rate of convergence to v for following sequences.

Definition 1. We call $\frac{x^k - x^0}{\alpha_k}$ with proper scaling factor $\alpha_k > 0$ a *normalized iterate*, and call $x^k - \mathbb{T}x^k$ a *fixed-point residual*.

Normalized iterate of Picard iteration converges to $-v$ (Applegate et al., 2021b), and fixed-point residual of Picard iteration with averaged operator converges to v (Ryu et al., 2019). Following lemma states that when v^k is either normalized iterate or fixed-point residual at iteration k, strong (norm) convergence of v^k to v is equivalent to the convergence of $\|v^k\|$ to $\|v\|$. Therefore, we measure the rate of convergence for both $\|v^k - v\|^2 \to 0$ and $\|v^k\| - \|v\| \to 0$.

Lemma 2. *Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator and v be its infimal displacement vector. If v^k for $k \in \mathbb{N}$ is either $-\frac{x^k - x^0}{\alpha_k}$ or $x^k - \mathbb{T}x^k$ with assumption that $\alpha_k > 0$ satisfies $-\frac{x^k-x^0}{\alpha_k} \in \overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}$ for all $k \in \mathbb{N}$, then*

$$\langle v^k,\, v\rangle \geq \|v\|^2, \quad k = 1, 2, \dots$$

and

$$\lim_{k\to\infty} v^k = v \quad \Leftrightarrow \quad \lim_{k\to\infty} \|v^k\| = \|v\|.$$

Proof of Lemma 2 is deferred to Appendix A.

2.1. Comparison of two optimality measures

In Section 3, we show upper bounds on the two optimality measures $\|v^k - v\|^2$ and $(\|v^k\| - \|v\|)^2$. Since

$$\|v^k - v\| \geq \|v^k\| - \|v\|$$

by the triangle inequality, the former is the more rigorous optimality measure in the sense that it is no easier to reduce. This makes intuitive sense as $\|v^k - v\|^2$ corresponds to characterizing the rate of $v^k \to v$, which is the convergence of both the magnitude and direction of the vectors, while $(\|v^k\| - \|v\|)^2$ corresponds to characterizing the rate of $\|v^k\| \to \|v\|$, which is the convergence of only the magnitude of the vectors.

The relative difference $\|v^k - v\| \geq \|v^k\| - \|v\|$ do manifest in terms of different constants. For both optimality measures $\|v^k - v\|^2$ and $(\|v^k\| - \|v\|)^2$, the best known upper bound, presented in Corollary 5, is $\frac{4}{k^2}\|x^0 - x_\star\|^2$. On the other hand, the best lower bound for $\|v^k - v\|^2$ is $\frac{4}{k^2}\|x^0 - x_\star\|^2$, while for $(\|v^k\| - \|v\|)^2$ it is $\frac{1}{2k^2}\|x^0 - x_\star\|^2$. So a conclusion of this work is that the two optimality measures are equivalent (up to a constant factor of at most 8) in their optimal worst-case computational complexity.

3. Rate of convergence to v

We study the rate of convergence to v for normalized iterate and fixed-point residual of (KM) and (Halpern). In the last part, we deal with the normalized iterate of general Mann iteration.

3.1. Convergence of KM iteration

Consider the *Krasnosel'skiĭ-Mann iteration* (KM)

$$x^{k+1} = \lambda_{k+1}x^k + (1-\lambda_{k+1})\mathbb{T}x^k, \quad k = 0, 1, \dots, \tag{KM}$$

where $x^0 \in \mathcal{H}$ is a starting point and $\lambda_{k+1} \in [0, 1)$.

Theorem 3 (Convergence rate of normalized iterate)**.** *Let $\{x^k\}_{k\in\mathbb{N}}$ be the iterates of* (KM) *starting from $x^0 \in \mathcal{H}$. For any $\varepsilon > 0$ and $x_\varepsilon \in \mathcal{H}$ such that $\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \leq \min\left\{\frac{\varepsilon^2}{2\|v\|+1},\, 1,\, \varepsilon\right\}$,*

$$\left\|\frac{x^k - x^0}{\sum_{i=1}^k (1-\lambda_i)} + v\right\|^2 \leq \left(\frac{2}{\sum_{i=1}^k(1-\lambda_i)}\|x^0 - x_\varepsilon\| + \varepsilon\right)^2$$

for all $k = 1, 2, \dots$. If we further assume that $v \in \mathcal{R}(\mathbb{I}-\mathbb{T})$, then there exists $x_\star \in \mathcal{H}$ such that $x_\star - \mathbb{T}x_\star = v$ and

$$\left\|\frac{x^k - x^0}{\sum_{i=1}^k (1-\lambda_i)} + v\right\|^2 \leq \frac{4}{\left(\sum_{i=1}^k(1-\lambda_i)\right)^2}\|x^0 - x_\star\|^2$$

for all $k = 1, 2, \dots$.

Theorem 4 (Convergence rate of fixed-point residual)**.** *Let $\{x^k\}_{k\in\mathbb{N}}$ be the iterates of* (KM) *starting from $x^0 \in \mathcal{H}$ and $k_0 = \min\{i \in \mathbb{N} \mid \lambda_i > 0\}$. For any $\varepsilon > 0$ and $x_\varepsilon \in \mathcal{H}$ such that $\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \leq \min\left\{\frac{\varepsilon^2}{2\|v\|+1},\, 1,\, \varepsilon\right\}$,*

$$\left(\sum_{i=0}^k \frac{\lambda_{i+1}(1-\lambda_{i+1})}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\|x^i - \mathbb{T}x^i - v\|\right)^2 \leq \left(\frac{1}{\sqrt{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}}\|x^0 - x_\varepsilon\| + \varepsilon\right)^2$$

and

$$\left(\|x^k - \mathbb{T}x^k\| - \|v\|\right)^2 \leq \left(\frac{1}{\sqrt{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}}\|x^0 - x_\varepsilon\| + \varepsilon\right)^2$$

for $k \geq k_0$. If we further assume that $v \in \mathcal{R}(\mathbb{I}-\mathbb{T})$, then there exists $x_\star \in \mathcal{H}$ such that $x_\star - \mathbb{T}x_\star = v$,

$$\left(\sum_{i=0}^k \frac{\lambda_{i+1}(1-\lambda_{i+1})}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\|x^i - \mathbb{T}x^i - v\|\right)^2 \leq \frac{1}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\|x^0 - x_\star\|^2,$$

and

$$\left(\|x^k - \mathbb{T}x^k\| - \|v\|\right)^2 \leq \frac{1}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\|x^0 - x_\star\|^2$$

for $k \geq k_0$.

We defer the proofs to Appendix B.1. Note that Theorems 3 and 4 imply the convergence of normalized iterate and fixed-point residual to v respectively when $\sum_{k=1}^\infty (1-\lambda_k) = \infty$ and $\sum_{k=1}^\infty \lambda_k(1-\lambda_k) = \infty$. We also point out that the bound on the Cesàro mean in Theorem 4 is practically useful when we use the randomized iterate selection technique of Ghadimi & Lan (2013; 2016): choosing $\bar{k} \in \{1, 2, \dots, k\}$ with probability proportional to $\lambda_{\bar{k}+1}(1-\lambda_{\bar{k}+1})$, fixed-point residual $x^{\bar{k}} - \mathbb{T}x^{\bar{k}}$ of $\bar{k}$-th iterate will yield the same rate of convergence as Theorem 4.

Corollary 5. *Let $\{x^k\}_{k\in\mathbb{N}}$ be the iterates of* (KM) *starting from $x^0 \in \mathcal{H}$. Assume that $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$. The bound of Theorem 3 is optimized at $\lambda_k = 0$ for all $k \in \mathbb{N}$ with*

$$\left\| \frac{x^k - x^0}{k} + v \right\|^2 \leq \frac{4}{k^2} \|x^0 - x_\star\|^2.$$

The bound of Theorem 4 is optimized at $\lambda_k = \frac{1}{2}$ for all $k \in \mathbb{N}$ with

$$\frac{1}{k+1} \sum_{i=0}^{k} \|x^i - \mathbb{T}x^i - v\|^2 \leq \frac{4}{k+1} \|x^0 - x_\star\|^2$$

and

$$(\|x^k - \mathbb{T}x^k\| - \|v\|)^2 \leq \frac{4}{k+1} \|x^0 - x_\star\|^2.$$

Corollary 5 recovers the rates of (Liu et al., 2019, Theorem 3) and (Applegate et al., 2021b, Theorem 3). To clarify, we view the results of Sections 3.2 and 5 to be the major contributions of this work. Our contribution of Section 3.1, presented in Theorems 3 and 4, is to generalize the results of (Liu et al., 2019; Applegate et al., 2021b) to the KM iteration with $\{\lambda_k\}_{k\in\mathbb{N}}$ that varies with k.

Counterexample. Theorems 3 and 4 show that convergence of normalized iterates requires $\sum_{k=1}^{\infty}(1 - \lambda_k) = \infty$, while convergence of fixed-point residual requires the stronger condition $\sum_{k=1}^{\infty} \lambda_k(1 - \lambda_k) = \infty$. The following demonstrates that it is possible for the normalized iterates to converge while the fixed-point residual diverges.

Define $\mathbb{T} \colon \mathbb{R}^3 \to \mathbb{R}^3$ as $\mathbb{T}(x, y, z) = (-y, x, z - 1)$. Then $\mathcal{R}(\mathbb{I} - \mathbb{T}) = \mathbb{R}^2 \times \{1\}$ and $v = (0, 0, 1)$. Let $\{(x^k, y^k, z^k)\}_{k\in\mathbb{N}\cup\{0\}}$ be a sequence of iterates generated by (KM) with $\mathbb{T}$ and $\lambda_k = 0$ for all $k \in \mathbb{N}$ starting from $(x^0, y^0, z^0) = (1, 0, 0)$. Then

$$(x^k, y^k, z^k) = \left(\cos\frac{k\pi}{2}, \sin\frac{k\pi}{2}, -k\right),$$

and the normalized iterates converge to $-v$. However, $\left\|(x^k, y^k, z^k) - \mathbb{T}(x^k, y^k, z^k) - v\right\| = \sqrt{2}$ for all $k \in \mathbb{N}$, so the fixed-point residual does not converge to v.

3.2. Convergence of Halpern iteration

Consider the *Halpern iteration* (Halpern)

$$x^{k+1} = \lambda_{k+1} x^0 + (1 - \lambda_{k+1}) \mathbb{T}x^k, \quad k = 0, 1, \dots, \qquad \text{(Halpern)}$$

where $x^0 \in \mathcal{H}$ is a starting point and $\lambda_{k+1} \in [0, 1)$. Note that (Picard) corresponds $\lambda_k \equiv 0$ and OHM (Lieder, 2021) corresponds to $\lambda_k = \frac{1}{k+1}$. Define $\theta_0 = 0$ and

$$\theta_k = \sum_{n=1}^{k} (1 - \lambda_k)(1 - \lambda_{k-1}) \cdots (1 - \lambda_{k-n+1})$$

for $k = 1, 2, \dots$.

Lemma 6. *For $k = 0, 1, \dots,$*

$$\theta_{k+1} = (1 - \lambda_{k+1})(1 + \theta_k).$$

If $\lambda_k \equiv 0$, then $\theta_k = k$. If $\lambda_k = \frac{1}{k+1}$, then $\theta_k = \frac{k}{2}$.

Theorem 7 (Convergence rate of normalized iterate)**.** *Let $\{x^k\}_{k\in\mathbb{N}}$ be the iterates of* (Halpern) *starting from $x^0 \in \mathcal{H}$. For any $\varepsilon > 0$ and $x_\varepsilon \in \mathcal{H}$ such that $\|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2 \leq \varepsilon^2$,*

$$\left\| \frac{x^k - x^0}{\theta_k} + v \right\|^2 \leq \left(\frac{2}{\theta_k} \|x^0 - x_\varepsilon\| + \varepsilon \right)^2$$

for $k = 1, 2, \dots$. If we further assume that $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, then there exists $x_\star \in \mathcal{H}$ such that $x_\star - \mathbb{T}x_\star = v$ and

$$\left\| \frac{x^k - x^0}{\theta_k} + v \right\|^2 \leq \frac{4}{\theta_k^2} \|x^0 - x_\star\|^2$$

for $k = 1, 2, \dots$.

We defer the proofs to Appendix B.2. Note that the normalized iterates converge to $-v$ if $\theta_k \to \infty$, which, in particular, happens if $\lambda_k \to 0$. See Lemma 23.

Theorem 8 (Convergence rate of fixed-point residual)**.** *Let $\{x^k\}_{k\in\mathbb{N}}$ be the iterates of* (Halpern) *starting from $x^0 \in \mathcal{H}$ with $\lambda_k = \frac{1}{k+1}$. For any $\varepsilon > 0$ and $x_\varepsilon \in \mathcal{H}$ such that $\|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2 \leq \mathcal{O}(\varepsilon^2)$, we have*

$$\left(\|x^k - \mathbb{T}x^k\| - \|v\|\right)^2 \leq \left(\frac{4}{k} \|x^0 - x_\varepsilon\| + \varepsilon \right)^2$$

and

$$\begin{aligned} &\|x^k - \mathbb{T}x^k - v\|^2 \\ &\leq \left(\frac{\sqrt{\sum_{n=1}^{k} \frac{1}{n} + 4} + 1}{k+1} \right)^2 \|x^0 - x_\varepsilon\|^2 + \varepsilon \end{aligned}$$

for $k = 1, 2, \dots$. If we further assume that $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, then there exists $x_\star \in \mathcal{H}$ such that $x_\star - \mathbb{T}x_\star = v$,

$$\left(\|x^k - \mathbb{T}x^k\| - \|v\|\right)^2 \leq \frac{16}{k^2} \|x^0 - x_\star\|^2,$$

and

$$\|x^k - \mathbb{T}x^k - v\|^2 \leq \left(\frac{\sqrt{\sum_{n=1}^{k} \frac{1}{n} + 4} + 1}{k+1} \right)^2 \|x^0 - x_\star\|^2$$

for $k = 1, 2, \dots$.

Proof outline. Consider a potential function V^k defined as

$$\begin{aligned} &V^k \\ &= (k+1)\left\{k\|x^k - \mathbb{T}x^k\|^2 + 2\langle x^k - \mathbb{T}x^k, x^k - x^0\rangle\right\} \\ &+ k(k+1)\left\langle -\frac{2}{k}(x^k - x^0) - (x_\varepsilon - \mathbb{T}x_\varepsilon), x_\varepsilon - \mathbb{T}x_\varepsilon\right\rangle \\ &+ \frac{2(k+1)}{k}\left\|x^k - x_\varepsilon + \frac{k}{2}(x_\varepsilon - \mathbb{T}x_\varepsilon)\right\|^2 \\ &- \left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2 \end{aligned}$$

for all $k \in \mathbb{N}$. We can show $V^k \le V^{k-1} \le \cdots \le V^1$. From $V^k \le V^1 \le 3\|x^0 - x_\varepsilon\|^2$, we obtain the desired convergence rate. When there exists $x_\star$ such that $v = x_\star - \mathbb{T}x_\star$, use $x_\star$ instead of x_ε. The detailed proof is deferred to Appendix B.2. ☐

The precise form of the $\mathcal{O}(\varepsilon^2)$-term in Theorem 8 is stated in the proof, which is deferred to Appendix B.2. Note that in V^k, the first term, written as $(k+1)\{\cdots\}$, is the potential function that was used in prior work (Diakonikolas, 2020; Park & Ryu, 2022) to analyze the convergence of consistent fixed-point iterations. So the first term is known to be nonincreasing, and the three additional terms are required to adapt the proof to the inconsistent case.

Corollary 9. *Let $\{x^k\}_{k\in\mathbb{N}}$ be the iterates of* (Halpern) *starting from $x^0 \in \mathcal{H}$ with $\lambda_k = \frac{1}{k+1}$. For any $\varepsilon > 0$, there exists $x_\varepsilon \in \mathcal{H}$ such that $\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| < \varepsilon$,*

$$\left\|\frac{2(x^k - x^0)}{k} + v\right\|^2 \le \left(\frac{4}{k}\|x^0 - x_\varepsilon\| + \varepsilon\right)^2,$$

$$\left(\|x^k - \mathbb{T}x^k\| - \|v\|\right)^2 \le \left(\frac{4}{k}\|x^0 - x_\varepsilon\| + \varepsilon\right)^2,$$

and

$$\begin{aligned} &\|x^k - \mathbb{T}x^k - v\|^2 \\ &\le \left(\frac{\sqrt{\sum_{n=1}^{k}\frac{1}{n} + 4} + 1}{k+1}\right)^2 \|x^0 - x_\varepsilon\|^2 + \varepsilon \end{aligned}$$

for $k = 1, 2, \ldots$. If we further assume that $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, then there exists $x_\star \in \mathcal{H}$ such that $x_\star - \mathbb{T}x_\star = v$,

$$\left\|\frac{2(x^k - x^0)}{k} + v\right\|^2 \le \frac{16}{k^2}\|x^0 - x_\star\|^2,$$

$$\left(\|x^k - \mathbb{T}x^k\| - \|v\|\right)^2 \le \frac{16}{k^2}\|x^0 - x_\star\|^2,$$

and

$$\|x^k - \mathbb{T}x^k - v\|^2 \le \left(\frac{\sqrt{\sum_{n=1}^{k}\frac{1}{n} + 4} + 1}{k+1}\right)^2 \|x^0 - x_\star\|^2$$

for $k = 1, 2, \ldots$.

An observation we point out is that when $\mathbb{T}$ is an affine operator, the normalized iterate $-\frac{x^{k+1} - x^0}{k+1}$ of Picard iteration coincides with the fixed-point residual $x^k - \mathbb{T}x^k$ of (Halpern) with $\lambda_k = \frac{1}{k+1}$. See Lemma 32.

3.3. Convergence of Mann iteration

The *Mann iteration* (Mann)

$$x^k = \sum_{i=0}^{k-1}\nu_i^k\,\mathbb{T}x^{i-1}, \qquad \text{(Mann)}$$

where $\nu_i^k \ge 0$ for $i = 0, \ldots, k$ and $k = 1, 2, \ldots$, $\sum_{i=0}^{k}\nu_i^k = 1$ for $k = 1, 2, \ldots$, and $\mathbb{T}x^{-1} := x^0$, is a further general class of iterations including (KM) and (Halpern). Lemma 33 of Appendix B.3 shows that there exists positive sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ that depends on $\{\nu_i^k\}_{0\le i\le k, k\in\mathbb{N}}$ such that

$$-\frac{x^k - x^0}{\alpha_k} \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}, \quad k = 1, 2, \ldots.$$

Furthermore, Theorem 36 of Appendix B.3 shows that

$$\left\|\frac{x^k - x^0}{\alpha_k} + v\right\|^2 \le \left(\frac{2}{\alpha_k}\|x^0 - x_\varepsilon\| + \varepsilon\right)^2$$

and the normalized iterate converges to $-v$ if $\alpha_k \to \infty$. This result generalizes the convergence results of Theorems 3 and 7 respectively for (KM) and (Halpern).

4. PEP with possibly infeasible operators

Instrumental in the discovery of the results of Section 3 was the use of the performance estimation problem (PEP) (Drori & Teboulle, 2014; Taylor et al., 2017). Loosely speaking, the PEP is a computer-assisted methodology for finding optimal methods by numerically solving semidefinite programs (Drori & Teboulle, 2014; 2016; Kim & Fessler, 2016; Taylor et al., 2018b; Drori & Taylor, 2020; Kim & Fessler, 2021; Kim, 2021; Park & Ryu, 2022). In prior work, PEP had been utilized in the analysis of *consistent* monotone inclusion and fixed-point problems (Ryu et al., 2020; Kim, 2021; Park & Ryu, 2022). In this section, we describe how to apply the PEP methodology in the analysis of algorithms for *inconsistent* problems.

4.1. Interpolation result

The performance estimation problem framework relies on certain interpolation results. The following result strengthens the prior interpolation result of (Ryu et al., 2020, Fact 2) by additionally restricting the range of the extension and thereby allows us to control the infimal displacement vector of the interpolation.

Theorem 10 (Interpolability)**.** *Let $\{(x_i, y_i)\}_{i\in I} \subset \mathcal{H}\times\mathcal{H}$ be a set of vectors with index set I such that*

$$\|y_i - y_j\| \le \|x_i - x_j\|, \quad \forall\, i,j \in I.$$

Let $C = \overline{\mathrm{conv}}\,\{x_i - y_i\}_{i\in I} \subseteq \mathcal{H}$, where $\overline{\mathrm{conv}}$ denotes the closure of the convex hull.

(i) There exists a nonexpansive $\widetilde{\mathbb{T}}\colon \mathcal{H}\to\mathcal{H}$ such that

$$y_i = \widetilde{\mathbb{T}} x_i, \quad \forall\, i \in I$$

and $v = \Pi_C(0)$ is its infimal displacement vector.

(ii) If we further assume that $v = x_\star - y_\star$, $\star \in I$ and

$$\langle x_i - y_i,\, v\rangle \ge \|v\|^2, \quad \forall\, i \in I,$$

then there exists a nonexpansive $\widetilde{\mathbb{T}}\colon \mathcal{H}\to\mathcal{H}$ such that

$$y_i = \widetilde{\mathbb{T}} x_i, \quad \forall\, i \in I$$

and v is its infimal displacement vector.

We defer the proof to Appendix C.1. The key insight is to use the range/domain-restricting extension of (Reich & Simons, 2005; Bauschke, 2007), construction of which, in turn, relies on the Fitzpatrick function (Fitzpatrick, 1988).

4.2. PEP formulation

We now describe the PEP formulation with inconsistent operators through an example. Consider (Halpern) with $\lambda_k = \frac{1}{k+1}$, which we refer to as the optimized Halpern method (OHM) of (Lieder, 2021). Let $k\in\mathbb{N}$ and define the index set $I = \{0,1,\dots,k,\star\}$. We consider nonexpansive operators $\mathbb{T}$ that have an infimal displacement vector v and a point $x_\star\in\mathcal{H}$ such that $v = x_\star - \mathbb{T}x_\star$. The goal is to find the worst-case instance of $\mathbb{T}$ such that $\|x^k - \mathbb{T}x^k - v\|^2$ is maximized.

We start from the infinite-dimensional performance estimation problem

$$\begin{array}{ll} \underset{\mathbb{T}}{\text{maximize}} & \|x^k - \mathbb{T}x^k - v\|^2 \\ \text{subject to} & \mathbb{T}\colon \mathcal{H}\to\mathcal{H} \text{ is nonexpansive} \\ & v = \Pi_{\overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}}(0) = x_\star - \mathbb{T}x_\star \\ & x^{n+1} = \frac{n+1}{n+2}\mathbb{T}x^n + \frac{1}{n+2}x^0 \\ & \|x^0 - x_\star\|^2 \le R^2 \end{array}$$

where $n = 0,1,\dots,k-1$. Using Theorem 10 and scaling by R, we get the equivalent non-convex finite-dimensional problem

$$\begin{array}{ll} \underset{(x^i,\,y^i)_{i\in I}}{\text{maximize}} & \|x^k - y^k - v\|^2 \\ \text{subject to} & \|y^i - y^j\|^2 \le \|x^i - x^j\|^2, \quad \forall i,j\in I,\, i\ne j \\ & v = x_\star - y_\star \\ & \left\langle x^i - y^i,\, v\right\rangle \ge \|v\|^2, \quad \forall i \in I \\ & x^{n+1} = \frac{n+1}{n+2}y^n + \frac{1}{n+2}x^0 \\ & \|x^0 - x_\star\|^2 \le 1 \end{array}$$

where $n = 0,1,\dots,k-1$. Next, consider the following Gram matrix $Z = G^\intercal G \in \mathbb{S}_+^{k+3}$, where

$$G = \begin{bmatrix} v^0 & \cdots & v^k & v & x^0 - x_\star \end{bmatrix} \tag{1}$$

with $v^i = x^i - y^i$ for $i = 0,1,\dots,k$. We finally obtain the following equivalent (convex) semidefinite program,

$$\begin{array}{ll} \underset{Z\in\mathbb{S}_+^{k+3}}{\text{maximize}} & \mathrm{tr}(C_k Z) \\ \text{subject to} & \mathrm{tr}(A_{i,j}Z) \ge 0, \quad \forall\, i,j \in I\setminus\{\star\},\, i\ne j \\ & \mathrm{tr}(A_{i,\star}Z) \ge 0, \quad \forall\, i \in I\setminus\{\star\} \\ & \mathrm{tr}(B_i Z) \le 0, \quad \forall\, i\in I\setminus\{\star\} \\ & \mathrm{tr}(D_0 Z) \le 1, \end{array}$$

where $A_{i,j}$, $A_{i,\star}$, and B_i for $i,j\in I\setminus\{\star\}$, C_k, and D_0 in $\mathbb{S}^{k+3}$ are all defined in Appendix C.2. The details and the subtleties of deriving the SDP representation are also further discussed in Appendix C.2.

5. Complexity lower bound

In this section, we establish a lower bound on the computational complexity of approximating the infimal displacement vector v. Following the information-based complexity framework (Nemirovski, 1992), we begin by considering algorithms satisfying the linear span condition

$$x^{k+1} = x^0 + \mathrm{span}\{x^0 - \mathbb{T}x^0, x^1 - \mathbb{T}x^1, \dots, x^k - \mathbb{T}x^k\}, \tag{span}$$

which covers a broad range of fixed-point iterations including (KM), (Halpern), (Mann), Anderson acceleration and many more. We then remove the linear span assumption and expand the class of algorithm to all "deterministic fixed-point iterations."

Theorem 11. *Let $k\in\mathbb{N}$, $x^0 = 0\in\mathcal{H}$, and $v\in\mathcal{H}$, where $\dim\mathcal{H}\ge k+1$. Then, there exists a nonexpansive operator $\mathbb{T}\colon\mathcal{H}\to\mathcal{H}$ and $x_\star\in\mathcal{H}$ such that $v = x_\star - \mathbb{T}x_\star$, v becomes the infimal displacement vector of $\mathbb{T}$, and*

$$\left(\left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{T}x^i)\right\| - \|v\|\right)^2 \ge \frac{1}{2k^2}\|x^0 - x_\star\|^2$$

and

$$\left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{T}x^i) - v\right\|^2 \ge \frac{4}{k^2}\|x^0 - x_\star\|^2$$

hold for any iterates $\{x^n\}_{n=0}^{k-1}$ satisfying (span) *and any choice of real numbers $\{\nu_i\}_{i=0}^{k-1}$ such that $\sum_i \nu_i = 1$.*

Proof outline. We construct a nonexpansive operator $\mathbb{T}\colon\mathbb{R}^{k+1}\to\mathbb{R}^{k+1}$ with its infimal displacement $\tilde{v} = (0,\dots,0,\|v\|)$. Then, we choose an orthogonal matrix $U\in\mathbb{R}^{(k+1)\times(k+1)}$ such that $U^\intercal U = \mathbb{I}$ and $U^\intercal\tilde{v} = v$

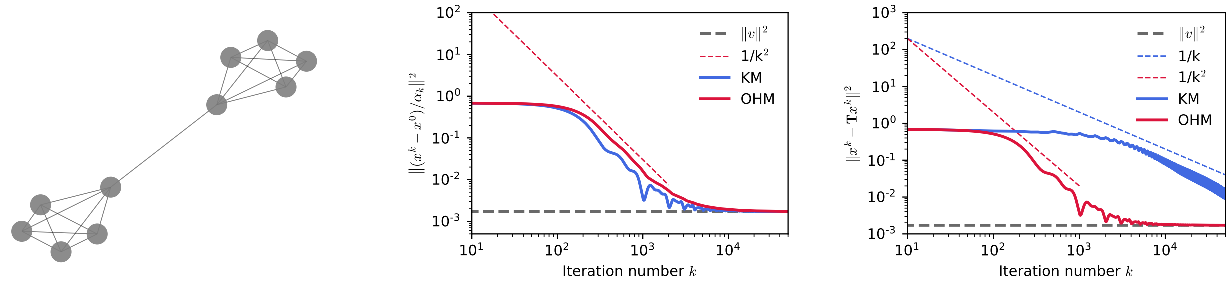

Figure 1. Solving SDP with $50,000$ iterations of PG-EXTRA (Picard) and OHM with PG-EXTRA (Halpern). We use an infeasible instance, whose setups are described in Appendix E. Parameters are $n = 10$, $m = 11$, $p = 10$ with $\alpha = \beta = 0.01$. (Left) Network graph. (Middle) Squared norm of normalized iterate $\|(x^k - x^0)/\alpha_k\|^2$. (Right) Squared norm of fixed-point residual $\|x^k - \mathbb{T}x^k\|^2$.

to construct a nonexpansive operator $\mathbb{T}_U = U\mathbb{T}U^\intercal$ whose infimal displacement vector is v. Our specific construction, inspired by Park & Ryu (2022), is

$$x - \mathbb{T}x = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 1 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}}_{\in \mathbb{R}^{(k+1)\times(k+1)}} x + \begin{bmatrix} \alpha \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \|v\| \end{bmatrix}$$

for all $x \in \mathbb{R}^{k+1}$ with $\alpha \neq 0$. We provide the detailed proof in Appendix D. □

Matching upper and lower bounds. The $\frac{4}{k^2}\|x^0 - x_\star\|^2$ upper bound on (Picard) for $-\frac{x^k - x^0}{k} \to v$ of Corollary 5 exactly matches the $\frac{4}{k^2}\|x^0 - x_\star\|^2$ lower bound of Theorem 11. The upper bounds on (KM) with $\lambda_k \equiv \lambda \in (0,1)$ and (Halpern) with $\lambda_k = \frac{1}{k+1}$ of Corollary 9 match the lower bound up to a constant.

The $\mathcal{O}(\frac{\log k}{k^2})$ upper bound of (Halpern) for $x^k - \mathbb{T}x^k \to v$ matches the lower bound up to logarithmic factors, and this is the fastest known rate for the convergence of the fixed-point residual $x^k - \mathbb{T}x^k$ to v. However, the $\mathcal{O}(1/k^2)$ upper bound of (Halpern) for $\|x^k - \mathbb{T}x^k\| \to \|v\|$ does match the lower bound up to a constant.

Finally, the upper bound

$$(\|v^k\| - \|v\|)^2 \leq (\|v^k - v\|)^2 \leq \frac{4}{k^2}\|x^0 - x_\star\|^2$$

of Corollary 5 and the $\frac{1}{2k^2}\|x^0 - x_\star\|^2$ lower bound of $\|v^k\| \to \|v\|$ of Theorem 11 match only up to a constant factor 8 (where v^k are as defined in Theorem 11 and Corollary 5). Reducing this gap may be an interesting direction of future work.

Lower bounds for deterministic iterations. Finally, we use the resisting oracle technique of Nemirovski & Yudin (1983) to extend the complexity lower bound to general deterministic fixed-point iterations, an algorithm class we formally define in Appendix D. The following result no longer requires the linear span assumption (span).

Theorem 12. *Let $k \in \mathbb{N}$, $x^0 \in \mathcal{H}$, and $v \in \mathcal{H}$, where $\dim \mathcal{H} \geq 2k-1$. Then, there exists a nonexpansive operator $\mathbb{T} \colon \mathcal{H} \to \mathcal{H}$ and $x_\star \in \mathcal{H}$ such that $v = x_\star - \mathbb{T}x_\star$, the infimal displacement vector of $\mathbb{T}$ is v, and*

$$\left(\left\|\sum_{i=0}^{k-1} \nu_i(x^i - \mathbb{T}x^i)\right\| - \|v\|\right)^2 \geq \frac{1}{2k^2}\|x^0 - x_\star\|^2$$

and

$$\left\|\sum_{i=0}^{k-1} \nu_i(x^i - \mathbb{T}x^i) - v\right\|^2 \geq \frac{4}{k^2}\|x^0 - x_\star\|^2$$

hold for iterates $\{x^n\}_{n=0}^{k-1}$ generated by any deterministic fixed-point iteration and any choice of real numbers $\{\nu_i\}_{i=0}^{k-1}$ such that $\sum_i \nu_i = 1$.

The proof of Theorem 12 is deferred to Appendix D.

6. Experiments

Consider an infeasible semidefinite problem (SDP)

$$\begin{array}{ll} \underset{x \in \mathbb{R}^d}{\text{minimize}} & \sum_{i=1}^p c_i^\intercal x \\ \text{subject to} & \mathcal{A}_i[x] = \sum_{j=1}^d A_i^j x_j \preceq B_i, \quad 1 \leq i \leq p, \end{array}$$

where A_i^j, $B_i \in \mathbb{S}^n$ and $\mathcal{A}_i \colon \mathbb{R}^d \to \mathbb{S}^n$ is a linear operator defined by $\mathcal{A}_i[x] = \sum_{j=1}^d A_i^j x_j$.

Consider a setup where each objective function $c_i^\intercal x$ and i-th constraint $\mathcal{A}_i[x] \preceq B_i$ are private to the local agent $i \in \{1, \dots, p\}$. Assume that they communicate only with their neighbors, which are represented in the graph as connected nodes. This SDP can be solved in decentralized manner with PG-EXTRA of Shi et al. (2015). See Appendix E for the details of infeasible SDP instance, derivation of PG-EXTRA for SDP, and the choices of parameters.

Figure 1 compares the results of PG-EXTRA and PG-EXTRA combined with OHM. Both algorithms' normalized iterates and fixed-point residuals converged to v, but OHM is faster for fixed-point residual, as our theory suggests.

7. Conclusions

In this work, we analyzed the convergence rates of fixed-point iterations towards the infimal displacement vector. By providing matching upper and lower bounds, we established the optimal accelerated complexity to be $\mathcal{O}(1/k^2)$. The discovery of our upper bounds was assisted by the performance estimation problem (PEP) methodology, which we extended to accommodate inconsistent problem setups.

In our view, the analysis of optimization algorithms applied to inconsistent problems is a necessary step in designing robust general-purpose solvers. Carrying out similar analyses for different algorithms under different inconsistent problems is an interesting direction of future work, and we expect our newly extended PEP methodology to be broadly useful in such endeavors.

Acknowledgements

JP and EKR were supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [NO.2021-0-01343, Artificial Intelligence Graduate School Program (Seoul National University)] and the Samsung Science and Technology Foundation (Project Number SSTF-BA2101-02). We thank Uijeong Jang for providing careful reviews and valuable feedback. We thank Bartolomeo Stellato and Wotao Yin for the discussion on practical aspects of infeasibility detection in conic program solvers.

References

Abbaszadehpeivasti, H., de Klerk, E., and Zamani, M. On the rate of convergence of the difference-of-convex algorithm (DCA). *arXiv preprint arXiv:2109.13566*, 2021.

Abbaszadehpeivasti, H., de Klerk, E., and Zamani, M. Conditions for linear convergence of the gradient method for non-convex optimization. *arXiv preprint arXiv:2204.00647*, 2022a.

Abbaszadehpeivasti, H., de Klerk, E., and Zamani, M. Convergence rate analysis of randomized and cyclic coordinate descent for convex optimization through semidefinite programming. *arXiv preprint arXiv:2212.12384*, 2022b.

Abbaszadehpeivasti, H., de Klerk, E., and Zamani, M. The exact worst-case convergence rate of the gradient method with fixed step lengths for l-smooth functions. *Optimization Letters*, 16(6):1649–1661, 2022c.

Anderson, D. G. Iterative procedures for nonlinear integral equations. *Journal of the Association for Computing Machinery*, 12(4):547–560, 1965.

Applegate, D., Díaz, M., Hinder, O., Lu, H., Lubin, M., O'Donoghue, B., and Schudy, W. Practical large-scale linear programming using primal-dual hybrid gradient. *Neural Information Processing Systems*, 2021a.

Applegate, D., Díaz, M., Lu, H., and Lubin, M. Infeasibility detection with primal-dual hybrid gradient for large-scale linear programming. *arXiv preprint arXiv:2102.04592*, 2021b.

ApS, M. Mosek optimization toolbox for Matlab. *User's Guide and Reference Manual, Version*, 4, 2019.

Aragón Artacho, F. J., Borwein, J. M., and Tam, M. K. Global behavior of the Douglas–Rachford method for a nonconvex feasibility problem. *Journal of Global Optimization*, 65(2):309–327, 2016.

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

Artacho, F. J. A., Borwein, J. M., and Tam, M. K. Douglas–Rachford feasibility methods for matrix completion problems. *The Australian and New Zealand Industrial and Applied Mathematics Journal*, 55(4):299–326, 2014.

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

Banjac, G. On the minimal displacement vector of the Douglas–Rachford operator. *Operations Research Letters*, 49(2):197–200, 2021.

Banjac, G. and Lygeros, J. On the asymptotic behavior of the Douglas–Rachford and proximal-point algorithms for convex optimization. *Optimization Letters*, 15(8): 2719–2732, 2021.

Banjac, G., Goulart, P., Stellato, B., and Boyd, S. Infeasibility detection in the alternating direction method of multipliers for convex optimization. *Journal of Optimization Theory and Applications*, 183(2):490–519, 2019.

Barré, M., Taylor, A., and d'Aspremont, A. Complexity guarantees for Polyak steps with momentum. In *Conference on Learning Theory*, 2020.

Barré, M., Taylor, A., and d'Aspremont, A. Convergence of constrained vector extrapolation scheme. *SIAM Journal on Mathematics of Data Science*, 4(3):979–1002, 2022a.

Barré, M., Taylor, A. B., and Bach, F. Principled analyses and design of first-order methods with inexact proximal operators. *Mathematical Programming*, pp. 1–46, 2022b.

Bauschke, H. Fenchel duality, Fitzpatrick functions and the extension of firmly nonexpansive mappings. *Proceedings of the American Mathematical Society*, 135(1):135–139, 2007.

Bauschke, H. H. and Borwein, J. M. Dykstra's alternating projection algorithm for two sets. *Journal of Approximation Theory*, 79(3):418–443, 1994.

Bauschke, H. H. and Combettes, P. L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, second edition, 2017.

Bauschke, H. H. and Moursi, W. M. The magnitude of the minimal displacement vector for compositions and convex combinations of firmly nonexpansive mappings. *Optimization Letters*, 12(7):1465–1474, 2018.

Bauschke, H. H. and Moursi, W. M. On the behavior of the Douglas–Rachford algorithm for minimizing a convex function subject to a linear constraint. *SIAM Journal on Optimization*, 30(3):2559–2576, 2020a.

Bauschke, H. H. and Moursi, W. M. On the minimal displacement vector of compositions and convex combinations of nonexpansive mappings. *Foundations of Computational Mathematics*, 20(6):1653–1666, 2020b.

Bauschke, H. H. and Moursi, W. M. On the Douglas-Rachford algorithm for solving possibly inconsistent optimization problems. *arXiv preprint arXiv:2106.11547*, 2021.

Bauschke, H. H., Borwein, J. M., and Lewis, A. S. The method of cyclic projections for closed convex sets in Hilbert space. *Contemporary Mathematics*, 204:1–38, 1997.

Bauschke, H. H., Combettes, P. L., and Luke, D. R. Finding best approximation pairs relative to two closed convex sets in Hilbert spaces. *Journal of Approximation theory*, 127(2):178–192, 2004.

Bauschke, H. H., Hare, W. L., and Moursi, W. M. Generalized solutions for the sum of two maximally monotone operators. *SIAM Journal on Control and Optimization*, 52(2):1034–1047, 2014.

Bauschke, H. H., Douglas, G. R., and Moursi, W. M. On a result of Pazy concerning the asymptotic behaviour of nonexpansive mappings. *Journal of Fixed Point Theory and Applications*, 18(2):297–307, 2016.

Borwein, J. M. and Tam, M. K. The cyclic Douglas–Rachford method for inconsistent feasibility problems. *Journal of Nonlinear and Convex Analysis*, 16(4):573–584, 2015.

Boyle, J. P. and Dykstra, R. L. A method for finding projections onto the intersection of convex sets in Hilbert spaces. In *Advances in Order Restricted Statistical Inference*, pp. 28–47. Springer, 1986.

Browder, F. E. and Petryshyn, W. The solution by iteration of nonlinear functional equations in Banach spaces. *Bulletin of the American Mathematical Society*, 72(3): 571–575, 1966.

Bruck Jr, R. E. 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(1):159–164, 1977.

Carmon, Y., Duchi, J. C., Hinder, O., and Sidford, A. Lower bounds for finding stationary points I. *Mathematical Programming*, 184:71–120, 2020.

Carmon, Y., Duchi, J. C., Hinder, O., and Sidford, A. Lower bounds for finding stationary points II: first-order methods. *Mathematical Programming*, 185:315–355, 2021.

Chambolle, A. and Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

Colla, S. and Hendrickx, J. M. Automated worst-case performance analysis of decentralized gradient descent. *Conference on Decision and Control*, 2021.

Colla, S. and Hendrickx, J. M. Automated performance estimation for decentralized optimization via network size independent problems. *Conference on Decision and Control*, 2022a.

Colla, S. and Hendrickx, J. M. Automatic performance estimation for decentralized optimization. *arXiv preprint arXiv:2203.05963*, 2022b.

Davis, D. Convergence rate analysis of primal-dual splitting schemes. *SIAM Journal on Optimization*, 25(3):1912–1943, 2015.

De Klerk, E., Glineur, F., and Taylor, A. B. On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions. *Optimization Letters*, 11(7):1185–1199, 2017.

De Klerk, E., Glineur, F., and Taylor, A. B. Worst-case convergence analysis of inexact gradient and Newton methods through semidefinite programming performance estimation. *SIAM Journal on Optimization*, 30(3):2053–2082, 2020.

De Loera, J. A., Malkin, P. N., and Parrilo, P. A. Computation with polynomial equations and inequalities arising in combinatorial optimization. In Lee, J. and Leyffer, S. (eds.), *Mixed Integer Nonlinear Programming*, pp. 447–481. Springer, 2012.

Diakonikolas, J. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. *Conference on Learning Theory*, 2020.

Dong, Q., Yuan, H., Cho, Y., and Rassias, T. M. Modified inertial Mann algorithm and inertial CQ-algorithm for nonexpansive mappings. *Optimization Letters*, 12(1): 87–102, 2018.

Dragomir, R.-A., Taylor, A. B., d'Aspremont, A., and Bolte, J. Optimal complexity and certification of Bregman first-order methods. *Mathematical Programming*, 194(1):41–83, 2022.

Drori, Y. The exact information-based complexity of smooth convex minimization. *Journal of Complexity*, 39:1–16, 2017.

Drori, Y. and Shamir, O. The complexity of finding stationary points with stochastic gradient descent. *International Conference on Machine Learning*, 2020.

Drori, Y. and Taylor, A. On the oracle complexity of smooth strongly convex minimization. *Journal of Complexity*, 68, 2022.

Drori, Y. and Taylor, A. B. Efficient first-order methods for convex minimization: a constructive approach. *Mathematical Programming*, 184(1–2):183–220, 2020.

Drori, Y. and Teboulle, M. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145(1–2):451–482, 2014.

Drori, Y. and Teboulle, M. An optimal variant of Kelley's cutting-plane method. *Mathematical Programming*, 160 (1–2):321–351, 2016.

Dykstra, R. L. An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, 78(384):837–842, 1983.

Eckstein, J. and Bertsekas, D. P. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.

Fitzpatrick, S. Representing monotone operators by convex functions. In Fitzpatrick, S. and Giles, J. (eds.), *Proceedings of the Centre for Mathematics and its Applications*, pp. 59–65. Australian National University, Mathematical Sciences Institute, 1988.

Gabay, D. and Mercier, B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.

Garstka, M., Cannon, M., and Goulart, P. COSMO: A conic operator splitting method for convex conic problems. *Journal of Optimization Theory and Applications*, 190(3):779–810, 2021.

Ghadimi, S. and Lan, G. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Ghadimi, S. and Lan, G. Accelerated gradient methods for nonconvex nonlinear and stochastic programming. *Mathematical Programming*, 156(1-2):59–99, 2016.

Glowinski, R. and Marroco, A. 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 (R2):41–76, 1975.

Gorbunov, E., Taylor, A., and Gidel, G. Last-iterate convergence of optimistic gradient method for monotone variational inequalities. *Neural Information Processing Systems*, 2022.

Gu, G. and Yang, J. Tight sublinear convergence rate of the proximal point algorithm for maximal monotone inclusion problems. *SIAM Journal on Optimization*, 30(3): 1905–1921, 2020.

Gu, G. and Yang, J. Tight ergodic sublinear convergence rate of the relaxed proximal point algorithm for monotone variational inequalities. *Journal of Optimization Theory and Applications*, 2022.

Gu, G. and Yang, J. Tight convergence rate in subgradient norm of the proximal point algorithm. *arXiv preprint arXiv:2301.03175*, 2023.

Gupta, S. D., Freund, R. M., Sun, X. A., and Taylor, A. Nonlinear conjugate gradient methods: worst-case convergence rates via computer-assisted analyses. *arXiv preprint arXiv:2301.01530*, 2023.

Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2023. URL `https://www.gurobi.com`.

Halpern, B. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73(6):957–961, 1967.

Ishikawa, S. Fixed points and iteration of a nonexpansive mapping in a Banach space. *Proceedings of the American Mathematical Society*, 59(1):65–71, 1976.

Kamri, Y., Hendrickx, J. M., and Glineur, F. On the worst-case analysis of cyclic coordinate-wise algorithms on smooth convex functions. *arXiv preprint arXiv:2211.17018*, 2022.

Kim, D. Accelerated proximal point method for maximally monotone operators. *Mathematical Programming*, 190 (1–2):57–87, 2021.

Kim, D. and Fessler, J. A. Optimized first-order methods for smooth convex minimization. *Mathematical programming*, 159(1):81–107, 2016.

Kim, D. and Fessler, J. A. On the convergence analysis of the optimized gradient method. *Journal of Optimization Theory and Applications*, 172(1):187–205, 2017.

Kim, D. and Fessler, J. A. Another look at the fast iterative shrinkage/thresholding algorithm (FISTA). *SIAM Journal on Optimization*, 28(1):223–250, 2018.

Kim, D. and Fessler, J. A. Optimizing the efficiency of first-order methods for decreasing the gradient of smooth convex functions. *Journal of Optimization Theory and Applications*, 188(1):192–219, 2021.

Krasnosel'skiĭ, M. A. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10:123–127, 1955.

Lieder, F. On the convergence rate of the Halpern-iteration. *Optimization Letters*, 15(2):405–418, 2021.

Lions, P.-L. and Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

Liu, Y., Ryu, E. K., and Yin, W. A new use of Douglas–Rachford splitting for identifying infeasible, unbounded, and pathological conic programs. *Mathematical Programming*, 177(1):225–253, 2019.

Maingé, P.-E. Convergence theorems for inertial KM-type algorithms. *Journal of Computational and Applied Mathematics*, 219(1):223–236, 2008.

Mann, W. R. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4(3):506–510, 1953.

Mattingley, J. and Boyd, S. CVXGEN: A code generator for embedded convex optimization. *Optimization and Engineering*, 13(1):1–27, 2012.

Moucer, C., Taylor, A., and Bach, F. A systematic approach to Lyapunov analyses of continuous-time models in convex optimization. *arXiv preprint arXiv:2205.12772*, 2022.

Moursi, W. M. and Saurette, M. 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.

Naik, V. V. and Bemporad, A. Embedded mixed-integer quadratic optimization using accelerated dual gradient projection. *IFAC-PapersOnLine*, 50(1):10723–10728, 2017.

Nemirovski, A. S. Information-based complexity of linear operator equations. *Journal of Complexity*, 8(2):153–175, 1992.

Nemirovski, A. S. and Yudin, D. B. *Problem Complexity and Method Efficiency in Optimization*. Wiley-Interscience, 1983.

Nesterov, Y. *Introductory Lectures on Convex Optimization: A Basic Course*. Springer, 2004.

Nesterov, Y. and Nemirovskii, A. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

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

O'Donoghue, B. Operator splitting for a homogeneous embedding of the linear complementarity problem. *SIAM Journal on Optimization*, 31(3):1999–2023, 2021.

O'Donoghue, B., Chu, E., Parikh, N., and Boyd, S. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, 2016.

Park, J. and Ryu, E. K. Exact optimal accelerated complexity for fixed-point iterations. *International Conference on Machine Learning*, 2022.

Pazy, A. Asymptotic behavior of contractions in Hilbert space. *Israel Journal of Mathematics*, 9(2):235–240, 1971.

Plant, A. T. and Reich, S. The asymptotics of nonexpansive iterations. *Journal of functional analysis*, 54(3):308–319, 1983.

Raghunathan, A. U. and Di Cairano, S. Infeasibility detection in alternating direction method of multipliers for convex quadratic programs. *Conference on Decision and Control*, 2014.

Reich, S. Asymptotic behavior of contractions in Banach spaces. *Journal of Mathematical Analysis and Applications*, 44(1):57–70, 1973.

Reich, S. On the asymptotic behavior of nonlinear semigroups and the range of accretive operators. I. *Journal of Mathematical Analysis and Applications*, 79(1):113–126, 1981.

Reich, S. On the asymptotic behavior of nonlinear semigroups and the range of accretive operators. II. *Journal of Mathematical Analysis and Applications*, 87(1):134–146, 1982.

Reich, S. and Shafrir, I. The asymptotic behavior of firmly nonexpansive mappings. *Proceedings of the American Mathematical Society*, 10(2):246–250, 1987.

Reich, S. and Simons, S. Fenchel duality, Fitzpatrick functions and the Kirszbraun–Valentine extension theorem. *Proceedings of the American Mathematical Society*, 133 (9):2657–2660, 2005.

Reich, S., Thong, D. V., Cholamjiak, P., and Van Long, L. Inertial projection-type methods for solving pseudomonotone variational inequality problems in Hilbert space. *Numerical Algorithms*, 88(2):813–835, 2021.

Rotaru, T., Glineur, F., and Patrinos, P. Tight convergence rates of the gradient method on hypoconvex functions. *arXiv preprint arXiv:2203.00775*, 2022.

Ryu, E. K. Cosmic divergence, weak cosmic convergence, and fixed points at infinity. *Journal of Fixed Point Theory and Applications*, 20(3):1–16, 2018.

Ryu, E. K. and Yin, W. *Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators*. Cambridge University Press, 2022.

Ryu, E. K., Liu, Y., and Yin, W. Douglas–Rachford splitting and ADMM for pathological convex optimization. *Computational Optimization and Applications*, 74(3):747–778, 2019.

Ryu, E. K., Taylor, A. B., Bergeling, C., and Giselsson, P. Operator splitting performance estimation: Tight contraction factors and optimal parameter selection. *SIAM Journal on Optimization*, 30(3):2251–2271, 2020.

Sabach, S. and Shtern, S. A first order method for solving convex bilevel optimization problems. *SIAM Journal on Optimization*, 27(2):640–660, 2017.

Scieur, D., d'Aspremont, A., and Bach, F. Regularized nonlinear acceleration. *Mathematical Programming*, 179 (1–2):47–83, 2020.

Shehu, Y. Convergence rate analysis of inertial Krasnoselskii–Mann type iteration with applications. *Numerical Functional Analysis and Optimization*, 39(10): 1077–1091, 2018.

Shehu, Y. and Gibali, A. Inertial Krasnosel'skiĭ–Mann method in Banach spaces. *Mathematics*, 8(4):638, 2020.

Shehu, Y., Iyiola, O. S., and Ogbuisi, F. U. Iterative method with inertial terms for nonexpansive mappings: applications to compressed sensing. *Numerical Algorithms*, 83 (4):1321–1347, 2020.

Shi, W., Ling, Q., Wu, G., and Yin, W. A proximal gradient algorithm for decentralized composite optimization. *IEEE Transactions on Signal Processing*, 63(22):6013–6023, 2015.

Sopasakis, P., Menounou, K., and Patrinos, P. SuperSCS: fast and accurate large-scale conic optimization. In *2019 18th European Control Conference (ECC)*, pp. 1500–1505. IEEE, 2019.

Stellato, B., Banjac, G., Goulart, P., Bemporad, A., and Boyd, S. OSQP: An operator splitting solver for quadratic programs. *Mathematical Programming Computation*, 12 (4):637–672, 2020.

Sturm, J. F. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999.

Taylor, A. and Drori, Y. An optimal gradient method for smooth strongly convex minimization. *Mathematical Programming*, 2022.

Taylor, A., Van Scoy, B., and Lessard, L. Lyapunov functions for first-order methods: Tight automated convergence guarantees. *International Conference on Machine Learning*, 2018a.

Taylor, A. B., Hendrickx, J. M., and Glineur, F. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161(1–2):307–345, 2017.

Taylor, A. B., Hendrickx, J. M., and Glineur, F. Exact worst-case convergence rates of the proximal gradient method for composite convex minimization. *Journal of Optimization Theory and Applications*, 178(2):455–476, 2018b.

Themelis, A. and Patrinos, P. SuperMann: a superlinearly convergent algorithm for finding fixed points of nonexpansive operators. *IEEE Transactions on Automatic Control*, 64(12):4875–4890, 2019.

Von Neumann, J. *Functional operators: The geometry of orthogonal spaces*, volume 2. Princeton University Press, 1951.

Walker, H. F. and Ni, P. Anderson acceleration for fixed-point iterations. *SIAM Journal on Numerical Analysis*, 49(4):1715–1735, 2011.

Xu, H.-K. Viscosity approximation methods for nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 298(1):279–291, 2004.

Yoon, T. and Ryu, E. K. Accelerated algorithms for smooth convex-concave minimax problems with $\mathcal{O}(1/k^2)$ rate on squared gradient norm. *International Conference on Machine Learning*, 2021.

Zamani, M., Abbaszadehpeivasti, H., and de Klerk, E. Convergence rate analysis of the gradient descent-ascent method for convex-concave saddle-point problems. *arXiv preprint arXiv:2209.01272*, 2022.

Zhang, J., O'Donoghue, B., and Boyd, S. Globally convergent type-I Anderson acceleration for nonsmooth fixed-point iterations. *SIAM Journal on Optimization*, 30(4): 3170–3197, 2020.

A. Omitted proof of Section 2

Proof of Lemma 2. $x^k - \mathbb{T}x^k \in \overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}$, so $v^k \in \overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}$. From the property of the projection, as $v^k \in \overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}$,

$$\langle v^k, v\rangle \geq \|v\|^2, \quad \forall k \in \mathbb{N}.$$

Then we have

$$\|v^k - v\|^2 = \|v^k\|^2 - 2\langle v^k, v\rangle + \|v\|^2 \leq \|v^k\|^2 - \|v\|^2.$$

If $\lim_{k\to\infty} v^k = v$, then obviously, $\lim_{k\to\infty} \|v^k\| = \|v\|$. If $\lim_{k\to\infty} \|v^k\| = \|v\|$, then $\lim_{k\to\infty} \|v^k - v\|^2 = 0$ from above inequality, so $\lim_{k\to\infty} v^k = v$. □

B. Omitted proofs of Section 3

B.1. Omitted proofs of Section 3.1

Following lemmas will be used in the proof of Theorem 3 and Theorem 4.

Lemma 13. *If $\{x^k\}_{k\in\mathbb{N}}$ and $\{y^k\}_{k\in\mathbb{N}}$ are sequences of iterates generated by* (KM) *starting from $x^0 \in \mathcal{H}$ and $y^0 \in \mathcal{H}$ respectively, for any $k \in \mathbb{N}\cup\{0\}$,*

$$\|x^{k+1} - \mathbb{T}x^{k+1}\| \leq \|x^k - \mathbb{T}x^k\|$$

and

$$\|x^{k+1} - y^{k+1}\| \leq \|x^k - y^k\|.$$

Proof.

$$\begin{aligned}\|x^{k+1} - \mathbb{T}x^{k+1}\| &= \|x^{k+1} - \mathbb{T}x^k + \mathbb{T}x^k - \mathbb{T}x^{k+1}\| \\ &\leq \|x^{k+1} - \mathbb{T}x^k\| + \|x^k - x^{k+1}\| \\ &= \lambda_{k+1}\|x^k - \mathbb{T}x^k\| + (1-\lambda_{k+1})\|x^k - \mathbb{T}x^k\| \\ &= \|x^k - \mathbb{T}x^k\|\end{aligned}$$

and

$$\begin{aligned}\|x^{k+1} - y^{k+1}\| &= \|(1-\lambda_{k+1})(\mathbb{T}x^k - \mathbb{T}y^k) + \lambda_{k+1}(x^k - y^k)\| \\ &\leq (1-\lambda_{k+1})\|x^k - y^k\| + \lambda_{k+1}\|x^k - y^k\| \\ &= \|x^k - y^k\|.\end{aligned}$$

□

Lemma 14. *For any $\varepsilon > 0$, there exists $x_\varepsilon \in \mathcal{H}$ such that*

$$\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \leq \varepsilon.$$

And for any $k \in \mathbb{N}\cup\{0\}$,

$$\|x_\varepsilon^k - \mathbb{T}x_\varepsilon^k\| - \|v\| \leq \varepsilon.$$

Proof. Since $v \in \overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}$, for any $\varepsilon > 0$, we may choose $y_\varepsilon \in \mathcal{R}(\mathbb{I}-\mathbb{T})$ such that $\|y_\varepsilon - v\| \leq \varepsilon$. As $y_\varepsilon \in \mathcal{R}(\mathbb{I}-\mathbb{T})$, there exists $x_\varepsilon \in \mathcal{H}$ such that $y_\varepsilon = x_\varepsilon - \mathbb{T}x_\varepsilon$, so

$$\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \leq \varepsilon.$$

We know that from Lemma 13 that for any $k \in \mathbb{N}$,

$$\|x_\varepsilon^k - \mathbb{T}x_\varepsilon^k\| \leq \|x_\varepsilon^{k-1} - \mathbb{T}x_\varepsilon^{k-1}\|.$$

Therefore,

$$\begin{aligned}\|x_\varepsilon^k - \mathbb{T}x_\varepsilon^k\| - \|v\| &\leq \|x_\varepsilon - \mathbb{T}x_\varepsilon\| - \|v\| \\ &\leq \|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \leq \varepsilon.\end{aligned}$$

□

We now prove our main results of this section.

Proof of Theorem 3. For $\varepsilon > 0$, define $\tilde{\varepsilon}$ as

$$\tilde{\varepsilon} = \min\left\{\frac{\varepsilon^2}{2\|v\|+1}, 1, \varepsilon\right\}$$

and let $x_\varepsilon \in \mathcal{H}$ be a vector in $\mathcal{H}$ such that

$$\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \leq \tilde{\varepsilon},$$

whose existence is guaranteed from Lemma 14.

Now let $\{x_\varepsilon^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by (KM) starting from x_ε. Expanding the x^k term, we get

$$\begin{aligned}
&\frac{x^k - x^0}{\sum_{i=1}^k (1-\lambda_i)} + v \\
&= \frac{1}{\sum_{i=1}^k (1-\lambda_i)} \left\{ (x^k - x_\varepsilon^k) - (x^0 - x_\varepsilon) - \left(x_\varepsilon - x_\varepsilon^k - \left(\sum_{i=1}^k (1-\lambda_i) \right) v \right) \right\} \\
&= \frac{1}{\sum_{i=1}^k (1-\lambda_i)} \left\{ (x^k - x_\varepsilon^k) - (x^0 - x_\varepsilon) - \sum_{i=1}^k (1-\lambda_i) \left(x_\varepsilon^{i-1} - \mathbb{T}x_\varepsilon^{i-1} - v \right) \right\}
\end{aligned}$$

and taking its norm,

$$\begin{aligned}
&\left\| \frac{x^k - x^0}{\sum_{i=1}^k (1-\lambda_i)} + v \right\| \\
&\leq \frac{1}{\sum_{i=1}^k (1-\lambda_i)} \left(\|x^k - x_\varepsilon^k\| + \|x^0 - x_\varepsilon\| \right) + \sum_{i=1}^k \frac{(1-\lambda_i)}{\sum_{i=1}^k (1-\lambda_i)} \left\| x_\varepsilon^{i-1} - \mathbb{T}x_\varepsilon^{i-1} - v \right\| \\
&\leq \frac{2}{\sum_{i=1}^k (1-\lambda_i)} \|x^0 - x_\varepsilon\| + \sum_{i=1}^k \frac{(1-\lambda_i)}{\sum_{i=1}^k (1-\lambda_i)} \|x_\varepsilon^{i-1} - \mathbb{T}x_\varepsilon^{i-1} - v\| && (\because \text{Lemma 13})
\end{aligned}$$

Since

$$\|x - \mathbb{T}x - v\|^2 = \|x - \mathbb{T}x\|^2 - 2\langle x - \mathbb{T}x, v\rangle + \|v\|^2 \leq \|x - \mathbb{T}x\|^2 - \|v\|^2, \quad \forall x \in \mathcal{H},$$

we get

$$\begin{aligned}
\|x_\varepsilon^i - \mathbb{T}x_\varepsilon^i - v\|^2 &\leq \|x_\varepsilon^i - \mathbb{T}x_\varepsilon^i\|^2 - \|v\|^2 \\
&\leq \|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2 && (\because \text{Lemma 13}) \\
&= (\|x_\varepsilon - \mathbb{T}x_\varepsilon\| - \|v\|)(\|x_\varepsilon - \mathbb{T}x_\varepsilon\| + \|v\|) \\
&\leq \tilde{\varepsilon}(2\|v\| + \tilde{\varepsilon}) && (\because \text{Lemma 14}) \\
&\leq \tilde{\varepsilon}(2\|v\| + 1) \leq \varepsilon^2
\end{aligned}$$

for any $i \in \mathbb{N} \cup \{0\}$. Gathering all facts above, we get

$$\left\| \frac{x^k - x^0}{\sum_{i=1}^k (1-\lambda_i)} + v \right\| \leq \frac{2}{\sum_{i=1}^k (1-\lambda_i)} \|x^0 - x_\varepsilon\| + \varepsilon$$

for any $k \in \mathbb{N}$.

If $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, there exists $x_\star \in \mathcal{H}$ such that $v = x_\star - \mathbb{T}x_\star$. The proof above applies well with $\varepsilon = 0$ and $x_\varepsilon = x_\star$, so we are done. ☐

According to Theorem 3, the normalized iterate of (KM) converges to $-v$ when $\sum_{i=1}^\infty (1-\lambda_i) = \infty$.

Corollary 15. *Let* $\{x^k\}_{k\in\mathbb{N}}$ *be the iterates of* (KM) *starting from* $x^0 \in \mathcal{H}$. *If* $\sum_{i=1}^{\infty}(1-\lambda_i) = \infty$, *then*

$$\lim_{k\to\infty} \frac{x^k - x^0}{\sum_{i=1}^{k}(1-\lambda_i)} = -v.$$

Proof. According to the first claim, for any $\varepsilon > 0$, there exists $x_\varepsilon \in \mathcal{H}$ such that

$$\left\| \frac{x^k - x^0}{\sum_{i=1}^{k}(1-\lambda_i)} + v \right\| \le \frac{2}{\sum_{i=1}^{k}(1-\lambda_i)} \|x^0 - x_\varepsilon\| + \varepsilon.$$

Therefore, given $\sum_{i=1}^{\infty}(1-\lambda_i) = \infty$,

$$0 \le \limsup_{k\to\infty} \left\| \frac{x^k - x^0}{\sum_{i=1}^{k}(1-\lambda_i)} + v \right\| \le \varepsilon$$

for any $\varepsilon > 0$. We may conclude that $\frac{x^k - x^0}{\sum_{i=1}^{k}(1-\lambda_i)}$ converges to $-v$ in norm. □

Convergence of the fixed-point residual $x^k - \mathbb{T}x^k$ to v requires a stronger assumption, which is $\sum_{k=0}^{\infty} \lambda_k(1-\lambda_k) = \infty$. This is a stronger condition than that of Theorem 3 in a sense that

$$\sum_{k=0}^{\infty} \lambda_i(1-\lambda_i) = \infty \quad \Longrightarrow \quad \sum_{k=0}^{\infty} \lambda_i = \infty.$$

In case of $\operatorname{Fix}\mathbb{T} \neq \emptyset$, The iterates $\{x^k\}$ generated by (KM) exhibits Fejer-monotonicity with respect to $\operatorname{Fix}\mathbb{T}$ (Bauschke & Combettes, 2017, Chapter 5), which is a useful concept in proving the convergence of (KM) in terms of $x^k - \mathbb{T}x^k \to 0$ and $x^k \to x_\star$. However, when $\operatorname{Fix}\mathbb{T} = \emptyset$, such analysis is impossible.

Consider a sequence $\{\lambda_k\}_{k\in\mathbb{N}\cup\{0\}}$ of stepsizes to (KM). Define $\mathbb{T}_k \colon \mathcal{H} \to \mathcal{H}$, for each $k \in \mathbb{N}$ as

$$\mathbb{T}_k := (1-\lambda_k)\mathbb{T} + \lambda_k \mathbb{I}.$$

Then if $\{x^k\}_{k\in\mathbb{N}}$ is a sequence of iterates generated by (KM) with $\{\lambda_k\}_{k\in\mathbb{N}\cup\{0\}}$ starting from $x^0 \in \mathcal{H}$,

$$x^{k+1} = \mathbb{T}_{k+1}x^k.$$

Lemma 16. *If* $\{x^k\}_{k\in\mathbb{N}}$ *and* $\{y^k\}_{k\in\mathbb{N}}$ *are sequences of iterates generated by* (KM) *starting from* $x^0 \in \mathcal{H}$ *and* $y^0 \in \mathcal{H}$ *respectively, for any* $k \in \mathbb{N}\cup\{0\}$,

$$\|x^k - y^k\|^2 - \|x^{k+1} - y^{k+1}\|^2 \ge \lambda_{k+1}(1-\lambda_{k+1})\|(x^k - \mathbb{T}x^k) - (y^k - \mathbb{T}y^k)\|^2$$

Proof. First of all, if $\lambda_{k+1} = 0$ or 1, the theorem trivially holds from the fact that $\mathbb{T}_{k+1}$ is a nonexpansive operator.

Suppose $\lambda_{k+1} \in (0,1)$.

$$\begin{aligned}
&\|(x^k - x^{k+1}) - (y^k - y^{k+1})\|^2 \\
&= \|x^k - y^k\|^2 + \|x^{k+1} - y^{k+1}\|^2 - 2\langle x^{k+1} - y^{k+1}, x^k - y^k\rangle \\
&= \|x^k - y^k\|^2 + \|x^{k+1} - y^{k+1}\|^2 - 2\langle \mathbb{T}_{k+1}x^k - \mathbb{T}_{k+1}y^k, x^k - y^k\rangle.
\end{aligned}$$

From $(1-\lambda_{k+1})$-averagedness of $\mathbb{T}_{k+1}$, (Bauschke & Combettes, 2017, Proposition 4.35(iv)) gives us

$$\|\mathbb{T}_{k+1}x^k - \mathbb{T}_{k+1}y^k\|^2 + (2\lambda_{k+1} - 1)\|x^k - y^k\|^2 \le 2\lambda_{k+1}\langle \mathbb{T}_{k+1}x^k - \mathbb{T}_{k+1}y^k, x^k - y^k\rangle.$$

Then

$$\begin{aligned}
&\lambda_{k+1}\|(x^k - x^{k+1}) - (y^k - y^{k+1})\|^2 \\
&= \lambda_{k+1}\|x^k - y^k\|^2 + \lambda_{k+1}\|x^{k+1} - y^{k+1}\|^2 - 2\lambda_{k+1}\langle \mathbb{T}_{k+1}x^k - \mathbb{T}_{k+1}y^k,\, x^k - y^k\rangle \\
&\le \lambda_{k+1}\|x^k - y^k\|^2 + \lambda_{k+1}\|x^{k+1} - y^{k+1}\|^2 - \left\{(2\lambda_{k+1} - 1)\|x^k - y^k\|^2 + \|x^{k+1} - y^{k+1}\|^2\right\} \\
&= (1-\lambda_{k+1})\left\{\|x^k - y^k\|^2 - \|x^{k+1} - y^{k+1}\|^2\right\}.
\end{aligned}$$

As

$$x^k - x^{k+1} = x^k - \{(1-\lambda_{k+1})\mathbb{T}x^k + \lambda_{k+1}x^k\} = (1-\lambda_{k+1})(x^k - \mathbb{T}x^k),$$

combining this fact with above inequality and dividing by $1-\lambda_{k+1} > 0$.

$$\|x^k - y^k\|^2 - \|x^{k+1} - y^{k+1}\|^2 \geq \lambda_{k+1}(1-\lambda_{k+1})\|(x^k - \mathbb{T}x^k) - (y^k - \mathbb{T}y^k)\|^2.$$

□

We now prove the second main result of this section.

Proof of Theorem 4. Given $\varepsilon > 0$, there exists $x_\varepsilon \in \mathcal{H}$ such that

$$\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \leq \tilde{\varepsilon} = \min\left\{\frac{\varepsilon^2}{2\|v\|+1}, 1, \varepsilon\right\}$$

by Lemma 14. Let $\{x_\varepsilon^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by (KM) starting from x_ε. With $y^0 = x_\varepsilon$, summing up the inequality in Lemma 16 and removing the telescoping terms, we get

$$\|x^0 - x_\varepsilon\|^2 - \|x^{k+1} - x_\varepsilon^{k+1}\|^2 \geq \sum_{i=0}^{k} \lambda_{i+1}(1-\lambda_{i+1})\|(x^i - \mathbb{T}x^i) - (x_\varepsilon^i - \mathbb{T}x_\varepsilon^i)\|^2$$

for any $k \in \mathbb{N}$. Therefore,

$$\begin{aligned}
&\frac{1}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\|x^0 - x_\varepsilon\|^2 \\
&\geq \sum_{i=0}^{k}\left(\frac{\lambda_{i+1}(1-\lambda_{i+1})}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\right)\left\|(x^i - \mathbb{T}x^i) - (x_\varepsilon^i - \mathbb{T}x_\varepsilon^i)\right\|^2 \\
&= \left\{\sum_{i=0}^{k}\left(\frac{\lambda_{i+1}(1-\lambda_{i+1})}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\right)\right\}\left\{\sum_{i=0}^{k}\left(\frac{\lambda_{i+1}(1-\lambda_{i+1})}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\right)\left\|(x^i - \mathbb{T}x^i) - (x_\varepsilon^i - \mathbb{T}x_\varepsilon^i)\right\|^2\right\} \\
&\geq \left\{\sum_{i=0}^{k}\left(\frac{\lambda_{i+1}(1-\lambda_{i+1})}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\right)\left\|(x^i - \mathbb{T}x^i) - (x_\varepsilon^i - \mathbb{T}x_\varepsilon^i)\right\|\right\}^2 \qquad \text{(Cauchy-Schwarz)}
\end{aligned}$$

or equivalently,

$$\sum_{i=0}^{k}\left(\frac{\lambda_{i+1}(1-\lambda_{i+1})}{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}\right)\left\|(x^i - \mathbb{T}x^i) - (x_\varepsilon^i - \mathbb{T}x_\varepsilon^i)\right\| \leq \frac{1}{\sqrt{\sum_{i=0}^k \lambda_{i+1}(1-\lambda_{i+1})}}\|x^0 - x_\varepsilon\|.$$

Note that for any $x \in \mathcal{H}$,

$$\|x - \mathbb{T}x - v\|^2 = \|x - \mathbb{T}x\|^2 - 2\underbrace{\langle x - \mathbb{T}x, v\rangle}_{\geq \|v\|^2} + \|v\|^2 \leq \|x - \mathbb{T}x\|^2 - \|v\|^2.$$

For any $i \in \mathbb{N}$,

$$\begin{aligned}
\|x_\varepsilon^i - \mathbb{T}x_\varepsilon^i - v\|^2 &\leq (\|x_\varepsilon^i - \mathbb{T}x_\varepsilon^i\| - \|v\|)(\|x_\varepsilon^i - \mathbb{T}x_\varepsilon^i\| + \|v\|) \\
&\leq (\|x_\varepsilon - \mathbb{T}x_\varepsilon\| - \|v\|)(\|x_\varepsilon - \mathbb{T}x_\varepsilon\| + \|v\|) \\
&\leq \tilde{\varepsilon}(2\|v\| + \tilde{\varepsilon}) \leq \varepsilon^2,
\end{aligned}$$

so

$$\begin{aligned}
\left\|(x^i - \mathbb{T}x^i) - (x_\varepsilon^i - \mathbb{T}x_\varepsilon^i)\right\| &\leq \left\|x^i - \mathbb{T}x^i - v\right\| - \left\|x_\varepsilon^i - \mathbb{T}x_\varepsilon^i - v\right\| \\
&\leq \left\|x^i - \mathbb{T}x^i - v\right\| - \varepsilon.
\end{aligned}$$

Therefore, we get

$$\sum_{i=0}^{k}\left(\frac{\lambda_{i+1}(1-\lambda_{i+1})}{\sum_{i=0}^{k}\lambda_{i+1}(1-\lambda_{i+1})}\right)\left\|x^i-\mathbb{T}x^i-v\right\| \leq \frac{1}{\sqrt{\sum_{i=0}^{k}\lambda_{i+1}(1-\lambda_{i+1})}}\|x^0-x_\varepsilon\|+\varepsilon.$$

Also, note that for any i such that $0 \leq i \leq k-1$,

$$\left\|x^i-\mathbb{T}x^i-v\right\| \geq \left\|x^i-\mathbb{T}x^i\right\| - \|v\| \geq \left\|x^k-\mathbb{T}x^k\right\| - \|v\|$$

where the first inequality comes from triangular inequality, and the last inequality comes from Lemma 13. Hence we get

$$\|x^k-\mathbb{T}x^k\| - \|v\| \leq \frac{1}{\sqrt{\sum_{i=0}^{k}\lambda_{i+1}(1-\lambda_{i+1})}}\|x^0-x_\varepsilon\|+\varepsilon.$$

If $v \in \mathcal{R}(\mathbb{I}-\mathbb{T})$, there exists $x_\star \in \mathcal{H}$ such that $v = x_\star - \mathbb{T}x_\star$. The proof above applies well with $\varepsilon = 0$ and $x_\varepsilon = x_\star$, so we are done. □

According to Theorem 4, the fixed-point residual of (KM) converges to v if $\sum_{i=1}^{\infty}\lambda_i(1-\lambda_i) = \infty$.

Corollary 17. *Let* $\{x^k\}_{k\in\mathbb{N}}$ *be the iterates of* (KM) *starting from* $x^0 \in \mathcal{H}$. *If* $\sum_{i=1}^{\infty}\lambda_i(1-\lambda_i) = \infty$, *then*

$$\lim_{k\to\infty}\left(x^k-\mathbb{T}x^k\right) = v.$$

Proof. Given $\sum_{i=1}^{\infty}\lambda_i(1-\lambda_i) = \infty$,

$$0 \leq \limsup_{k\to\infty}\|x^k-\mathbb{T}x^k\| - \|v\| \leq \varepsilon.$$

Since above inequality holds for any choice of $\varepsilon > 0$, $\lim_{k\to\infty}\|x^k-\mathbb{T}x^k\| = \|v\|$. Since

$$\|x^k-\mathbb{T}x^k-v\|^2 \leq \|x^k-\mathbb{T}x^k\|^2 - \|v\|^2,$$

taking limit on both sides, we get

$$0 \leq \limsup_{k\to\infty}\|x^k-\mathbb{T}x^k-v\|^2 \leq \lim_{k\to\infty}\|x^k-\mathbb{T}x^k\|^2 - \|v\|^2 = \|v\|^2 - \|v\|^2 = 0.$$

□

B.2. Omitted proofs of Section 3.2

Following lemmas will be used in the proof of Theorem 7 and Theorem 8.

We first prove Lemma 6.

Proof of Lemma 6. If $k = 0$, then

$$\theta_1 = (1-\lambda_1) = (1-\lambda_1)(1+\underbrace{\theta_0}_{=0}).$$

Suppose $k \geq 1$.

$$\begin{aligned}
\theta_{k+1} &= \sum_{n=1}^{k+1}(1-\lambda_{k+1})(1-\lambda_k)\cdots(1-\lambda_{k-n+2}) \\
&= (1-\lambda_{k+1}) + (1-\lambda_{k+1})\sum_{n=2}^{k+1}(1-\lambda_k)\cdots(1-\lambda_{k-n+2} \\
&= (1-\lambda_{k+1}) + (1-\lambda_{k+1})\sum_{n=1}^{k}(1-\lambda_k)\cdots(1-\lambda_{k-n+1} \\
&= (1-\lambda_{k+1})(1+\theta_k).
\end{aligned}$$

Suppose $\lambda_k \equiv 0$. Then

$$\theta_k = 1 + \theta_{k-1} = 2 + \theta_{k-2} = \cdots = k + \theta_0 = k.$$

If $\lambda_k = \frac{1}{k+1}$ for all $k \in \mathbb{N}$, then from $\theta_0 = 0$, suppose $\theta_{k-1} = \frac{k-1}{2}$. Then as

$$\theta_k = \left(1 - \frac{1}{k+1}\right)(1 + \theta_{k-1}) = \frac{k}{k+1}\frac{k+1}{2} = \frac{k}{2},$$

the induction holds. □

Remark 18. Let $\{x^k\}_{k\in\mathbb{N}}$ be the iterates of (Halpern) starting from $x^0 \in \mathcal{H}$. Then the k-th iterate x^k of (Halpern) can be expressed as

$$x^k - x^0 = -\sum_{i=0}^{k-1} \{(1-\lambda_k)\cdots(1-\lambda_{i+1})\}(x^i - \mathbb{T}x^i).$$

If $\lambda_k = \frac{1}{k+1}$ for $k \in \mathbb{N}$, the k-th iterate x^k of (Halpern) can be expressed as

$$x^k - x^0 = -\sum_{i=0}^{k-1} \frac{i+1}{k+1}(x^i - \mathbb{T}x^i)$$

The sequence $\{\theta_k\}$ refers to the sum of all linear coefficients to $\{x^i - \mathbb{T}x^i\}_{i=0,1,\dots,k-1}$ used in the x^k-update of (Halpern).

Following lemma refers to the property that two independent iterates $\{x^k\}$ and $\{y^k\}$ generated by (Halpern) cannot be further than the distance between initial points $\|x^0 - y^0\|$.

Lemma 19. *Let $\{x^k\}_{k\in\mathbb{N}}$ and $\{y^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by* (Halpern) *starting from $x^0 \in \mathcal{H}$ and $y^0 \in \mathcal{H}$, respectively. Then*

$$\|x^k - y^k\| \le \|x^0 - y^0\|, \quad k = 0, 1, \dots.$$

Proof. We prove by induction on k. If $k = 0$, the claim automatically holds. Suppose $k \ge 1$ and $\|x^{k-1} - y^{k-1}\| \le \|x^0 - y^0\|$. Then

$$\begin{aligned}
\|x^k - y^k\| &\le (1-\lambda_k)\|\mathbb{T}x^{k-1} - \mathbb{T}y^{k-1}\| + \lambda_k\|x^0 - y^0\| \\
&\le (1-\lambda_k)\|x^{k-1} - y^{k-1}\| + \lambda_k\|x^0 - y^0\| \\
&\le (1-\lambda_k)\|x^0 - y^0\| + \lambda_k\|x^0 - y^0\| = \|x^0 - y^0\|.
\end{aligned}$$

□

Lemma 20. *If $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by* (Halpern) *starting from $x^0 \in \mathcal{H}$, then*

$$\frac{\|x^k - x^0\|}{\theta_k} \le \|x^0 - \mathbb{T}x^0\|, \quad k = 1, 2, \dots.$$

Proof. We prove by induction on k.

(i) $k = 1$. First of all,

$$x^1 - x^0 = -(1-\lambda_1)(x^0 - \mathbb{T}x^0)$$

so from $\theta_1 = 1 - \lambda_1$,

$$\frac{\|x^1 - x^0\|}{\theta_1} = \|x^0 - \mathbb{T}x^0\|.$$

(ii) $k \ge 2$. Suppose that the claim holds true for all n such that $n < k$.

$$\begin{aligned}
x^k - x^0 &= (1-\lambda_k)(\mathbb{T}x^{k-1} - x^0) \\
&= (1-\lambda_k)(\mathbb{T}x^{k-1} - \mathbb{T}x^0) + (1-\lambda_k)(\mathbb{T}x^0 - x^0) \\
\|x^k - x^0\| &\le (1-\lambda_k)\|\mathbb{T}x^{k-1} - \mathbb{T}x^0\| + (1-\lambda_k)\|x^0 - \mathbb{T}x^0\| \\
&\le (1-\lambda_k)\|x^{k-1} - x^0\| + (1-\lambda_k)\|x^0 - \mathbb{T}x^0\|
\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{\|x^k - x^0\|}{\theta_k} &\leq \frac{1-\lambda_k}{\theta_k}\|x^{k-1} - x^0\| + \frac{1-\lambda_k}{\theta_k}\|x^0 - \mathbb{T}x^0\| \\ &= \frac{\theta_{k-1}}{1+\theta_{k-1}}\frac{\|x^{k-1} - x^0\|}{\theta_{k-1}} + \frac{1}{1+\theta_{k-1}}\|x^0 - \mathbb{T}x^0\| && (\because \text{Lemma 6}) \\ &\leq \frac{\theta_{k-1}}{1+\theta_{k-1}}\|x^0 - \mathbb{T}x^0\| + \frac{1}{1+\theta_{k-1}}\|x^0 - \mathbb{T}x^0\| \\ &= \|x^0 - \mathbb{T}x^0\|.\end{aligned}$$

□

Following lemma identifies the proper averaging of x^k that resides in the closure of the range of $\mathbb{I} - \mathbb{T}$, which becomes the candidate for the sequence $\{v^k\}_{k\in\mathbb{N}}$ converging to v.

Lemma 21. *If* $\{x^k\}_{k\in\mathbb{N}}$ *be a sequence of iterates generated by* (Halpern) *starting from* $x^0 \in \mathcal{H}$*, then*

$$-\frac{x^k - x^0}{\theta_k} \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$$

for $k = 1, 2, \ldots$.

Proof. We prove by induction on k, using the convexity of $\overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$.

(i) $k = 1$.

$$-\frac{x^1 - x^0}{\theta_1} = x^0 - \mathbb{T}x^0 \in \overline{\mathcal{R}(I - T)}.$$

(ii) $k \geq 2$. Suppose that

$$-\frac{x^{k-1} - x^0}{\theta_{k-1}} \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}.$$

As

$$\begin{aligned}-\frac{x^k - x^0}{\theta_k} &= -\frac{(1-\lambda_k)\theta_{k-1}}{\theta_k}\frac{x^{k-1} - x^0}{\theta_{k-1}} + \frac{1-\lambda_k}{\theta_k}(x^{k-1} - \mathbb{T}x^{k-1}) \\ &= \frac{\theta_{k-1}}{1+\theta_{k-1}}\left(-\frac{x^{k-1} - x^0}{\theta_{k-1}}\right) + \frac{1}{1+\theta_{k-1}}(x^{k-1} - \mathbb{T}x^{k-1}),\end{aligned}$$

$-\frac{x^k - x^0}{\theta_k}$ is a convex combination of vectors in a convex set $\overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$, so it is also an element of $\overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$.

□

We now prove Theorem 7.

Proof of Theorem 7. From $v \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$, we may choose a point x_ε in $\mathcal{H}$ such that

$$\|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2 \leq \varepsilon^2.$$

Let $k \geq 1$. From Lemma 19,

$$\begin{aligned}\left\|\frac{x^k - x^0}{\theta_k} - \frac{x^k_\varepsilon - x_\varepsilon}{\theta_k}\right\| &\leq \left\|\frac{x^k - x^k_\varepsilon}{\theta_k}\right\| + \left\|\frac{x^0 - x_\varepsilon}{\theta_k}\right\| \\ &\leq \frac{2}{\theta_k}\|x^0 - x_\varepsilon\|.\end{aligned}$$

Note that

$$\begin{aligned}\left\|\frac{x_\varepsilon^k - x_\varepsilon}{\theta_k} + v\right\|^2 &\leq \left\|\frac{x_\varepsilon^k - x_\varepsilon}{\theta_k}\right\|^2 - \|v\|^2 \\ &\leq \|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2 \leq \varepsilon^2, \qquad (\because \text{Lemma 20})\end{aligned}$$

and from this we have

$$\begin{aligned}\left\|\frac{x^k - x^0}{\theta_k} + v\right\| &\leq \left\|\frac{x^k - x^0}{\theta_k} - \frac{x_\varepsilon^k - x_\varepsilon}{\theta_k}\right\| + \left\|\frac{x_\varepsilon^k - x_\varepsilon}{\theta_k} + v\right\| \\ &\leq \frac{2}{\theta_k}\|x^0 - x_\varepsilon\| + \varepsilon.\end{aligned}$$

This result holds for any $k \geq 1$.

If $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, there exists $x_\star \in \mathcal{H}$ such that $v = x_\star - \mathbb{T}x_\star$. The proof above applies well with $\varepsilon = 0$ and $x_\varepsilon = x_\star$, so we are done. □

According to Theorem 7, the normalized iterate of (Halpern) converges to $-v$ when $\lim_{k\to\infty} \theta_k = \infty$.

Corollary 22. *Let $\{x^k\}_{k\in\mathbb{N}}$ be the iterates of* (Halpern) *starting form $x^0 \in \mathcal{H}$. If $\lim_{k\to\infty} \theta_k = \infty$, then*

$$\lim_{k\to\infty} \frac{x^k - x^0}{\theta_k} = -v.$$

Proof. Further assume that $\lim_{k\to\infty} \theta_k = \infty$. Using triangle inequality,

$$\left\|\frac{x^k - x^0}{\theta_k}\right\| - \|v\| \leq \left\|\frac{x^k - x^0}{\theta_k} + v\right\| \leq \frac{2}{\theta_k}\|x^0 - x_\varepsilon\| + \varepsilon.$$

From the fact that $-\frac{x^k - x^0}{\theta_k} \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$ by Lemma 21 and the fact that v is the minimum norm element in $\overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$,

$$\left\|\frac{x^k - x^0}{\theta_k}\right\| \geq \|v\|.$$

Then

$$\|v\| \leq \liminf_{k\to\infty} \left\|\frac{x^k - x^0}{\theta_k}\right\| \leq \limsup_{k\to\infty} \left\|\frac{x^k - x^0}{\theta_k}\right\| \leq \|v\| + \varepsilon$$

holds for any possible choice of $\varepsilon > 0$, so

$$\lim_{k\to\infty} \left\|\frac{x^k - x^0}{\theta_k}\right\| = \|v\|.$$

We may conclude that, by the uniqueness of v as a minimum norm element in $\overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$,

$$\lim_{k\to\infty} \frac{x^k - x^0}{\theta_k} = -v.$$

□

As in Section 3.2, we have a simpler condition for $\{\lambda_k\}_{k\in\mathbb{N}}$ to ensure the convergence of normalized iterate of Halpern iteration to $-v$.

Lemma 23. *If*

$$\lim_{k\to\infty} \lambda_k = 0,$$

then

$$\lim_{k\to\infty} \theta_k = \infty.$$

Proof.

$$\lim_{k\to\infty}\frac{\theta_{k+1}}{1+\theta_k}=\lim_{k\to\infty}(1-\lambda_{k+1})=1,$$

and $\lambda_{k+1}\in[0,1]$, so for any $0<\varepsilon<1$, there exists $N_\varepsilon\in\mathbb{N}$ such that

$$\frac{\theta_{k+1}}{1+\theta_k}\geq 1-\varepsilon,\quad \forall\, k\geq N_\varepsilon.$$

Then

$$\begin{aligned}\theta_{k+N_\varepsilon}&\geq(1-\varepsilon)\theta_{k+N_\varepsilon-1}+(1-\varepsilon)\\&\geq(1-\varepsilon)^k\theta_{N_\varepsilon}+(1-\varepsilon)+\cdots+(1-\varepsilon)^k\\&=(1-\varepsilon)^k\theta_{N_\varepsilon}+\left(\frac{1}{\varepsilon}-1\right)\left\{1-(1-\varepsilon)^k\right\}.\end{aligned}$$

As $k\to\infty$,

$$\liminf_{k\to\infty}\theta_k\geq\frac{1}{\varepsilon}-1$$

holds for all $\varepsilon\in(0,1)$. As $\varepsilon\to 0$, $\liminf_{k\to\infty}\theta_k=\infty$, so we are done. □

In order to prove Theorem 8, we use the following fact to construct Lyapunov function.

Lemma 24. *If $\{x^k\}_{k\in\mathbb{N}}$ is a sequence of iterates generated by* (Halpern) *starting from $x^0\in\mathcal{H}$ with $\lambda_k=\frac{1}{k+1}$, then*

$$\begin{aligned}&\|\mathbb{T}x^k-\mathbb{T}x^{k+1}\|^2\leq\|x^k-x^{k+1}\|^2\\&\Leftrightarrow\ (k+2)\left\{(k+1)\|x^{k+1}-\mathbb{T}x^{k+1}\|^2+2\langle x^{k+1}-\mathbb{T}x^{k+1},\,x^{k+1}-x^0\rangle\right\}\\&\qquad\leq(k+1)\left\{k\|x^k-\mathbb{T}x^k\|^2+2\langle x^k-\mathbb{T}x^k,\,x^k-x^0\rangle\right\}\end{aligned}$$

Proof. From

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

we have

$$\begin{aligned}&\|x^k-x^{k+1}\|^2-\|\mathbb{T}x^k-\mathbb{T}x^{k+1}\|^2\\&=\|(x^k-\mathbb{T}x^k)-(x^{k+1}-\mathbb{T}x^k)\|^2-\|(x^{k+1}-\mathbb{T}x^{k+1})-(x^{k+1}-\mathbb{T}x^k)\|^2\\&=\|x^k-\mathbb{T}x^k\|^2-\|x^{k+1}-\mathbb{T}x^{k+1}\|^2-2\langle x^k-\mathbb{T}x^k,\,x^{k+1}-\mathbb{T}x^k\rangle+2\langle x^{k+1}-\mathbb{T}x^{k+1},\,x^{k+1}-\mathbb{T}x^k\rangle\\&=\|x^k-\mathbb{T}x^k\|^2-\|x^{k+1}-\mathbb{T}x^{k+1}\|^2-2\left\langle x^k-\mathbb{T}x^k,\,\frac{1}{k+2}(x^k-\mathbb{T}x^k)-\frac{1}{k+2}(x^k-x^0)\right\rangle\\&\quad+2\left\langle x^{k+1}-\mathbb{T}x^{k+1},\,(x^{k+1}-x^0)-\frac{k+2}{k+1}(x^{k+1}-x^0)\right\rangle\\&=\frac{1}{k+2}\left\{k\|x^k-\mathbb{T}x^k\|^2+2\langle x^k-\mathbb{T}x^k,\,x^k-x^0\rangle\right\}\\&\quad-\frac{1}{k+1}\left\{(k+1)\|x^{k+1}-\mathbb{T}x^{k+1}\|^2+2\langle x^{k+1}-\mathbb{T}x^{k+1},\,x^{k+1}-x^0\rangle\right\}\end{aligned}$$

Equivalence follows immediately. □

We use the Lyapunov function V^k for $k=1,2,\ldots$ of the following form.

$$\begin{aligned}V^k&=(k+1)\left\{k\|x^k-\mathbb{T}x^k\|^2+2\langle x^k-\mathbb{T}x^k,x^k-x^0\rangle\right\}-\left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0-x_\varepsilon\|^2\\&\quad+k(k+1)\left\langle-\frac{2}{k}(x^k-x^0)-(x_\varepsilon-\mathbb{T}x_\varepsilon),x_\varepsilon-\mathbb{T}x_\varepsilon\right\rangle+\frac{2(k+1)}{k}\left\|x^k-x_\varepsilon+\frac{k}{2}(x_\varepsilon-\mathbb{T}x_\varepsilon)\right\|^2\end{aligned}\tag{Lyapunov function}$$

$x_\varepsilon \in \mathcal{H}$ is chosen to be the point which makes $x_\varepsilon - \mathbb{T}x_\varepsilon$ very close to v. In particular, if $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, choose x_ε such that $v = x_\varepsilon - \mathbb{T}x_\varepsilon$.

Now, we show the monotonicity of $\{V^k\}_k$ in k.

Lemma 25. *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by* (Halpern) *starting from $x^0 \in \mathcal{H}$ with $\lambda_k = \frac{1}{k+1}$, and define $\{V^k\}_{k\in\mathbb{N}\cup\{0\}}$ as* (Lyapunov function). *For any $k \in \mathbb{N}$,*

$$V^k \geq V^{k+1}.$$

Proof. From Lemma 24,

$$\begin{aligned}
&V^k - V^{k+1} \\
&\geq \frac{1}{k+1}\|x^0 - x_\varepsilon\|^2 + k(k+1)\left\langle -\frac{2}{k}(x^k - x^0) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x_\varepsilon - \mathbb{T}x_\varepsilon \right\rangle \\
&\quad - (k+1)(k+2)\left\langle -\frac{2}{k+1}(x^{k+1} - x^0) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x_\varepsilon - \mathbb{T}x_\varepsilon \right\rangle \\
&\quad + \frac{2(k+1)}{k}\left\| x^k - x_\varepsilon + \frac{k}{2}(x_\varepsilon - \mathbb{T}x_\varepsilon)\right\|^2 - \frac{2(k+2)}{k+1}\left\| x^{k+1} - x_\varepsilon + \frac{k+1}{2}(x_\varepsilon - \mathbb{T}x_\varepsilon)\right\|^2 \\
&= \frac{1}{k+1}\|x^0 - x_\varepsilon\|^2 + \left\{ -k(k+1) + (k+1)(k+2) + \frac{k(k+1)}{2} - \frac{(k+1)(k+2)}{2} \right\}\|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 \\
&\quad + \left\langle x_\varepsilon - \mathbb{T}x_\varepsilon,\, -2(k+1)(x^k - x^0) + 2(k+2)(x^{k+1} - x^0) + 2(k+1)(x^k - x_\varepsilon) - 2(k+2)(x^{k+1} - x_\varepsilon) \right\rangle \\
&\quad + \frac{2(k+1)}{k}\|x^k - x_\varepsilon\|^2 - \frac{2(k+2)}{(k+1)}\|x^{k+1} - x_\varepsilon\|^2 \\
&= \frac{1}{k+1}\|x^0 - x_\varepsilon\|^2 + (k+1)\|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - 2\left\langle x_\varepsilon - \mathbb{T}x_\varepsilon,\, x^0 - x_\varepsilon \right\rangle \\
&\quad + \frac{2(k+1)}{k}\|x^k - x_\varepsilon\|^2 - \frac{2(k+2)}{(k+1)}\|x^{k+1} - x_\varepsilon\|^2.
\end{aligned}$$

Using

$$\|x^k - x_\varepsilon\|^2 \geq \|\mathbb{T}x^k - \mathbb{T}x_\varepsilon\|^2,$$

we get

$$\begin{aligned}
&V^k - V^{k+1} \\
&\geq \frac{1}{k+1}\|x^0 - x_\varepsilon\|^2 + (k+1)\|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - 2\left\langle x_\varepsilon - \mathbb{T}x_\varepsilon,\, x^0 - x_\varepsilon \right\rangle \\
&\quad + \frac{2(k+1)}{k}\| \underbrace{\mathbb{T}x^k - \mathbb{T}x_\varepsilon}_{(\mathbb{T}x^k - x_\varepsilon) + (x_\varepsilon - \mathbb{T}x_\varepsilon)} \|^2 - \frac{2(k+2)}{k+1}\| \underbrace{x^{k+1} - x_\varepsilon}_{= \frac{k+1}{k+2}(\mathbb{T}x^k - x_\varepsilon) + \frac{1}{k+2}(x^0 - x_\varepsilon)} \|^2 \\
&= \frac{k}{(k+1)(k+2)}\|x^0 - x_\varepsilon\|^2 + \frac{(k+1)(k+2)}{k}\|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - 2\langle x_\varepsilon - \mathbb{T}x_\varepsilon, x^0 - x_\varepsilon\rangle \\
&\quad + \frac{4(k+1)}{k(k+2)}\|\mathbb{T}x^k - x_\varepsilon\|^2 + \frac{4(k+1)}{k}\langle x_\varepsilon - \mathbb{T}x_\varepsilon, \mathbb{T}x^k - x_\varepsilon\rangle - \frac{4}{k+2}\langle \mathbb{T}x^k - x_\varepsilon, x^0 - x_\varepsilon\rangle \\
&= \frac{1}{k(k+1)(k+2)}\left\| 2(k+1)(\mathbb{T}x^k - x_\varepsilon) - k(x^0 - x_\varepsilon) + (k+1)(k+2)(x_\varepsilon - \mathbb{T}x_\varepsilon)\right\|^2 \\
&\geq 0.
\end{aligned}$$

□

Lemma 26. *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by* (Halpern) *starting from $x^0 \in \mathcal{H}$ with $\lambda_k = \frac{1}{k+1}$ and*

$\{V^k\}_{k\in\mathbb{N}\cup\{0\}}$ *be defined as* (Lyapunov function). *For* $k \geq 1$*,*

$$V^k \geq (k+1)^2\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 + 2k(k+1)\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x_\varepsilon - \mathbb{T}x_\varepsilon\rangle \\ - 2(k+1)\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x^0 - x_\varepsilon\rangle - \left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2.$$

Proof.

$$\begin{aligned}
V^k &= (k+1)\left\{k\|x^k - \mathbb{T}x^k\|^2 + 2\langle x^k - \mathbb{T}x^k, x^k - x^0\rangle\right\} - \left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2 \\
&\quad + k(k+1)\left\langle -\frac{2}{k}(x^k - x^0) - (x_\varepsilon - \mathbb{T}x_\varepsilon), x_\varepsilon - \mathbb{T}x_\varepsilon\right\rangle + \frac{2(k+1)}{k}\left\| x^k - x_\varepsilon + \frac{k}{2}(x_\varepsilon - \mathbb{T}x_\varepsilon)\right\|^2 \\
&\geq (k+1)\left\{k\|x^k - \mathbb{T}x^k\|^2 + 2\langle x^k - \mathbb{T}x^k, x^k - x^0\rangle\right\} - \left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2 \\
&\quad + k(k+1)\left\langle -\frac{2}{k}(x^k - x^0) - (x_\varepsilon - \mathbb{T}x_\varepsilon), x_\varepsilon - \mathbb{T}x_\varepsilon\right\rangle \\
&= k(k+1)\left(\|x^k - \mathbb{T}x^k\|^2 - \|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2\right) + 2(k+1)\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x^k - x^0\rangle \\
&\quad - \left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2 \\
&= k(k+1)\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 + 2k(k+1)\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x_\varepsilon - \mathbb{T}x_\varepsilon\rangle \\
&\quad + 2(k+1)\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x^k - x_\varepsilon\rangle - 2(k+1)\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x^0 - x_\varepsilon\rangle \\
&\quad - \left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2.
\end{aligned}$$

$\mathbb{T}$ is nonexpansive, from

$$\|\mathbb{T}x^k - \mathbb{T}x_\varepsilon\|^2 \leq \|x^k - x_\varepsilon\|^2,$$

we get

$$\begin{aligned}
\|x^k - x_\varepsilon\|^2 - \|\mathbb{T}x^k - \mathbb{T}x_\varepsilon\|^2 &= \left\langle (x^k - x_\varepsilon) - (\mathbb{T}x^k - \mathbb{T}x_\varepsilon),\, (x^k - x_\varepsilon) + (\mathbb{T}x^k - \mathbb{T}x_\varepsilon)\right\rangle \\
&= \left\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, 2(x^k - x_\varepsilon) - \{(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\}\right\rangle \\
&= 2\left\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x^k - x_\varepsilon\right\rangle - \left\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\right\|^2 \\
&\geq 0
\end{aligned}$$

so

$$\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x^k - x_\varepsilon\rangle \geq \frac{1}{2}\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2.$$

From this, we get

$$V^k \geq (k+1)^2\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 + 2k(k+1)\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x_\varepsilon - \mathbb{T}x_\varepsilon\rangle \\ - 2(k+1)\langle (x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x^0 - x_\varepsilon\rangle - \left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2.$$

□

Lemma 27. *Let* $\{x^k\}_{k\in\mathbb{N}}$ *be a sequence of iterates generated by* (Halpern) *starting from* $x^0 \in \mathcal{H}$ *with* $\lambda_k = \frac{1}{k+1}$ *and* $\{V^k\}_{k\in\mathbb{N}\cup\{0\}}$ *be defined as* (Lyapunov function). *Then*

$$V^1 \leq 3\|x^0 - x_\varepsilon\|^2.$$

Proof.

$$
\begin{aligned}
&V^1 \\
&= 2\left\{\|x^1-\mathbb{T}x^1\|^2+2\langle x^1-\mathbb{T}x^1,\, x^1-x^0\rangle\right\}-\|x^0-x_\varepsilon\|^2 \\
&\quad+2\left\langle -2(x^1-x^0)-(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x_\varepsilon-\mathbb{T}x_\varepsilon\right\rangle+4\left\|x^1-x_\varepsilon+\frac{1}{2}(x_\varepsilon-\mathbb{T}x_\varepsilon)\right\|^2 \\
&\le 0-2\left\langle 2(x^1-x^0)+(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x_\varepsilon-\mathbb{T}x_\varepsilon\right\rangle+4\left\|(x^1-x^0)+(x^0-x_\varepsilon)+\frac{1}{2}(x_\varepsilon-\mathbb{T}x_\varepsilon)\right\|^2-\|x^0-x_\varepsilon\|^2 \\
&= \left\|\{2(x^1-x^0)+(x_\varepsilon-\mathbb{T}x_\varepsilon)\}+2(x^0-x_\varepsilon)\right\|^2-2\langle 2(x^1-x^0)+(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x_\varepsilon-\mathbb{T}x_\varepsilon\rangle-\|x^0-x_\varepsilon\|^2 \\
&= \|-\{(x^0-\mathbb{T}x^0)-(x_\varepsilon-\mathbb{T}x_\varepsilon)\}+2(x^0-x_\varepsilon)\|^2+2\langle (x^0-\mathbb{T}x^0)-(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x_\varepsilon-\mathbb{T}x_\varepsilon\rangle-\|x^0-x_\varepsilon\|^2 \\
&= \|(x^0-\mathbb{T}x^0)-(x_\varepsilon-\mathbb{T}x_\varepsilon)\|^2-4\langle (x^0-\mathbb{T}x^0)-(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x^0-x_\varepsilon\rangle \\
&\quad+3\|x^0-x_\varepsilon\|^2+2\langle (x^0-\mathbb{T}x^0)-(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x_\varepsilon-\mathbb{T}x_\varepsilon\rangle \\
&\le 2\langle (x^0-\mathbb{T}x^0)-(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x_\varepsilon-\mathbb{T}x_\varepsilon\rangle+3\|x^0-x_\varepsilon\|^2-\|(x^0-\mathbb{T}x^0)-(x_\varepsilon-\mathbb{T}x_\varepsilon)\|^2 \\
&= 3\|x^0-x_\varepsilon\|^2-\|x^0-\mathbb{T}x^0\|^2-3\|x_\varepsilon-\mathbb{T}x_\varepsilon\|^2 \\
&\le 3\|x^0-x_\varepsilon\|^2.
\end{aligned}
$$

First inequality comes from Lemma 24 with $k=0$. ◻

Theorem 28. *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by* (Halpern) *starting from $x^0\in\mathcal{H}$ with $\lambda_k=\frac{1}{k+1}$ and $\{V^k\}_{k\in\mathbb{N}\cup\{0\}}$ be defined as* (Lyapunov function)*. For any $k\ge 1$,*

$$
\begin{aligned}
&(k+1)^2\|(x^k-\mathbb{T}x^k)-(x_\varepsilon-\mathbb{T}x_\varepsilon)\|^2+2k(k+1)\langle (x^k-\mathbb{T}x^k)-(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x_\varepsilon-\mathbb{T}x_\varepsilon\rangle \\
&-2(k+1)\langle (x^k-\mathbb{T}x^k)-(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x^0-x_\varepsilon\rangle-\left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0-x_\varepsilon\|^2 \\
&\le 3\|x^0-x_\varepsilon\|^2.
\end{aligned}
$$

Proof. Direct application of Lemma 25, Lemma 26 and Lemma 27. ◻

Lemma 29. *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by* (Halpern) *starting from $x^0\in\mathcal{H}$ with $\lambda_k=\frac{1}{k+1}$. For any $k\in\mathbb{N}$,*

$$
\|x^k-\mathbb{T}x^k\|\le\|x^0-\mathbb{T}x^0\|.
$$

Proof. We use Lemma 24, the definition of x^{k+1}-update and and that $\theta_k=\frac{k}{2}$, which is from Lemma 6. Dividing by $\frac{(k+1)(k+2)}{2}$, we have

$$
\begin{aligned}
&\frac{2k}{k+2}\|x^k-\mathbb{T}x^k\|^2+4\langle x^k-\mathbb{T}x^k,\, x^k-x^0\rangle \\
&\ge 2\|x^{k+1}-\mathbb{T}x^{k+1}\|^2+\frac{4}{k+1}\langle x^{k+1}-\mathbb{T}x^{k+1},\, x^{k+1}-x^0\rangle \\
&= \|x^{k+1}-\mathbb{T}x^{k+1}\|^2+\left\|(x^{k+1}-\mathbb{T}x^{k+1})+\frac{x^{k+1}-x^0}{\theta_{k+1}}\right\|^2-\left\|\frac{x^{k+1}-x^0}{\theta_{k+1}}\right\|^2.
\end{aligned}
$$

Since

$$
\frac{x^{k+1}-x^0}{\theta_{k+1}}=\frac{k}{k+2}\left(\frac{x^k-x^0}{\theta_k}\right)-\frac{2}{k+2}(x^k-\mathbb{T}x^k),
$$

we have

$$\begin{aligned}
&\|x^{k+1} - \mathbb{T}x^{k+1}\|^2 + \left\|(x^{k+1} - \mathbb{T}x^{k+1}) + \frac{x^{k+1} - x^0}{\theta_{k+1}}\right\|^2 \\
&\le \frac{2k}{k+2}\|x^k - \mathbb{T}x^k\|^2 + 4\langle x^k - \mathbb{T}x^k,\, x^k - x^0\rangle + \left\|\frac{k}{k+2}\left(\frac{x^k - x^0}{\theta_k}\right) - \frac{2}{k+2}(x^k - \mathbb{T}x^k)\right\|^2 \\
&= \|x^k - \mathbb{T}x^k\|^2 + \left(\frac{k}{k+2}\right)^2 \left\|(x^k - \mathbb{T}x^k) + \frac{x^k - x^0}{\theta_k}\right\|^2
\end{aligned}$$

hold for all $k = 0, 1, \ldots$. Therefore, for any $k \in \mathbb{N}$, we get

$$\begin{aligned}
\|x^0 - \mathbb{T}x^0\|^2 &\ge \|x^1 - \mathbb{T}x^1\|^2 + \left\|(x^1 - \mathbb{T}x^1) + \frac{x^1 - x^0}{\theta_1}\right\|^2 \\
&\ge \|x^1 - \mathbb{T}x^1\|^2 + \left(\frac{1}{1+2}\right)^2 \left\|(x^1 - \mathbb{T}x^1) + \frac{x^1 - x^0}{\theta_1}\right\|^2 \\
&\ge \|x^2 - \mathbb{T}x^2\|^2 + \left\|(x^2 - \mathbb{T}x^2) + \frac{x^2 - x^0}{\theta_2}\right\|^2 \\
&\ge \cdots \\
&\ge \|x^k - \mathbb{T}x^k\|^2 + \left\|(x^k - \mathbb{T}x^k) + \frac{x^k - x^0}{\theta_k}\right\|^2 \\
&\ge \|x^k - \mathbb{T}x^k\|^2.
\end{aligned}$$

□

Now we find some relation between $x_\varepsilon - \mathbb{T}x_\varepsilon$ and v.

Lemma 30. *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by* (Halpern) *starting from $x^0 \in \mathcal{H}$ with $\lambda_k = \frac{1}{k+1}$. For any $\varepsilon > 0$, there exists $x_\varepsilon \in \operatorname{dom}\mathbb{T}$ such that*

$$\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \le \varepsilon,$$

and from this,

$$\|x^k - \mathbb{T}x^k - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 \ge \|x^k - \mathbb{T}x^k - v\|^2 - 2\|x^0 - \mathbb{T}x^0\|\varepsilon$$

and

$$\langle x^k - \mathbb{T}x^k - (x_\varepsilon - \mathbb{T}x_\varepsilon), x_\varepsilon - \mathbb{T}x_\varepsilon\rangle \ge \langle x^k - \mathbb{T}x^k - v, v\rangle - \left\{\|x^0 - \mathbb{T}x^0\| + 2\|v\| + \varepsilon\right\}\varepsilon.$$

Furthermore, if $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, then there exists $x_\star \in \operatorname{dom}\mathbb{T}$ such that $x_\star - \mathbb{T}x_\star = v$.

Proof.

$$\begin{aligned}
&\|x^k - \mathbb{T}x^k - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 - \|x^k - \mathbb{T}x^k - v\|^2 \\
&= -2\langle x^k - \mathbb{T}x^k, x_\varepsilon - \mathbb{T}x_\varepsilon - v\rangle + \underbrace{\|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2}_{\ge 0} \\
&\ge -2\|x^k - \mathbb{T}x^k\|\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \\
&\ge -2\|x^0 - \mathbb{T}x^0\|\varepsilon
\end{aligned}$$

where the last inequality comes from Lemma 29. Also,

$$\begin{aligned}
&\langle x^k - \mathbb{T}x^k - (x_\varepsilon - \mathbb{T}x_\varepsilon), x_\varepsilon - \mathbb{T}x_\varepsilon\rangle \\
&= \langle (x^k - \mathbb{T}x^k - v) - (x_\varepsilon - \mathbb{T}x_\varepsilon - v), (x_\varepsilon - \mathbb{T}x_\varepsilon - v) + v\rangle \\
&= \langle x^k - \mathbb{T}x^k - v, v\rangle + \langle x^k - \mathbb{T}x^k - v, x_\varepsilon - \mathbb{T}x_\varepsilon - v\rangle - \|x_\varepsilon - \mathbb{T}x_\varepsilon - v\|^2 \\
&\quad - \langle x_\varepsilon - \mathbb{T}x_\varepsilon - v, v\rangle \\
&\ge \langle x^k - \mathbb{T}x^k - v, v\rangle - \|x^k - \mathbb{T}x^k - v\|\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \\
&\quad - \|x_\varepsilon - \mathbb{T}x_\varepsilon - v\|^2 - \|x_\varepsilon - \mathbb{T}x_\varepsilon - v\|\|v\|.
\end{aligned}$$

Then

$$\begin{aligned}
&\langle x^k - \mathbb{T}x^k - (x_\varepsilon - \mathbb{T}x_\varepsilon), x_\varepsilon - \mathbb{T}x_\varepsilon\rangle - \langle x^k - \mathbb{T}x^k - v, v\rangle \\
&\geq -\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \left\{\|x^k - \mathbb{T}x^k - v\| + \|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| + \|v\|\right\} \\
&\geq -\varepsilon \left\{(\|x^k - \mathbb{T}x^k\| + \|v\|) + \varepsilon + \|v\|\right\} \\
&= -\varepsilon \left\{\|x^0 - \mathbb{T}x^0\| + 2\|v\| + \varepsilon\right\}
\end{aligned}$$

where the last inequality comes from Lemma 29. □

We now prove the convergence rate result of (Halpern) with $\lambda_k = \frac{1}{k+1}$.

Theorem 31. *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by* (Halpern) *starting from $x^0 \in \mathcal{H}$ with $\lambda_k = \frac{1}{k+1}$. For any $\varepsilon > 0$ and $0 < \alpha < 1$, there exists $x_\varepsilon \in \operatorname{dom}\mathbb{T}$ such that*

$$\|x^k - \mathbb{T}x^k - v\|^2 \leq \frac{1}{(1-\alpha)(k+1)^2}\left(\sum_{n=1}^{k}\frac{1}{n} + 3 + \frac{1}{\alpha}\right)\|x^0 - x_\varepsilon\|^2 + \varepsilon.$$

If we further assume that $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, there exists $x_\star \in \mathcal{H}$ such that $v = x_\star - \mathbb{T}x_\star$ and

$$\|x^k - \mathbb{T}x^k - v\|^2 \leq \frac{1}{(1-\alpha)(k+1)^2}\left(\sum_{n=1}^{k}\frac{1}{n} + 3 + \frac{1}{\alpha}\right)\|x^0 - x_\star\|^2.$$

Proof. For $\varepsilon > 0$ and $0 < \alpha < 1$, consider $x_\varepsilon \in \operatorname{dom}\mathbb{T}$ such that

$$\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| \leq \tilde{\varepsilon}$$

where

$$\tilde{\varepsilon} = \min\left\{\left(2\|x^0 - \mathbb{T}x^0\| + \frac{2}{1-\alpha}(\|x^0 - \mathbb{T}x^0\| + 2\|v\| + 1)\right)^{-1}\varepsilon, 1, \varepsilon\right\}.$$

According to Theorem 28,

$$\begin{aligned}
&(k+1)^2\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 + 2k(k+1)\langle(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x_\varepsilon - \mathbb{T}x_\varepsilon\rangle \\
&- 2(k+1)\langle(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x^0 - x_\varepsilon\rangle - \left(\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2 \\
&\leq 3\|x^0 - x_\varepsilon\|^2.
\end{aligned}$$

For any $\alpha \in (0,1)$,

$$\begin{aligned}
3\|x^0 - x_\varepsilon\|^2 &\geq (1-\alpha)(k+1)^2\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 + 2k(k+1)\langle(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x_\varepsilon - \mathbb{T}x_\varepsilon\rangle \\
&\quad + \alpha(k+1)^2\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 + 2(k+1)\langle(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x^0 - x_\varepsilon\rangle \\
&\quad + \frac{1}{\alpha}\|x^0 - x_\varepsilon\|^2 - \left(\frac{1}{\alpha} + \sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2 \\
&= (1-\alpha)(k+1)^2\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 + 2k(k+1)\langle(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x_\varepsilon - \mathbb{T}x_\varepsilon\rangle \\
&\quad + \frac{1}{\alpha}\left\|\alpha(k+1)\{(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\} + (x^0 - x_\varepsilon)\right\|^2 - \left(\frac{1}{\alpha} + \sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2 \\
&\geq (1-\alpha)(k+1)^2\|(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon)\|^2 - \left(\frac{1}{\alpha} + \sum_{n=1}^{k}\frac{1}{n}\right)\|x^0 - x_\varepsilon\|^2 \\
&\quad + 2k(k+1)\langle(x^k - \mathbb{T}x^k) - (x_\varepsilon - \mathbb{T}x_\varepsilon),\, x_\varepsilon - \mathbb{T}x_\varepsilon\rangle.
\end{aligned}$$

Rearranging the terms, we get

$$\begin{aligned}
&\frac{1}{(1-\alpha)(k+1)^2}\left(3+\frac{1}{\alpha}+\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0-x_\varepsilon\|^2\\
&\geq \|(x^k-\mathbb{T}x^k)-(x_\varepsilon-\mathbb{T}x_\varepsilon)\|^2+\frac{2}{1-\alpha}\frac{k}{k+1}\langle(x^k-\mathbb{T}x^k)-(x_\varepsilon-\mathbb{T}x_\varepsilon),\, x_\varepsilon-\mathbb{T}x_\varepsilon\rangle\\
&\geq \|x^k-\mathbb{T}x^k-v\|^2-2\|x^0-\mathbb{T}x^0\|\tilde{\varepsilon}+\frac{2}{1-\alpha}\frac{k}{k+1}\langle x^k-\mathbb{T}x^k-v,\, v\rangle\\
&\quad -\frac{2}{1-\alpha}\frac{k}{k+1}\left\{\|x^0-\mathbb{T}x^0\|+2\|v\|+\tilde{\varepsilon}\right\}\tilde{\varepsilon}\\
&\geq \|x^k-\mathbb{T}x^k-v\|^2-\frac{2}{1-\alpha}\left\{(2-\alpha)\|x^0-\mathbb{T}x^0\|+2\|v\|+\tilde{\varepsilon}\right\}\tilde{\varepsilon}\\
&\geq \|x^k-\mathbb{T}x^k-v\|^2-\varepsilon.
\end{aligned}$$

The second inequality comes from Lemma 30, the third inequality comes from Lemma 2, and the last inequality comes from the definition of $\tilde{\varepsilon}>0$. □

Proof of Theorem 8. With $\lambda_k=\frac{1}{k+1}$, we have $\theta_k=\frac{k}{2}$ from Lemma 6. Using Lemma 24, we get

$$\begin{aligned}
&(k+2)\left\{(k+1)\|x^{k+1}-\mathbb{T}x^{k+1}\|^2+2\langle x^{k+1}-\mathbb{T}x^{k+1},x^{k+1}-x^0\rangle\right\}\\
&\leq (k+1)\left\{k\|x^k-\mathbb{T}x^k\|^2+2\langle x^k-\mathbb{T}x^k,x^k-x^0\rangle\right\}
\end{aligned}$$

for all $k=0,1,\dots$. Therefore, for any $k\in\mathbb{N}$,

$$k\|x^k-\mathbb{T}x^k\|^2+2\langle x^k-\mathbb{T}x^k,x^k-x^0\rangle\leq 0.$$

Using the Cauchy-Schwarz inequality, we get

$$\|x^k-\mathbb{T}x^k\|\leq\left\|\frac{2}{k}(x^k-x^0)\right\|=\left\|\frac{x^k-x^0}{\theta_k}\right\|$$

for any $k\in\mathbb{N}$. Therefore, for any $\varepsilon>0$ and $x_\varepsilon\in\mathcal{H}$ such that $\|x_\varepsilon-\mathbb{T}x_\varepsilon\|^2-\|v\|^2\leq\varepsilon^2$, we have

$$\|x^k-\mathbb{T}x^k\|-\|v\|\leq\left\|\frac{2}{k}(x^k-x^0)\right\|-\|v\|\leq\left\|\frac{2}{k}(x^k-x^0)+v\right\|\leq\frac{4}{k}\|x^0-x_\varepsilon\|+\varepsilon$$

for any $k\in\mathbb{N}$, where the second from last inequality comes from Theorem 7.

From Theorem 31, given an arbitrary $\varepsilon>0$ and x_ε such that

$$\|x_\varepsilon-\mathbb{T}x_\varepsilon-v\|\leq\tilde{\varepsilon}$$

where

$$\tilde{\varepsilon}=\min\left\{\left(2\|x^0-\mathbb{T}x^0\|+\frac{2}{1-\alpha}(\|x^0-\mathbb{T}x^0\|+2\|v\|+1)\right)^{-1}\varepsilon,1,\varepsilon\right\}=\mathcal{O}(\varepsilon),$$

we get

$$\|x^k-\mathbb{T}x^k-v\|^2\leq\frac{1}{(1-\alpha)(k+1)^2}\left(3+\frac{1}{\alpha}+\sum_{n=1}^{k}\frac{1}{n}\right)\|x^0-x_\varepsilon\|^2+\varepsilon$$

for any $0<\alpha<1$. Now we find a minimizer $\alpha^\star$ of

$$\frac{1}{1-\alpha}\left(c_k+\frac{1}{\alpha}\right)$$

where

$$c_k = 3 + \sum_{n=1}^{k} \frac{1}{n}$$

is a positive constant.

$$\begin{aligned}
\frac{1}{1-\alpha}\left(c_k + \frac{1}{\alpha}\right) &= \frac{c_k}{1-\alpha} + \frac{1}{\alpha(1-\alpha)} \\
&= \frac{c_k+1}{1-\alpha} + \frac{1}{\alpha} \\
&= \left(\frac{c_k+1}{1-\alpha} + \frac{1}{\alpha}\right)((1-\alpha)+\alpha) \\
&\geq (\sqrt{c_k+1}+1)^2 && (\because \text{Cauchy-Schwarz.})
\end{aligned}$$

and the equality holds if and only if

$$\frac{c_k+1}{(1-\alpha)^2} = \frac{1}{\alpha^2} \quad \Leftrightarrow \quad \alpha = \frac{1}{\sqrt{c_k+1}+1} = \frac{\sqrt{c_k+1}-1}{c_k}.$$

With such α, we get

$$\frac{1}{1-\alpha}\left(c_k + \frac{1}{\alpha}\right) = (\sqrt{c_k+1}+1)^2.$$

Therefore,

$$\|x^k - \mathbb{T}x^k - v\|^2 \leq \left(\frac{\sqrt{\sum_{n=1}^{k}\frac{1}{n}+4}+1}{k+1}\right)^2 \|x^0 - x_\varepsilon\|^2 + \varepsilon.$$

If $v = x_\star - \mathbb{T}x_\star$, we follow the same steps and get

$$\|x^k - \mathbb{T}x^k - v\|^2 \leq \left(\frac{\sqrt{\sum_{n=1}^{k}\frac{1}{n}+4}+1}{k+1}\right)^2 \|x^0 - x_\star\|^2.$$

□

We now prove the equivalence of the normalized iterate $-\frac{x^{k+1}-x^0}{k+1}$ of Picard iteration and the fixed-point residual $x^k - \mathbb{T}x^k$ of (Halpern) with $\lambda_k = \frac{1}{k+1}$ for affine $\mathbb{T}$, which was discussed in the last part of Section 3.2. Let $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ be an affine operator, i.e., $\mathbb{T}x = Ax + b$ where $A\colon \mathcal{H} \to \mathcal{H}$ is a linear operator and $b \in \mathcal{H}$.

Lemma 32. *Suppose $\mathbb{T}\colon \mathcal{H} \to \mathcal{H}$ is an affine operator. Let $\{x^k\}_{k\in\mathbb{N}}$ and $\{y^k\}_{k\in\mathbb{N}}$ be the sequences of iterates generated by* (Halpern) *with $\lambda_k = \frac{1}{k+1}$ and Picard iteration with $\mathbb{T}$, respectively, starting from the same initial point $x^0 = y^0$. Then for any $k \in \mathbb{N}\cup\{0\}$,*

$$x^k - \mathbb{T}x^k = -\frac{y^{k+1}-y^0}{k+1}.$$

Proof. First, note that when $\mathbb{T}$ is an affine operator, i.e., $\mathbb{T}x = Ax + b$ for any $x \in \mathcal{H}$,

$$\mathbb{T}\left(\sum_{i=1}^{k} \nu_i \mathbb{T}x_i\right) = \sum_{i=1}^{k} \nu_i \mathbb{T}x_i$$

for any $x_i \in \mathcal{H}$ and $\nu_i \in [0,1]$ such that $\sum_{i=1}^{k} \nu_i = 1$.

We see that for Picard iteration,

$$-\frac{y^{k+1}-y^0}{k+1} = -\frac{\mathbb{T}^{k+1}y^0 - y^0}{k+1} = -\frac{\mathbb{T}^{k+1}x^0 - x^0}{k+1}.$$

Considering the (Halpern) iterates $\{x^k\}_{k\in\mathbb{N}}$ starting from x^0, we claim by induction on k that

$$x^k = \frac{1}{k+1}\sum_{i=0}^{k}\mathbb{T}^i x^0.$$

When $k=1$,

$$x^1 = \frac{1}{2}\mathbb{T}x^0 + \frac{1}{2}x^0.$$

Now, suppose the claim holds for $k=n$.

$$\begin{aligned}
x^{n+1} &= \frac{n+1}{n+2}\mathbb{T}x^n + \frac{1}{n+2}x^0\\
&= \frac{n+1}{n+2}\mathbb{T}\left(\frac{1}{n+1}\sum_{i=0}^{n}\mathbb{T}^i x^0\right) + \frac{1}{n+2}x^0\\
&= \frac{n+1}{n+2}\left(\frac{1}{n+1}\sum_{i=0}^{n}\mathbb{T}^{i+1} x^0\right) + \frac{1}{n+2}x^0\\
&= \frac{1}{n+2}\sum_{i=0}^{n+1}\mathbb{T}^i x^0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
x^k - \mathbb{T}x^k &= \frac{1}{k+1}\sum_{i=0}^{k}\mathbb{T}^i x^0 - \mathbb{T}\left(\frac{1}{k+1}\sum_{i=0}^{k}\mathbb{T}^i x^0\right)\\
&= \frac{1}{k+1}\sum_{i=0}^{k}\mathbb{T}^i x^0 - \frac{1}{k+1}\sum_{i=0}^{k}\mathbb{T}^{i+1} x^0\\
&= \frac{1}{k+1}x^0 - \frac{1}{k+1}\mathbb{T}^{k+1}x^0\\
&= -\frac{\mathbb{T}^{k+1}x^0 - x^0}{k+1}.
\end{aligned}$$

□

Due to Lemma 32, when $\mathbb{T}$ is an affine nonexpansive operator, (Halpern) with $\lambda_k = \frac{1}{k+1}$ is exactly optimal with matching lower bound, for fixed-point residual.

B.3. Omitted proofs of Section 3.3

Consider a *Mann iteration*

$$x^k = \sum_{i=0}^{k}\nu_i^k \mathbb{T}x^{i-1} \tag{Mann}$$

where $\nu_i \geq 0$, $\sum_{i=0}^{k}\nu_i^k = 1$ and $\mathbb{T}x^{-1} := x^0$.

Lemma 33. *Let $\alpha_0 = 0$ and $\{x^k\}_{k\in\mathbb{N}}$ be a sequence of iterates generated by* (Mann) *starting from $x^0 \in \mathcal{H}$. If the sequence of real numbers $\{\alpha_k\}_{k\in\mathbb{N}\cup\{0\}}$ is defined recursively from the equation*

$$\alpha_k = (1-\nu_0^k) + \sum_{i=1}^{k}\nu_i^k\alpha_{i-1}, \quad k = 1, \ldots,$$

and $\alpha_k > 0$ for all $k\in\mathbb{N}$, then

$$-\frac{x^k - x^0}{\alpha_k} \in \overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}.$$

Proof. Note that α_k for $k \geq 1$ can also be written as

$$\alpha_k = \sum_{i=1}^{k} \nu_i^k + \sum_{i=1}^{k} \nu_i^k \alpha_{i-1}$$

since $\sum_{i=0}^{k} \nu_i^k = 1$.

Let $k = 0$. Then from the definition of (Mann), $\nu_0^0 = 1$. If $k = 1$, $\alpha_1 = \nu_1^1$ and

$$x^1 = \nu_0^1 \mathbb{T} x^{-1} + \nu_1^1 \mathbb{T} x^0 = x^0 - \nu_1^1 (x^0 - \mathbb{T} x^0).$$

so

$$-\frac{x^1 - x^0}{\alpha_1} = -\frac{x^1 - x^0}{\nu_1^1} = x^0 - \mathbb{T} x^0 \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}.$$

Now, fix $k > 1$ and suppose that

$$-\frac{x^i - x^0}{\alpha_i} \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}, \quad \forall\, i < k.$$

Then from

$$\begin{aligned} x^k - x^0 &= \sum_{i=0}^{k} \nu_i^k (\mathbb{T} x^{i-1} - x^0) \\ &= \sum_{i=1}^{k} \nu_i^k (\mathbb{T} x^{i-1} - x^0) \\ &= -\sum_{i=1}^{k} \nu_i^k (x^i - \mathbb{T} x^i) + \sum_{i=1}^{k} \nu_i^k (x^{i-1} - x^0), \end{aligned}$$

we get

$$-\frac{x^k - x^0}{\alpha_k} = \frac{\sum_{i=1}^{k} \nu_i^k (x^{i-1} - \mathbb{T} x^{i-1}) + \sum_{i=1}^{k} \nu_i^k \alpha_{i-1} \left(-\frac{x^{i-1} - x^0}{\alpha_{i-1}} \right)}{\sum_{i=1}^{k} \nu_i^k + \sum_{i=1}^{k} \nu_i^k \alpha_{i-1}}.$$

Since $\overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$ is a closed convex set, it is closed under convex combination. Therefore, $-\frac{x^k - x^0}{\alpha_k} \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$. $\square$

Remark 34. Note that $\{\alpha_k\}_{k \in \mathbb{N} \cup \{0\}}$ of (33) recovers $\sum_{i=1}^{k} (1 - \lambda_i)$ of (KM) and θ_k of (Halpern).

First of all, (KM) is defined as

$$x^k = (1 - \lambda_k) \mathbb{T} x^{k-1} + \lambda_k x^{k-1},$$

so $\nu_k^k = 1 - \lambda_k$. From recursively applying the same identity as above, we get

$$\nu_i^k = \begin{cases} 1 - \lambda_k & \text{if } i = k \\ \lambda_k \cdots \lambda_{i+1} (1 - \lambda_i) & \text{if } 1 \leq i < k \\ \lambda_{k+1} \lambda_k \cdots \lambda_1 & \text{if } i = 0 \end{cases}$$

From Lemma 33, as

$$\alpha_k = \sum_{i=1}^{k} \nu_i^k + \sum_{i=1}^{k} \nu_i^k \alpha_{i-1}$$

with $\alpha_0 = 0$, we get

$$\alpha_k = \sum_{i=1}^{k} (1 - \lambda_i)$$

from plugging ν_i^k above.

Next, (Halpern) is defined as

$$x^k = (1-\lambda_k)\mathbb{T}x^k + \lambda_k x^0$$

so

$$\nu_i^k = \begin{cases} 1-\lambda_k & \text{if } i = k \\ 0 & \text{if } 1 \le i < k \\ \lambda_k & \text{if } i = 0 \end{cases}$$

Then

$$\alpha_k = \sum_{i=1}^{k} \nu_i^k + \sum_{i=1}^{k} \nu_i^k \alpha_{i-1} = (1-\lambda_k) + (1-\lambda_k)\alpha_{k-1} = (1-\lambda_k)(1+\alpha_{k-1}).$$

This recursive formula is exactly the same as the recursive formula in Lemma 6, so $\alpha_k = \theta_k$.

We elaborate on some properties of (Mann) which will be used in our main result, Theorem 36.

Lemma 35. *Let $\{x^k\}_{k\in\mathbb{N}}$ and $\{y^k\}_{k\in\mathbb{N}}$ be the sequences of iterates generated by* (Mann) *starting from $x^0 \in \mathcal{H}$ and $y^0 \in \mathcal{H}$, respectively. Define $\{\alpha_k\}_{k\in\mathbb{N}\cup\{0\}}$ as in Lemma 33. Then*

$$\left\| \frac{x^k - x^0}{\alpha_k} \right\| \le \|x^0 - \mathbb{T}x^0\|, \quad k = 1, 2, \dots$$

and

$$\|x^k - y^k\| \le \|x^0 - y^0\|, \quad k = 1, 2, \dots.$$

Proof. For $k = 1$, $-\frac{x^1 - x^0}{\alpha_1} = x^0 - \mathbb{T}x^0$ so the claim is trivial. Now, let

$$\left\| \frac{x^i - x^0}{\alpha_i} \right\| \le \|x^0 - \mathbb{T}x^0\|$$

for all $i < k$. Then

$$\begin{aligned} x^k - x^0 &= \sum_{i=1}^{k} \nu_i^k (\mathbb{T}x^{i-1} - x^0) \\ &= -\sum_{i=1}^{k} \nu_i^k (x^0 - \mathbb{T}x^0) + \sum_{i=1}^{k} \nu_i^k (\mathbb{T}x^{i-1} - \mathbb{T}x^0) \\ &= -(1-\nu_0^k)(x^0 - \mathbb{T}x^0) + \sum_{i=1}^{k} \nu_i^k (\mathbb{T}x^{i-1} - \mathbb{T}x^0), \end{aligned}$$

so

$$\begin{aligned} \left\| \frac{x^k - x^0}{\alpha_k} \right\| &= \frac{1}{\alpha_k} \left\| (1-\nu_0^k)(x^0 - \mathbb{T}x^0) - \sum_{i=2}^{k} \nu_i^k \alpha_{i-1} \left(\frac{\mathbb{T}x^{i-1} - \mathbb{T}x^0}{\alpha_{i-1}} \right) \right\| \\ &\le \frac{1}{\alpha_k} \left\{ (1-\nu_0^k)\|x^0 - \mathbb{T}x^0\| + \sum_{i=2}^{k} \nu_i^k \alpha_{i-1} \left\| \frac{\mathbb{T}x^{i-1} - \mathbb{T}x^0}{\alpha_{i-1}} \right\| \right\} \\ &\le \frac{1}{\alpha_k} \left\{ (1-\nu_0^k)\|x^0 - \mathbb{T}x^0\| + \sum_{i=2}^{k} \nu_i^k \alpha_{i-1} \left\| \frac{x^{i-1} - x^0}{\alpha_{i-1}} \right\| \right\} \\ &\le \frac{1}{\alpha_k} \left\{ (1-\nu_0^k)\|x^0 - \mathbb{T}x^0\| + \sum_{i=2}^{k} \nu_i^k \alpha_{i-1} \|x^0 - \mathbb{T}x^0\| \right\} \\ &= \frac{1}{\alpha_k} \left\{ (1-\nu_k^k) + \sum_{i=2}^{k} \nu_i^k \alpha_{i-1} \right\} \|x^0 - \mathbb{T}x^0\| = \|x^0 - \mathbb{T}x^0\|. \end{aligned}$$

Now we prove the second claim. First of all,

$$\begin{aligned}\|x^1 - y^1\| &= \|\nu_0^1(x^0 - y^0) + \nu_1^1(\mathbb{T}x^0 - \mathbb{T}y^0)\| \\ &\le \nu_0^1\|x^0 - y^0\| + \nu_1^1\|\mathbb{T}x^0 - \mathbb{T}y^0\| \\ &\le \nu_0^1\|x^0 - y^0\| + \nu_1^1\|x^0 - y^0\| = \|x^0 - y^0\|.\end{aligned}$$

Suppose $\|x^i - y^i\| \le \|x^0 - y^0\|$ for all $i < k$. Then

$$\begin{aligned}\|x^k - y^k\| &= \left\|\sum_{i=0}^{k} \nu_i^k(\mathbb{T}x^{i-1} - \mathbb{T}y^{i-1})\right\| \\ &\le \sum_{i=0}^{k} \nu_i^k\|\mathbb{T}x^{i-1} - \mathbb{T}y^{i-1}\| \\ &\le \nu_0^0\|x^0 - y^0\| + \sum_{i=1}^{k} \nu_i^k\|x^{i-1} - y^{i-1}\| \\ &\le \nu_0^0\|x^0 - y_0\| + \sum_{i=1}^{k} \nu_i^k\|x^0 - y^0\| = \|x^0 - y^0\|.\end{aligned}$$

□

We can extend Theorem 3 and Theorem 7 to cover the case of general Mann iteration.

Theorem 36. *Let $\{x^k\}_{k\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ generated by* (Mann) *starting from $x^0 \in \mathcal{H}$ and $\{\alpha_k\}_{k\in\mathbb{N}\cup\{0\}}$ be a sequence of positive numbers defined in Lemma 33. Then for any $\varepsilon > 0$, there exists $x_\varepsilon \in \mathcal{H}$ such that $\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\| < \varepsilon$ and*

$$\left\|\frac{x^k - x^0}{\alpha_k} + v\right\| \le \frac{2}{\alpha_k}\|x^0 - x_\varepsilon\| + \varepsilon.$$

If we further assume that $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, then there exists $x_\star \in \mathcal{H}$ such that

$$\left\|\frac{x^k - x^0}{\alpha_k} + v\right\| \le \frac{2}{\alpha_k}\|x^0 - x_\star\|.$$

Therefore, if $\lim_{k\to\infty} \alpha_k = \infty$, then

$$\lim_{k\to\infty} \frac{x^k - x^0}{\alpha_k} = -v.$$

Proof. Fix $\varepsilon > 0$. Let $x_\varepsilon \in \mathcal{H}$ be a point such that $\|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2 < \varepsilon^2$. Then

$$\begin{aligned}\|x_\varepsilon - \mathbb{T}x_\varepsilon - v\|^2 &= \|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2 - 2\underbrace{\langle x_\varepsilon - \mathbb{T}x_\varepsilon - v,\, v\rangle}_{\ge 0} \\ &\le \|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2 \\ &< \varepsilon^2.\end{aligned}$$

Suppose $\{x_\varepsilon^k\}_{k\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ generated by (Mann) starting from x_ε. Since $-\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k} \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$ by Lemma 33,

$$\left\langle -\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k},\, v\right\rangle \ge \|v\|^2$$

for any $k \in \mathbb{N}$, so we get

$$\begin{aligned}\left\|\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k} + v\right\|^2 &= \left\|\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k}\right\|^2 + \|v\|^2 - 2\left\langle -\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k},\, v\right\rangle \\ &\le \left\|\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k}\right\|^2 - \|v\|^2 \\ &\le \|x_\varepsilon - \mathbb{T}x_\varepsilon\|^2 - \|v\|^2 \le \varepsilon^2\end{aligned}$$

or $\left\|\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k} + v\right\| \le \varepsilon$.

$$\begin{aligned}\left\|\frac{x^k - x^0}{\alpha_k} + v\right\| &\le \left\|\frac{x^k - x^0}{\alpha_k} - \frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k}\right\| + \left\|\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k} + v\right\| \\ &\le \frac{\|x^k - x_\varepsilon^k\|}{\alpha_k} + \frac{\|x^0 - x_\varepsilon\|}{\alpha_k} + \left\|\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k} + v\right\| \\ &\le \frac{2}{\alpha_k}\|x^0 - x_\varepsilon\| + \left\|\frac{x_\varepsilon^k - x_\varepsilon}{\alpha_k} + v\right\| \\ &\le \frac{2}{\alpha_k}\|x^0 - x_\varepsilon\| + \varepsilon\end{aligned}$$

holds, where the third inequality comes from Lemma 35. If $v \in \mathcal{R}(\mathbb{I} - \mathbb{T})$, there exists $x_\star \in \mathcal{H}$ such that $v = x_\star - \mathbb{T}x_\star$, and the above proof stil holds with $\varepsilon = 0$ and $x_\varepsilon = x_\star$. Therefore,

$$\left\|\frac{x^k - x^0}{\alpha_k} + v\right\| \le \frac{2}{\alpha_k}\|x^0 - x_\star\|.$$

If $\lim_{k\to\infty} \alpha_k = \infty$, for any $\varepsilon > 0$,

$$\limsup_{k\to\infty} \left\|\frac{x^k - x^0}{\alpha_k} + v\right\| \le \varepsilon,$$

so we get $\lim_{k\to\infty} \frac{x^k - x^0}{\alpha_k} = -v$. □

Remark 37. By obtaining the upper bound to α_k, we may optimize the upper bound of Theorem 36. From the definition of $\{\alpha_k\}_{k\in\mathbb{N}\cup\{0\}}$,

$$\alpha_k = (1 - \nu_0^k) + \sum_{i=2}^{k} \nu_i^k \alpha_{i-1} = \sum_{i=1}^{k} \nu_i^k (1 + \alpha_{i-1})$$

with $\alpha_0 = 0$.

Consider an extreme case of (Picard), which corresponds to the choice of $\{\nu_i^k\}_{i=1,\dots,k}$ for $k \in \mathbb{N}\cup\{0\}$ as

$$\nu_i^k = \begin{cases} 0 & (0 \le i \le k-1) \\ 1 & (i = k) \end{cases}$$

In this case, $\alpha_k = k$. We claim that this is the biggest possible value for α_k for any $k \in \mathbb{N}$, using induction. First, $\alpha_0 = 0$. Suppose $\alpha_i \le i$ for all i such that $0 \le i < k$. Then

$$\begin{aligned}\alpha_k &= \sum_{i=1}^{k} \nu_i^k (1 + \alpha_{i-1}) \\ &\le \sum_{i=1}^{k} \nu_i^k \{1 + (i-1)\} \\ &\le \sum_{i=1}^{k} \nu_i^k \{1 + (k-1)\} \\ &= k \sum_{i=1}^{k} \nu_i^k \le k.\end{aligned}$$

Therefore, $\alpha_k \le k$ for all $k \in \mathbb{N}$. Hence (Picard) yields optimal upper bound, which is the same optimal upper bound as in Theorem 36.

C. Omitted proofs of Section 4

C.1. Omitted proofs of Section 4.1

Below results will be used to prove Theorem 10.

Theorem 38 (Projection theorem, Theorem 3.16, Bauschke & Combettes (2017))**.** *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then for every x and p in $\mathcal{H}$,*

$$p = \Pi_C x \;\Leftrightarrow\; \Big[\langle y - p,\, x - p\rangle \leq 0, \quad \forall\, y \in C \Big]$$

Theorem 39 (Corollary 5, Bauschke (2007))**.** *Let D be a nonempty subset of $\mathcal{H}$ and let $\mathbb{T}\colon D \to \mathcal{H}$ be firmly-nonexpansive operator. Then there exists a firmly-nonexpansive operator $\widetilde{\mathbb{T}}\colon \mathcal{H} \to \mathcal{H}$ such that $\widetilde{\mathbb{T}}\,|_D = \mathbb{T}$ and $\mathcal{R}(\widetilde{\mathbb{T}}) \subset \overline{\mathrm{conv}}\,\mathcal{R}(\mathbb{T})$.*

Lemma 40. *Let R be a nonempty set in $\mathcal{H}$. Suppose that $v \in R$ is a vector such that*

$$\langle x - v,\, v\rangle \geq 0, \quad \forall\, x \in R.$$

Then

$$\langle x - v,\, v\rangle \geq 0, \quad \forall\, x \in \overline{\mathrm{conv}}\, R.$$

Proof. Let $x \in \overline{\mathrm{conv}}\, R$. Then there exists $\{x_k\}_{k\in\mathbb{N}}$ such that $x_k \in \mathrm{conv}\, R$ for all k and $\lim_{k\to\infty} x_k = x$.

Since $x_k \in \mathrm{conv}\, R$, for each k, there exist $n_k \in \mathbb{N}$, $\alpha_i^k \in (0, 1]$ and $x_i^k \in R$ for $i = 1, \dots, n_k$ such that

$$x_k = \sum_{i=1}^{n_k} \alpha_i^k x_i^k$$

and $\sum_{i=1}^{n_k} \alpha_i^k = 1$.

$$\langle x_k,\, v\rangle = \left\langle \sum_{i=1}^{n_k} \alpha_i^k x_i^k,\, v \right\rangle = \sum_{i=1}^{n_k} \alpha_i^k \langle x_i^k,\, v\rangle \geq \sum_{i=1}^{n_k} \alpha_i^k \|v\|^2 = \|v\|^2.$$

So $\langle x_k - v,\, v\rangle \geq 0$ for all $x_k \in \mathrm{conv}\, R$. Then

$$\langle x,\, v\rangle = \left\langle \lim_{k\to\infty} x_k,\, v \right\rangle = \lim_{k\to\infty} \langle x_k,\, v\rangle \geq \|v\|^2.$$

□

Lemma 41. *Let $\{(x_i,\, y_i)\}_{i\in I} \subset \mathcal{H} \times \mathcal{H}$ be a set of vectors with index set I such that*

$$\|y_i - y_j\| \leq \|x_i - x_j\|, \quad \forall\, i, j \in I$$

and define $D = \{x_i\}_{i\in I} \subset \mathcal{H}$. Then there exists a nonexpansive operator $\widetilde{\mathbb{T}}\colon \mathcal{H} \to \mathcal{H}$ such that $\widetilde{\mathbb{T}}\,|_D = \mathbb{T}$ and

$$\overline{\mathcal{R}(\mathbb{I} - \widetilde{\mathbb{T}})} = \overline{\mathrm{conv}}\,\mathcal{R}(\mathbb{I} - \mathbb{T}).$$

Proof. Define an operator $\mathbb{T}\colon D \to \mathcal{H}$ as

$$\mathbb{T}x_i = y_i, \quad i \in I$$

where $D = \{x_i\}_{i\in I} \subset \mathcal{H}$. Then $\mathbb{S}\colon D \to \mathcal{H}$ defined as $\mathbb{S} = \frac{\mathbb{I}-\mathbb{T}}{2}$ is a firmly-nonexpansive operator. According to Theorem 39, there exists a firmly-nonexpansive extension $\widetilde{\mathbb{S}}\colon \mathcal{H} \to \mathcal{H}$ of $\mathbb{S}$ such that $\widetilde{\mathbb{S}}\,|_D = \mathbb{S}$ and $\mathcal{R}(\widetilde{\mathbb{S}}) \subset \overline{\mathrm{conv}}\,\mathcal{R}(\mathbb{S})$. If $\widetilde{\mathbb{T}} = \mathbb{I} - 2\widetilde{\mathbb{S}}$, then $\widetilde{\mathbb{T}}\colon \mathcal{H} \to \mathcal{H}$ becomes a nonexpansive extension of $\mathbb{T}$ such that $\widetilde{\mathbb{T}}\,|_D = \mathbb{T}$ and $\mathcal{R}(\mathbb{I} - \widetilde{\mathbb{T}}) = 2\mathcal{R}(\widetilde{\mathbb{S}}) \subset 2\overline{\mathrm{conv}}\,\mathcal{R}(\mathbb{S}) = \overline{\mathrm{conv}}\,\mathcal{R}(2\mathbb{S}) = \overline{\mathrm{conv}}\,\mathcal{R}(\mathbb{I} - \mathbb{T})$. Obviously,

$$\mathcal{R}(\mathbb{I} - \mathbb{T}) \subseteq \mathcal{R}(\mathbb{I} - \widetilde{\mathbb{T}}) \subseteq \overline{\mathrm{conv}}\,\mathcal{R}(\mathbb{I} - \mathbb{T}).$$

Since $\overline{\mathcal{R}(\mathbb{I}-\widetilde{\mathbb{T}})}$ is a convex set,

$$\operatorname{conv}\mathcal{R}(\mathbb{I}-\mathbb{T}) \subseteq \overline{\mathcal{R}(\mathbb{I}-\widetilde{\mathbb{T}})} \subseteq \overline{\operatorname{conv}}\,\mathcal{R}(\mathbb{I}-\mathbb{T}),$$

and as it is also a closed set,

$$\overline{\mathcal{R}(\mathbb{I}-\widetilde{\mathbb{T}})} = \overline{\operatorname{conv}}\,\mathcal{R}(\mathbb{I}-\mathbb{T}).$$

□

We now prove Theorem 10.

Proof of Theorem 10. Let $C = \overline{\operatorname{conv}}\,\{x_i - y_i\}_{i\in I}$.

(i) From Lemma 41, there exists a nonexpansive operator $\widetilde{\mathbb{T}}\colon \mathcal{H}\to\mathcal{H}$ such that $y_i = \widetilde{\mathbb{T}}x_i,\ \forall\, i\in I$ and $\overline{\mathcal{R}(\mathbb{I}-\widetilde{\mathbb{T}})} = C$. Then $v = \Pi_C(0) = \Pi_{\overline{\mathcal{R}(\mathbb{I}-\widetilde{\mathbb{T}})}}(0)$ is an infimal displacement vector of $\widetilde{\mathbb{T}}$.

(ii) Further assume that $v = x_\star - y_\star$ with $\star\in I$ and

$$\langle x_i - y_i,\, v\rangle \geq \|v\|^2, \quad \forall\, i\in I.$$

Then Lemma 41 asserts that there exists a nonexpansive operator $\widetilde{\mathbb{T}}\colon \mathcal{H}\to\mathcal{H}$ such that $y_i = \widetilde{\mathbb{T}}x_i,\ \forall\, i\in I$ and $\overline{\mathcal{R}(\mathbb{I}-\widetilde{\mathbb{T}})} = C$. According to Lemma 40, $\langle z,\, v\rangle \geq \|v\|^2$ for all $z\in C$. Then from Theorem 38, $v = \Pi_C(0) = \Pi_{\overline{\mathcal{R}(\mathbb{I}-\widetilde{\mathbb{T}})}}(0)$ so it is an infimal displacement vector of $\widetilde{\mathbb{T}}$.

□

C.2. Omitted proofs of Section 4.2

Problem we want to solve is a maximization problem in following form, given $k\in\mathbb{N}$ and an index set $I=\{0,1,\dots,k,\star\}$. As pointed out, we restrict the choice of nonexpansive operator $\mathbb{T}$ to be the ones where v actually lies in the range of $\mathbb{I}-\mathbb{T}$.

$$\begin{array}{ll}
\underset{\mathbb{T}}{\text{maximize}} & \|x^k - \mathbb{T}x^k - v\|^2 \\
\text{subject to} & \mathbb{T}\colon \mathcal{H}\to\mathcal{H} \text{ is nonexpansive} \\
& v = \Pi_{\overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}}(0) = x_\star - \mathbb{T}x_\star \\
& x^{n+1} = \dfrac{n+1}{n+2}\mathbb{T}x^n + \dfrac{1}{n+2}x^0, \quad n = 0,1,\dots,k-1 \\
& \|x^0 - x_\star\|^2 \leq R^2
\end{array}$$

Without loss of generality, we may only consider the case of $R=1$, which can be rescaled by R to obtain original problem.

$$\begin{array}{ll}
\underset{\mathbb{T}}{\text{maximize}} & \|x^k - \mathbb{T}x^k - v\|^2 \\
\text{subject to} & \mathbb{T}\colon \mathcal{H}\to\mathcal{H} \text{ is nonexpansive} \\
& v = \Pi_{\overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}}(0) = x_\star - \mathbb{T}x_\star \\
& x^{n+1} = \dfrac{n+1}{n+2}\mathbb{T}x^n + \dfrac{1}{n+2}x^0, \quad n = 0,1,\dots,k-1 \\
& \|x^0 - x_\star\|^2 \leq 1
\end{array}$$

Above problem is an infinite-dimensional problem, which is possibly an intractable problem. Such dimensionality stems from the variable of this problem, $\mathbb{T}$, lying in a function space which cannot be finite-dimensional.

We use Theorem 10 reduce the problem dimension by not considering the whole function space of nonexpansive operators any more. According to Theorem 10, the existence of nonexpansive $\mathbb{T}$ with $v = x_\star - \mathbb{T}x_\star$ is equivalent to the existence of iterates $\{(x^i,\, y^i)\}_{i\in I}$ satisfying the following inequalities.

$$\begin{aligned} \|y^i - y^j\|^2 &\le \|x^i - x^j\|^2, \quad \forall\, i,j \in I,\, i \neq j \\ \langle x^i - y^i,\, v\rangle &\ge \|v\|^2, \quad \forall\, i \in I \end{aligned}$$

Therefore, the problem can be reformulated as

$$\begin{array}{ll} \underset{\{(x^i,\,y^i)\}_{i\in I}}{\text{maximize}} & \|x^k - y^k - v\|^2 \\ \text{subject to} & (\exists \mathcal{H})\ \big(x^i,\, y^i \in \mathcal{H}, \quad \forall\, i \in I\big) \\ & \|y^i - y^j\|^2 \le \|x^i - x^j\|^2, \quad \forall\, i,j \in I,\, i\neq j \\ & v = x_\star - y_\star \\ & \langle x^i - y^i,\, v\rangle \ge \|v\|^2, \quad \forall\, i \in I \\ & x^{n+1} = \dfrac{n+1}{n+2} y^n + \dfrac{1}{n+2} x^0, \quad n = 0,1,\dots,k-1 \\ & \|x^0 - x_\star\|^2 \le 1 \end{array}$$

However, this problem is still intractable in a sense that the iterates $\{(x^i,\, y^i)\}_{i\in I}$ needs to be searched within any choice of real Hilbert space $\mathcal{H}$. We remove such dependency using the semidefinite formulation of PEP.

Consider a gram matrix $Z \in \mathbb{S}^{k+3}$ defined as

$$\begin{aligned} Z &= \begin{bmatrix} \|v^0\|^2 & \langle v^0,\, v^1\rangle & \cdots & \langle v^0,\, v^k\rangle & \langle v^0,\, v\rangle & \langle v^0,\, x^0 - x_\star\rangle \\ \langle v^1,\, v^0\rangle & \|v^1\|^2 & \cdots & \langle v^1,\, v^k\rangle & \langle v^1,\, v\rangle & \langle v^1,\, x^0 - x_\star\rangle \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \langle v^k,\, v^0\rangle & \langle v^k,\, v^1\rangle & \cdots & \|v^k\|^2 & \langle v^k,\, v\rangle & \langle v^k,\, x^0 - x_\star\rangle \\ \langle v,\, v^0\rangle & \langle v,\, v^1\rangle & \cdots & \langle v,\, v^k\rangle & \|v\|^2 & \langle v,\, x^0 - x_\star\rangle \\ \langle x^0 - x_\star, v^0\rangle & \langle x^0 - x_\star, v^1\rangle & \cdots & \langle x^0 - x_\star,\, v^k\rangle & \langle x^0 - x_\star,\, v\rangle & \|x^0 - x_\star\|^2 \end{bmatrix} \\ &= \begin{bmatrix} v^0 & v^1 & \cdots & v^k & v & x^0 - x_\star \end{bmatrix}^\intercal \begin{bmatrix} v^0 & v^1 & \cdots & v^k & v & x^0 - x_\star \end{bmatrix} \end{aligned}$$

where $v^i = x^i - y^i$ for $i \in I \setminus \{\star\}$ and $v = x_\star - y_\star$. Let G denote the horizontal stack of vectors

$$G = \begin{bmatrix} v^0 & v^1 & \cdots & v^k & v & x_\star & x^0 - x_\star \end{bmatrix},$$

then $Z = G^\intercal G$. From

$$x^{n+1} = \frac{n+1}{n+2} y^n + \frac{1}{n+2} x^0, \quad n = 0,1,\dots,k-1$$

being equivalent to

$$x^{n+1} = x^0 - \sum_{i=0}^{n} \frac{i+1}{n+2}\, v^i, \quad n = 0,1,\dots,k-1, \tag{2}$$

and we use this fact for our semidefinite PEP formulation.

For notational simplicity, let $\mathbf{e}_i \in \mathbb{R}^{k+3}$ denote the i-th canonical basis vector, i.e., only the i-th entry of $(k+3)$-dimensional real vector is 1 and all the other entries are 0, and let $a \odot b = \frac{1}{2}(ab^\intercal + ba^\intercal)$.

(i) Objective function.

$$\begin{aligned} \|x^k - y^k - v\|^2 &= \|v^k - v\|^2 \\ &= (G(\mathbf{e}_{k+1} - \mathbf{e}_{k+2}))^\intercal\, (G(\mathbf{e}_{k+1} - \mathbf{e}_{k+2})) \\ &= \operatorname{tr}\left((\mathbf{e}_{k+1} - \mathbf{e}_{k+2})(\mathbf{e}_{k+1} - \mathbf{e}_{k+2})^\intercal Z\right) \end{aligned}$$

Letting $C_k = (\mathbf{e}_{k+1} - \mathbf{e}_{k+2})(\mathbf{e}_{k+1} - \mathbf{e}_{k+2})^\intercal$, $\|x^k - y^k - v\|^2 = \operatorname{tr}(C_k Z)$.

(ii) Interpolation condition on nonexpansiveness.

Using (2),

$$
\begin{aligned}
&\|x^i - x^j\|^2 - \|y^i - y^j\|^2 \\
&= 2\langle (x^i - y^i) - (x^j - y^j),\, x^i - x^j\rangle - \|(x^i - y^i) - (x^j - y^j)\|^2 \\
&= 2\langle v^i - v^j,\, x^i - x^j\rangle - \|v^i - v^j\|^2 \\
&= 2\left\langle G(\mathbf{e}_{i+1} - \mathbf{e}_{j+1}),\, G\left(-\sum_{l=0}^{i-1}\frac{l+1}{i+1}\mathbf{e}_{l+1} + \sum_{m=0}^{j-1}\frac{m+1}{j+1}\mathbf{e}_{j+1}\right)\right\rangle - (G(\mathbf{e}_{i+1} - \mathbf{e}_{j+1}))^\intercal\,(G(\mathbf{e}_{i+1} - \mathbf{e}_{j+1})) \\
&= \mathrm{tr}\left[\left\{-2(\mathbf{e}_{i+1} - \mathbf{e}_{j+1}) \odot \left(\sum_{l=0}^{i-1}\frac{l+1}{i+1}\mathbf{e}_{l+1} - \sum_{m=0}^{j-1}\frac{m+1}{j+1}\mathbf{e}_{j+1}\right) + (\mathbf{e}_{i+1} - \mathbf{e}_{j+1})(\mathbf{e}_{i+1} - \mathbf{e}_{j+1})^\intercal\right\} Z\right]
\end{aligned}
$$

Letting

$$
A_{i,j} = -2(\mathbf{e}_{i+1} - \mathbf{e}_{j+1}) \odot \left(\sum_{l=0}^{i-1}\frac{l+1}{i+1}\mathbf{e}_{l+1} - \sum_{m=0}^{j-1}\frac{m+1}{j+1}\mathbf{e}_{j+1}\right) + (\mathbf{e}_{i+1} - \mathbf{e}_{j+1})(\mathbf{e}_{i+1} - \mathbf{e}_{j+1})^\intercal,
$$

the inequality condition $\|y^i - y^j\|^2 \le \|x^i - x^j\|^2$ for $i, j \in I \setminus \{\star\}$ is equivalent to $\mathrm{tr}(A_{i,j}Z) \ge 0$.

Also,

$$
\begin{aligned}
&\|x^i - x_\star\|^2 - \|y^i - y_\star\|^2 \\
&= 2\langle (x^i - y^i) - (x_\star - y_\star),\, x^i - y^i\rangle - \|(x^i - y^i) - (x_\star - y_\star)\|^2 \\
&= 2\langle v^i - v,\, x^i - x_\star\rangle - \|v^i - v\|^2 \\
&= 2\left\langle G(\mathbf{e}_{i+1} - \mathbf{e}_{k+2}),\, G\left(\mathbf{e}_{k+3} - \sum_{l=0}^{i-1}\frac{l+1}{i+1}\mathbf{e}_{l+1}\right)\right\rangle - (G(\mathbf{e}_{i+1} - \mathbf{e}_{k+2}))^\intercal\,(G(\mathbf{e}_{i+1} - \mathbf{e}_{k+2})) \\
&= \mathrm{tr}\left[\left\{-2(\mathbf{e}_{i+1} - \mathbf{e}_{k+2}) \odot \left(\sum_{l=0}^{i-1}\frac{l+1}{i+1}\mathbf{e}_{l+1} - \mathbf{e}_{k+3}\right) - (\mathbf{e}_{i+1} - \mathbf{e}_{k+2})(\mathbf{e}_{i+1} - \mathbf{e}_{k+2})^\intercal\right\} Z\right].
\end{aligned}
$$

Letting

$$
A_{i,\star} = -2(\mathbf{e}_{i+1} - \mathbf{e}_{k+2}) \odot \left(\sum_{l=0}^{i-1}\frac{l+1}{i+1}\mathbf{e}_{l+1} - \mathbf{e}_{k+3}\right) - (\mathbf{e}_{i+1} - \mathbf{e}_{k+2})(\mathbf{e}_{i+1} - \mathbf{e}_{k+2})^\intercal,
$$

the inequality condition $\|y^i - y_\star\|^2 \le \|x^i - x_\star\|^2$ is equivalent to $\mathrm{tr}(A_{i,\star}Z) \ge 0$.

(iii) Interpolation condition on infimal displacement vector.

$$
\begin{aligned}
\langle v^i,\, v\rangle - \|v\|^2 &= \langle v^i - v,\, v\rangle \\
&= \langle G(\mathbf{e}_{i+1} - \mathbf{e}_{k+2}),\, G\mathbf{e}_{k+2}\rangle \\
&= \mathrm{tr}\left[\left((\mathbf{e}_{i+1} - \mathbf{e}_{k+2}) \odot \mathbf{e}_{k+2}\right) Z\right]
\end{aligned}
$$

Therefore, letting $B_i = (\mathbf{e}_{i+1} - \mathbf{e}_{k+2}) \odot \mathbf{e}_{k+2}$, the inequality condition $\langle v^i,\, v\rangle \ge \|v\|^2$ is equivalent to $\mathrm{tr}(B_iZ) \ge 0$.

(iv) Initial point condition.

$$
\begin{aligned}
\|x^0 - x_\star\|^2 &= (G\mathbf{e}_{k+3})^\intercal\,(G\mathbf{e}_{k+3}) \\
&= \mathrm{tr}\left(\mathbf{e}_{k+3}\mathbf{e}_{k+3}{}^\intercal Z\right)
\end{aligned}
$$

so if $D_0 = \mathbf{e}_{k+3}\mathbf{e}_{k+3}{}^\intercal$, the inequality condition $\|x^0 - x_\star\|^2 \le 1$ is equivalent to $\mathrm{tr}(D_0Z) \le 1$.

Gathering all these facts, the problem at hand can be reformulated into the semidefinite program

$$\begin{array}{ll}\underset{Z\in\mathbb{S}_+^{k+3}}{\text{maximize}} & \operatorname{tr}(C_k Z)\\ \text{subject to} & \operatorname{tr}(A_{i,j}Z)\geq 0,\quad \forall\, i,j\in I\setminus\{\star\},\ i\neq j\\ & \operatorname{tr}(A_{i,\star}Z)\geq 0,\quad \forall\, i\in I\setminus\{\star\}\\ & \operatorname{tr}(B_i Z)\leq 0,\quad \forall\, i\in I\setminus\{\star\}\\ & \operatorname{tr}(D_0 Z)\leq 1\end{array}$$

Here, the condition on which real Hilbert space $\mathcal{H}$ and that the iterates x^i's and y^i's must be defined can be ignored, and this problem indeed can be solved with numerical solvers. The equivalence of the last reformulation comes from Lemma 42.

Lemma 42. *If $\dim\mathcal{H}\geq k+3$, Z is a positive-semidefinite $(k+3)\times(k+3)$ matrix if and only if there exist $x^0-x_\star$, v, and $v^i=x^i-y^i$ for $i=0,1,\dots,k$ in $\mathcal{H}$ such that G is defined as in (1) and $Z=G^\intercal G$.*

C.3. Numerical result of PEP

We numerically solved the SDP formulated in Section 4.2 to obtain a numerical guarantee on the rate of convergence to $\|x^k-\mathbb{T}x^k-v\|^2$ for (Halpern) with $\lambda_k=\frac{1}{k+1}$. We used MOSEK (ApS, 2019) with $k=1,2,\dots,100$. We observe that the numerical solution of PEP to indicate an optimal rate of $\tilde{\mathcal{O}}(1/k^2)$ but not $\mathcal{O}(1/k^2)$.

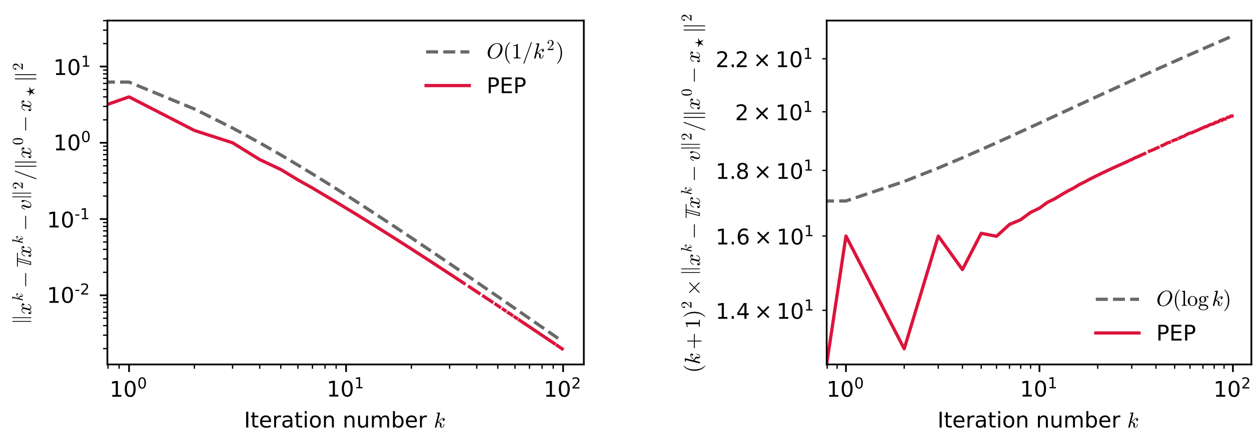

Figure 2. We solved the problem for iteration numbers $k=1$ through $k=100$. (Left) Plot of $\|x^k-\mathbb{T}x^k-v\|^2/\|x^0-x_\star\|^2$ for $k=1,\dots,100$. (Right) Plot of $(k+1)^2\cdot\|x^k-\mathbb{T}x^k-v\|^2/\|x^0-x_\star\|^2$ for $k=1,\dots,100$.

D. Omitted proofs of Section 5

We use following lemma in the proof of Theorem 11 and Theorem 12.

Lemma 43. *Consider any orthogonal matrix $U\colon\mathbb{R}^m\to\mathbb{R}^n$ such that $m\leq n$ and $U^\intercal U=I_m$. For any nonexpansive operator $\mathbb{T}\colon\mathbb{R}^m\to\mathbb{R}^m$ and any $x_0\in\mathbb{R}^n$, define $\mathbb{T}_U\colon\mathbb{R}^n\to\mathbb{R}^n$ as $\mathbb{T}_U(\cdot)=U\mathbb{T}U^\intercal(\cdot-x_0)+x_0$. Then,*

(i) $\|Ux\|=\|x\|$ *for any* $x\in\mathbb{R}^m$ *and* $\|U^\intercal x\|\leq\|x\|$ *for any* $x\in\mathbb{R}^n$.

(ii) $\mathbb{T}_U\colon\mathbb{R}^n\to\mathbb{R}^n$ *is a nonexpansive operator.*

(iii) $U^\intercal\mathcal{R}(\mathbb{I}-\mathbb{T}_U)=U^\intercal U\mathcal{R}(\mathbb{I}-\mathbb{T})=\mathcal{R}(\mathbb{I}-\mathbb{T})$

(iv) $\tilde{v}=\Pi_{\overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}}(0)$ *if and only if* $v=U\tilde{v}=\Pi_{\overline{\mathcal{R}(\mathbb{I}-\mathbb{T}_U)}}(0)$. *If there exists* $x_\star\in\mathcal{H}_1$ *such that* $\tilde{v}=x_\star-\mathbb{T}x_\star$, *then* $y_\star=x_0+Ux_\star$ *implies* $v=y_\star-\mathbb{T}_Uy_\star$. *If there exists* $y_\star\in\mathcal{H}_2$ *such that* $v=y_\star-\mathbb{T}_Uy_\star$, *then* $x_\star=U^\intercal(y_\star-x_0)$ *implies* $x_\star-\mathbb{T}x_\star=\tilde{v}$. *This implies* $\operatorname{Fix}\mathbb{T}=\emptyset$ *if and only if* $\operatorname{Fix}\mathbb{T}_U=\emptyset$ *as well.*

Proof. From orthogonality of U, $U^\intercal U = I_m$ and $UU^\intercal$ is an orthogonal projection onto the range of U.

$$\|Ux\|^2 = \langle Ux, Ux\rangle = \langle x, U^\intercal Ux\rangle = \langle x, x\rangle = \|x\|^2, \qquad \forall\, x \in \mathbb{R}^m$$

and

$$\|U^\intercal x\|^2 = \|UU^\intercal x\|^2 \le \|x\|^2, \qquad \forall\, x \in \mathbb{R}^n.$$

Also,

$$\begin{aligned}\|\mathbb{T}_U y - \mathbb{T}_U z\| &= \|U(\mathbb{T}U^\intercal(y - x_0) - \mathbb{T}U^\intercal(z - x_0))\| \\ &= \|\mathbb{T}U^\intercal(y - x_0) - \mathbb{T}U^\intercal(z - x_0)\| \\ &\le \|U^\intercal(y - x_0) - U^\intercal(z - x_0)\| \\ &\le \|y - z\|, \quad \forall\, y, z \in \mathbb{R}^n\end{aligned}$$

so $\mathbb{T}_U$ is a nonexpansive operator.

Finally,

$$U(x - \mathbb{T}x) = Ux - U\mathbb{T}x = Ux - U\mathbb{T}U^\intercal Ux = (\mathbb{I} - \mathbb{T}_U)(Ux + x_0), \quad \forall\, x \in \mathbb{R}^m,$$

so $U\mathcal{R}(\mathbb{I} - \mathbb{T}) \subseteq \mathcal{R}(\mathbb{I} - \mathbb{T}_U)$, and $U^\intercal U\mathcal{R}(\mathbb{I} - \mathbb{T}) = \mathcal{R}(\mathbb{I} - \mathbb{T}) \subseteq U^\intercal\mathcal{R}(\mathbb{I} - \mathbb{T}_U)$.

$$U^\intercal(y - \mathbb{T}_U y) = U^\intercal y - U^\intercal U\mathbb{T}U^\intercal y = (\mathbb{I} - \mathbb{T})(U^\intercal y + x_0), \quad \forall\, y \in \mathbb{R}^n,$$

so $U^\intercal\mathcal{R}(\mathbb{I} - \mathbb{T}_U) \subseteq \mathcal{R}(\mathbb{I} - \mathbb{T})$ and (iii) holds true.

In order to prove (iv), note that from (i), $\|\tilde{v}\|^2 = \|U\tilde{v}\|^2$. Suppose $\tilde{v} = \Pi_{\overline{\mathcal{R}(\mathbb{I} - \mathbb{T}_U)}}(0)$. Then from (iii), $U^\intercal v \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$. $\tilde{v}$ is the minimum norm element of $\overline{\mathcal{R}(\mathbb{I} - \mathbb{T})}$, so

$$\|U\tilde{v}\| = \|\tilde{v}\| \le \|U^\intercal v\| \le \|v\|.$$

$U\tilde{v} \in \overline{\mathcal{R}(\mathbb{I} - \mathbb{T}_U)}$ and its norm is smaller than or equal to that of v. Therefore, $U\tilde{v}$ must also be the minimum norm element of $\overline{\mathcal{R}(\mathbb{I} - \mathbb{T}_U)}$, which leads to $U\tilde{v} = v$ by the uniqueness of such element.

If $x_\star \in \mathbb{R}^m$ is a point where $\tilde{v} = x_\star - \mathbb{T}x_\star$, then setting $y_\star = x_0 + Ux_\star$ leads to

$$v = U\tilde{v} = Ux_\star - U\mathbb{T}x_\star = (Ux_\star + x_0) - \mathbb{T}_U(Ux_\star + x_0) = y_\star - \mathbb{T}_U y_\star.$$

Now, if $y_\star \in \mathbb{R}^n$ is a point where $v = y_\star - \mathbb{T}_U y_\star$, then setting $x_\star = U^\intercal(y_\star - x_0)$ leads to

$$\begin{aligned}x_\star - \mathbb{T}x_\star &= U^\intercal(y_\star - x_0) - U^\intercal U\mathbb{T}U^\intercal(y_\star - x_0) \\ &= U^\intercal(y_\star - \mathbb{T}_U y_\star) = U^\intercal\tilde{v} = U^\intercal U\tilde{v} = \tilde{v}.\end{aligned}$$

□

D.1. Fixed-Point Iteration with Span Assumption

Proof of Theorem 11. Let $\mathbf{e}_i \in \mathbb{R}^{k+1}$ denote an i-th canonical basis vector whose i-th entry is 1 and the other entries are all zero. As we discussed in the outline of proof, we only consider the case $\tilde{v} = (0, \dots, 0, \|v\|) = \|v\|\mathbf{e}_{k+1}$. Define $\mathbb{T}\colon \mathbb{R}^{k+1} \to \mathbb{R}^{k+1}$ as

$$x - \mathbb{T}x = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 1 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}}_{=M \in \mathbb{R}^{(k+1)\times(k+1)}} x + \alpha\mathbf{e}_1 - \|v\|\mathbf{e}_{k+1}, \quad \forall x \in \mathcal{H}_1$$

for $\alpha \neq 0$. Note that M is an invertible matrix. If $v = 0$, $\mathbb{T}$ has fixed-points of the form

$$-\frac{\alpha}{2}\sum_{i=1}^{k}\mathbf{e}_i + (x_\star)_{k+1}\mathbf{e}_{k+1}, \quad (x_\star)_{k+1} \in \mathbb{R}.$$

If $v \neq 0$, $\mathbb{T}$ does not have a fixed point, and its infimal displacement vector $\tilde{v} = \|v\|\mathbf{e}_{k+1} \neq 0$.

Let the iterates $\{x^n\}_{n=0}^{k}$ satisfy the linear span assumption (span). Since $x^0 = 0$, $x^0 - \mathbb{T}x^0 \in \operatorname{span}\{\mathbf{e}_1, \tilde{v}\}$ and

$$x^1 \in x^0 + \operatorname{span}\{x^0 - \mathbb{T}x^0\} \subseteq \operatorname{span}\{\mathbf{e}_1, \tilde{v}\}.$$

Then $x^1 - \mathbb{T}x^1 \in \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \tilde{v}\}$ and we also have

$$x^2 = x^0 + \operatorname{span}\{x^0 - \mathbb{T}x^0, x^1 - \mathbb{T}x^1\} \in \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \tilde{v}\}.$$

From the observation above, we claim that

$$\begin{aligned} x^n &\in \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n, \tilde{v}\} \\ x^n - \mathbb{T}x^n &\in \alpha\mathbf{e}_1 + \tilde{v} + \operatorname{span}\{M\mathbf{e}_1, M\mathbf{e}_2, \dots, M\mathbf{e}_n\} \subseteq \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n+1}, \tilde{v}\} \end{aligned}$$

for $n = 1, \dots, k-1$.

We have already proven the case of $n = 1$. Let $n < k-1$ and assume that the claim above hold for all m such that $m \leq n$. Then

$$x^{n+1} \in x^0 + \operatorname{span}\{x^0 - \mathbb{T}x^0, \dots, x^n - \mathbb{T}x^n\} \subseteq \operatorname{span}\{\mathbf{e}_1, \dots, \mathbf{e}_{n+1}, \tilde{v}\}.$$

Also,

$$\begin{aligned} x^{n+1} - \mathbb{T}x^{n+1} &= Mx^{n+1} + \alpha\mathbf{e}_1 + \tilde{v} \\ &\in \alpha\mathbf{e}_1 + \tilde{v} + \operatorname{span}\{M\mathbf{e}_1, \dots, M\mathbf{e}_{n+1}\} \\ &\subseteq \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n+2}, \tilde{v}\}. \end{aligned}$$

The claim holds for $n+1$ as well.

From above claim and its proof, we have

$$\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{T}x^i) - \tilde{v} \in \alpha\mathbf{e}_1 + \underbrace{\operatorname{span}\{M\mathbf{e}_1, \dots, M\mathbf{e}_{k-1}\}}_{=:R_{k-1}},$$

so

$$\begin{aligned} \left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{T}x^i) - \tilde{v}\right\|^2 &\geq \|\alpha\mathbf{e}_1\|^2 - \|\Pi_{R_{k-1}}(\alpha\mathbf{e}_1)\|^2 \\ &= \alpha^2\left\|\Pi_{R_{k-1}^\perp}(\mathbf{e}_1)\right\|^2. \end{aligned}$$

Since $R_{k-1}^\perp = \operatorname{span}\left\{\sum_{i=1}^{k}\mathbf{e}_i,\, \tilde{v}\right\}$,

$$\begin{aligned} \left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{T}x^i) - \tilde{v}\right\|^2 &\geq \alpha^2\left\|\Pi_{\operatorname{span}\{\sum_{i=1}^{k}\mathbf{e}_i,\, \tilde{v}\}}(\mathbf{e}_1)\right\|^2 \\ &= \alpha^2\left\|\Pi_{\operatorname{span}\{\sum_{i=1}^{k}\mathbf{e}_i\}}(\mathbf{e}_1)\right\|^2 \\ &= \alpha^2\left\|\left\langle\mathbf{e}_1, \frac{\sum_{i=1}^{k}\mathbf{e}_i}{\|\sum_{i=1}^{k}\mathbf{e}_i\|}\right\rangle\frac{\sum_{i=1}^{k}\mathbf{e}_i}{\|\sum_{i=1}^{k}\mathbf{e}_i\|}\right\|^2 = \frac{\alpha^2}{k} \end{aligned}$$

We know that the set of possible choice of $x_\star \in \mathcal{H}$ is

$$\left\{ x_\star \in \mathbb{R}^{k+1} \;\middle|\; x_\star = -\frac{\alpha}{2}\sum_{i=1}^{k} \mathbf{e}_i + (x_\star)_{k+1}\mathbf{e}_{k+1},\, \forall (x_\star)_{k+1} \in \mathbb{R} \right\}.$$

As $x^0 = 0$, $\|x^0 - x_\star\|^2 \geq \frac{k\alpha^2}{4}$ and equality holds when $(x_\star)_{k+1} = 0$.

Gathering the facts above, we may conclude that there exists $\mathbb{T}\colon \mathbb{R}^{k+1} \to \mathbb{R}^{k+1}$ with infimal displacement vector $\tilde{v}$ and corresponding $x_\star \in \mathbb{R}^{k+1}$ such that $x_\star - \mathbb{T}x_\star = \tilde{v}$ and

$$\left\| \sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i) - \tilde{v} \right\|^2 \geq \frac{4}{k^2}\|x^0 - x_\star\|^2$$

for any iterates $\{x^n\}_{n=0}^{k-1}$ satisfying the linear span assumption, starting from $x^0 = 0$.

Now, we may use the same operator $\mathbb{T}$ to prove that

$$\left(\left\| \sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i) \right\| - \|\tilde{v}\| \right)^2 \geq \frac{1}{2k^2}\|x^0 - x_\star\|^2$$

for any iterates $\{x^n\}_{n=0}^{k}$ satisfying the linear span assumption, starting from $x^0 = 0$.

If $v = 0$, then $\tilde{v} = 0$ so

$$\begin{aligned} \left\| \sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i) \right\| - \|\tilde{v}\| &= \left\| \sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i) - \tilde{v} \right\| \\ &\geq \frac{2}{k}\|x^0 - x_\star\|^2 \\ &\geq \frac{1}{\sqrt{2}k}\|x^0 - x_\star\|^2 \end{aligned}$$

so the desired inequality

$$\left(\left\| \sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i) \right\| - \|\tilde{v}\| \right)^2 \geq \frac{1}{2k^2}\|x^0 - x_\star\|^2$$

holds true.

Now suppose $v \neq 0$. Following the calculations above,

$$\begin{aligned} \left\| \sum_{i=0}^{k} \nu_i (x^i - \mathbb{T}x^i) \right\| - \|\tilde{v}\| &= \sqrt{\|\tilde{v}\|^2 + \|\alpha\mathbf{e}_1\|^2 - \|\Pi_{R_{k-1}}(\alpha\mathbf{e}_1)\|^2} - \|\tilde{v}\| \\ &= \sqrt{\|\tilde{v}\|^2 + \alpha^2 \left\| \Pi_{R_{k-1}^\perp}(\mathbf{e}_1) \right\|^2} - \|\tilde{v}\|. \end{aligned}$$

With a real function $f(t) = \sqrt{t}$ defined on $[0, \infty)$, we may use the fact that f is a concave function, therefore

$$f(t+h) - f(t) \geq hf'(t+h) = \frac{h}{2\sqrt{t+h}}, \quad \forall\, t, h > 0.$$

Substituting t and h by $\|v\|^2 > 0$ and $\alpha^2\|\Pi_{R_{k-1}^\perp}(\mathbf{e}_1)\|^2 > 0$,

$$\begin{aligned} \left\| \sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i) \right\| - \|\tilde{v}\| &\geq \frac{\alpha^2\|\Pi_{R_{k-1}^\perp}(\mathbf{e}_1)\|^2}{2\sqrt{\|\tilde{v}\|^2 + \alpha^2\|\Pi_{R_{k-1}^\perp}(\mathbf{e}_1)\|^2}} \\ &= \frac{\alpha\|\Pi_{R_{k-1}^\perp}(\mathbf{e}_1)\|}{2\sqrt{1 + \frac{\|\tilde{v}\|^2}{\alpha^2\|\Pi_{R_{k-1}^\perp}(\mathbf{e}_1)\|^2}}}. \end{aligned}$$

$\alpha > 0$ is a positive real number which was left unspecified, so we may just assign any positive value for the calculation. Let $\alpha = \frac{\|\tilde{v}\|}{\|\Pi_{R_{k-1}^{\perp}}(\mathbf{e}_1)\|} = \sqrt{k}\|\tilde{v}\| > 0$, then as $\|x^0 - x_\star\| = \frac{\alpha}{2}\sqrt{k} = \frac{k\|\tilde{v}\|}{2}$, we get

$$\left(\left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{T}x^i)\right\| - \|\tilde{v}\|\right)/\|x^0 - x_\star\| \geq \frac{\|\tilde{v}\|}{2\sqrt{2}}\frac{2}{k\|\tilde{v}\|} = \frac{1}{\sqrt{2}k}.$$

so

$$\left(\left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{T}x^i)\right\| - \|\tilde{v}\|\right)^2 \geq \frac{1}{2k^2}\|x^0 - x_\star\|^2.$$

We now extend the result to the arbitrarily given v, not $\tilde{v}$. The case $v = 0$ is trivial, so suppose $v \neq 0$. Let $U \in \mathbb{R}^{(k+1)\times(k+1)}$ be an orthogonal matrix such that $U^\intercal U = I_{k+1}$ and $U\tilde{v} = v$. Since U is a square matrix, $U^\intercal U = UU^\intercal = I_{k+1}$ so $\|U^\intercal x\| = \|x\|$ for all $x \in \mathbb{R}^{k+1}$. According to Lemma 43, $\mathbb{T}_U$ defined as $\mathbb{T}_U(\cdot) = U\mathbb{T}U^\intercal(\cdot - y^0) + y^0$ is a nonexpansive operator with $v = U\tilde{v} = \Pi_{\overline{\mathcal{R}(\mathbb{I}-\mathbb{T}_U)}}(0)$. Let $\{y^n\}_{n=0}^{k}$ is a sequence of iterates satisfying the linear span assumption (span) with $\mathbb{T}_U$.

$$\begin{aligned} U^\intercal(y^n - \mathbb{T}_U y^n) &= U^\intercal(y^n - y^0) - U^\intercal U\mathbb{T}U^\intercal(y^n - y^0) \\ &= U^\intercal(y^n - y^0) - \mathbb{T}U^\intercal(y^n - y^0) \end{aligned}$$

and as

$$y^{n+1} \in y^0 + \operatorname{span}\{y^0 - \mathbb{T}_U y^0, \dots, y^n - \mathbb{T}_U y^n\}$$

implies

$$\begin{aligned} U^\intercal(y^{n+1} - y^0) &\in \operatorname{span}\{U^\intercal y^0 - U^\intercal\mathbb{T}_U y^0, \dots, U^\intercal y^n - U^\intercal\mathbb{T}_U y^n\} \\ &= \operatorname{span}\{U^\intercal(y^0 - y^0) - \mathbb{T}U^\intercal(y^0 - y^0), \dots, U^\intercal(y^n - y^0) - \mathbb{T}U^\intercal(y^n - y^0)\}, \end{aligned}$$

$\left\{U^\intercal(y^n - y^0)\right\}_{n=0}^{k-1}$ satisfies linear span assumption (span) with $\mathbb{T}$ where $\tilde{v} = U^\intercal v$ is an infimal displacement vector of $\mathbb{T}$. Therefore,

$$\begin{aligned} \left\|\sum_{i=0}^{k-1}\nu_i(y^i - \mathbb{T}_U y^i) - v\right\| &\geq \left\|U^\intercal\sum_{i=0}^{k-1}\nu_i(y^i - \mathbb{T}_U y^i) - U^\intercal v\right\| \\ &= \left\|\sum_{i=0}^{k-1}\nu_i\left\{U^\intercal(y^n - y^0) - \mathbb{T}U^\intercal(y^n - y^0)\right\} - \tilde{v}\right\| && (\because U^\intercal v = U^\intercal U\tilde{v} = \tilde{v}) \\ &\geq \frac{4}{k^2}\|U^\intercal(y^0 - y^0) - U^\intercal(y_\star - y^0)\|^2 \\ &= \frac{4}{k^2}\|y^0 - y_\star\|^2. \end{aligned}$$

Similarly,

$$\begin{aligned} \left\|\sum_{i=0}^{k-1}\nu_i(y^i - \mathbb{T}_U y^i)\right\| - \|v\| &\geq \left\|U^\intercal\sum_{i=0}^{k-1}\nu_i(y^i - \mathbb{T}_U y^i)\right\| - \|v\| \\ &= \left\|\sum_{i=0}^{k-1}\nu_i\left\{U^\intercal(y^n - y^0) - \mathbb{T}U^\intercal(y^n - y^0)\right\}\right\| - \|\tilde{v}\| && (\because v = U\tilde{v}) \\ &\geq \frac{1}{\sqrt{2}k}\|U^\intercal(y^0 - y^0) - U^\intercal(y_\star - y^0)\| \\ &= \frac{1}{\sqrt{2}k}\|y^0 - y_\star\|. \end{aligned}$$

Proof with real Hilbert space $\mathcal{H}$ can be done in the same manner as the proof of $\mathcal{H} = \mathbb{R}^{k+1}$. Set $\{\mathbf{e}_i\}_{i=1}^{k+1}$ as a set of orthonormal basis vectors of $\mathcal{H}$ where $\mathbf{e}_{k+1} = \frac{v}{\|v\|}$ if $v \neq 0$ and arbitrary if $v = 0$, and proceed with the same proof as above. □

Remark 44. We calculate the complexity lower bound of various quantities including fixed-point residual $x^k - \mathbb{T}x^k$ and normalized iterate $(x^k - x^0)/\alpha_k$ by appropriately choosing the convex coefficients ν_i's. In order to measure the lower bound of fixed-point residual $x^{k-1} - \mathbb{T}x^{k-1}$ converging to v, choose $\nu_{k-1} = 1$ and choose all other ν_i's as 0. For normalized iterate, we use the fact that KM and Halpern choose the iterate x^k to be in form of $x^0 + \sum_{i=0}^{k-1} \lambda_i^k (x^i - \mathbb{T}x^i)$. Therefore, if we choose $\nu_i = \frac{\lambda_i^k}{\sum_{i=0}^{k-1} \lambda_i^k}$, we obtain normalized iterates $\frac{x^k - x^0}{\sum_{i=1}^{k}(1-\lambda_i)}$ for KM and $\frac{x^k - x^0}{\theta_k}$ for Halpern. This scheme can be extended to calculate the lower bound of Mann iteration as well.

D.2. General Fixed-Point Iterations

We follow the general complexity lower bound result of Park & Ryu (2022) for operators *with* fixed points, and extend their result to the case where fixed point might not exist.

Definition 45 (Section D.2,& D.4, Park & Ryu (2022))**.** Let $\mathcal{H}$ be a real Hilbert space and $\mathbb{T} \colon \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator. Let $\{\mathbf{e}_i\}_{i \in I}$ with index set I denote a set of orthonormal basis of $\mathcal{H}$.

A *deterministic fixed-point iteration* $\mathbb{F}$ is defined as a mapping of the point $x^0 \in \mathcal{H}$ and an operator $\mathbb{T}$ to sequences of iterates $\{x^k\}_{k\in\mathbb{N}}$ and $\{\bar{x}^k\}_{k\in\mathbb{N}}$. Here, x^k is the k-th *query point* and $\bar{x}^k$ is the k-th *approximation point*, and we consider the setup with $x^k = \bar{x}^k$ so we omit $\bar{x}^k$. Actually, $\mathbb{F}$ is defined as a sequence of mappings $\{\mathbb{F}_k\}_{k\in\mathbb{N}}$, where x^k is an output of $\mathbb{F}_k$ given the point x^0 and the operator $\mathbb{T}$ where

$$x^k = \mathbb{F}_k[x^0, \mathbb{T}] = \mathbb{F}_k[x^0, \mathcal{O}_{\mathbb{T}}(x^0), \mathcal{O}_{\mathbb{T}}(x^1), \dots, \mathcal{O}_{\mathbb{T}}(x^{k-1})]$$

for any $k \in \mathbb{N}$. Here, $\{x^k\}_{k\in\mathbb{N}}$ only depends on x^0 and $\mathbb{T}$ via the *fixed-point residual oracle* $\mathcal{O}_{\mathbb{T}}(x) = x - \mathbb{T}x$. $\mathbb{F}$ is deterministic in a sense that, when provided with the same point x^0 and the same oracle queries $\mathcal{O}_{\mathbb{T}}(x^k) = x^k - \mathbb{T}x^k$ for $k \in \mathbb{N}$, $\mathbb{F}$ will give the same sequence of iterates $\{x^k\}$ as an output of $\mathbb{F}$.

For $z \in \mathcal{H}$, denote by $\operatorname{supp}\{z\}$ the *support* of z, i.e.,

$$\operatorname{supp}\{z\} = \{i \in I \mid \langle z, \mathbf{e}_i \rangle \neq 0\}.$$

We say a sequence $\{z^t\}_{t\in\mathbb{N}\cup\{0\}}$ is *zero-respecting with respect to* $\mathbb{T}$ if

$$\operatorname{supp}\{z^t\} \subseteq \bigcup_{s<t} \operatorname{supp}\{z^s - \mathbb{T}z^s\}$$

If $\{z^t\}_{t\in\mathbb{N}\cup\{0\}}$ is a zero-respecting sequence with respect to $\mathbb{T}$, then by the definition, $\operatorname{supp}\{z^0\} \subseteq \emptyset$ so $z^0 = 0$.

As we did for fixed-point iterations with linear span assumption (span), from now on, we prove the result for Euclidean spaces, since the proof naturally extends to Hilbert spaces $\mathcal{H}$ with $\dim \mathcal{H} \geq 2k-1$ and its set of orthonormal vectors $\{\mathbf{e}_i\}_{i=1}^{2k-1}$.

Lemma 46. *Given $k \in \mathbb{N}$ and $v \in \mathbb{R}^{k+1}$, let $\mathbb{T} \colon \mathbb{R}^{k+1} \to \mathbb{R}^{k+1}$ be the worst-case operator defined in the proof of Theorem 11, along with its infimal displacement vector v and $x_\star$. If the iterates $\{x^n\}_{n=0}^{k-1}$ are zero-respecting with respect to $\mathbb{T}$,*

$$\left\| \sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i) - v \right\| \geq \frac{2}{k+1} \|x^0 - x_\star\|$$

and

$$\left\| \sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i) \right\| - \|v\| \geq \frac{1}{\sqrt{2}(k+1)} \|x^0 - x_\star\|$$

for any choice of real numbers $\{\nu_i\}_{i=0}^{k-1}$ such that $\sum_{i=0}^{k-1} \nu_i = 1$.

Proof. We claim that any zero respecting sequence $\{x^n\}_{n=0}^{k-1}$ satisfies

$$x^n \in \operatorname{span}\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$$
$$x^n - \mathbb{T}x^n \in \operatorname{span}\{\mathbf{e}_1, \dots, \mathbf{e}_{n+1}\}$$

for $n = 0, \dots, k$. If this holds, then the proof of Theorem 11 is still applicable to zero-respecting sequences, leading to the desired result.

If $n = 0$, $x^0 = 0$ and $x^0 - \mathbb{T}x^0 \in \operatorname{span}\{\mathbf{e}_1\}$, so the case of $n = 0$ holds true. Now, suppose that $n < k-1$ and the claim above holds for all m such that $m \le n$. Since the iterates form a zero-respecting sequence with respect to $\mathbb{T}$, $\operatorname{supp}\{x^{n+1}\} \subseteq \cup_{m\le n}\operatorname{supp}\{x^m - \mathbb{T}x^m\}$ and therefore $x^{n+1} \in \operatorname{span}\{\mathbf{e}_1, \dots, \mathbf{e}_{n+1}\}$. Using this fact, $x^{n+1} - \mathbb{T}x^{n+1} \in \operatorname{span}\{\mathbf{e}_1, \dots, \mathbf{e}_{n+2}\}$ easily follows. □

Lemma 47. *Let $\mathbb{F}$ be a general deterministic fixed-point iteration and $\mathbb{T}\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be a nonexpansive operator defined as in the proof of Theorem 11. For any arbitrary $x^0 \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$ with $d \ge n+N$, there exists an orthogonal matrix $U \in \mathbb{R}^{d\times(n+1)}$ and the iterates $\{x^k\}_{k=1}^N = \mathbb{F}[x^0, \mathbb{T}_U]$ such that $x^{(k)} = U^\intercal(x^k - x^0)$, $\left\{x^{(k)}\right\}_{k=0}^N$ is zero-respecting with respect to $\mathbb{T}$, and v becomes an infimal displacement vector of $\mathbb{T}_U$.*

Proof. We prove the existence of an orthogonal matrix $U \in \mathbb{R}^{d\times n}$ such that $\left\{x^k\right\}_{k=1}^N = \mathbb{F}[x^0,\, \mathbb{T}_U]$, $x^{(k)} = U^\intercal(x^k - x^0)$ and v becomes an infimal displacement vector of $\mathbb{T}_U$. Constructing such orthogonal matrix is equivalent to choosing appropriate set of orthonormal vectors $\{u_i\}_{i=1}^{n+1}$, whose i-th vector u_i becomes an i-th column of matrix U, i.e.,

$$U = \begin{bmatrix} | & \dots & | \\ u_1 & \dots & u_n \\ | & \dots & | \end{bmatrix}.$$

We modify the proof of (Park & Ryu, 2022, Lemma D.4(i)) to cover the case when $v \neq 0$, as the original proof is restricted to the case where the fixed point exists, or in other words, the case where $v = 0$.

We provide the scheme which inductively finds the column u_i's, given an arbitrary nonzero vector $v \in \mathbb{R}^d$. Define the set of indices S_t for $t \in \{1, \dots, N\}$ as

$$S_t = \bigcup_{s<t} \operatorname{supp}\left\{x^{(s)} - \mathbb{T}x^{(s)}\right\} \subseteq \{1, 2, \dots, n+1\}$$

and note that $S_0 = \emptyset \subseteq S_1 \subseteq \dots \subseteq S_t$. As $v \neq 0$, $x^{(0)} - \mathbb{T}x^{(0)} \neq 0$, and $S_1 \neq \emptyset$. Choose a set of vectors $\{u_i\}_{i\in S_1}$ to be any unit vectors which are orthonormal to each other. The precise choice of $\{u_i\}_{i\in S_1}$ will be later specified, and it will make v to be an infimal displacement vector of $\mathbb{T}_U$.

Now, suppose that for $t \ge 2$, $\{u_i\}_{i\in S_{t-1}}$ is already chosen. Choose a set of unit vectors $\{u_i\}_{i\in S_t\setminus S_{t-1}}$ from the orthogonal complement of

$$W_t := \operatorname{span}\left(\{x^1 - x^0,\, \cdots,\, x^{t-1} - x^0\} \cup \{u_i\}_{i\in S_{t-1}}\right) \subseteq \mathbb{R}^d$$

and let them be orthogonal to each other. When $t = N$ and $S_N \neq \{1, \dots, n+1\}$, choose any $\{u_i\}_{i\in\{1,\dots,n+1\}\setminus S_N}$ which makes U orthogonal. Above scheme is well-defined if

$$\dim W_t^\perp \ge |S_t \setminus S_{t-1}|,$$

and this is guaranteed from the fact that $d \ge n+N$ and $t \le N$ since

$$\dim W_t^\perp = d - \dim W_t \ge d - \{(t-1) + |S_{t-1}|\} \ge (n+1) - |S_{t-1}| \ge |S_t \setminus S_{t-1}|.$$

Since $\langle u_i,\, y_t - y_0\rangle$ for $i \notin S_t$ for $t = 1, \dots, N$,

$$x^{(t)} = U^\intercal(x^t - x^0) \in \operatorname{span}\{e_i\}_{i\in S_t}$$

leads to $\operatorname{supp}\{x^{(t)}\} \subseteq S_t$. This proves that there exists an orthogonal matrix $U \in \mathbb{R}^{d\times(n+1)}$ such that $\{x^k\}_{k=1}^N = \mathbb{F}[x^0, \mathbb{T}_U]$ and $x^{(k)} = U^\intercal(x^k - x^0)$ for all $k = 1, \dots, N$.

Now, it remains to show that certain choice of $\{u_i\}_{i\in S_1}$ implies that $\mathbb{T}_U$ has v as its infimal displacement vector. First, observe that for any arbitrary choice of $\{u_i\}_{i\in S_1}$,

$$S_1 = \operatorname{supp}\left\{x^{(0)} - \mathbb{T}x^{(0)}\right\}$$

and

$$x^{(0)} - \mathbb{T}x^{(0)} = 0 - \mathbb{T}0 = -\alpha \mathbf{e}_1 + \|v\|\mathbf{e}_n$$

so $S_1 = \{1, n\}$. Note that the infimal displacement vector of $\mathbb{T}$ is $\tilde{v} = \|v\|\mathbf{e}_{n+1}$. From Lemma 43, $U\tilde{v} = \|v\| \cdot U\mathbf{e}_{n+1} = \|v\|u_n$ is an infimal displacement vector of $\mathbb{T}_U$. As $n+1 \in S_1$, we may choose $u_{n+1} = \frac{v}{\|v\|}$ so that v is an infimal displacement vector of $\mathbb{T}_U$. □

Lemma 48. *Consider the setup of Lemma 47 with $U \in \mathbb{R}^{m\times n}$ and the iterates $\{x^k\}_{k=1}^N = \mathbb{F}[x^0, \mathbb{T}_U]$. Then*

$$\left\|\sum_{i=0}^{k} \nu_i (x^{(i)} - \mathbb{T}x^{(i)}) - \tilde{v}\right\| \leq \left\|\sum_{i=0}^{k} \nu_i (x^i - \mathbb{T}_U x^i) - v\right\|$$

and

$$\left\|\sum_{i=0}^{k} \nu_i (x^{(i)} - \mathbb{T}x^{(i)})\right\| - \|\tilde{v}\| \leq \left\|\sum_{i=0}^{k} \nu_i (x^i - \mathbb{T}_U x^i)\right\| - \|v\|$$

for any $\nu_i \in \mathbb{R}$ such that $\sum_{i=0}^{k} \nu_i = 1$, where v is an infimal displacement vector of $\mathbb{T}_U$ and $\tilde{v}$ is an infimal displacement vector of $\mathbb{T}$.

Proof. According to Lemma 43, $v = U\tilde{v}$.

$$\begin{aligned}
\left\|\sum_{i=0}^{k} \nu_i (x^{(i)} - \mathbb{T}x^{(i)}) - \tilde{v}\right\| &= \left\|\sum_{i=0}^{k} \nu_i \left(U^\intercal (x^i - x^0) - \mathbb{T}U^\intercal (x^i - x^0)\right) - U^\intercal v\right\| \\
&= \left\|U^\intercal \sum_{i=0}^{k} \nu_i \left((x^i - x^0) - U\mathbb{T}U^\intercal (x^i - x^0)\right) - U^\intercal v\right\| \\
&= \left\|U^\intercal \sum_{i=0}^{k} \nu_i (x^i - \mathbb{T}_U x^i) - U^\intercal v\right\| \\
&\leq \left\|\sum_{i=0}^{k} \nu_i (x^i - \mathbb{T}_U x^i) - v\right\|.
\end{aligned}$$

Note that $\|\tilde{v}\| = \|Uv\| = \|v\|$. Therefore,

$$\begin{aligned}
\left\|\sum_{i=0}^{k} \nu_i (x^{(i)} - \mathbb{T}x^{(i)})\right\| - \|\tilde{v}\| &= \left\|U^\intercal \sum_{i=0}^{k} \nu_i (x^i - \mathbb{T}_U x^i)\right\| - \|v\| \\
&\leq \left\|\sum_{i=0}^{k} \nu_i (x^i - \mathbb{T}_U x^i)\right\| - \|v\|.
\end{aligned}$$

□

We now prove the main result.

Theorem 49. *Let $d \geq 2k-1$ for $k \in \mathbb{N}$. For any deterministic fixed-point iteration $\mathbb{F}$, any initial point $x^0 \in \mathbb{R}^d$ and any vector $v \in \mathbb{R}^d$, there exists a nonexpansive operator $\mathbb{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ such that*

$$\left\|\sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i) - v\right\|^2 \geq \frac{4}{k^2}\|x^0 - x_\star\|^2$$

and

$$\left(\left\|\sum_{i=0}^{k-1} \nu_i (x^i - \mathbb{T}x^i)\right\| - \|v\|\right)^2 \geq \frac{1}{2k^2}\|x^0 - x_\star\|^2$$

where $v = x_\star - \mathbb{T}x_\star = \Pi_{\overline{\mathcal{R}(\mathbb{I}-\mathbb{T})}}(0)$ and $\nu_i \in \mathbb{R}$ with $\sum_{i=0}^{k-1} \nu_i = 1$.

Proof. Let $\mathbb{S}\colon \mathbb{R}^{k+1} \to \mathbb{R}^{k+1}$ be a worst-case operator in the proof of Theorem 11. From Lemma 47, there exists an orthogonal matrix $U \in \mathbb{R}^{d\times(k+1)}$ such that $d \geq k + (k-1) = 2k-1$, $\{x^{(t)}\}_{t=0}^{k-1}$ a sequence of iterates defined as $x^{(t)} = U^\intercal(x^t - x^0)$ is zero-respecting with respect to $\mathbb{S}$,

$$\left\|\sum_{i=0}^{k-1}\nu_i(x^{(i)} - \mathbb{S}x^{(i)}) - \tilde{v}\right\| \leq \left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{S}_U x^i) - v\right\|,$$

$$\left\|\sum_{i=0}^{k-1}\nu_i(x^{(i)} - \mathbb{S}x^{(i)})\right\| - \|\tilde{v}\| \leq \left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{S}_U x^i)\right\| - \|v\|$$

for any $\nu_i \in \mathbb{R}$ such that $\sum_{i=0}^{k-1}\nu_i = 1$, and $v = U\tilde{v}$ or $\tilde{v} = U^\intercal v$ by Lemma 43.

Since $\{x^{(t)}\}_{t=0}^{k-1}$ is zero-respecting with respect to $\mathbb{T}$, Lemma 46 implies

$$\left\|\sum_{i=0}^{k-1}\nu_i(x^{(i)} - \mathbb{S}x^{(i)}) - \tilde{v}\right\|^2 \geq \frac{4}{k^2}\|x^{(0)} - x^{(\star)}\|^2$$

and

$$\left(\left\|\sum_{i=0}^{k-1}\nu_i(x^{(i)} - \mathbb{S}x^{(i)})\right\| - \|\tilde{v}\|\right)^2 \geq \frac{1}{2k^2}\|x^{(0)} - x^{(\star)}\|^2$$

where $x^{(\star)} \in \mathcal{H}_0$ is a point such that $x^{(\star)} - \mathbb{S}x^{(\star)} = v$. If $x_\star = x^0 + Ux^{(\star)}$,

$$\|x^{(0)} - x^{(\star)}\|^2 = \|-x^{(\star)}\|^2 = \|-Ux^{(\star)}\|^2 = \|x^0 - x_\star\|^2$$

so we may conclude that

$$\left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{S}_U x^i) - v\right\|^2 \geq \frac{4}{k^2}\|x^0 - x_\star\|^2$$

and

$$\left(\left\|\sum_{i=0}^{k-1}\nu_i(x^i - \mathbb{S}_U x^i)\right\| - \|v\|\right)^2 \geq \frac{1}{2k^2}\|x^0 - x_\star\|^2.$$

Therefore, $\mathbb{T} = \mathbb{S}_U$ is our desired worst-case operator. □

Proof of Theorem 12. Use the worst-case nonexpansive operator of Theorem 49 and construct the nonexpansive operator with the orthonormal basis set $\{\mathbf{e}_i\}_{i=1}^{2k-1}$ of $\mathcal{H}$ with $\dim\mathcal{H} = 2k-1$. □

E. Details of experiment in Section 6

Consider a semidefinite problem (SDP)

$$\begin{array}{ll} \underset{x\in\mathbb{R}^d}{\text{minimize}} & \sum_{i=1}^p c_i^\intercal x \\ \text{subject to} & \mathcal{A}_i[x] = \sum_{j=1}^d A_i^j x_j \preceq B_i, \quad 1 \leq i \leq p, \end{array}$$

and PG-EXTRA applied on this problem.

$$\begin{aligned} U_i^{k+1} &= \Pi_{-\mathbb{S}_+^n}\left(U_i^k + \beta(B_i - \mathcal{A}_i[x^k])\right) \\ \mathbf{w}^{k+1} &= \mathbf{w}^k + \frac{1}{2}(I - W)\mathbf{x}^k \\ x_i^{k+1} &= x_i^k - \alpha\beta(2w_i^{k+1} - w_i^k) + \alpha\left(\mathcal{A}_i^*[2U_i^{k+1} - U_i^k] - c_i\right) \end{aligned} \tag{PG-EXTRA}$$

E.1. Deriving PG-EXTRA for SDP

Consider

$$\begin{array}{ll} \underset{x \in \mathbb{R}^m}{\text{minimize}} & \sum_{i=1}^{p} \langle c_i, x \rangle_{\mathbb{R}^m} \\ \text{subject to} & \sum_{j=1}^{m} x_j A_i^j \preceq_{S_+^n} B_i, \quad i = 1, \cdots, p \end{array}$$

or in other words,

$$\begin{array}{ll} \underset{x_i \in \mathbb{R}^n}{\text{minimize}} & \sum_{i=1}^{p} \langle c_i, x_i \rangle_{\mathbb{R}^m} \\ \text{subject to} & L_i(x_i) := B_i - \sum_{j=1}^{m} (x_i)_j A_i^j \succeq_{S_+^n} 0, \quad i = 1, \cdots, p \ \cdot \\ & (I - W)\mathbf{x} = 0 \ \Leftrightarrow \ U\mathbf{x} = 0 \end{array}$$

Defining $L\,\mathbf{x} = (L_1(x_1) - B_1, \cdots, L_p(x_p) - B_p, U\mathbf{x})$, which is a linear map from $\mathbb{R}^{p \times m}$ to $(S^n)^p \times \mathbb{R}^{p \times m}$,

$$\underset{x_i \in \mathbb{R}^n}{\text{minimize}} \quad \sum_{i=1}^{p} \langle c_i, x_i \rangle_{\mathbb{R}^m} + \delta_{(S_+^n)^p \times \{0\}}(L\,\mathbf{x} + \mathbf{B})$$

Corresponding Lagrangian is

$$\mathbf{L}(\mathbf{x}, \mathbf{u}) = \underbrace{\sum_{i=1}^{p} \langle c_i,\, x_i \rangle_{\mathbb{R}^m}}_{:= \langle \mathbf{c}, \mathbf{x} \rangle_{\mathbb{R}^{p \times m}}} + \langle \mathbf{u}, L\,\mathbf{x} + \mathbf{B} \rangle_{(S^n)^p \times \mathbb{R}^{p \times m}} - \left(\delta_{(S_+^n)^p \times \{0\}} \right)^* (\mathbf{u})$$

where $\mathbf{u} = (u_1, \cdots, u_p, \mathbf{y}) \in (S^n)^p \times \mathbb{R}^{p \times m}$ and its saddle subdifferential is

$$\partial \mathbf{L}(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} \mathbf{c} + L^* \mathbf{u} \\ -L\,\mathbf{x} - \mathbf{B} + \partial \left(\delta_{(S_+^n)^p \times \{0\}} \right)^* (\mathbf{u}) \end{bmatrix}$$

where

$$\mathbf{c} = \begin{bmatrix} c_1^\intercal \\ \vdots \\ c_p^\intercal \end{bmatrix} \in \mathbb{R}^{p \times m}.$$

Note that from

$$L\,\mathbf{x} = \left(-\sum_{j=1}^{m} (x_1)_j A_1^j, \, \cdots, \, -\sum_{j=1}^{m} (x_p)_j A_p^j, \, U\mathbf{x} \right),$$

$L^* \colon (S^n)^p \times \mathbb{R}^{p \times m} \to \mathbb{R}^{p \times m}$ is

$$L^*(y_1, \cdots, y_p, \mathbf{z}) = U^\intercal \mathbf{z} - \begin{bmatrix} (A_1^* y_1)^\intercal \\ \vdots \\ (A_p^* y_p)^\intercal \end{bmatrix}.$$

Then

$$\begin{aligned} \partial \mathbf{L}(\mathbf{x}, \mathbf{u}) &= \begin{bmatrix} \mathbf{c} - U\mathbf{y} + \begin{bmatrix} (A_1^* u_1)^\intercal \\ \vdots \\ (A_p^* u_p)^\intercal \end{bmatrix} \\ -L\,\mathbf{x} - \mathbf{B} + \partial \left(\delta_{(S_+^n)^p \times \{0\}} \right)^* (\mathbf{u}) \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \mathbf{c} \\ -\mathbf{B} \end{bmatrix}}_{=:\mathbb{H}(\mathbf{x}, \mathbf{u})} + \underbrace{\begin{bmatrix} 0 & L^* \\ -L & \partial \left(\delta_{(S_+^n)^p \times \{0\}} \right)^* \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix}}_{=:\mathbb{F}(\mathbf{x}, \mathbf{u})}. \end{aligned}$$

Note that

$$\delta_{\{0\}}^*(x) = \sup_y \left(\langle x, y \rangle - \delta_{\{0\}}(y) \right) = 0.$$

and

$$\delta^*_{S^n_+}(X) = \sup_Y \Big(\langle X, Y\rangle - \delta_{S^n_+}(Y) \Big) = \sup_{Y\in S^n_+} \langle X, Y\rangle = \begin{cases} 0 & -X \in S^n_+ \\ \infty & \text{o.w.} \end{cases} = \delta_{-S^n_+}(X).$$

Let

$$M = \begin{bmatrix} (1/\alpha)\mathbb{I} & L^* \\ L & (1/\beta)\mathbb{I} \end{bmatrix},$$

then the FPI of forward-backward splitting $(\mathbf{x}^{k+1},\, \mathbf{u}^{k+1}) = (M+\mathbb{F})^{-1}(M-\mathbb{H})(\mathbf{x}^k,\, \mathbf{u}^k)$ is

$$\begin{aligned}
&\begin{bmatrix} (1/\alpha)\mathbb{I} & 2L^* \\ 0 & (1/\beta)\mathbb{I} + \partial\left(\delta_{(S^n_+)^p\times\{0\}}\right)^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^{k+1} \\ \mathbf{u}^{k+1} \end{bmatrix} \ni \begin{bmatrix} (1/\alpha)\mathbf{x}^k + L^*\mathbf{u}^k - \mathbf{c} \\ L\mathbf{x}^k + (1/\beta)\mathbf{u}^k + \mathbf{B} \end{bmatrix} \\
&\Leftrightarrow \mathbf{x}^{k+1} + 2\alpha L^*\mathbf{u}^{k+1} = \mathbf{x}^k + \alpha(L^*\mathbf{u}^k - \mathbf{c}) \\
&\qquad \mathbf{u}^{k+1} = \mathrm{Prox}_{\beta(\delta_{(S^n_+)^p\times\{0\}})^*}\left(\mathbf{u}^k + \beta(L\mathbf{x}^k + \mathbf{B})\right) \\
&\Leftrightarrow \mathbf{u}^{k+1} = \mathrm{Prox}_{\beta(\delta_{(S^n_+)^p\times\{0\}})^*}\left(\mathbf{u}^k + \beta(L\mathbf{x}^k + \mathbf{B})\right) \\
&\qquad \mathbf{x}^{k+1} = \mathbf{x}^k - \alpha(2L^*\mathbf{u}^{k+1} - L^*\mathbf{u}^k + \mathbf{c}) \\
&\Leftrightarrow u_i^{k+1} = \Pi_{-S^n_+}\left(u_i^k + \beta(B_i - \textstyle\sum_{j=1}^n (x_i^k)_j A_i^j)\right) \\
&\qquad \mathbf{y}^{k+1} = \mathbf{y}^k + \beta U\mathbf{x}^k \\
&\qquad \mathbf{x}^{k+1} = \mathbf{x}^k - \alpha(2L^*\mathbf{u}^{k+1} - L^*\mathbf{u}^k + \mathbf{c})
\end{aligned}$$

Note that y_i and $\mathbf{y}$ are not related to each other. If we let $\mathbf{w}^0 = 0$, $\mathbf{x}^0$ arbitrary, and

$$\mathbf{w}^k = \frac{1}{\beta}U\mathbf{y}^k = \frac{1}{2}(I-W)\sum_{j=0}^{k-1}\mathbf{x}^j,$$

then

$$\begin{aligned}
u_i^{k+1} &= \Pi_{-S^n_+}\left(u_i^k + \beta(B_i - \textstyle\sum_{j=1}^n (x_i^k)_j A_i^j)\right) \\
\mathbf{x}^{k+1} &= \mathbf{x}^k - \alpha\beta(2\mathbf{w}^{k+1} - \mathbf{w}^k) + \alpha\left(2\begin{bmatrix} (A_1^* u_1^{k+1})^\intercal \\ \vdots \\ (A_p^* u_p^{k+1})^\intercal \end{bmatrix} - \begin{bmatrix} (A_1^* u_1^k)^\intercal \\ \vdots \\ (A_p^* u_p^k)^\intercal \end{bmatrix} - \begin{bmatrix} c_1^\intercal \\ \vdots \\ c_p^\intercal \end{bmatrix}\right) \\
\mathbf{w}^{k+1} &= \mathbf{w}^k + \frac{1}{2}(I-W)\mathbf{x}^k
\end{aligned}$$

or

$$\begin{aligned}
u_i^{k+1} &= \Pi_{-S^n_+}\left(u_i^k + \beta(B_i - \textstyle\sum_{j=1}^n (x_i^k)_j A_i^j)\right) \\
\mathbf{w}^{k+1} &= \mathbf{w}^k + \frac{1}{2}(I-W)\mathbf{x}^k \\
x_i^{k+1} &= x_i^k - \alpha\beta(2w_i^{k+1} - w_i^k) + \alpha\left(A_i^*(2u_i^{k+1} - u_i^k) - c_i\right)
\end{aligned}$$

where U and u_i are irrelevant and

$$A^*u = \begin{bmatrix} \mathrm{tr}(A_1 u_1) \\ \vdots \\ \mathrm{tr}(A_n u_n) \end{bmatrix}.$$

Above solves decentralized semidefinite problem, when $\alpha, \beta > 0$ are chosen to define a metric on $\mathbb{R}^{n\times p} \times (S^n)^p \times \mathbb{R}^{n\times p}$.

$$M = \begin{bmatrix} (1/\alpha)\mathbb{I} & L^* \\ L & (1/\beta)\mathbb{I} \end{bmatrix}$$

E.2. Measuring fixed-point residual in M-norm

Although the algorithm itself does not keep the iterate $\mathbf{y}^k$ such that

$$\mathbf{w}^k = \frac{1}{\beta} U \mathbf{y}^k,$$

we need $\mathbf{y}^k$-iterates in order to calculate the fixed-point residual $\|(\mathbf{x}^k, \mathbf{u}^k)\|_M^2$ where $M\colon \mathbb{R}^{p\times m} \times ((S^n)^p \times \mathbb{R}^{p\times m}) \to \mathbb{R}^{p\times m} \times ((S^n)^p \times \mathbb{R}^{p\times m})$ is a linear map defined as

$$M = \begin{bmatrix} (1/\alpha)\mathbb{I} & L^* \\ L & (1/\beta)\mathbb{I} \end{bmatrix}.$$

Then for any $\mathbf{x} \in \mathbb{R}^{p\times m}$ and $\mathbf{u} = (u_1, \cdots, u_p, \mathbf{y}) \in (S^n)^p \times \mathbb{R}^{p\times m}$,

$$\begin{aligned} \|(\mathbf{x}, \mathbf{u})\|_M^2 &= \frac{1}{\alpha}\|\mathbf{x}\|^2_{\mathbb{R}^{p\times m}} + \frac{1}{\beta}\|\mathbf{u}\|^2_{(S^n)^p\times\mathbb{R}^{p\times m}} + \langle \mathbf{x},\, L^*\mathbf{u}\rangle_{\mathbb{R}^{p\times m}} + \langle L\mathbf{x},\, \mathbf{u}\rangle_{(S^n)^p\times\mathbb{R}^{p\times m}} \\ &= \frac{1}{\alpha}\|\mathbf{x}\|^2_{\mathbb{R}^{p\times m}} + \frac{1}{\beta}\sum_{i=1}^{p}\|u_i\|^2_{S^n} + \frac{1}{\beta}\|\mathbf{y}\|^2_{\mathbb{R}^{p\times m}} + 2\underbrace{\langle L\mathbf{x},\, \mathbf{u}\rangle_{(S^n)^p\times\mathbb{R}^{p\times m}}}_{(\star)}. \end{aligned}$$

Then

$$\begin{aligned} (\star) &= \left\langle \left(-\sum_{j=1}^{m}(x_1)_j A_1^j,\, -\sum_{j=1}^{m}(x_2)_j A_2^j, \cdots, -\sum_{j=1}^{m}(x_p)^j A_p^j,\, U\mathbf{x} \right), (u_1, u_2, \cdots, u_p, \mathbf{y}) \right\rangle \\ &= \langle \mathbf{x},\, \beta\mathbf{w}\rangle - \sum_{k=1}^{p}\sum_{j=1}^{m}(x_k)_j \mathrm{tr}(A_k^j u_k), \end{aligned}$$

so

$$\|(\mathbf{x}, \mathbf{u})\|_M^2 = \frac{1}{\alpha}\|\mathbf{x}\|^2_{\mathbb{R}^{p\times m}} + \frac{1}{\beta}\sum_{i=1}^{p}\|u_i\|^2_{S^n} + \frac{1}{\beta}\|\mathbf{y}\|^2_{\mathbb{R}^{p\times m}} + 2\beta\langle \mathbf{x},\, \mathbf{w}\rangle - 2\sum_{k=1}^{p}\sum_{j=1}^{m}(x_k)_j \mathrm{tr}(A_k^j u_k).$$

Now, $\mathbf{y}$ can be calculated as follows. Consider a eigenvalue decomposition $(I - W) = V\Sigma V^\intercal$. Let v_i be the i-th column of V, σ_i be the i-th eigenvalue corresponding to v_i. Suppose $\sigma_p = 0$ with $v_p = \mathbf{1}$. As $\mathbf{y} \perp \mathbf{1}$, $\mathbf{y} = \sum_{i=1}^{p-1} y_i v_i$. Then

$$\beta\mathbf{w} = U\mathbf{y} = U\sum_{i=1}^{p-1} y_i v_i.$$

As $U = V\Sigma^{1/2}V^\intercal$,

$$\beta\mathbf{w} = V\Sigma^{1/2}V^\intercal \sum_{i=1}^{p=1} y_i v_i = V\Sigma^{1/2}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{p-1} \\ y_p = 0 \end{bmatrix} = V\begin{bmatrix} \sqrt{\sigma_1}y_1 \\ \sqrt{\sigma_2}y_2 \\ \vdots \\ \sqrt{\sigma_{p-1}}y_{p-1} \\ 0 \end{bmatrix} = \sum_{i=1}^{p-1}\sqrt{\sigma_i}y_i v_i$$

Calculate y_i from taking inner product of $\beta\mathbf{w}$ and v_i.

E.3. Experiment settings and additional plots

In this experiment, we use the parameters $\alpha = \beta = 0.01$ with $n = 10$, $m = 11$, $p = 10$, and $\varepsilon = 0.5$. These numbers come from the infeasible linear matrix inequality (LMI) designed for this experiment, which we state below.

$$\begin{bmatrix} x_1 & x_2 \\ x_2 & \varepsilon \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} x_2 & x_3 \\ x_3 & \varepsilon \end{bmatrix} \succeq 0$$

$$\vdots$$

$$\begin{bmatrix} x_k & x_{k+1} \\ x_{k+1} & \varepsilon \end{bmatrix} \succeq 0,$$

with $\varepsilon > 0$. Then the set of inequalities above is a subset of

$$\left\{ (x_1, x_2, \ldots, x_k, x_{k+1}) \in \mathbb{R}^{k+1} \mid \frac{x_1}{\varepsilon} \geq \left(\frac{x_{k+1}}{\varepsilon}\right)^{2^k} \right\}.$$

If we add another LMI

$$\begin{bmatrix} -x_1 & x_2 \\ x_2 & \varepsilon \end{bmatrix} \succeq 0,$$

The feasible region is also a subset of

$$\left\{ (x_1, x_2, \ldots, x_k, x_{k+1}) \in \mathbb{R}^{k+1} \mid \frac{x_1}{\varepsilon} \leq -\left(\frac{x_{k+1}}{\varepsilon}\right)^{2^k} \right\}.$$

Reversing the sign of the first $(1,1)$-entry of each LMI results in the only feasible region $\{(0, \ldots, 0)\}$. Then, if we additionally impose an LMI such as

$$\begin{bmatrix} x_1 & 0 \\ 0 & x_{k+1} \end{bmatrix} \succeq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

Then the origin $\{(0, \ldots, 0)\}$ is never in a feasible region of the set of all LMIs, so the SDP becomes infeasible. The value of $\|v\|^2$ has been numerically calculated using the normalized iterate of Picard iteration after $200,000$ iterations.

Additionally, we draw plots of the difference of fixed-point residual or normalized iterate between v and $-v$, respectively.

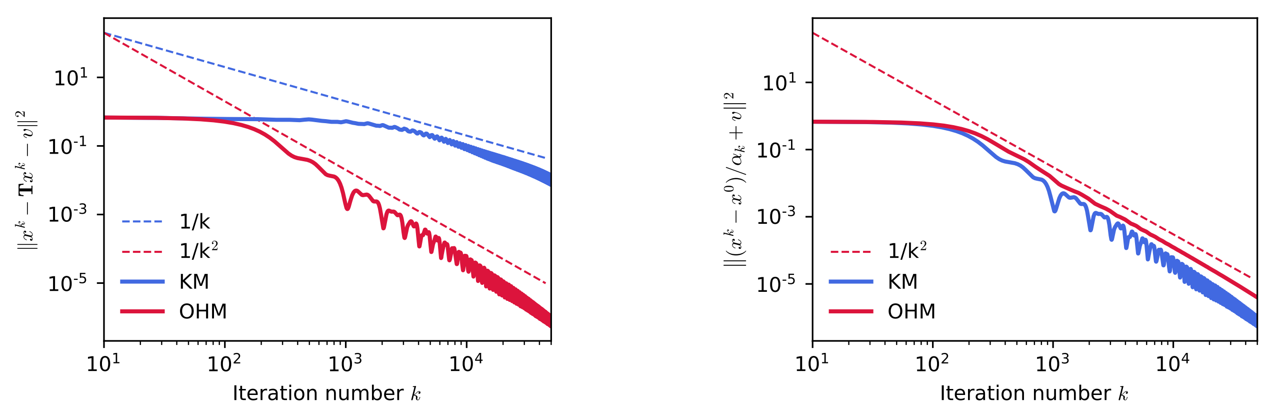

Figure 3. (Left) Squared difference between fixed-point residual and v, $\|x^k - \mathbb{T}x^k - v\|^2$. (Right) Squared difference between normalized iterate and $-v$, $\|(x^k - x^0)/\alpha_k + v\|^2$.