Fast Krasnosel'skiĭ-Mann algorithm with a convergence rate of the fixed point iteration of $o\left(1/k\right)$

Radu Ioan Boţ[*]   Dang-Khoa Nguyen[†]

August 22, 2023

Abstract

The Krasnosel'skiĭ-Mann (KM) algorithm is the most fundamental iterative scheme designed to find a fixed point of an averaged operator in the framework of a real Hilbert space, since it lies at the heart of various numerical algorithms for solving monotone inclusions and convex optimization problems. We enhance the Krasnosel'skiĭ-Mann algorithm with Nesterov's momentum updates and show that the resulting numerical method exhibits a convergence rate for the fixed point residual of $o\left(\frac{1}{k}\right)$ while preserving the weak convergence of the iterates to a fixed point of the operator. Numerical experiments illustrate the superiority of the resulting so-called Fast KM algorithm over various fixed point iterative schemes, and also its oscillatory behavior, which is a specific of Nesterov's momentum optimization algorithms.

Key Words. nonexpansive operator, averaged operator, Krasnosel'skiĭ-Mann iteration, Nesterov's momentum, Lyapunov analysis, convergence rates, convergence of iterates

AMS subject classification. 47J20, 47H05, 65K15, 65Y20

1 Introduction

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$. In this paper we are interested in formulating a fast numerical method for solving the fixed point problem

$$\text{Find } x \in \mathcal{H} \text{ such that } x = T\left(x\right), \tag{1}$$

where $T\colon \mathcal{H} \to \mathcal{H}$ is a θ-averaged operator with $\theta \in (0,1]$. Recall that an operator $R\colon \mathcal{H} \to \mathcal{H}$ is *nonexpansive* if it is 1-Lipschitz continuous, that is

$$\|R\left(x\right) - R\left(y\right)\| \leqslant \|x - y\| \quad \forall x, y \in \mathcal{H}.$$

Then T is said to be *averaged with constant θ* or *θ-averaged* if there exists a nonexpansive operator $R\colon \mathcal{H} \to \mathcal{H}$ such that

$$T = (1 - \theta)\,\text{Id} + \theta R,$$

where $\text{Id}\colon \mathcal{H} \to \mathcal{H}$ denotes the identity mapping on $\mathcal{H}$. Obviously, an operator T is nonexpansive if it is at least 1-averaged. We denote the set of all *fixed points* of T by $\text{Fix}\,T := \{x \in \mathcal{H} \mid x = T\left(x\right)\}$.

The most naive approach when looking for a fixed point of T is the following process, also called *Banach-Picard iteration*,

$$x_{k+1} := T\left(x_k\right) \quad \forall k \geqslant 0, \tag{2}$$

where $x_0 \in \mathcal{H}$ is a starting point.

According to the Banach-Picard fixed point theorem, if T is a *contraction*, namely, T is Lipschitz continuous with modulus $\delta \in [0, 1)$, then the sequence $(x_k)_{k \geqslant 0}$ generated by (2) converges strongly to the unique fixed point of T with linear convergence rate. If T is just nonexpansive, then this statement is no longer true. To illustrate this, it is enough to choose $T = -\mathrm{Id}$ and $x_0 \neq 0$, in which case the Banach-Picard iteration not only fails approach a fixed point of T, but also generates a sequence that does not satisfy the *asymptotic regularity property*. We say that the sequence $(x_k)_{k \geqslant 0}$ satisfies the asymptotic regularity property if the difference $x_k - T(x_k)$ converges strongly to 0 as k tends to $+\infty$. This property is crucial for guaranteeing the convergence of the iterates, as we will see later.

In order to overcome the restrictive contraction assumption on T, Krasnosel'skiĭ proposed in [26] to apply the Banach-Picard iteration (2) to the operator $\frac{1}{2}\mathrm{Id} + \frac{1}{2}T$ instead of T. Following on the idea of using convex combinations the so-called *Krasnosel'skiĭ-Mann (KM) iteration*

$$x_{k+1} := (1 - s_k) x_k + s_k T(x_k) \quad \forall k \geqslant 0, \tag{3}$$

where $(s_k)_{k \geqslant 0}$ is a sequence in $(0, 1]$, emerged. It turned out that a fundamental step in proving the convergence of the iterates of (3) is to show that $x_k - T(x_k) \to 0$ as $k \to +\infty$, as it was done by Browder and Petryshyn in [11] in the constant case $s_k \equiv s \in (0, 1)$. The extension to nonconstant sequences was achieved by Groetsch in [19] who proved that, if $\sum_{k \geqslant 0} s_k (1 - s_k) = +\infty$, then the asymptotic regularity property is satisfied. The weak convergence of the iterates was then studied in various settings in [19, 24, 37, 4, 3]. Tikhonov regularization based techniques to improve the convergence of the iterates from weak to strong have been recently studied in [6, 8].

By considering convex combinations with a fixed so-called *anchor point* $x_0 \in \mathcal{H}$ one obtains the *Halpern iteration* [21]

$$x_{k+1} := (1 - s_k) x_0 + s_k T(x_k) \quad \forall k \geqslant 0, \tag{4}$$

a method that has recently attracted a lot of interest [29, 43, 36, 35]. The asymptotic regularity property of this iterative scheme has been studied in [41, 42].

Despite having ubiquitous applications in various fields, the study of the computational complexity of fixed point iteration schemes is still limited. One natural measure to quantify this is by means of the rate of convergence of the *fixed point residual* $\|x_k - T(x_k)\|$. Notice that the asymptotic regularity property does not automatically provide an explicit convergence rate.

Sabach and Shtern proved in [38] for a general form of the Halpern iteration that the rate of convergence of the fixed point residual is of $O\left(\frac{1}{k}\right)$. Lieder tightened this results in [29] by a constant factor, for the Halpern iteration with $s_k := 1 - \frac{1}{k+2}$ for every $k \geqslant 0$, whereas Park and Ryu proved in [35] that the convergence rate of $O\left(\frac{1}{k}\right)$ is optimal for this iterative scheme, which means that it can not be improved in general. On the other hand, the convergence of the Krasnosel'skiĭ-Mann iteration expressed in terms of the fixed point residual was in the nineties proved to be of $o\left(\frac{1}{\sqrt{k}}\right)$ in [2] in the case of a constant sequence $(s_k)_{k \geqslant 0}$. Later on, in the case of a nonconstant sequence, it was proved to be of $O\left(\frac{1}{\sqrt{k}}\right)$ in [12, 28], and of $o\left(\frac{1}{\sqrt{k}}\right)$ in [9, 15, 32], whereas in [5] it was shown that the asymptotic rate of convergence for the fixed point residual of the continuous time counterpart of the Krasnosel'skiĭ-Mann iteration is of $o\left(\frac{1}{\sqrt{t}}\right)$. Recently, Fierro, Maulén and Peypouquet proved in [18] that the rate of convergence of the fixed point residual of a general *inertial Krasnosel'skiĭ-Mann algorithm* is also of $o\left(\frac{1}{\sqrt{k}}\right)$. Noticeably, Contreras and Cominetti showed in [13] that in the *Banach space setting* the lower bound of the Krasnosel'skiĭ-Mann iteration is $O\left(\frac{1}{\sqrt{k}}\right)$, which does not say anything about the lower bound in the Hilbert space setting.

For a family of general approaches aimed to "accelerate" the convergence of sequences relying on Shanks transformation and including Anderson acceleration, which can be applied also to fixed point problems, we refer to [10].

In this paper we introduce an iterative method for solving the fixed point problem (1) which exhibits a rate of convergence for the fixed point residual of $o\left(\frac{1}{k}\right)$ and guarantees the weak convergence of the iterates to a fixed point of T. The method is obtained by enhancing the Krasnosel'skiĭ-Mann iteration with Nesterov's momentum updates and follows via the temporal discretization of the second order dynamical system with vanishing damping term proposed in [7] for solving monotone equations.

The iterative scheme exploits the coercivity of the operator $\mathrm{Id} - T$ and has consequently a much more simple formulation than the Fast OGDA algorithm introduced in [7] for solving monotone equations, which requires the construction of auxiliary sequences. Numerical experiments show that the resulting so-called Fast KM algorithm outperforms various fixed point iterative schemes including recently introduced ones using anchoring. The numerical experiments also illustrate the oscillatory behavior of the method, which is a specific of algorithms with Nesterov's momentum updates.

As a by-product of our proposed approach we obtain several fast splitting methods for solving monotone inclusions. It is well-known that some of the most prominent splitting schemes result as particular instances of the Krasnosel'skiĭ-Mann iteration, since they can be reduced to the solving of a fixed point problem governed by an average operator. This is the case for the Douglas-Rachford splitting [17, 30], the forward-backward splitting [30], and the three operator splitting [16, 14]. For a comprehensive study of operator splitting schemes we refer to [3]. Recent contributions to the acceleration of the convergence of splitting methods have been made in [25, 43, 27, 40, 35].

2 A fast Krasnosel'skiĭ-Mann iteration

In our approach, we rely on the simple observation that

$$x_* \in \mathsf{Fix}T \Leftrightarrow (\mathrm{Id} - T)(x_*) = 0,$$

which allows us to benefit from the recent development on a continuous fast method for solving monotone equations in [7]. To be more specific, we have that T is θ-averaged if and only if $\mathrm{Id} - T$ is $\frac{1}{2\theta}$-cocoercive [3, Proposition 4.39], that is

$$\langle x - y, (\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y)\rangle \geqslant \frac{1}{2\theta}\|(\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y)\|^2 \geqslant 0 \quad \forall x, y \in \mathcal{H}. \quad (5)$$

From here one can immediately see that it follows immediately that $\mathrm{Id} - T$ is monotone. Furthermore, from the Cauchy-Schwarz inequality we can see that $\mathrm{Id} - T$ is at most 2θ-Lipschitz continuous.

As a direct consequence of (5) we have that for every $x_* \in \mathsf{Fix}T$ it holds

$$\langle x - x_*, x - T(x)\rangle \geqslant \frac{1}{2\theta}\|x - T(x)\|^2 \geqslant 0 \quad \forall x \in \mathcal{H}.$$

The dynamical system studied in [7], formulated for the monotone equation $(\mathrm{Id} - T)(x) = 0$ and for constant time scaling term $\beta(t) \equiv 1$, reads

$$\begin{cases} \ddot{x}(t) + \dfrac{\alpha}{t}\dot{x}(t) + \dfrac{d}{dt}(\mathrm{Id} - T)(x(t)) + \dfrac{\alpha}{2t}(\mathrm{Id} - T)(x(t)) = 0 \\ x(t_0) = x_0 \text{ and } \dot{x}(t_0) = \dot{x}_0, \end{cases} \quad (6)$$

where $\alpha \geqslant 2$, $(x_0, \dot{x}_0) \in \mathcal{H} \times \mathcal{H}$. The Lipschitz continuity of $\mathrm{Id} - T$ guarantees the existence and uniqueness of a strong global solution $x : [t_0, +\infty] \to \mathcal{H}$, which means that x and $\dot{x}$ are locally absolutely continuous, $x(t_0) = x_0$, $\dot{x}(t_0) = \dot{x}_0$ and x fulfills the first equation in (6) almost everywhere.

We set the time scaling term equal to one since our aim is to derive via temporal discretization an explicit iterative fixed point scheme, whereas nonconstant time scaling terms are known to positively impact the convergence rates of implicit numerical algorithms; see [7, Remark 2] for a detailed discussion on this issue.

Consider the first-order formulation of the first equation in (6)

$$\begin{cases} \dot{u}(t) &= (2 - \alpha)(\mathrm{Id} - T)(x(t)) \\ u(t) &= 2(\alpha - 1)x(t) + 2t\dot{x}(t) + 2t(\mathrm{Id} - T)(x(t)). \end{cases} \quad (7)$$

We fix a time step $s > 0$, set $s_k := s(k + 1)$ for every $k \geqslant 1$, and approximate $x(s_k) \approx x_{k+1}$, and $u(s_k) \approx u_{k+1}$. The explicit finite-difference scheme for (7) at time $t := s_k$ gives for every $k \geqslant 1$

$$\begin{cases} \dfrac{u_{k+1} - u_k}{s} &= (2 - \alpha)(\mathrm{Id} - T)(x_k) \\ u_{k+1} &= 2(\alpha - 1)x_{k+1} + 2(k + 1)(x_{k+1} - x_k) + 2s(k + 1)(\mathrm{Id} - T)(x_k), \end{cases} \quad (8)$$

3

with the initialization $u_0 := x_0 - s\dot{x}_0$ and $u_1 := x_0$. Different to [7], where for the discretization of the argument of $\mathrm{Id} - T$ we used an auxiliary sequence, this time we can use $(x_k)_{k \geqslant 0}$. This is thanks to the stronger property of cocoercivity the operator $\mathrm{Id} - T$ is enhanced with and which will be reflected in the convergence analysis. We will see that this allows us not only to design a simpler algorithm, but also to consider larger step sizes than for the one proposed in [7].

Next we will simplify the sequence $(u_k)_{k \geqslant 0}$. The second equation in (8) gives for every $k \geqslant 1$

$$u_k = 2(\alpha - 1)x_k + 2k(x_k - x_{k-1}) + 2sk(\mathrm{Id} - T)(x_{k-1}).$$

Taking the difference we obtain for every $k \geqslant 1$

$$
\begin{aligned}
u_{k+1} - u_k &= 2(k + \alpha)(x_{k+1} - x_k) - 2k(x_k - x_{k-1}) + 2s(\mathrm{Id} - T)(x_k) \\
&\quad + 2sk((\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})) \\
&= (2 - \alpha)s(\mathrm{Id} - T)(x_k),
\end{aligned}
\tag{9}
$$

where the last relation comes from the first equation in (8). After rearranging (9), we obtain for every $k \geqslant 1$

$$2(k + \alpha)(x_{k+1} - x_k) = 2k(x_k - x_{k-1}) - \alpha s(\mathrm{Id} - T)(x_k) - 2sk((\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})).
\tag{10}$$

From here, we deduce that for every $k \geqslant 1$

$$
\begin{aligned}
x_{k+1} &= x_k + \frac{k}{k + \alpha}(x_k - x_{k-1}) - \frac{s\alpha}{2(k + \alpha)}(\mathrm{Id} - T)(x_k) - \frac{sk}{k + \alpha}((\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})) \\
&= \left(1 - \frac{s\alpha}{2(k + \alpha)}\right)x_k + \frac{(1 - s)k}{k + \alpha}(x_k - x_{k-1}) + \frac{s\alpha}{2(k + \alpha)}T(x_k) + \frac{sk}{k + \alpha}(T(x_k) - T(x_{k-1})).
\end{aligned}
$$

Summing up, the algorithm we propose in this paper for solving (1) has the following formulation.

Algorithm 1. (Fast KM algorithm) *Let $\alpha > 2, x_0, x_1 \in \mathcal{H}$ and $0 < s \leqslant \frac{1}{\theta}$. For every $k \geqslant 1$ we set*

$$x_{k+1} := \left(1 - \frac{s\alpha}{2(k + \alpha)}\right)x_k + \frac{(1 - s)k}{k + \alpha}(x_k - x_{k-1}) + \frac{s\alpha}{2(k + \alpha)}T(x_k) + \frac{sk}{k + \alpha}(T(x_k) - T(x_{k-1})).
\tag{11}$$

Remark 1. For the step size choice $s := 1$ which is allowed for every θ-averaged operator T with $\theta \in (0, 1]$, our iterative scheme becomes

$$x_{k+1} := \left(1 - \frac{\alpha}{2(k + \alpha)}\right)x_k + \frac{\alpha}{2(k + \alpha)}T(x_k) + \frac{k}{k + \alpha}(T(x_k) - T(x_{k-1})) \quad \forall k \geqslant 1.
\tag{12}$$

Notice that for a *nonexpansive* operator T, which corresponds to the case $\theta = 1$, the value $s := 1$ is the largest step size that can be taken.

The numerical algorithm (12) can be interpreted as a Krasnosel'skiĭ-Mann iteration enhanced with the extrapolation term $\frac{k}{k + \alpha}(T(x_k) - T(x_{k-1}))$ which proves to have an accelerating effect on the convergence of the fixed point residual. We learn form here that, in order to improve the convergence rate while preserving the convergence of the iterates, one must address iterative schemes that go beyond the classical *Mann iteration* [31]. The latter allows in the update rule only nonnegative coefficients for both the previous iterates and the operator evaluations at the previous iterates.

Remark 2. A direct application of the explicit Fast OGDA method in [7] to the solving of the monotone equation $(\mathrm{Id} - T)(x) = 0$ leads for given $\alpha > 2, x_0, x_1, y_0 \in \mathcal{H}, 0 < s < \max\left\{\frac{1}{4\theta}, \frac{1-\theta}{2\theta}\right\}$ to the following iterative scheme: for every $k \geqslant 1$ set

$$y_k := x_k + \left(1 - \frac{\alpha}{k + \alpha}\right)(x_k - x_{k-1}) - \frac{\alpha s}{2(k + \alpha)}(\mathrm{Id} - T)(y_{k-1})
\tag{13a}$$

$$x_{k+1} := y_k - \frac{s}{2}\left(1 + \frac{k}{k + \alpha}\right)((\mathrm{Id} - T)(y_k) - (\mathrm{Id} - T)(y_{k-1})).
\tag{13b}$$

If T is θ-averaged, then $\mathrm{Id} - T$ is L-Lipschitz with $L := \min\left\{2\theta, \frac{\theta}{1-\theta}\right\}$ (here we make the convention $\frac{1}{0} := +\infty$), thus the upper bound of the step size is $\frac{1}{2L} = \max\left\{\frac{1}{4\theta}, \frac{1-\theta}{2\theta}\right\}$. Indeed, we already noticed that, since $\mathrm{Id} - T$ is $\frac{1}{2\theta}$-cocoercive, it is at most 2θ-Lipschitz continuous. On the other hand, T is θ-averaged if and only if (cf. [3, Proposition 4.35])

$$\|T(x) - T(y)\|^2 \leqslant \|x - y\|^2 - \frac{1-\theta}{\theta}\|(\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y)\|^2 \quad \forall x, y \in \mathcal{H},$$

which implies that $\mathrm{Id} - T$ is Lipschitz continuous with modulus at most $\frac{\theta}{1-\theta}$.

Noticeably, the step size upper bound for algorithm (13) is more restrictive compared to the one in the Fast KM iteration (11). In addition, (11) has an obviously simpler formulation than (13). Even if one rewrites (13) in terms of a single sequence $(y_k)_{k \geqslant 0}$, it would require y_k, y_{k-1} and y_{k-2} to compute y_{k+1}. In comparison, (11) requires x_k and x_{k-1} to compute x_{k+1}.

3 Convergence analysis

The fundamental tool of the convergence analysis is the following discrete energy function which, for fixed $x_* \in \mathsf{Fix}T$ and $0 \leqslant \lambda \leqslant \alpha - 1$, is defined for every $k \geqslant 1$ as

$$\mathcal{E}_{\lambda,k} := \frac{1}{2}\left\|2\lambda(x_k - x_*) + 2k(x_k - x_{k-1}) + \frac{1}{2(\alpha-1)}(3\alpha-2)sk(\mathrm{Id}-T)(x_{k-1})\right\|^2$$

$$+ 2\lambda(\alpha-1-\lambda)\|x_k - x_*\|^2 + \frac{1}{\alpha-1}(\alpha-2)\lambda sk\langle x_k - x_*, (\mathrm{Id}-T)(x_{k-1})\rangle$$

$$+ \frac{1}{8(\alpha-1)^2}(\alpha-2)(3\alpha-2)s^2k^2\|(\mathrm{Id}-T)(x_{k-1})\|^2.$$

The discrete energy function is defined in analogy with the Lyapunov energy functions used in the study of continuous time dynamical systems associated with convex minimization problems and monotone equations ([1, 7]). While in convex minimization, the distance from the objective function at the current iterate to its minimal objective value plays the prominent role in the definition of the discrete energy function, in the present setting, this role is taken by $\|(\mathrm{Id}-T)(x_{k-1})\|^2$. The coefficient k^2 in front of this term suggests the rate at which we expect this term to converge, provided the energy sequence converges as $k \to +\infty$. The same reasoning applies to the third summand in the discrete energy function, while the first two summands will play an important role in proving the convergence of the iterates.

The properties of the discrete energy function are presented in the following lemma, with the proof deferred to the Appendix.

Lemma 3. *Let $x_* \in \mathsf{Fix}T$ and $(x_k)_{k \geqslant 0}$ be the sequence generated by Algorithm 1. Then the following statements are true:*

(i) *for $0 \leqslant \lambda \leqslant \alpha - 1$ and every $k \geqslant 1$ it holds*

$$\mathcal{E}_{\lambda,k+1} - \mathcal{E}_{\lambda,k} \leqslant 2(2-\alpha)\lambda s\langle x_k - x_*, (\mathrm{Id}-T)(x_k)\rangle + \omega_1 k\|x_{k+1} - x_k\|^2$$

$$+ s(\omega_2 k + \omega_3)\langle x_{k+1} - x_k, (\mathrm{Id}-T)(x_k)\rangle + \omega_4 s^2 k\|(\mathrm{Id}-T)(x_k)\|^2$$

$$+ \frac{1}{(\alpha-1)}(\alpha-2)\left(s - \frac{1}{\theta}\right)sk^2\|(\mathrm{Id}-T)(x_k) - (\mathrm{Id}-T)(x_{k-1})\|^2, \qquad (14)$$

where

$$\omega_1 := 4(\lambda + 1 - \alpha) \leqslant 0, \qquad (15a)$$

$$\omega_2 := \frac{1}{\alpha-1}\left(4(\alpha-1)(\lambda+1-\alpha) + \alpha(2-\alpha)\right) \leqslant 0, \qquad (15b)$$

$$\omega_3 := \frac{1}{\alpha-1}\left(2\alpha(\alpha-1)(\lambda+1-\alpha) + \alpha - 2(\alpha-1)^2 + 2(2-\alpha)(\alpha-1)\right), \qquad (15c)$$

$$\omega_4 := \frac{1}{2(\alpha-1)}(2-\alpha)(3\alpha-2) \leqslant 0; \qquad (15d)$$

5

(ii) *for $0 \leqslant \lambda \leqslant \frac{3\alpha}{4} - \frac{1}{2}$ the sequence $(\mathcal{E}_{\lambda,k})_{k \geqslant 1}$ is nonnegative.*

In the following lemma, we demonstrate that there exist infinitely many choices for the parameter λ (depending on α) for which an essential quantity in the expression on the right-hand side of (14) becomes non-positive after a finite number of iterations. As we will see in Proposition 5, this behaviour will lead to the convergence of the corresponding discrete energy function $\mathcal{E}_{\lambda,k}$ as $k \to +\infty$. The proof of Lemma 4 can also be found in the Appendix.

Lemma 4. *Let*

$$\underline{\lambda}(\alpha) := \frac{1}{8(\alpha-1)}\alpha^2 + \frac{1}{2}(\alpha-1) - \frac{1}{8(\alpha-1)}(\alpha-2)\sqrt{(\alpha-2)(5\alpha-2)} > 0, \tag{16a}$$

$$\overline{\lambda}(\alpha) := \min\left\{\frac{3\alpha}{4} - \frac{1}{2}, \frac{1}{8(\alpha-1)}\alpha^2 + \frac{1}{2}(\alpha-1) + \frac{1}{8(\alpha-1)}(\alpha-2)\sqrt{(\alpha-2)(5\alpha-2)}\right\}. \tag{16b}$$

Then for every λ satisfying $\underline{\lambda}(\alpha) < \lambda < \overline{\lambda}(\alpha)$ one can find an integer $k(\lambda) \geqslant 1$ with the property that the following inequality holds for every $k \geqslant k(\lambda)$

$$R_k := \sqrt{\frac{5\alpha-2}{2(3\alpha-2)}}\omega_1 k\|x_{k+1} - x_k\|^2 + s(\omega_2 k + \omega_3)\langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k)\rangle$$

$$+ \sqrt{\frac{5\alpha-2}{2(3\alpha-2)}}\omega_4 s^2 k\|(\mathrm{Id} - T)(x_k)\|^2 \leqslant 0, \tag{17}$$

where ω_1, ω_2, ω_3 and ω_4 are the constants defined in (15).

Proposition 5. *Let $x_* \in \mathsf{Fix}T$ and $(x_k)_{k\geqslant 0}$ be the sequence generated by Algorithm 1. Then it holds*

$$\sum_{k\geqslant 1}\langle x_k - x_*, (\mathrm{Id} - T)(x_k)\rangle < +\infty, \tag{18a}$$

$$\sum_{k\geqslant 1} k\|x_{k+1} - x_k\|^2 < +\infty, \tag{18b}$$

$$\sum_{k\geqslant 1} k\|(\mathrm{Id} - T)(x_k)\|^2 < +\infty, \tag{18c}$$

$$\left(\frac{1}{\theta} - s\right)\sum_{k\geqslant 1} k^2\|(\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\|^2 < +\infty. \tag{18d}$$

In addition, the sequence $(\mathcal{E}_{\lambda,k})_{k\geqslant 1}$ converges for every $\underline{\lambda}(\alpha) < \lambda < \overline{\lambda}(\alpha)$, where the pair $\left(\underline{\lambda}(\alpha), \overline{\lambda}(\alpha)\right)$ is defined in (16). Consequently, the sequence $(x_k)_{k\geqslant 0}$ is bounded.

Proof. Let $\left(\underline{\lambda}(\alpha), \overline{\lambda}(\alpha)\right)$ be the pair defined in (16) and $\underline{\lambda}(\alpha) < \lambda < \overline{\lambda}(\alpha)$. By Lemma 4, there exists an integer $k(\lambda) \geqslant 1$ such that for every $k \geqslant k(\lambda)$ it holds

$$s(\omega_2 k + \omega_3)\langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k)\rangle$$

$$\leqslant -\sqrt{\frac{5\alpha-2}{2(3\alpha-2)}}\omega_1 k\|x_{k+1} - x_k\|^2 - \sqrt{\frac{5\alpha-2}{2(3\alpha-2)}}\omega_4 s^2 k\|(\mathrm{Id} - T)(x_k)\|^2.$$

By plugging this inequality into (14) it follows that for every $k \geqslant k(\lambda)$ it holds

$$\mathcal{E}_{\lambda,k+1} - \mathcal{E}_{\lambda,k} \leqslant 2(2-\alpha)\lambda s\langle x_k - x_*, (\mathrm{Id} - T)(x_k)\rangle + \left(1 - \sqrt{\frac{5\alpha-2}{2(3\alpha-2)}}\right)\omega_1 k\|x_{k+1} - x_k\|^2$$

$$+ \left(1 - \sqrt{\frac{5\alpha-2}{2(3\alpha-2)}}\right)\omega_4 s^2 k\|(\mathrm{Id} - T)(x_k)\|^2$$

$$+ \frac{1}{(\alpha-1)}(\alpha-2)\left(s - \frac{1}{\theta}\right)sk^2\|(\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\|^2.$$

6

Taking into account that $\omega_1, \omega_4 \leqslant 0$ (see (15)), and $\frac{5\alpha-2}{2(3\alpha-2)} < 1$, we can apply Lemma 13 to obtain the summability results in (18) as well as the fact that the sequence $(\mathcal{E}_{\lambda,k})_{k\geqslant 1}$ is convergent. Since $0 < \underline{\lambda}(\alpha) < \lambda < \overline{\lambda}(\alpha) < \alpha - 1$, the boundedness of $(x_k)_{k\geqslant 0}$ will then follow from the definition of the discrete energy function $(\mathcal{E}_{\lambda,k})_{k\geqslant 1}$. $\qquad\square$

Next we will show the convergence of the sequence of iterates. The proof relies on the Opial Lemma (see Lemma 11) and the *demiclosedness principle* for nonexpansive operators. According to this principle, if $(z_k)_{k\geqslant 0} \subseteq \mathcal{H}$ is a sequence which converges weakly to $z \in \mathcal{H}$ such that $z_k - T(z_k)$ converges strongly to 0 as $k \to +\infty$, then $z \in \mathsf{Fix}T$ (see [3, Corollary 4.28]).

Theorem 6. *Let $(x_k)_{k\geqslant 0}$ be the sequence generated by Algorithm 1. Then $(x_k)_{k\geqslant 0}$ converges weakly to an element in $\mathsf{Fix}T$ as $k \to +\infty$.*

Proof. Let $x_* \in \mathsf{Fix}T$, $\left(\underline{\lambda}(\alpha), \overline{\lambda}(\alpha)\right)$ be the pair defined in (16) and $\underline{\lambda}(\alpha) < \lambda < \overline{\lambda}(\alpha)$. By the definition we have for every $k \geqslant 1$

$$
\begin{aligned}
\mathcal{E}_{\lambda,k} = {} & 2\lambda \left\langle x_k - x_*, 2k(x_k - x_{k-1}) + \frac{1}{2(\alpha-1)}(3\alpha - 2)sk(\mathrm{Id} - T)(x_{k-1}) \right\rangle \\
& + 2\lambda(\alpha - 1)\|x_k - x_*\|^2 + \frac{1}{\alpha - 1}(\alpha - 2)\lambda sk\langle x_k - x_*, (\mathrm{Id} - T)(x_{k-1})\rangle \\
& + \frac{k^2}{2}\left\| 2(x_k - x_{k-1}) + \frac{1}{2(\alpha-1)}(3\alpha - 2)s(\mathrm{Id} - T)(x_{k-1}) \right\|^2 \\
& + \frac{1}{8(\alpha-1)^2}(\alpha - 2)(3\alpha - 2)s^2 k^2 \|(\mathrm{Id} - T)(x_{k-1})\|^2,
\end{aligned}
\tag{19}
$$

which implies for every $\underline{\lambda}(\alpha) < \lambda_1 < \lambda_2 < \overline{\lambda}(\alpha)$ and every $k \geqslant 1$

$$
\mathcal{E}_{\lambda_2,k} - \mathcal{E}_{\lambda_1,k} = 4(\lambda_2 - \lambda_1)\left(k\langle x_k - x_*, x_k - x_{k-1} + s(\mathrm{Id} - T)(x_{k-1})\rangle + \frac{1}{2}(\alpha - 1)\|x_k - x_*\|^2\right).
\tag{20}
$$

For every $k \geqslant 1$ we set

$$
p_k := \frac{1}{2}(\alpha - 1)\|x_k - x_*\|^2 + k\langle x_k - x_*, x_k - x_{k-1} + s(\mathrm{Id} - T)(x_{k-1})\rangle,
\tag{21}
$$

$$
q_k := \frac{1}{2}\|x_k - x_*\|^2 + s\sum_{i=1}^{k}\langle x_i - x_*, (\mathrm{Id} - T)(x_{i-1})\rangle.
\tag{22}
$$

We notice that for every $k \geqslant 2$

$$
q_k - q_{k-1} = \langle x_k - x_*, x_k - x_{k-1}\rangle - \frac{1}{2}\|x_k - x_{k-1}\|^2 + s\langle x_k - x_*, (\mathrm{Id} - T)(x_k)\rangle,
$$

and thus

$$
(\alpha - 1)q_k + k(q_k - q_{k-1}) = p_k + (\alpha - 1)s\sum_{i=1}^{k}\langle x_i - x_*, (\mathrm{Id} - T)(x_{i-1})\rangle - \frac{k}{2}\|x_k - x_{k-1}\|^2.
$$

Since the discrete energy function converges for every $\underline{\lambda}(\alpha) < \lambda_1 < \lambda_2 < \overline{\lambda}(\alpha)$, we obtain that $\lim_{k\to+\infty}(\mathcal{E}_{\lambda_2,k} - \mathcal{E}_{\lambda_1,k}) \in \mathbb{R}$ exists. This implies in view of (20) and (21) that

$$
\lim_{k\to+\infty} p_k \in \mathbb{R} \text{ exists.}
\tag{23}
$$

Moreover, thanks to the triangle inequality and the statements (18a) - (18c) in Proposition 5, we have

for every $k \geqslant 1$

$$\sum_{i=1}^{k} |\langle x_i - x_*, (\mathrm{Id} - T)(x_{i-1}) \rangle|$$

$$\leqslant \sum_{i=1}^{k} |\langle x_i - x_{i-1}, (\mathrm{Id} - T)(x_{i-1}) \rangle| + \sum_{i=1}^{k} \langle x_{i-1} - x_*, (\mathrm{Id} - T)(x_{i-1}) \rangle$$

$$\leqslant \frac{1}{2} \sum_{i=1}^{k} \|x_i - x_{i-1}\|^2 + \frac{1}{2} \sum_{i=1}^{k} \|(\mathrm{Id} - T)(x_{i-1})\|^2 + \sum_{i=1}^{k} \langle x_{i-1} - x_*, (\mathrm{Id} - T)(x_{i-1}) \rangle$$

$$\leqslant \frac{1}{2} \sum_{i \geqslant 1} \|x_i - x_{i-1}\|^2 + \frac{1}{2} \sum_{i \geqslant 1} \|(\mathrm{Id} - T)(x_{i-1})\|^2 + \sum_{i \geqslant 1} \langle x_{i-1} - x_*, (\mathrm{Id} - T)(x_{i-1}) \rangle < +\infty.$$

This means that the series $\sum_{i=1}^{k} \langle x_i - x_*, (\mathrm{Id} - T)(x_{i-1}) \rangle$ is absolutely convergent. In addition, due to (18b),

$$\lim_{k \to +\infty} k \|x_k - x_{k-1}\|^2 = 0,$$

which implies that

$$\lim_{k \to +\infty} ((\alpha - 1) q_k + k (q_k - q_{k-1})) \in \mathbb{R} \text{ exists.}$$

From Proposition 5, we have that $(x_k)_{k \geqslant 0}$ is bounded, hence $(q_k)_{k \geqslant 1}$ is also bounded. This allows us to apply Lemma 12 to conclude that $\lim_{k \to +\infty} q_k \in \mathbb{R}$ also exists. Once again, by the definition of q_k in (22) and the fact that the sequence $\left(\sum_{i=1}^{k} \langle x_{i-1} - x_*, (\mathrm{Id} - T)(x_{i-1}) \rangle \right)_{k \geqslant 1}$ converges, it follows that $\lim_{k \to +\infty} \|x_k - x_*\| \in \mathbb{R}$ exists. In other words, the hypothesis (i) in Opial Lemma (see Lemma 11) is fulfilled.

Now let $\bar{x}$ be a weak sequential cluster point of $(x_k)_{k \geqslant 0}$, meaning that there exists a subsequence $(x_{k_n})_{n \geqslant 0}$ such that

$$x_{k_n} \text{ converges weakly to } \bar{x} \text{ as } n \to +\infty.$$

On the other hand, according to (18c),

$$(\mathrm{Id} - T)(x_{k_n}) \text{ converges strongly to } 0 \text{ as } n \to +\infty.$$

Due to the demiclosedness principle we conclude from here that $\bar{x} \in \mathsf{Fix}\, T$. This shows that the hypothesis (ii) in Opial Lemma is also fulfilled, and completes the proof. $\square$

The following results proves the convergence rate of the Fast KM algorithm in terms of the discrete velocity and fixed point residual.

Theorem 7. *Let $(x_k)_{k \geqslant 0}$ be the sequence generated by Algorithm 1. Then it holds*

$$\|x_k - x_{k-1}\| = o\left(\frac{1}{k}\right) \qquad \text{and} \qquad \|x_{k-1} - T(x_{k-1})\| = o\left(\frac{1}{k}\right) \text{ as } k \to +\infty.$$

Proof. Let $x_* \in \mathsf{Fix}\, T$, $\left(\underline{\lambda}(\alpha), \overline{\lambda}(\alpha) \right)$ be the pair defined in (16) and $\underline{\lambda}(\alpha) < \lambda < \overline{\lambda}(\alpha)$. According to Proposition 5, the sequence $(\mathcal{E}_{\lambda,k})_{k \geqslant 1}$ converges.

From (19) and (21) we have that for every $k \geqslant 1$

$$\mathcal{E}_{\lambda,k} = 4\lambda p_k + \frac{k^2}{2} \left\| 2(x_k - x_{k-1}) + \frac{1}{2(\alpha - 1)} (3\alpha - 2) s (\mathrm{Id} - T)(x_{k-1}) \right\|^2$$

$$+ \frac{1}{8(\alpha - 1)^2} (\alpha - 2)(3\alpha - 2) s^2 k^2 \|(\mathrm{Id} - T)(x_{k-1})\|^2.$$

We set for every $k \geqslant 1$

$$h_k := \frac{k^2}{2} \left(\left\| 2(x_k - x_{k-1}) + \frac{1}{2(\alpha-1)}(3\alpha-2) s \, (\text{Id} - T)(x_{k-1}) \right\|^2 \right.$$

$$\left. + \frac{1}{4(\alpha-1)^2}(\alpha-2)(3\alpha-2) s^2 \left\| (\text{Id} - T)(x_{k-1}) \right\|^2 \right),$$

so that $\mathcal{E}_{\lambda,k} = 4\lambda p_k + h_k$. Furthermore, since $\lim_{k\to+\infty} \mathcal{E}_{\lambda,k} \in \mathbb{R}$ and $\lim_{k\to+\infty} p_k \in \mathbb{R}$ (see also (23)), it holds

$$\lim_{k\to+\infty} h_k \in \mathbb{R} \text{ exists.}$$

On the other hand, in view of (18b) and (18c) in Proposition 5 we have

$$\sum_{k\geqslant 1} \frac{1}{k} h_k \leqslant 4 \sum_{k\geqslant 1} k \|x_k - x_{k-1}\|^2 + \frac{1}{8(\alpha-1)^2}(3\alpha-2)(7\alpha-6) s^2 \sum_{k\geqslant 1} k \|(\text{Id}-T)(x_{k-1})\|^2 < +\infty.$$

Consequently, $\lim_{k\to+\infty} h_k = 0$, which yields

$$\lim_{k\to\infty} k \left\| 2(x_k - x_{k-1}) + \frac{1}{2(\alpha-1)}(3\alpha-2) s \, (\text{Id}-T)(x_{k-1}) \right\| = \lim_{k\to\infty} k \left\| (\text{Id}-T)(x_{k-1}) \right\| = 0.$$

This immediately implies $\lim_{k\to+\infty} k \|x_k - x_{k-1}\| = 0$. $\qquad\square$

Remark 8. In [35], Park and Ryu established a fixed point residual lower bound of $O(\frac{1}{k})$ for various fixed point iterations designed to find a fixed point of a 1-Lipschitz continuous operator. In the following, we will explain that this statement is not in contradiction with the convergence rate statement in Theorem 7.

According to [35, Theorem 4.6], for given $K \geqslant 2$, $d \geqslant K$ and every initial point $x_0 \in \mathbb{R}^d$, there exists an 1-Lipschitz continuous operator $T : \mathbb{R}^d \to \mathbb{R}^d$ with $x_* \in \text{Fix}\,T$ such that the inequality

$$\|x_{K-1} - T(x_{K-1})\| \geqslant \frac{2\|x_0 - x_*\|}{K}, \tag{24}$$

holds for every iterates $(x_k)_{k=0,\dots,K-1}$ satisfying

$$x_k \in x_0 + \text{span}\{x_0 - T(x_0), \dots, x_{k-1} - T(x_{k-1})\} \tag{25}$$

for $k = 1, \dots, K-1$. It is evident that the sequence generated by the Fast KM algorithm with $x_1 := x_0$ fulfills (25). Consequently, there exists such an 1-Lipschitz continuous operator $T : \mathbb{R}^d \to \mathbb{R}^d$ that fulfills (24) for the sequence of iterates generated by the Fast KM algorithm with $x_1 := x_0$. In contrast, according to Theorem 7, for the same operator (as it is the case for every 1-Lipschitz continuous operator), it holds $k\|x_{k-1} - T(x_{k-1})\| \to 0$ as $k \to +\infty$, which means that there exists $k_0 > K$ such that

$$\|x_{k-1} - T(x_{k-1})\| < \frac{2\|x_0 - x_*\|}{k}$$

for every $k \geqslant k_0$.

Remark 9. In [13], Contreras and Cominetti presented an example of a 1-Lipschitz continuous operator defined on a Banach space with the property that the fixed point residual of the *Mann iteration* is bounded from below by $O(\frac{1}{k})$. For given initial points y_0 and x_0, the Mann iterates are defined for every $k \geqslant 0$ recursively as follows

$$x_{k+1} := s_0^{k+1} y_0 + \sum_{i=0}^{k} s_{i+1}^{k+1} T(x_i),$$

where $s_i^{k+1} \geqslant 0$ for every $i = 0, \dots, k+1$ and $\sum_{i=0}^{k+1} s_i^{k+1} = 1$.

As observed in Remark 1, the Fast KM algorithm is not a Mann iteration since not all the weights are nonnegative, although they do sum up to 1. This observation suggests that in order to obtain a fixed point residual rate of $o(\frac{1}{k})$ in Banach spaces, one might have to go beyond general Mann iterations and allow, for instance, negative weights in the iterative scheme.

4 Application to several splitting algorithms

In the light of the fact that fixed point methods lie at the heart of important splitting iterative schemes for monotone inclusions, we will discuss in this section how the Fast KM algorithm impacts the latter. In addition, we will review some recent acceleration approaches of splitting algorithms from the literature.

Consider the following monotone inclusion problem

$$\text{Find } x \in \mathcal{H} \text{ such that } 0 \in A(x) + B(x) + C(x), \tag{26}$$

where $A, B : \mathcal{H} \to 2^{\mathcal{H}}$ are set-valued maximally monotone operators and $C : \mathcal{H} \to \mathcal{H}$ is a β-cocoercive operator with $\beta > 0$.

Davis and Yin introduced in [16] the following operator

$$T_{DY} : \mathcal{H} \to \mathcal{H}, \quad T_{DY} := J_{\gamma A} \circ (2J_{\gamma B} - \mathrm{Id} - \gamma C \circ J_{\gamma B}) + \mathrm{Id} - J_{\gamma B}, \tag{27}$$

where $0 < \gamma \leqslant 2\beta$ and $J_{\gamma A} := (\mathrm{Id} + \gamma A)^{-1}$ denotes the *resolvent operator* of γA with constant γ. The set of zeros of $A + B + C$, denoted by $\mathrm{Zer}(A + B + C)$, can be characterized in terms of T_{DY} by (see [16, Lemma 2.2])

$$\mathrm{Zer}(A + B + C) = J_{\gamma B}(\mathsf{Fix}T_{DY}).$$

The Krasnosel'skiĭ-Mann iteration applied to T_{DY} gives rise to the *three-operator splitting method*

$$x_{k+1} := (1 - s_k) x_k + s_k T_{DY}(x_k) \quad \forall k \geqslant 0,$$

where $x_0 \in \mathcal{H}$ and $(s_k)_{k \geqslant 0} \subseteq \left(0, 2 - \frac{\gamma}{2\beta}\right]$. The operator T_{DY} is $\frac{2\beta}{4\beta - \gamma}$-averaged, here we use the convention $\beta := +\infty$ whenever $C \equiv 0$, in which case the operator is $\frac{1}{2}$-averaged. According to [23], this constant is tight. It has been shown in [14] that the convergence rate of three-operator splitting method is, as expected, of $O\left(\frac{1}{\sqrt{k}}\right)$.

The theoretical statements of the previous section applied to this particular setting lead to the following result.

Corollary 10. *Let $\alpha > 2, x_0, x_1 \in \mathcal{H}$, $0 < \gamma \leqslant 2\beta$ and $0 < s \leqslant 2 - \frac{\gamma}{2\beta}$. For every $k \geqslant 1$ we set*

$$x_{k+1} := \left(1 - \frac{s\alpha}{2(k + \alpha)}\right) x_k + \frac{(1 - s)k}{k + \alpha}(x_k - x_{k-1})$$

$$+ \frac{s\alpha}{2(k + \alpha)} T_{DY}(x_k) + \frac{sk}{k + \alpha}(T_{DY}(x_k) - T_{DY}(x_{k-1})).$$

Then the following statements are true:

(i) $(x_k)_{k \geqslant 0}$ converges weakly to an element x_* in $\mathsf{Fix}T_{DY}$ such that $J_{\gamma B}(x_*)$ is a solution of (26);

(ii) it holds

$$\|x_k - x_{k-1}\| = o\left(\frac{1}{k}\right) \qquad \text{and} \qquad \|x_{k-1} - T_{DY}(x_{k-1})\| = o\left(\frac{1}{k}\right) \text{ as } k \to +\infty.$$

In the following we will revisit some of the particular formulations of (26) and of the corresponding underlying operator T_{DY} also in order to emphasize the broad applicability of Corollary 10.

Resolvent operator. For $B \equiv C \equiv 0$, the problem (26) reduces to

$$\text{Find } x \in \mathcal{H} \text{ such that } 0 \in A(x),$$

and, for $\gamma > 0$,

$$T_{DY} = J_{\gamma A},$$

which is $\frac{1}{2}$-averaged.

The fixed point residual of the classical proximal point algorithm

$$x_{k+1} = J_{\gamma A}(x_k) \quad \forall k \geqslant 0,$$

is known to be in general of $O\left(\frac{1}{\sqrt{k}}\right)$, whereas Gu and Yang have shown in [20] that, for $\mathcal{H} := \mathbb{R}^n$, it can be tightened to

$$\|J_{\gamma A}(x_k) - x_k\| = \begin{cases} O\left(\dfrac{1}{k}\right), & \text{if } n = 1, \\[3mm] O\left(\dfrac{1}{\sqrt{\left(1 + \frac{1}{k}\right)^k k}}\right), & \text{if } n \geqslant 2. \end{cases}$$

In the same setting of finite-dimensional Hilbert spaces, Kim proposed in [25] (see also [35]) the following *accelerated proximal point method*, which, given $y_1 = x_0 = x_1 \in \mathcal{H}$ and $\gamma > 0$, reads: for every $k \geqslant 1$ set

$$\begin{aligned} y_{k+1} &:= J_{\gamma A}(x_k), \\ x_{k+1} &:= y_{k+1} + \frac{k}{k+2}(y_{k+1} - y_k) - \frac{k}{k+2}(y_k - x_{k-1}). \end{aligned} \tag{28}$$

Using the performance estimation approach, the the author proved that the method exhibits a convergence rate of the fixed point residual of $O\left(\frac{1}{k}\right)$.

By comparison, the algorithm in Corollary 10 for $\gamma > 0$, $0 < s \leqslant 2$ and $T_{DY} = J_{\gamma A}$ exhibits a convergence rate of the fixed point residual of $o\left(\frac{1}{k}\right)$.

Forward-backward operator. For $B \equiv 0$, the problem (26) reduces to

$$\text{Find } x \in \mathcal{H} \text{ such that } 0 \in A(x) + C(x),$$

and, for $0 < \gamma \leqslant 2\beta$,

$$T_{DY} := T_{FB} = J_{\gamma A} \circ (\mathrm{Id} - \gamma C),$$

which is $\frac{2\beta}{4\beta - \gamma}$-averaged.

The Krasnosel'skiĭ-Mann iteration gives rise in this case to the classical *forward-backward* algorithm, which is known to exhibit a convergence rate of the fixed point residual of $O\left(\frac{1}{\sqrt{k}}\right)$. By comparison, the algorithm in Corollary 10 for $T_{DY} = T_{FB}$ exhibits a convergence rate of the fixed point residual of $o\left(\frac{1}{k}\right)$.

Douglas-Rachford operator. For $C \equiv 0$, the problem (26) reduces to

$$\text{Find } x \in \mathcal{H} \text{ such that } 0 \in A(x) + B(x),$$

and, for $\gamma > 0$,

$$T_{DY} := T_{DR} = J_{\gamma A} \circ (2J_{\gamma B} - \mathrm{Id}) + \mathrm{Id} - J_{\gamma B},$$

which is $\frac{1}{2}$-averaged.

The Krasnosel'skiĭ-Mann iteration gives rise in this case to the classical *Douglas-Rachford* algorithm (see [17], [30]), which is known to exhibit a convergence rate of the fixed point residual of $O\left(\frac{1}{\sqrt{k}}\right)$ (see [22]). By comparison, the algorithm in Corollary 10 for $\gamma > 0$, $0 < s \leqslant 2$ and $T_{DY} = T_{DR}$ exhibits a convergence rate of the fixed point residual of $o\left(\frac{1}{k}\right)$.

Recent contributions to acceleration approaches. The idea of the Halpern iteration of considering in the iterative schemes convex combinations with an anchor point has been recently extensively exploited as it led to convergence rate improvements of numerical methods. This has been first done for fixed point iterations ([38], [29]), then for algorithms for solving monotone equations and minimax problems ([43], [27]), and later on for variants of splitting algorithms like the forward-backward, the Douglas-Rachford and the three-operator splitting method ([36, 40, 39, 44]).

Our method, however, relies on the continuous time approach from [7] and uses the idea of Nesterov's momentum updates ([33]).

5 Numerical experiments

5.1 Proximal point type methods

In order to illustrate the numerical performances of the Fast KM algorithm by comparison to other iterative schemes we consider first, for $n \geqslant 1$, the fixed point problem

$$\text{Find } x \in \mathbb{R}^{2n} \text{ such that } J_A(x) = x,$$

where $J_A : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is the resolvent of the maximally monotone operator given by the matrix

$$A = \frac{1}{M-1} \begin{pmatrix} \mathbb{O} & \mathbb{I} \\ -\mathbb{I} & \mathbb{O} \end{pmatrix} \in \mathbb{R}^{2n \times 2n},$$

where M is a positive constant, $\mathbb{I}$ and $\mathbb{O}$ denote the identity and the all-zeros matrix in $\mathbb{R}^{n \times n}$, respectively. This operator has been used in the literature to illustrate the worse-case performance of the proximal point method and of the Banach-Picard iteration (see [20, 35]). Notice that $x = 0$ is the unique fixed point of J_A and that J_A is $\frac{1}{2}$-averaged, therefore we take as step size $s := 2$.
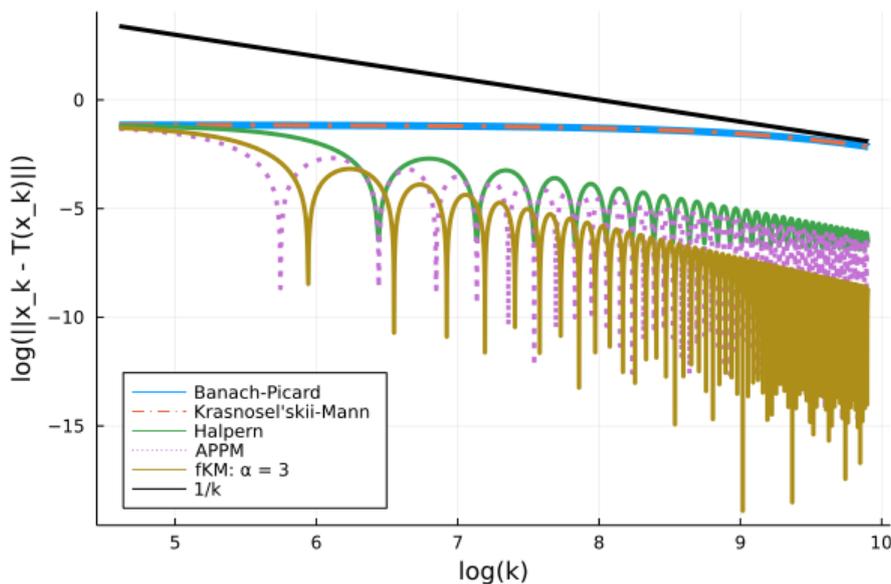

Figure 1: Comparison of the behaviour of the fixed point residual for different methods

We solve the fixed point problem with the Banach-Picard iteration, which corresponds to the proximal point algorithm, the Krasnosel'skiĭ-Mann iteration (3), which corresponds to the relaxed proximal point algorithm, the Halpern iteration (4), the accelerated proximal point method (APPM) (28), and the Fast KM algorithm (11) with $\alpha = 3$.

For all iterative methods we consider as starting $\begin{pmatrix} \mathbf{1}_n \\ \mathbf{0}_n \end{pmatrix} \in \mathbb{R}^{2n}$, where $\mathbf{1}_n$ and $\mathbf{0}_n$ denote then all ones and all zeros vector in $\mathbb{R}^n$, respectively.
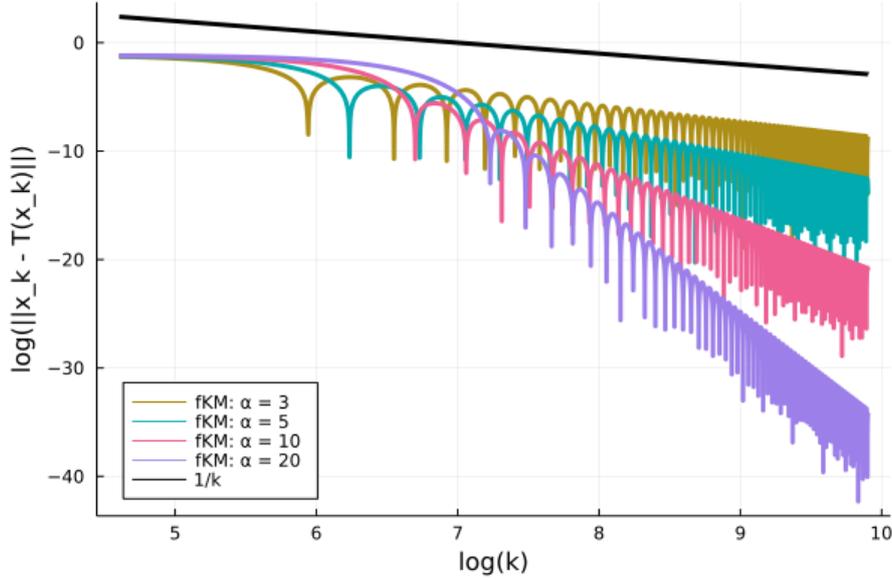
Figure 2: The parameter α influences the convergence behaviour of the Fast KM algorithm

In a first experiment, we run all these methods in case $n := 5000$. The values of the corresponding fixed point residuals are plotted in Figure 1 in logarithmic scale. It is obvious that the Fast KM algorithm outperforms all other numerical algorithms.

In a second experiment, we run the Fast KM algorithm for different values of α in $\{3, 5, 10, 20\}$. The values of the corresponding fixed point residuals are plotted in Figure 2 in logarithmic scale. As there is almost no difference between the methods in the first 100 iterations, one can notice that the speed of convergence of the fixed point residual increases with increasing α and it is consistently faster than $o\,(1/k)$. This phenomenon seems to be common for algorithms enhanced with Nesterov's momentum update (see also [7]). However, whereas in the case of Nesterov's acceleration algorithms for minimizing smooth and convex function the values of α are correlated with the speed of convergence of the objective function values, here this applies to the fixed point residual.

In addition, the plots of the fixed point residual exhibit a strong oscillatory behaviour, very similar to the behaviour of the objective function values of Nesterov's acceleration algorithms for convex minimization. This is another evidence that Nesterov's momentum improves the convergence behaviour of numerical algorithms beyond the optimization setting (see also [7]).
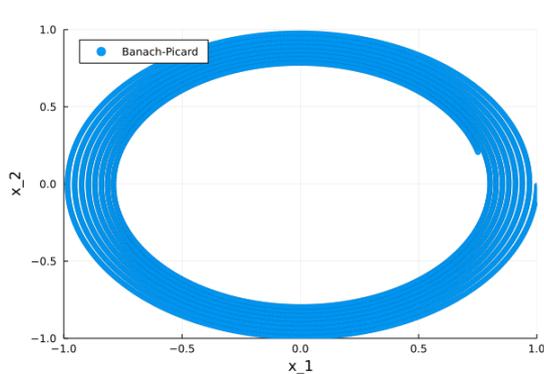

Figure 3: The trajectory of the Banach-Picard iteration/proximal point algorithm
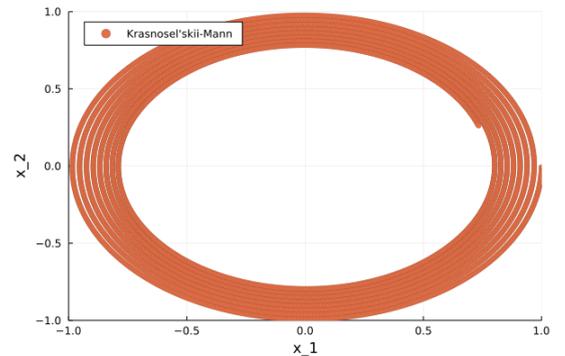

Figure 4: The trajectory of the Krasnosel'skiĭ-Mann iteration/relaxed proximal point algorithm

In the figures 3 - 10 we plot the trajectories generated by all methods considered in the numerical experiments in case $n := 1$. From the convergence analysis we know that they all converge to the unique fixed point of the operator, however, as the plots show, after spiralling around it. It is also
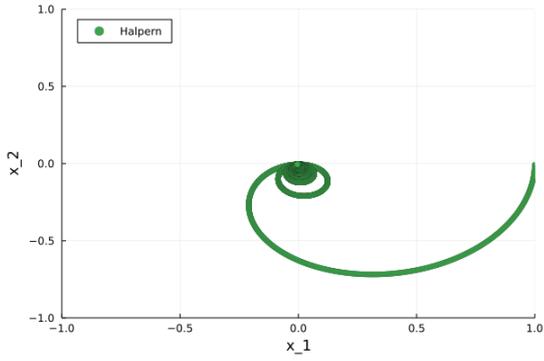
Figure 5: The trajectory of Halpern iteration
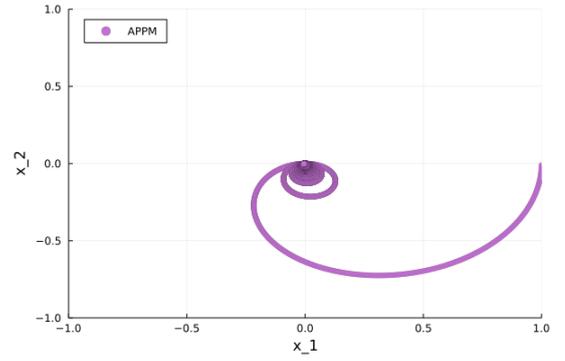
Figure 6: The trajectory of the accelerated proximal point method

obvious that the spiralling effect in case of the Fast KM algorithm is less pronounced than for the other algorithms.
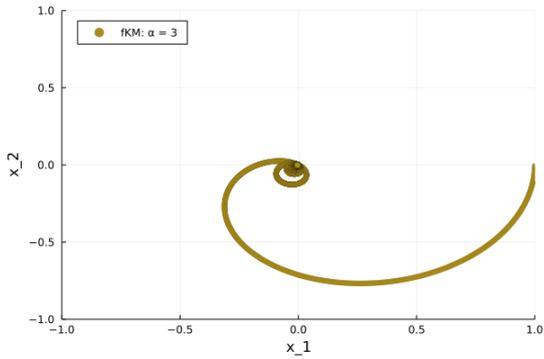

Figure 7: The trajectory of the Fast KM algorithm for $\alpha = 3$
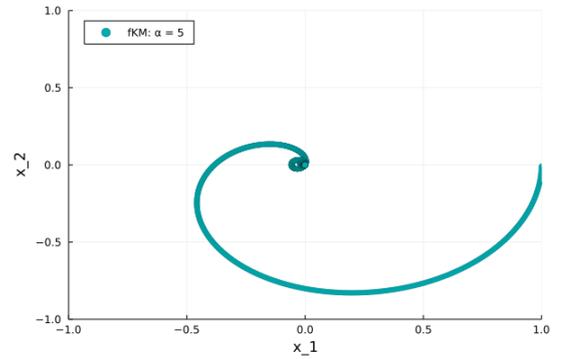
Figure 8: The trajectory of the Fast KM algorithm for $\alpha = 5$
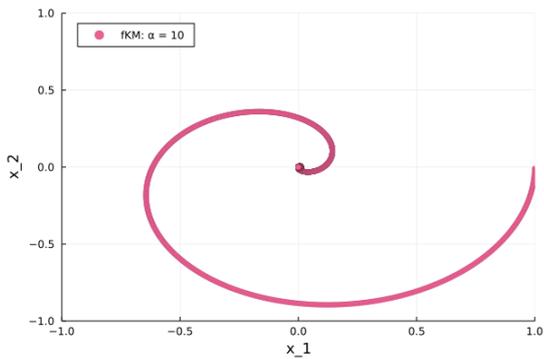
Figure 9: The trajectory of the Fast KM algorithm for $\alpha = 10$
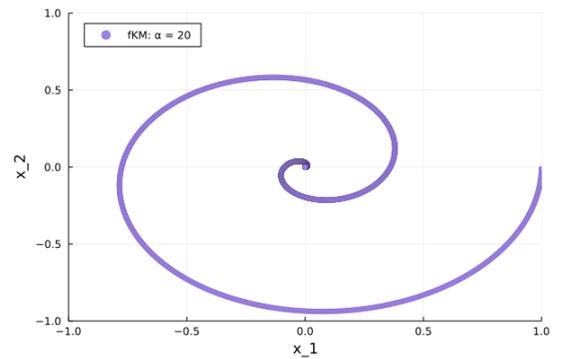
Figure 10: The trajectory of the Fast KM algorithm for $\alpha = 20$

5.2 Douglas-Rachford type methods

For the second family of numerical experiments we consider the following feasibility problem

$$\text{Find } x \in \mathbb{R}^{2n} \text{ such that } x \in \mathbb{R}^{2n}_+ \cap H_{u,\nu},$$

where $\mathbb{R}^{2n}_+$ is the set of vectors in $\mathcal{R}^{2n}$ with nonnegative entries and, for a given vector $u \in \mathbb{R}^{2n}$ and real number $\nu \in \mathbb{R}$,

$$H_{u,\nu} := \left\{ x \in \mathbb{R}^{2n} \colon \langle x, u \rangle = \nu \right\}.$$

If $\mathbb{R}^{2n}_+ \cap H_{u,\nu} \neq \varnothing$, then (see, for instance, [3, Corollary 27.6])

$$\mathbb{R}^{2n}_+ \cap H_{u,\nu} = \text{Zer}\left(N_{\mathbb{R}^{2n}_+} + N_{H_{u,\nu}}\right) \neq \varnothing,$$

where $N_D : \mathbb{R}^{2n} \rightrightarrows \mathbb{R}^{2n}$ denotes the normal cone operator of a nonempty closed convex set $D \subseteq \mathcal{R}^n$. The normal cone operator N_D is a maximally monotone and its resolvent J_{N_D} is nothing else than the projection Proj_D onto the set D.

The Douglas-Rachford (DR) algorithm is known as one of the most successful numerical method for solving such feasibility problems. In this concrete case it reads

$$x_{k+1} := (1 - s_k)x_k + s_k T_{DR}(x_k) \quad \forall k \geqslant 0,$$

where $x_0 \in \mathcal{R}^{2n}, (s_k)_{k \geqslant 0} \subseteq (0, 2]$ and $T_{DR} := \text{Proj}_{\mathbb{R}^{2n}_+} \circ (2\text{Proj}_{H_{u,\nu}} - \text{Id}) + \text{Id} - \text{Proj}_{H_{u,\nu}}$.

In the following we compare the performances of the DR algorithm for various choices for $(s_k)_{k \geqslant 0} \subseteq (0, 2]$ with the ones of the Halpern algorithm ([21]) and of the Fast KM algorithm, which make also use of the Douglas-Rachford operator T_{DR}. For the Fast KM algorithm we consider $\alpha \in \{5, 10, 30, 100, 500\}$ and as a step size $s := 2$.

In the numerical experiments we generate for different values for the dimension $n \geqslant 1$ a number of $\texttt{N}_{\texttt{test}}$ pairs $(u, \nu) \in \mathbb{R}^{2n}_+ \times \mathbb{R}_+$, such that the intersection of $\mathbb{R}^{2n}_+$ and $H_{u,\nu}$ is nonempty, and a number of $\texttt{N}_{\texttt{init}}$ normally distributed starting points $x_0 \in \mathbb{R}^{2n}$, which we scale then by 100. For each generated hyperplane $H_{u,\nu}$ and starting point x_0 we run several variants of the DR algorithm, the Fast KM method and the Halpern algorithm. The algorithms terminate either after $\texttt{k}_{\max}$ iterations or once the following condition is fulfilled

$$\left\| \text{Proj}_{H_{u,\nu}}(x_k) - \text{Proj}_{\mathbb{R}^{2n}_+}\left(\text{Proj}_{H_{u,\nu}}(x_k)\right) \right\| \leqslant \texttt{Tol}, \tag{29}$$

where $\texttt{Tol}$ denotes the tolerance error. This condition is motivated by the fact that for the Douglas-Rachford methods the so-called shadow sequence is the one that converges to a solution; in other words, (29) guarantees that $\text{Proj}_{H_{u,\nu}}(x_k)$ is close to the intersection $\mathbb{R}^{2n}_+ \cap H_{u,\nu}$. A trial fulfilling (29) before $\texttt{k}_{\max}$ iterations will be counted as a successful attempt. In the tables 1 and 2 we report the ratio of successfully solved problems and the average number of iterations the algorithms need until termination.

Table 1 shows the results for three settings determined by three choices for the triple $(n, \texttt{N}_{\texttt{test}}, \texttt{N}_{\texttt{init}})$ with $\texttt{Tol} := 10^{-16}$ for the first two and $\texttt{Tol} := 10^{-12}$ for the third one, and $\texttt{k}_{\max} := 100$. For each setting we write in boldface the best values for the ratios and the average number of iterates for the DR algorithms. It is evident that the Fast KM algorithm outperforms in both criteria the best performing variants of the Douglas-Rachford algorithm and the Halpern algorithm already for $\alpha = 30$, and even more so for larger values of α.

Table 2 shows the results obtained for the Fast KM algorithm for larger values of n, $\texttt{Tol} := 10^{-8}$ and $\texttt{k}_{\max} := 200$. It emphasizes once more that the numerical performances of our method become better when α takes larger values. Different from the class of DR algorithms, the growing dimension seems to less affect the number of iterations needed by Fast KM to provide a solution.

To further illustrate the behaviour of the considered numerical methods, we plot below some generated trajectories in case $n := 1$, and for $u = (1, 5)^T$ and $\nu := 6$. The generated sequences by the different methods may converge to different solutions for the same starting point which is plotted as a black square in the figures. However, the way the iterates of the DR algorithms and the Halpern algorithm, on the one hand, an the Fast KM algorithm, on the other hand, tend to their limits through are totally different. While the iterates of the DR algorithms and the Halpern algorithm move along a curve above the hyperplane $H_{u,\nu}$, the ones generated by the Fast KM algorithm approach in a more straight manner the solution.

In the figures 17 and 18 we plot the trajectory generated by the Halpern algorithm for two different starting points in order to emphasize their pronounced spiralling around the limit point. This also explains why (see also Figure 14), even if the algorithm finds a solution in less than $\texttt{k}_{\max}$ steps, it requires more iterations than the DR algorithms and significantly more than the Fast KM algorithm.

$(n, \mathtt{N_{test}}, \mathtt{N_{init}})$	$(1, 10^2, 10^4)$		$(5, 10^2, 10^4)$		$(50, 10^2, 10^3)$	
method	**ratio**	**iterations**	**ratio**	**iterations**	**ratio**	**iterations**
DR : $s_k = 1 - \frac{1}{k+2}$	0.9939	7.6181 ± 6.26	0.9887	14.5431 ± 12.54	0.8456	42.823 ± 20.99
DR : $s_k \equiv 1$	0.9940	$\mathbf{4.8877 \pm 6.00}$	0.9907	$\mathbf{11.6671 \pm 12.17}$	0.8700	39.1163 ± 21.08
DR : $s_k = 1 + \frac{1}{k+2}$	0.9944	5.9398 ± 6.23	0.9923	12.0513 ± 10.52	0.8949	36.6431 ± 20.41
DR : $s_k \equiv \frac{7}{5}$	0.9976	6.8541 ± 6.61	0.9952	13.291 ± 7.85	0.9428	$\mathbf{33.5245 \pm 17.25}$
DR : $s_k \equiv \frac{3}{2}$	0.9980	7.6111 ± 6.54	0.9957	15.0055 ± 7.06	0.9514	34.1237 ± 15.8
DR : $s_k \equiv \frac{7}{4}$	0.9992	12.1969 ± 7.59	0.9965	27.535 ± 5.69	0.9644	46.8346 ± 10.99
DR : $s_k = \frac{9}{5} - \frac{1}{k+2}$	0.9992	9.3479 ± 6.96	$\mathbf{0.9968}$	23.3639 ± 6.36	0.9648	47.6916 ± 11.91
DR : $s_k \equiv \frac{9}{5}$	0.9994	14.5185 ± 8.64	0.9966	34.2205 ± 6.08	$\mathbf{0.9649}$	54.9974 ± 10.09
DR : $s_k = \frac{9}{5} + \frac{1}{k+2}$	$\mathbf{0.9996}$	23.1938 ± 11.67	0.9963	47.9954 ± 6.10	0.9598	64.4231 ± 7.85
Halpern	0.2400	32.8750 ± 22.87	0.0000	$-//-$	0.0000	$-//-$
Fast KM : $\alpha = 5$	0.9690	22.4536 ± 18.26	0.4973	70.4548 ± 16.42	0.0000	$-//-$
Fast KM : $\alpha = 10$	1.0000	9.7566 ± 6.83	0.9996	27.5816 ± 11.78	0.8753	65.5713 ± 14.46
Fast KM : $\alpha = 30$	1.0000	4.9323 ± 2.23	1.0000	10.0186 ± 2.41	1.0000	17.6134 ± 3.3
Fast KM : $\alpha = 100$	1.0000	3.5014 ± 1.35	1.0000	6.2383 ± 1.19	1.0000	9.5427 ± 1.43
Fast KM : $\alpha = 500$	1.0000	2.6151 ± 1.00	1.0000	4.3118 ± 0.73	1.0000	6.2944 ± 0.75

Table 1: The ratio of successfully solved problems and the average number of iterations for several algorithms

$(n, \mathtt{N_{test}}, \mathtt{N_{init}})$	$(500; 100; 500)$		$(5000, 50, 100)$	
method	**ratio**	**iterations**	**ratio**	**iterations**
Fast KM : $\alpha = 10$	0.6076	154.5383 ± 24.31	0.0000	$-//-$
Fast KM : $\alpha = 30$	1.0000	29.3096 ± 5.14	1.0000	40.7248 ± 3.56
Fast KM : $\alpha = 100$	1.0000	13.8564 ± 1.77	1.0000	17.4264 ± 1.18
Fast KM : $\alpha = 500$	1.0000	8.5773 ± 0.94	1.0000	10.282 ± 0.70

Table 2: The ratio of successfully solved problems and the average number of iterations for the Fast KM algorithm

A   Appendix

In the appendix, we have compiled some auxiliary results and provided the proofs of the two technical lemmas used in the convergence analysis of the Fast KM algorithm.

A.1   Auxiliary results

The Opial Lemma ([34]) is used in the proof of the convergence of the iterates.
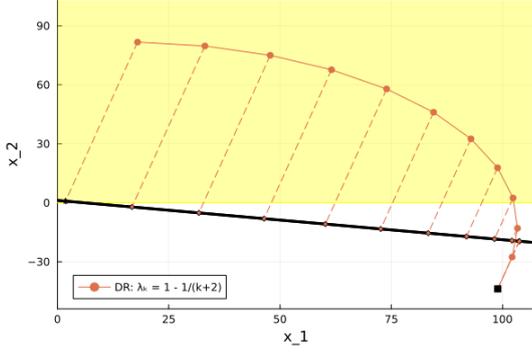
Figure 11: The trajectory of the Douglas-Rachford algorithm for $s_k := 1 - \frac{1}{k+2}$
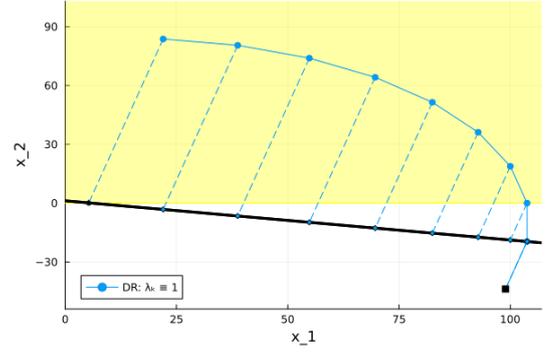

Figure 12: The trajectory of the Douglas-Rachford algorithm for $s_k := 1$
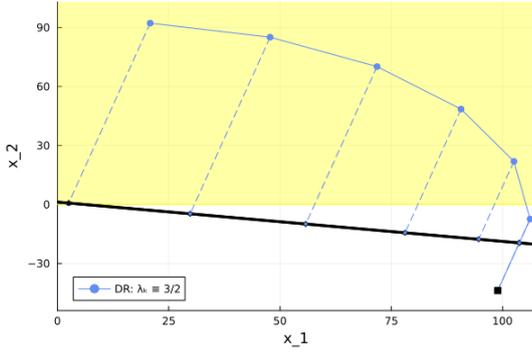

Figure 13: The trajectory of the Douglas-Rachford algorithm for $s_k := \frac{3}{2}$
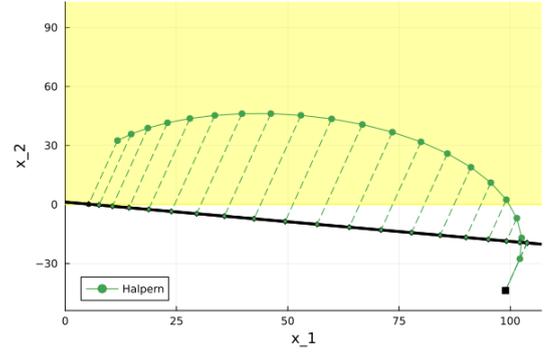
Figure 14: The trajectory of the Halpern algorithm
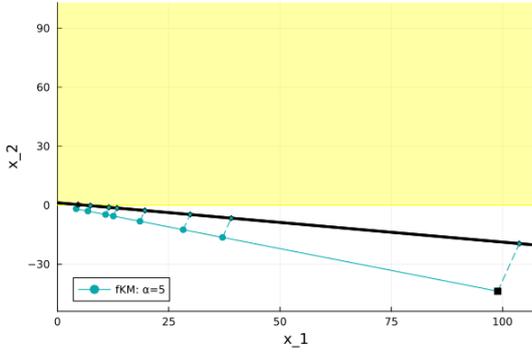
Figure 15: The trajectory of the Fast KM algorithm for $\alpha = 5$
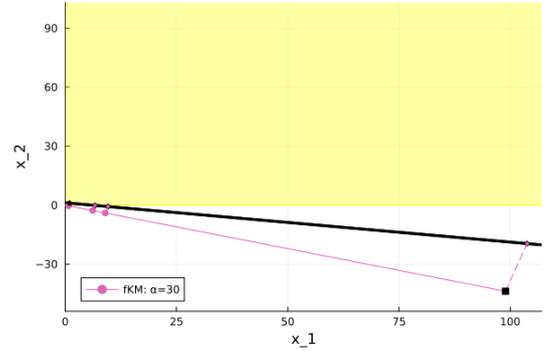
Figure 16: The trajectory of the Fast KM algorithm for $\alpha = 30$

Lemma 11. *Let $\mathcal{S}$ be a nonempty subset of $\mathcal{H}$ and $(x_k)_{k \geqslant 0}$ be a sequence in $\mathcal{H}$. Assume that*

(i) *for every $x_* \in \mathcal{S}$, $\lim\limits_{k \to +\infty} \|x_k - x_*\|$ exists;*

(ii) *every weak sequential cluster point of the sequence $(x_k)_{k \geqslant 0}$ as $k \to +\infty$ belongs to $\mathcal{S}$.*

Then $(x_k)_{k \geqslant 0}$ converges weakly to a point in $\mathcal{S}$ as $k \to +\infty$.

For the proof of the following result, which is the discrete counterpart of [1, Lemma A.2], we refer to [7, Lemma 21].

Lemma 12. *Let $a \geqslant 1$ and $(q_k)_{k \geqslant 0}$ be a bounded sequence in $\mathcal{H}$ such that*

$$\lim_{k \to +\infty} \left(q_{k+1} + \frac{k}{a} (q_{k+1} - q_k) \right) = l \in \mathcal{H}.$$

17
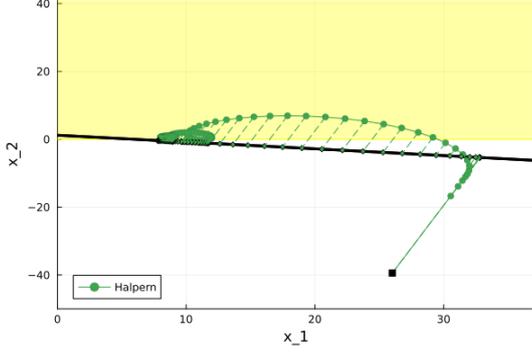
Figure 17: An instance for which the Halpern algorithm does not terminates in less than $\mathtt{k}_{\max}$ iterations
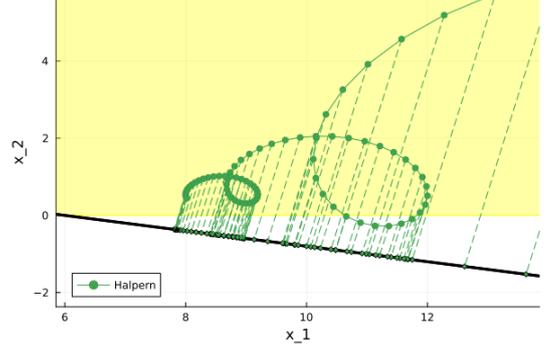
Figure 18: The spiral behavior of the Halpern algorithm

Then it holds $\lim_{k \to +\infty} q_k = l$.

The following result is a particular instance of [3, Lemma 5.31].

Lemma 13. *Let* $(a_k)_{k \geqslant 1}$, $(b_k)_{k \geqslant 1}$ *and* $(d_k)_{k \geqslant 1}$ *be sequences of real numbers. Assume that* $(a_k)_{k \geqslant 1}$ *is bounded from below, and* $(b_k)_{k \geqslant 1}$ *and* $(d_k)_{k \geqslant 1}$ *are nonnegative sequences such that* $\sum_{k \geqslant 1} d_k < +\infty$. *If*

$$a_{k+1} \leqslant a_k - b_k + d_k \quad \forall k \geqslant 1,$$

then the following statements are true:

(i) *the sequence* $(b_k)_{k \geqslant 1}$ *is summable, namely* $\sum_{k \geqslant 1} b_k < +\infty$;

(ii) *the sequence* $(a_k)_{k \geqslant 1}$ *is convergent.*

The following elementary result is used several times in the paper.

Lemma 14. *Let* $a, b, c \in \mathbb{R}$ *be such that* $a < 0$ *and* $b^2 - ac \leqslant 0$. *Then it holds*

$$a \|x\|^2 + 2b \langle x, y \rangle + c \|y\|^2 \leqslant 0 \quad \forall x, y \in \mathcal{H}.$$

A.2 Proofs of the technical lemmas used in the analysis of the Fast KM algorithm

In this subsection we provide the proofs of Lemma 3 and Lemma 4.

Proof of Lemma 3 . (i) Let $0 \leqslant \lambda \leqslant \alpha - 1$. First we will show that for every $k \geqslant 1$ the following identity holds

$$
\begin{aligned}
&\left(\mathcal{E}_{\lambda,k+1} + \frac{1}{4(\alpha-1)} (\alpha-2) \alpha s^2 (k+1) \|(\mathrm{Id} - T)(x_k)\|^2 \right) \\
&\quad - \left(\mathcal{E}_{\lambda,k} + \frac{1}{4(\alpha-1)} (\alpha-2) \alpha s^2 k \|(\mathrm{Id} - T)(x_{k-1})\|^2 \right) \\
&= 2(2-\alpha) \lambda s \langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k) \rangle + 2(\lambda+1-\alpha)(2k+\alpha+1) \|x_{k+1} - x_k\|^2 \\
&\quad + \frac{1}{4(\alpha-1)} (2-\alpha) s^2 \left(2(3\alpha-2)k + 2\alpha^2 + \alpha - 2 \right) \|(\mathrm{Id} - T)(x_k)\|^2 \\
&\quad + \frac{1}{\alpha-1} \Big(4(\alpha-1)(\lambda+1-\alpha) + \alpha(2-\alpha) \Big) sk \langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k) \rangle \\
&\quad + \frac{1}{\alpha-1} \left(2\alpha(\alpha-1)(\lambda+1-\alpha) + \alpha - 2(\alpha-1)^2 \right) s \langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k) \rangle \\
&\quad + \frac{1}{\alpha-1} (2-\alpha) s (k+\alpha) k \langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1}) \rangle \\
&\quad + \frac{1}{4(\alpha-1)} (2-\alpha) s^2 k \big(2k + \alpha \big) \|(\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\|^2 .
\end{aligned}
\tag{30}
$$

18

For brevity we denote for every $k \geqslant 0$

$$u_{\lambda,k+1} := 2\lambda \left(x_{k+1} - x_*\right) + 2\left(k+1\right)\left(x_{k+1} - x_k\right) + \frac{1}{2\left(\alpha - 1\right)}\left(3\alpha - 2\right)s\left(k+1\right)\left(\mathrm{Id} - T\right)\left(x_k\right), \quad (31)$$

which means that for every $k \geqslant 1$ it holds

$$u_{\lambda,k} = 2\lambda \left(x_k - x_*\right) + 2k\left(x_k - x_{k-1}\right) + \frac{1}{2\left(\alpha - 1\right)}\left(3\alpha - 2\right)sk\left(\mathrm{Id} - T\right)\left(x_{k-1}\right). \quad (32)$$

Subtracting (32) from (31) and then using (10) we obtain for every $k \geqslant 1$

$$u_{\lambda,k+1} - u_{\lambda,k}$$
$$= 2\left(\lambda + 1 - \alpha\right)\left(x_{k+1} - x_k\right) + 2\left(k + \alpha\right)\left(x_{k+1} - x_k\right) - 2k\left(x_k - x_{k-1}\right)$$
$$+ \frac{1}{2\left(\alpha - 1\right)}\left(3\alpha - 2\right)s\left(\mathrm{Id} - T\right)\left(x_k\right) + \frac{1}{2\left(\alpha - 1\right)}\left(3\alpha - 2\right)sk\Big(\left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\Big)$$
$$= 2\left(\lambda + 1 - \alpha\right)\left(x_{k+1} - x_k\right) + \frac{1}{2\left(\alpha - 1\right)}\left(\alpha - 2\left(\alpha - 1\right)^2\right)s\left(\mathrm{Id} - T\right)\left(x_k\right)$$
$$+ \frac{1}{2\left(\alpha - 1\right)}\left(2 - \alpha\right)sk\Big(\left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\Big). \quad (33)$$

In the following we want to use the identity

$$\frac{1}{2}\left(\|u_{\lambda,k+1}\|^2 - \|u_{\lambda,k}\|^2\right) = \langle u_{\lambda,k+1}, u_{\lambda,k+1} - u_{\lambda,k}\rangle - \frac{1}{2}\|u_{\lambda,k+1} - u_{\lambda,k}\|^2 \quad \forall k \geqslant 1. \quad (34)$$

Thanks to the relations (31) and (33), we derive for every $k \geqslant 1$ that

$$\langle u_{\lambda,k+1}, u_{\lambda,k+1} - u_{\lambda,k}\rangle$$
$$= 4\lambda\left(\lambda + 1 - \alpha\right)\langle x_{k+1} - x_*, x_{k+1} - x_k\rangle + \frac{1}{\alpha - 1}\left(\alpha - 2\left(\alpha - 1\right)^2\right)\lambda s\langle x_{k+1} - x_*, \left(\mathrm{Id} - T\right)\left(x_k\right)\rangle$$
$$+ \frac{1}{\alpha - 1}\left(2 - \alpha\right)\lambda sk\langle x_{k+1} - x_*, \left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\rangle$$
$$+ 4\left(\lambda + 1 - \alpha\right)\left(k+1\right)\|x_{k+1} - x_k\|^2$$
$$+ \frac{1}{\alpha - 1}\left(\alpha - 2\left(\alpha - 1\right)^2 + \left(3\alpha - 2\right)\left(\lambda + 1 - \alpha\right)\right)s\left(k+1\right)\langle x_{k+1} - x_k, \left(\mathrm{Id} - T\right)\left(x_k\right)\rangle$$
$$+ \frac{1}{\alpha - 1}\left(2 - \alpha\right)s\left(k+1\right)k\langle x_{k+1} - x_k, \left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\rangle$$
$$+ \frac{1}{4\left(\alpha - 1\right)^2}\left(3\alpha - 2\right)\left(\alpha - 2\left(\alpha - 1\right)^2\right)s^2\left(k+1\right)\|\left(\mathrm{Id} - T\right)\left(x_k\right)\|^2$$
$$+ \frac{1}{4\left(\alpha - 1\right)^2}\left(3\alpha - 2\right)\left(2 - \alpha\right)s^2\left(k+1\right)k\langle\left(\mathrm{Id} - T\right)\left(x_k\right), \left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\rangle, \quad (35)$$

and

$$-\frac{1}{2}\|u_{\lambda,k+1} - u_{\lambda,k}\|^2$$
$$= -2\left(\lambda + 1 - \alpha\right)^2\|x_{k+1} - x_k\|^2 - \frac{1}{8\left(\alpha - 1\right)^2}\left(\alpha - 2\left(\alpha - 1\right)^2\right)^2 s^2\|\left(\mathrm{Id} - T\right)\left(x_k\right)\|^2$$
$$- \frac{1}{8\left(\alpha - 1\right)^2}\left(2 - \alpha\right)^2 s^2 k^2\|\left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\|^2$$
$$- \frac{1}{\alpha - 1}\left(\alpha - 2\left(\alpha - 1\right)^2\right)\left(\lambda + 1 - \alpha\right)s\langle x_{k+1} - x_k, \left(\mathrm{Id} - T\right)\left(x_k\right)\rangle$$
$$- \frac{1}{\alpha - 1}\left(2 - \alpha\right)\left(\lambda + 1 - \alpha\right)sk\langle x_{k+1} - x_k, \left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\rangle$$
$$- \frac{1}{4\left(\alpha - 1\right)^2}\left(\alpha - 2\left(\alpha - 1\right)^2\right)\left(2 - \alpha\right)s^2 k\langle\left(\mathrm{Id} - T\right)\left(x_k\right), \left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\rangle. \quad (36)$$

19

A direct computation shows that for every $k \geqslant 0$

$$\left((3\alpha - 2)(\lambda + 1 - \alpha) + \alpha - 2(\alpha - 1)^2\right)(k+1) - (\lambda + 1 - \alpha)\left(\alpha - 2(\alpha - 1)^2\right)$$

$$= \left((3\alpha - 2)(\lambda + 1 - \alpha) + \alpha - 2(\alpha - 1)^2\right)k + 2\alpha(\alpha - 1)(\lambda + 1 - \alpha) + \alpha - 2(\alpha - 1)^2$$

$$= \left((3\alpha - 2)(\lambda + 1 - \alpha) + \alpha - 2(\alpha - 1)^2 - (2 - \alpha)\lambda\right)k + (2 - \alpha)\lambda k$$
$$\quad + 2\alpha(\alpha - 1)(\lambda + 1 - \alpha) + \alpha - 2(\alpha - 1)^2$$

$$= \left(4(\alpha - 1)(\lambda + 1 - \alpha) + \alpha(2 - \alpha)\right)k + (2 - \alpha)\lambda k + 2\alpha(\alpha - 1)(\lambda + 1 - \alpha) + \alpha - 2(\alpha - 1)^2.$$

Therefore, by plugging (35) and (36) into (34), we get for every $k \geqslant 1$

$$\frac{1}{2}\left(\|u_{\lambda,k+1}\|^2 - \|u_{\lambda,k}\|^2\right)$$

$$= 4\lambda(\lambda + 1 - \alpha)\langle x_{k+1} - x_*, x_{k+1} - x_k\rangle + \frac{1}{\alpha - 1}\left(\alpha - 2(\alpha - 1)^2\right)\lambda s\langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k)\rangle$$

$$\quad + \frac{1}{\alpha - 1}(2 - \alpha)\lambda s k\langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\rangle$$

$$\quad + 2(\lambda + 1 - \alpha)(2k + \alpha + 1 - \lambda)\|x_{k+1} - x_k\|^2$$

$$\quad + \frac{1}{8(\alpha - 1)^2}\left(\alpha - 2(\alpha - 1)^2\right)s^2\left(2(3\alpha - 2)k + 2\alpha^2 + \alpha - 2\right)\|(\mathrm{Id} - T)(x_k)\|^2$$

$$\quad + \frac{1}{\alpha - 1}\left(4(\alpha - 1)(\lambda + 1 - \alpha) + \alpha(2 - \alpha) + (2 - \alpha)\lambda\right)s k\langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k)\rangle$$

$$\quad + \frac{1}{\alpha - 1}\left(2\alpha(\alpha - 1)(\lambda + 1 - \alpha) + \alpha - 2(\alpha - 1)^2\right)s\langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k)\rangle$$

$$\quad + \frac{1}{\alpha - 1}(2 - \alpha)s(k + \alpha - \lambda)k\langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\rangle$$

$$\quad + \frac{1}{4(\alpha - 1)^2}(2 - \alpha)s^2\left((3\alpha - 2)k + 2(\alpha - 1)\alpha\right)k\langle(\mathrm{Id} - T)(x_k), (\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\rangle$$

$$\quad - \frac{1}{8(\alpha - 1)^2}(2 - \alpha)^2 s^2 k^2\|(\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\|^2. \tag{37}$$

Notice that by the definition of the discrete energy function we have for every $k \geqslant 1$

$$\mathcal{E}_{\lambda,k} + \frac{1}{4(\alpha - 1)}(\alpha - 2)\alpha s^2 k\|(\mathrm{Id} - T)(x_{k-1})\|^2$$

$$= \frac{1}{2}\|u_{\lambda,k}\|^2 + 2\lambda(\alpha - 1 - \lambda)\|x_k - x_*\|^2 + \frac{1}{\alpha - 1}(\alpha - 2)\lambda s k\langle x_k - x_*, (\mathrm{Id} - T)(x_{k-1})\rangle$$

$$\quad + \frac{1}{8(\alpha - 1)^2}(\alpha - 2)s^2 k\left((3\alpha - 2)k + 2(\alpha - 1)\alpha\right)\|(\mathrm{Id} - T)(x_{k-1})\|^2. \tag{38}$$

Later we will subtract the above identity at consecutive indices and to this end we will make use of the following identities which hold for every $k \geqslant 1$:

$$\|x_{k+1} - x_*\|^2 - \|x_k - x_*\|^2 = 2\langle x_{k+1} - x_*, x_{k+1} - x_k\rangle - \|x_{k+1} - x_k\|^2, \tag{39}$$

$$2\lambda s(k+1)\langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k)\rangle - 2\lambda s k\langle x_k - x_*, (\mathrm{Id} - T)(x_{k-1})\rangle$$

$$= 2\lambda s\langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k)\rangle + 2\lambda s k\left(\langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k)\rangle - \langle x_k - x_*, (\mathrm{Id} - T)(x_{k-1})\rangle\right)$$

$$= 2\lambda s\langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k)\rangle + 2\lambda s k\langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\rangle$$
$$\quad + 2\lambda s k\langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_{k-1})\rangle$$

$$= 2\lambda s\langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k)\rangle + 2\lambda s k\langle x_{k+1} - x_*, (\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\rangle$$
$$\quad - 2\lambda s k\langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k) - (\mathrm{Id} - T)(x_{k-1})\rangle + 2\lambda s k\langle x_{k+1} - x_k, (\mathrm{Id} - T)(x_k)\rangle, \tag{40}$$

and

$$\frac{1}{4\,(\alpha-1)}s^2\,(k+1)\,((3\alpha-2)\,(k+1)+2\,(\alpha-1)\,\alpha)\,\|(\mathrm{Id}-T)\,(x_k)\|^2$$

$$-\frac{1}{4\,(\alpha-1)}s^2k\big((3\alpha-2)\,k+2\,(\alpha-1)\,\alpha\big)\,\|(\mathrm{Id}-T)\,(x_{k-1})\|^2$$

$$=\frac{1}{4\,(\alpha-1)}s^2\big(2\,(3\alpha-2)\,k+2\alpha^2+\alpha-2\big)\,\|(\mathrm{Id}-T)\,(x_k)\|^2$$

$$+\frac{1}{4\,(\alpha-1)}s^2k\,((3\alpha-2)\,k+2\,(\alpha-1)\,\alpha)\left(\|(\mathrm{Id}-T)\,(x_k)\|^2-\|(\mathrm{Id}-T)\,(x_{k-1})\|^2\right)$$

$$=\frac{1}{4\,(\alpha-1)}s^2\big(2\,(3\alpha-2)\,k+2\alpha^2+\alpha-2\big)\,\|(\mathrm{Id}-T)\,(x_k)\|^2$$

$$+\frac{1}{2\,(\alpha-1)}s^2k\big((3\alpha-2)\,k+2\,(\alpha-1)\,\alpha\big)\,\langle(\mathrm{Id}-T)\,(x_k)\,,(\mathrm{Id}-T)\,(x_k)-(\mathrm{Id}-T)\,(x_{k-1})\rangle$$

$$-\frac{1}{4\,(\alpha-1)}s^2k\big((3\alpha-2)\,k+2\,(\alpha-1)\,\alpha\big)\,\|(\mathrm{Id}-T)\,(x_k)-(\mathrm{Id}-T)\,(x_{k-1})\|^2\,. \tag{41}$$

Therefore, by multiplying (40) and (41) by $\frac{1}{2(\alpha-1)}\,(\alpha-2)$, relation (38) gives for every $k\geqslant 1$

$$\left(\mathcal{E}_{\lambda,k+1}+\frac{1}{4\,(\alpha-1)}\,(\alpha-2)\,\alpha s^2\,(k+1)\,\|(\mathrm{Id}-T)\,(x_k)\|^2\right)$$

$$-\left(\mathcal{E}_{\lambda,k}+\frac{1}{4\,(\alpha-1)}\,(\alpha-2)\,\alpha s^2k\,\|(\mathrm{Id}-T)\,(x_{k-1})\|^2\right)$$

$$=\frac{1}{2}\left(\|u_{\lambda,k+1}\|^2-\|u_{\lambda,k}\|^2\right)+2\lambda(\alpha-1-\lambda)\left(\|x_{k+1}-x_*\|^2-\|x_k-x_*\|^2\right)$$

$$+\frac{1}{\alpha-1}\,(\alpha-2)\,\lambda s\,\langle x_{k+1}-x_*,(\mathrm{Id}-T)\,(x_k)\rangle$$

$$+\frac{1}{\alpha-1}\,(\alpha-2)\,\lambda sk\,\langle x_{k+1}-x_*,(\mathrm{Id}-T)\,(x_k)-(\mathrm{Id}-T)\,(x_{k-1})\rangle$$

$$-\frac{1}{\alpha-1}\,(\alpha-2)\,\lambda sk\,\langle x_{k+1}-x_k,(\mathrm{Id}-T)\,(x_k)-(\mathrm{Id}-T)\,(x_{k-1})\rangle$$

$$+\frac{1}{\alpha-1}\,(\alpha-2)\,\lambda sk\,\langle x_{k+1}-x_k,(\mathrm{Id}-T)\,(x_k)\rangle$$

$$+\frac{1}{8\,(\alpha-1)^2}\,(\alpha-2)\,s^2\big(2\,(3\alpha-2)\,k+2\alpha^2+\alpha-2\big)\,\|(\mathrm{Id}-T)\,(x_k)\|^2$$

$$+\frac{1}{4\,(\alpha-1)^2}\,(\alpha-2)\,s^2k\big((3\alpha-2)\,k+2\,(\alpha-1)\,\alpha\big)\,\langle(\mathrm{Id}-T)\,(x_k)\,,(\mathrm{Id}-T)\,(x_k)-(\mathrm{Id}-T)\,(x_{k-1})\rangle$$

$$-\frac{1}{8\,(\alpha-1)^2}\,(\alpha-2)\,s^2k\big((3\alpha-2)\,k+2\,(\alpha-1)\,\alpha\big)\,\|(\mathrm{Id}-T)\,(x_k)-(\mathrm{Id}-T)\,(x_{k-1})\|^2$$

By multiplying (39) by $2\lambda(\alpha-1-\lambda)$ and by taking into consideration (37) and that

$$\alpha-2\,(\alpha-1)^2+\alpha-2=2\,(\alpha-1)\,(2-\alpha)\ \text{ and }\ -(\alpha-2)^2-(\alpha-2)\,(3\alpha-2)=-4\,(\alpha-2)\,(\alpha-1)\,,$$

we immediately obtain from here identity (30).

Next we will focus on the term $\langle x_{k+1}-x_k,(\mathrm{Id}-T)\,(x_k)-(\mathrm{Id}-T)\,(x_{k-1})\rangle$ for which we will provide an upper bound by exploiting the cocoercivity of $\mathrm{Id}-T$. Precisely, the relations (5) and (10)

guarantee that for every $k \geqslant 1$

$$
\begin{aligned}
&- 2sk\left(k + \alpha\right)\left\langle x_{k+1} - x_k, \left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\rangle \\
&= - 2sk^2\left\langle x_k - x_{k-1}, \left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\rangle + 2s^2k^2\left\|\left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 \\
&\quad + \alpha s^2 k\left\langle\left(\mathrm{Id} - T\right)\left(x_k\right), \left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\rangle \\
&\leqslant \left(2s - \frac{1}{\theta}\right)sk^2\left\|\left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 + \frac{1}{2}\alpha s^2 k\left\|\left(\mathrm{Id} - T\right)\left(x_k\right)\right\|^2 \\
&\quad + \frac{1}{2}\alpha s^2 k\left\|\left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 - \frac{1}{2}\alpha s^2 k\left\|\left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 \\
&\leqslant \left(\left(2s - \frac{1}{\theta}\right)sk^2 + \frac{1}{2}\alpha s^2 k\right)\left\|\left(\mathrm{Id} - T\right)\left(x_k\right) - \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 \\
&\quad + \frac{1}{2}\alpha s^2\left(k + 1\right)\left\|\left(\mathrm{Id} - T\right)\left(x_k\right)\right\|^2 - \frac{1}{2}\alpha s^2 k\left\|\left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2.
\end{aligned}
$$

After multiplying this inequality by $\frac{1}{2(\alpha-1)}\left(\alpha - 2\right) > 0$, adding it to (30), and using that

$$
\begin{aligned}
&2\left(2 - \alpha\right)\lambda s\left\langle x_{k+1} - x_*, \left(\mathrm{Id} - T\right)\left(x_k\right)\right\rangle \\
&= 2\left(2 - \alpha\right)\lambda s\left\langle x_{k+1} - x_k, \left(\mathrm{Id} - T\right)\left(x_k\right)\right\rangle + 2\left(2 - \alpha\right)\lambda s\left\langle x_k - x_*, \left(\mathrm{Id} - T\right)\left(x_k\right)\right\rangle,
\end{aligned}
$$

we deduce the desired inequality (14). In order to obtain the coefficients of $\left\|x_{k+1} - x_k\right\|^2$ and $\left\|\left(\mathrm{Id} - T\right)\left(x_k\right)\right\|^2$ as given in (15), one also has to take into consideration that $\lambda + 1 - \alpha \leqslant 0$ and $2\alpha^2 + \alpha - 2 > 0$, as $\alpha > 2$. The assumptions we made on α and λ immediately imply that ω_1, ω_2 and ω_4 are nonpositive numbers.

(ii) Since

$$
\begin{aligned}
&\frac{1}{\alpha - 1}\left(\alpha - 2\right)\lambda sk\left\langle x_k - x_*, \left(\mathrm{Id} - T\right)\left(x_k\right)\right\rangle + \frac{1}{8\left(\alpha - 1\right)^2}\left(\alpha - 2\right)\left(3\alpha - 2\right)s^2 k^2\left\|\left(\mathrm{Id} - T\right)\left(x_k\right)\right\|^2 \\
&= \frac{1}{3\alpha - 2}\left(\alpha - 2\right)\left(\frac{1}{\alpha - 1}\left(3\alpha - 2\right)\lambda sk\left\langle x_k - x_*, \left(\mathrm{Id} - T\right)\left(x_k\right)\right\rangle\right. \\
&\hspace{6cm}\left. + \frac{1}{8\left(\alpha - 1\right)^2}\left(3\alpha - 2\right)^2 s^2 k^2\left\|\left(\mathrm{Id} - T\right)\left(x_k\right)\right\|^2\right) \\
&= \frac{1}{3\alpha - 2}\left(\alpha - 2\right)\left(\frac{1}{2}\left\|2\lambda\left(x_k - x_*\right) + \frac{1}{2\left(\alpha - 1\right)}\left(3\alpha - 2\right)sk\left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 - 2\lambda^2\left\|x_k - x_*\right\|^2\right),
\end{aligned}
$$

we deduce that for every $k \geqslant 1$

$$
\begin{aligned}
\mathcal{E}_{\lambda,k} &= \frac{1}{2}\left\|2\lambda\left(x_k - x_*\right) + 2k\left(x_k - x_{k-1}\right) + \frac{1}{2\left(\alpha - 1\right)}\left(3\alpha - 2\right)sk\left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 \\
&\quad + 2\lambda\left(\alpha - 1 - \lambda\right)\left\|x_k - x_*\right\|^2 + \frac{1}{\alpha - 1}\left(\alpha - 2\right)\lambda sk\left\langle x_k - x_*, \left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\rangle \\
&\quad + \frac{1}{8\left(\alpha - 1\right)^2}\left(\alpha - 2\right)\left(3\alpha - 2\right)s^2 k^2\left\|\left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 \\
&= \frac{1}{2}\left\|2\lambda\left(x_k - x_*\right) + 2k\left(x_k - x_{k-1}\right) + \frac{1}{2\left(\alpha - 1\right)}\left(3\alpha - 2\right)sk\left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 \\
&\quad + \frac{1}{2\left(3\alpha - 2\right)}\left(\alpha - 2\right)\left\|2\lambda\left(x_k - x_*\right) + \frac{1}{2\left(\alpha - 1\right)}\left(3\alpha - 2\right)sk\left(\mathrm{Id} - T\right)\left(x_{k-1}\right)\right\|^2 \\
&\quad + 2\lambda\left(\alpha - 1\right)\left(1 - \frac{4\lambda}{3\alpha - 2}\right)\left\|x_k - x_*\right\|^2.
\end{aligned}
$$

Using the identity

$$
\|x\|^2 + \|y\|^2 = \frac{1}{2}\left(\|x + y\|^2 + \|x - y\|^2\right) \quad \forall x, y \in \mathcal{H},
$$

we obtain for every $k \geqslant 1$

$$
\mathcal{E}_{\lambda,k} = \frac{\alpha}{3\alpha - 2} \left\| 2\lambda \left(x_k - x_*\right) + 2k\left(x_k - x_{k-1}\right) + \frac{1}{2\left(\alpha - 1\right)} \left(3\alpha - 2\right) sk \left(\mathrm{Id} - T\right)\left(x_{k-1}\right) \right\|^2
$$

$$
+ \frac{1}{4\left(3\alpha - 2\right)} \left(\alpha - 2\right) \left\| 4\lambda \left(x_k - x_*\right) + 2k\left(x_k - x_{k-1}\right) + \frac{1}{\alpha - 1} \left(3\alpha - 2\right) sk \left(\mathrm{Id} - T\right)\left(x_{k-1}\right) \right\|^2
$$

$$
+ \frac{1}{3\alpha - 2} \left(\alpha - 2\right) k^2 \left\| x_k - x_{k-1} \right\|^2 + 2\lambda \left(\alpha - 1\right) \left(1 - \frac{4\lambda}{3\alpha - 2}\right) \left\| x_k - x_* \right\|^2, \tag{42}
$$

This shows that for $0 \leqslant \lambda \leqslant \frac{3\alpha}{4} - \frac{1}{2}$ all terms in the expression (42) are nonnegative, thus the sequence $(\mathcal{E}_{\lambda,k})_{k \geqslant 1}$ is nonnegative, too. ∎

Proof of Lemma 4 . For the quadratic expression in R_k we calculate

$$
\frac{\Delta_k}{s^2} := \left(\omega_2 k + \omega_3\right)^2 - \frac{2(5\alpha - 2)}{3\alpha - 2} \omega_1 \omega_4 k^2 = \left(\omega_2^2 - \frac{2(5\alpha - 2)}{3\alpha - 2} \omega_1 \omega_4\right) k^2 + 2\omega_2 \omega_3 k + \omega_3^2.
$$

It suffices to guarantee that $\omega_2^2 - \dfrac{2(5\alpha - 2)}{3\alpha - 2} \omega_1 \omega_4 < 0$ in order to be sure that there exits some integer $k\left(\lambda\right) \geqslant 1$ such that $\Delta_k \leqslant 0$ for every $k \geqslant k\left(\lambda\right)$ and to obtain from here, due to Lemma 14, that $R_k \leqslant 0$ for every $k \geqslant k\left(\lambda\right)$.

We will show that there exists a nonempty open interval contained in $[0, \alpha - 1]$ with the property that $\omega_2^2 - \dfrac{2(5\alpha - 2)}{3\alpha - 2} \omega_1 \omega_4 < 0$ holds when λ is chosen within this open interval. To this end we set $\xi := \lambda + 1 - \alpha \leqslant 0$ and get

$$
\omega_2 = \frac{1}{\alpha - 1} \left(4\left(\alpha - 1\right)\xi - \alpha\left(\alpha - 2\right)\right) \qquad \text{and} \qquad \omega_1 \omega_4 = -\frac{2}{\alpha - 1} \left(\alpha - 2\right)\left(3\alpha - 2\right)\xi.
$$

Written in terms of ξ, we have first to guarantee that

$$
\omega_2^2 - \frac{2(5\alpha - 2)}{3\alpha - 2} \omega_1 \omega_4 = \frac{1}{\left(\alpha - 1\right)^2} \left(\left(4\left(\alpha - 1\right)\xi - \alpha\left(\alpha - 2\right)\right)^2 + 4\left(5\alpha - 2\right)\left(\alpha - 1\right)\left(\alpha - 2\right)\xi\right)
$$

$$
= \frac{1}{\left(\alpha - 1\right)^2} \left(16\left(\alpha - 1\right)^2 \xi^2 + 4\left(\alpha - 1\right)\left(\alpha - 2\right)\left(3\alpha - 2\right)\xi + \alpha^2\left(\alpha - 2\right)^2\right) < 0. \tag{43}
$$

A direct computation shows that

$$
\Delta_\xi = 16\left(\alpha - 1\right)^2 \left(2 - \alpha\right)^2 \left(\left(3\alpha - 2\right)^2 - 4\alpha^2\right) = 16\left(\alpha - 1\right)^2 \left(\alpha - 2\right)^3 \left(5\alpha - 2\right) > 0.
$$

Hence, in order to get (43), we have to choose ξ between the two roots of the quadratic function arising in this formula, in other words

$$
\xi_1\left(\alpha\right) := \frac{1}{32\left(\alpha - 1\right)^2} \left(-4\left(\alpha - 1\right)\left(\alpha - 2\right)\left(3\alpha - 2\right) - \sqrt{\Delta_\xi}\right)
$$

$$
= -\frac{1}{8\left(\alpha - 1\right)} \left(\alpha - 2\right)\left(3\alpha - 2 + \sqrt{\left(\alpha - 2\right)\left(5\alpha - 2\right)}\right)
$$

$$
< \xi = \lambda + 1 - \alpha < \xi_2\left(\alpha\right) := \frac{1}{32\left(\alpha - 1\right)^2} \left(-4\left(\alpha - 1\right)\left(\alpha - 2\right)\left(3\alpha - 2\right) + \sqrt{\Delta_\xi}\right)
$$

$$
= -\frac{1}{8\left(\alpha - 1\right)} \left(\alpha - 2\right)\left(3\alpha - 2 - \sqrt{\left(\alpha - 2\right)\left(5\alpha - 2\right)}\right).
$$

Obviously $\xi_1\left(\alpha\right) < 0$ and from Viète's formula $\xi_1\left(\alpha\right) \cdot \xi_2\left(\alpha\right) = \frac{\alpha^2 (\alpha - 2)^2}{16(\alpha - 1)^2}$, it follows that we must have $\xi_2\left(\alpha\right) < 0$ as well.

Therefore, going back to λ, in order to be sure that $\omega_2^2 - \dfrac{2(5\alpha - 2)}{3\alpha - 2}\omega_1\omega_4 < 0$ this must be chosen such that

$$\alpha - 1 + \xi_1(\alpha) < \lambda < \alpha - 1 + \xi_2(\alpha).$$

Next we will show that

$$0 < \alpha - 1 - \frac{1}{8(\alpha - 1)}(\alpha - 2)(3\alpha - 2) < \frac{3\alpha}{4} - \frac{1}{2}. \tag{44}$$

Indeed, the inequality on the left-hand side follows immediately, since

$$\alpha - 1 - \frac{1}{8(\alpha - 1)}(\alpha - 2)(3\alpha - 2) = \frac{1}{8(\alpha - 1)}\left(5\alpha^2 - 8\alpha + 4\right) = \frac{1}{8(\alpha - 1)}\left(\alpha^2 + 4(\alpha - 1)^2\right) > 0.$$

Using this relation, one can notice that the inequality on the right-hand side of (44) can be equivalently written as

$$5\alpha^2 - 8\alpha + 4 < 2(\alpha - 1)(3\alpha - 2) \Leftrightarrow 0 < \alpha^2 - 2\alpha = \alpha(\alpha - 2),$$

which is true as $\alpha > 2$.

From (44) we immediately deduce that

$$0 < \alpha - 1 + \xi_2(\alpha) \quad \text{and} \quad \alpha - 1 + \xi_1(\alpha) < \frac{3\alpha}{4} - \frac{1}{2}.$$

This allows us to choose

$$\underline{\lambda}(\alpha) := \alpha - 1 + \xi_1(\alpha) = \frac{1}{8(\alpha - 1)}\alpha^2 + \frac{1}{2}(\alpha - 1) - \frac{1}{8(\alpha - 1)}(\alpha - 2)\sqrt{(\alpha - 2)(5\alpha - 2)}$$

$$< \overline{\lambda}(\alpha) := \min\left\{\frac{3\alpha}{4} - \frac{1}{2}, \alpha - 1 + \xi_2(\alpha)\right\}$$

$$:= \min\left\{\frac{3\alpha}{4} - \frac{1}{2}, \frac{1}{8(\alpha - 1)}\alpha^2 + \frac{1}{2}(\alpha - 1) + \frac{1}{8(\alpha - 1)}(\alpha - 2)\sqrt{(\alpha - 2)(5\alpha - 2)}\right\},$$

since

$$\frac{1}{8(\alpha - 1)}\alpha^2 + \frac{1}{2}(\alpha - 1) - \frac{1}{8(\alpha - 1)}(\alpha - 2)\sqrt{(\alpha - 2)(5\alpha - 2)} > 0.$$

Indeed, as $(\alpha - 1)\sqrt{\alpha - 1} > (\alpha - 2)\sqrt{\alpha - 2}$ and $4\sqrt{\alpha - 1} > \sqrt{5\alpha - 2}$ we can easily deduce that

$$\alpha^2 + 4(\alpha - 1)^2 > 4(\alpha - 1)^2 > (\alpha - 2)\sqrt{(\alpha - 2)(5\alpha - 2)}$$

and the claim follows.

In conclusion, choosing λ to satisfy $\underline{\lambda}(\alpha) < \lambda < \overline{\lambda}(\alpha)$, we have $\omega_2^2 - \dfrac{2(5\alpha - 2)}{3\alpha - 2}\omega_1\omega_4 < 0$ and therefore there exists some integer $k(\lambda) \geqslant 1$ such that $R_k \leqslant 0$ for every $k \geqslant k(\lambda)$. $\blacksquare$

References

[1] **H. Attouch, J. Peypouquet, P. Redont**. *Fast convex optimization via inertial dynamics with Hessian driven damping.* Journal of Differential Equations 261(10), 5734–5783 (2016)

[2] **J. B. Baillon, R. E. Bruck**. *The rate of asymptotic regularity is $o(1/n)$.* In: Theory and Applications of Nonlinear Operators of Accretive and Monotone Types, Lecture Notes in Pure and Applied Mathematics 178, 51–81. Dekker, New York (1996)

[3] **H. H. Bauschke, P. L. Combettes**. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics, Springer, New York (2017)

[4] **J. Borwein, S. Reich, I. Shafrir**. *Krasnoselski-Mann iterations in normed spaces*. Canadian Mathematical Bulletin 35(1), 21–28 (1992)

[5] **R. I. Boţ, E.R. Csetnek**. *A dynamical system associated with the fixed points set of a nonexpansive operator*. Journal of Dynamics and Differential Equations 29(1), 155–168 (2017)

[6] **R. I. Boţ, E.R. Csetnek, D. Meier**. *Inducing strong convergence into the asymptotic behaviour of proximal splitting algorithms in Hilbert spaces*. Optimization Methods and Software 34(3), 489–514 (2019)

[7] **R. I. Boţ, E. R. Csetnek, D.-K. Nguyen**. *Fast OGDA in continuous and discrete time*. `arXiv:2203.10947`

[8] **R. I. Boţ, D. Meier**. *A strongly convergent Krasnosel'skii-Mann-type algorithm for finding a common fixed point of a countably infinite family of nonexpansive operators in Hilbert spaces*. Journal of Computational and Applied Mathematics 395, Article number: 113589 (2021)

[9] **M. Bravo, R. Cominetti**. *Sharp convergence rates for averaged nonexpansive maps*. Israel Journal of Mathematics 227, 163–188 (2018)

[10] **C. Brezinski, M. Redivo-Zaglia, Y. Saad**. *Shanks sequence transformations and Anderson acceleration*. SIAM Review 60 (3), 646–669 (2018)

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

[12] **R. Cominetti, J. A. Soto, J. Vaisman**. *On the rate of convergence of Krasnosel'skiĭ-Mann iterations and their connection with sums of Bernoullis*. Israel Journal of Mathematics 199(2), 757–772 (2014)

[13] **J. P. Contreras, R. Cominetti**. *Optimal error bounds for nonexpansive fixed-point iterations in normed spaces*. Mathematical Programming 199, 343–374 (2023)

[14] **D. Davis**. *Convergence rate analysis of the forward-Douglas-Rachford splitting scheme*. SIAM Journal on Optimization 25(3), 1760–1786 (2015)

[15] **D. Davis, W. Yin**. *Convergence rate analysis of several splitting schemes*. In: **R. Glowinski, S. J. Osher, W. Yin**, Splitting Methods in Communication, Imaging, Science, and Engineering, 115–163. Springer (2016)

[16] **D. Davis, W. Yin**. *A three-operator splitting scheme and its optimization applications*. Set-Valued and Variational Analysis 25 (4), 829–858 (2017)

[17] **J. Douglas, H. H. Rachford**. *On the numerical solution of heat conduction problems in two and three space variables*. Transactions of the American Mathematical Society 82(2), 421–439 (1956)

[18] **I. Fierro, J.J. Maulén, J. Peypouquet**. *Inertial Krasnoselskii-Mann iterations*. `arXiv:2210.03791`

[19] **C. W. Groetsch**. *A note on segmenting Mann iterates*. Journal of Mathematical Analysis and Applications 40, 369–372 (1972)

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

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

[22] **B. He, X. Yuan**. *On the convergence rate of Douglas–Rachford operator splitting method*. Mathematical Programming 153, 715–722 (2015)

[23] **X. Huang, E. K. Ryu, W. Yin**. *Tight coefficients of averaged operators via scaled relative graph*. Journal of Mathematical Analysis and Applications 490, 124211 (2020)

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

[25] **D. Kim**. *Accelerated proximal point method for maximally monotone operators*. Mathematical Programming 190, 57–87 (2021)

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

[27] **S. Lee, D. Kim**. *Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems*. Advances in Neural Information Processing Systems 34 (NeurIPS 2021)

[28] **J. Liang, J. Fadili, G. Peyré**. Convergence rates with inexact non-expansive operators. Mathematical Programming 159(1–2), 403–434 (2016)

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

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

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

[32] **S.-Y. Matsushita**. *On the convergence rate of the Krasnosel'skiĭ-Mann iteration*. Bulletin of the Australian Mathematical Society 96(1), 162–170 (2017)

[33] **Y. Nesterov**. *A method of solving a convex programming problem with convergence rate $\mathcal{O}\left(1/k^2\right)$*. Soviet Mathematics Doklady 27, 372–376 (1983)

[34] **Z. Opial**. *Weak convergence of the sequence of successive approximations for nonexpansive mappings*. Bulletin of the American Mathematical Society 73, 591–597 (1967)

[35] **J. Park, E. K. Ryu**. *Exact optimal accelerated complexity for fixed-point iterations*. Proceedings of the 39th International Conference on Machine Learning 162, 17420–17457 (2022)

[36] **H. Qi, H.-K. Xu**. *Convergence of Halpern's iteration method with applications in optimization*. Numerical Functional Analysis and Optimization 42:15, 1839–1854 (2021)

[37] **S. Reich**. *Weak convergence theorems for nonexpansive mappings in Banach spaces*. Journal of Mathematical Analysis and Applications 67, 274–276 (1979)

[38] **S. Sabach, S. Shtern** *A first order method for solving convex bilevel optimization problems*. SIAM Journal on Optimization 27 (2), 640-660 (2017)

[39] **Q. Tran-Dinh**. *The connection between Nesterov's accelerated methods and Halpern fixed-point iterations*. `arXiv:2203.04869`

[40] **Q. Tran-Dinh, Y. Luo**. *Halpern-type accelerated and splitting algorithms for monotone inclusions*. `arXiv:2110.08150`

[41] **R. Wittmann**. *Approximation of fixed points of nonexpansive mappings*. Archiv der Mathematik 58(5), 486–491 (1992)

[42] **H.-K. Xu**. *Iterative algorithms for nonlinear operators*. Journal of the London Mathematical Society 66(1), 240–256 (2002)

[43] **T. H. Yoon, E. K. Ryu**. *Accelerated algorithms for smooth convex-concave minimax problems with $\mathcal{O}\left(1/k^2\right)$ rate on squared gradient norm*. Proceedings of the 38th International Conference on Machine Learning, 12098–12109 (2021)

[44] **T. H. Yoon, E. K. Ryu**. *Accelerated minimax algorithms flock together*. `arXiv:2205.11093`