



On the convergence rate of the Halpern-iteration

Felix Lieder¹

Received: 30 June 2019 / Accepted: 2 July 2020 / Published online: 9 July 2020

© The Author(s) 2020

Abstract

In this work, we give a tight estimate of the rate of convergence for the Halpern-iteration for approximating a fixed point of a nonexpansive mapping in a Hilbert space. Specifically, using semidefinite programming and duality we prove that the norm of the residuals is upper bounded by the distance of the initial iterate to the closest fixed point divided by the number of iterations plus one.

Keywords Halpern-iteration · Fixed point methods · First order methods · Semidefinite programming · Performance estimation · Proximal point

1 Introduction

Let H be a Hilbert space equipped with a symmetric inner product $\langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbb{R}$. Let $T : H \rightarrow H$ be a nonexpansive mapping and consider for fixed $x_0 \in H$ the Halpern-Iteration

$$x_{k+1} := \lambda_k x_0 + (1 - \lambda_k)T(x_k) \quad (1)$$

from [5] with $\lambda_k := \frac{1}{k+2}$ for approximating a fixed point of T . Let $\|x\| := \sqrt{\langle x, x \rangle}$ denote the induced norm and $\text{Fix}(T) := \{x \in H : x = T(x)\}$ the set of fixed points of T . It is well known that, if the set $\text{Fix}(T)$ is nonempty, then the sequence $\{x_k\}_{k \in \mathbb{N}_0}$ will converge to $x_* \in \text{Fix}(T)$ minimizing the distance to x_0 ; see [17] Theorem 2, and [18] for generalizations of this remarkable property. As a consequence the norm of the residuals $x_k - T(x_k)$ tends to zero, i.e. $\lim_{k \rightarrow \infty} \|x_k - T(x_k)\| = 0$. Our goal here is to quantify their rate of convergence. A first result of this type was generated via proof mining in [10] in normed spaces (see also [8] and [9] for further details on results in more general spaces.). Here, we improve the result for the setting of Hilbert spaces. Our proof technique is not based on proof mining, but on semidefinite programming, and is strongly motivated by the recent work of Taylor et al. [15] on

✉ Felix Lieder
lieder@opt.uni-duesseldorf.de

¹ Mathematisches Institut, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

worst case performance of first order minimization methods. Our methodology and focus here are, however, slightly different. We present two new proofs below. The first one is short and uses a parameter choice derived from the technique used in [15]. The second proof based on semidefinite programming is self-contained and adapts the framework of [15] to fixed point problems. The second approach can also be applied to other choices of parameters λ_k and to other fixed point methods. The rates are however in general not obvious. After the manuscript of this work was made public in late 2017, the author (see section 4.1 and following of [12]) and independently other authors (see e.g. [2–4,6,16]) have studied similar setups, which may all be categorized as part of the performance estimation framework. Let us briefly sketch how our setup here and in [12] may be applied in the context of proximal point algorithms (e.g. the setup of [2,6]) by defining the nonexpansive operator $T := 2J - I$, where I is the identity and J is a firmly nonexpansive operator (e.g. the resolvent operator of a maximal monotone operator). Because the fixed points of J and T are the same, one may now apply the Halpern-Iteration (1) to find a fixed point of J . The (tight) convergence rate is then implied by Theorem 2.1 below.

2 Main result

Theorem 2.1 *Let $x_0 \in H$ be arbitrary but fixed. If T has fixed points, i.e. $\text{Fix}(T) \neq \emptyset$, then the iterates defined in (1) satisfy*

$$\frac{1}{2} \|x_k - T(x_k)\| \leq \frac{\|x_0 - x_*\|}{k+1} \quad \forall k \in \mathbb{N}_0 \quad \forall x_* \in \text{Fix}(T) \quad (2)$$

This bound is tight.

Remark 2.1 A generalization of the Halpern-Iteration, the sequential averaging method (SAM), was analyzed in the recent paper [14], where for the first time a rate of convergence of order $O(1/k)$ could be established for SAM. The rate of convergence in (2) is even slightly faster than the one established for the more general framework in [14] (by a factor of 4). More importantly, however, as shown by Example 3.1 below, the estimate (2) is actually tight, in the sense that for every $k \in \mathbb{N}_0$ there exists a Hilbert space H and a nonexpansive operator T with some fixed point x_* such that the inequality (2) is not strict.

Estimate (2) refers to the step length $\lambda_k := 1/(k+2)$. The restriction to this choice is motivated by problem (17) below in the proof based on semidefinite programming; in numerical tests for small for dimensions k these coefficients provided a better worst-case complexity than any other choice of coefficients.

Next, an elementary direct proof of Theorem 2.1 is given.

Direct proof based on a weighted sum

The iteration (1) with $\lambda_k = 1/(k+2)$ implies for $1 \leq j \leq k$:

$$x_j = \frac{1}{j+1}x_0 + \frac{j}{j+1}T(x_{j-1}) \quad \text{or} \quad T(x_{j-1}) = \frac{j+1}{j}x_j - \frac{1}{j}x_0. \quad (3)$$

By nonexpansiveness the following inequalities hold:

$$\|T(x_k) - x_*\|^2 \leq \|x_k - x_*\|^2 \quad \text{for } x_* \in \text{Fix}(T) \tag{4}$$

and

$$\|T(x_j) - T(x_{j-1})\|^2 \leq \|x_j - x_{j-1}\|^2 \quad \text{for } j = 1, \dots, k. \tag{5}$$

Below we reformulate the following weighted sum of (5):

$$0 \geq \sum_{j=1}^k j(j+1) \left(\|T(x_j) - T(x_{j-1})\|^2 - \|x_j - x_{j-1}\|^2 \right). \tag{6}$$

Using the second relation in (3) the first terms in the summation (6) are

$$\begin{aligned} & j(j+1)\|T(x_j) - T(x_{j-1})\|^2 \\ &= j(j+1)\|x_j - T(x_j) + \frac{1}{j}(x_j - x_0)\|^2 \\ &= j(j+1)\|x_j - T(x_j)\|^2 + 2(j+1)\langle x_j - T(x_j), x_j - x_0 \rangle + \frac{j+1}{j}\|x_j - x_0\|^2, \end{aligned} \tag{7}$$

and using the first relation in (3) it follows for the second terms in (6)

$$\begin{aligned} & -j(j+1)\|x_j - x_{j-1}\|^2 \\ &= -j(j+1)\|\frac{1}{j+1}(x_0 - T(x_{j-1})) + T(x_{j-1}) - x_{j-1}\|^2 \\ &= -\frac{j}{j+1}\|x_0 - T(x_{j-1})\|^2 - 2j\langle x_0 - T(x_{j-1}), T(x_{j-1}) - x_{j-1} \rangle \\ & \quad -j(j+1)\|T(x_{j-1}) - x_{j-1}\|^2. \end{aligned} \tag{8}$$

Observe [using again the second relation in (3)] that the first term in (8)

$$\begin{aligned} -\frac{j}{j+1}\|x_0 - T(x_{j-1})\|^2 &= -\frac{j}{j+1}\|\frac{j+1}{j}x_0 - \frac{j+1}{j}x_j\|^2 \\ &= -\frac{j+1}{j}\|x_0 - x_j\|^2 \end{aligned} \tag{9}$$

cancels the third term in (7). Summing up the second terms in (8) for $j = 1, \dots, k$ we shift the summation index,

$$-\sum_{j=1}^k 2j\langle x_0 - T(x_{j-1}), T(x_{j-1}) - x_{j-1} \rangle = \sum_{j=0}^{k-1} 2(j+1)\langle x_j - T(x_j), x_0 - T(x_j) \rangle,$$

so that summing up the second terms in (7) and in (8) for $j = 1, \dots, k$ results in

$$2(k + 1)\langle x_k - T(x_k), x_k - x_0 \rangle + 2 \sum_{j=1}^{k-1} (j + 1)\langle x_j - T(x_j), x_j - T(x_j) \rangle + 2\|x_0 - T(x_0)\|^2. \tag{10}$$

Shifting again the index in the summation of the third terms in (8)

$$- \sum_{j=1}^k j(j + 1)\|x_{j-1} - T(x_{j-1})\|^2 = - \sum_{j=0}^{k-1} (j + 1)(j + 2)\|x_j - T(x_j)\|^2$$

and summing up the first terms in (7) and the third terms in (8) for $j = 1, \dots, k$ gives

$$k(k + 1)\|x_k - T(x_k)\|^2 - 2 \sum_{j=1}^{k-1} (j + 1)\|x_j - T(x_j)\|^2 - 2\|x_0 - T(x_0)\|^2 \tag{11}$$

where the sum in the middle cancels the sum in the middle of (10) and the terms $2\|x_0 - T(x_0)\|^2$ cancel as well. The only remaining terms are the first terms in (10) and (11).

Thus, inserting (9), (10), and (11) in (6) leads to

$$0 \geq k(k + 1)\|x_k - T(x_k)\|^2 + 2(k + 1)\langle x_k - T(x_k), x_k - x_0 \rangle. \tag{12}$$

Applying the Cauchy–Schwarz inequality to the second term in (12) leads to

$$\frac{1}{2}\|x_k - T(x_k)\| \leq \frac{1}{k}\|x_k - x_0\|$$

which may be interesting in its own right. To prove the theorem, (12) is divided by $k + 1$ and then (4) is added:

$$\begin{aligned} 0 &\geq k\|x_k - T(x_k)\|^2 + 2\langle x_k - T(x_k), x_k - x_0 \rangle + \|T(x_k) - x_*\|^2 - \|x_k - x_*\|^2 \\ &= \frac{k+1}{2}\|x_k - T(x_k)\|^2 - \frac{2}{k+1}\|x_0 - x_*\|^2 \\ &\quad + \frac{2}{k+1}\|x_0 - x_* - \frac{k+1}{2}(x_k - T(x_k))\|^2. \end{aligned} \tag{13}$$

To see the last equation, the last two terms in (13) can be combined, and then a straightforward but tedious multiplication of the terms $a := x_k - T(x_k)$, $b := x_k - x_0$, $c := T(x_k) - x_*$, $a + c = x_k - x_*$, and $a + c - b = x_0 - x_*$ reveals the identity.

Omitting the last term in (13) one obtains

$$\|x_k - T(x_k)\|^2 \leq \left(\frac{2}{k+1}\right)^2 \|x_0 - x_*\|^2$$

which proves the theorem when taking square roots on both sides. □

The above proof is somewhat unintuitive as the choice of the weights with which the inequalities (4) and (5) are added in (6) and (13) is far from obvious. In fact we owe the suggestion of these weights to an extremely helpful anonymous referee, who extracted it from a more complex construction in [15] which was also the basis for the initial proof of this paper based on semidefinite programming. We state this proof next since it offers a generalizable approach for analyzing fixed point iterations; it can be modified to the KM iteration, for example in the recent thesis [12]—though this modification is quite technical. The proof based on semidefinite programming also led to Example 3.1 below showing that the rate of convergence is tight.

Proof based on semidefinite programming Let $x_* \in \text{Fix}(T)$. The Halpern-Iteration was stated in the form (1) to comply with existing literature. For our proof however, it is more convenient to consider the shifted sequence $\bar{x}_1 := x_0$ and $\bar{x}_k := x_{k-1} \forall k \in \mathbb{N}_{\neq 0} := \{1, 2, 3, \dots\}$ and to show a shifted statement

$$\frac{1}{2} \|\bar{x}_k - T(\bar{x}_k)\| \leq \frac{\|\bar{x}_1 - x_*\|}{k} \quad \forall k \in \mathbb{N}_{\neq 0} \tag{14}$$

Let us define $g(x) := \frac{1}{2}(x - T(x))$. It is well known that g is firmly nonexpansive. For sake of completeness the argument is repeated here:

$$\begin{aligned} & \|g(x) - g(y)\|^2 - \langle g(x) - g(y), x - y \rangle \\ &= \|g(x) - g(y) - \frac{1}{2}(x - y)\|^2 - \frac{1}{4}\|x - y\|^2 \\ &= \frac{1}{4}\|T(x) - T(y)\|^2 - \frac{1}{4}\|x - y\|^2 \leq 0 \quad \forall x, y \in H. \end{aligned}$$

Nonexpansiveness and the Cauchy–Schwarz inequality also imply $\|g(x) - g(y)\| \leq \|x - y\| \forall x, y \in H$. For $k = 1$ the statement (14) follows immediately since $g(x_*) = 0$ and therefore $\frac{1}{2}\|\bar{x}_1 - T(\bar{x}_1)\| = \|g(\bar{x}_1)\| = \|g(\bar{x}_1) - g(x_*)\| \leq \frac{\|\bar{x}_1 - x_*\|}{1}$.

For fixed $k \geq 2$ we first consider the differences $\bar{x}_j - \bar{x}_1$ for $j \in \{2, \dots, k\}$

$$\begin{aligned} \bar{x}_j - \bar{x}_1 &= x_{j-1} - \bar{x}_1 \\ &= \lambda_{j-2}x_0 + (1 - \lambda_{j-2})T(x_{j-2}) - \bar{x}_1 \\ &= \frac{1}{j}x_0 + \left(1 - \frac{1}{j}\right)T(x_{j-2}) - \bar{x}_1 \\ &= \left(\frac{1}{j} - 1\right)\bar{x}_1 + \left(1 - \frac{1}{j}\right)T(\bar{x}_{j-1}) \\ &= \left(\frac{1}{j} - 1\right)\bar{x}_1 + \left(1 - \frac{1}{j}\right)(\bar{x}_{j-1} - 2g(\bar{x}_{j-1})) \\ &= \left(1 - \frac{1}{j}\right)(\bar{x}_{j-1} - \bar{x}_1) - 2\left(1 - \frac{1}{j}\right)g(\bar{x}_{j-1}) \\ &= \frac{j-1}{j}(\bar{x}_{j-1} - \bar{x}_1) - 2\frac{j-1}{j}g(\bar{x}_{j-1}) \end{aligned}$$

which inductively leads to

$$\bar{x}_j - \bar{x}_1 = -2 \sum_{l=1}^{j-1} \frac{l}{j} g(\bar{x}_l).$$

Let us shorten the notation slightly and define $g_i := g(\bar{x}_i)$, $R := \|\bar{x}_1 - x_*\| \geq 0$, the vector $b = (\langle g_i, \bar{x}_1 - x_* \rangle)_{i=1}^k$, the matrices $A := (\langle g_i, g_j \rangle)_{i,j=1}^k$ and

$$L := -2 \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{k} \\ 0 & 0 & \frac{2}{3} & \cdots & \frac{2}{k} \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \frac{k-1}{k} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{k \times k}.$$

Let b^T denote the transpose of b . Note that

$$\begin{pmatrix} R^2 & b^T \\ b & A \end{pmatrix} \in \mathbb{R}^{(k+1) \times (k+1)}$$

is a Gramian matrix formed from $\bar{x}_1 - x_*$, $g_1, \dots, g_k \in H$ and is therefore symmetric and positive semidefinite. We proceed by expressing the inequalities from firm nonexpansiveness in terms of the Gram-Matrix. Since L often is of much lower dimension than H , this is sometimes referred to as “kernel trick”. Keeping in mind that we can rewrite the differences $\bar{x}_j - \bar{x}_1 = -2 \sum_{l=1}^{j-1} \frac{l}{j} g_l$ for $j \in \{1, \dots, k\}$, we arrive at

$$AL = (\langle g_i, \bar{x}_j - \bar{x}_1 \rangle)_{i,j=1}^k.$$

Let $e \in \mathbb{R}^k$ denote the vector of all ones. Then

$$\text{diag}(AL)e^T - AL = (\langle g_i, \bar{x}_i - \bar{x}_j \rangle)_{i,j=1}^k,$$

where $\text{diag}(\cdot)$ denotes the diagonal of its matrix argument, holds true. Hence

$$\text{diag}(AL)e^T + e \text{diag}(AL)^T - AL - L^T A = (\langle g_i - g_j, \bar{x}_i - \bar{x}_j \rangle)_{i,j=1}^k$$

and

$$\begin{aligned} be^T + AL &= (\langle g_i, \bar{x}_j - x_* \rangle)_{i,j=1}^k, \\ \text{diag}(A)e^T + e \text{diag}(A)^T - 2A &= (\|g_i - g_j\|^2)_{i,j=1}^k. \end{aligned}$$

The firm nonexpansiveness inequalities $\|g_i - g_j\|^2 \leq \langle g_i - g_j, \bar{x}_i - \bar{x}_j \rangle$ are equivalent to the component-wise inequality

$$\text{diag}(A)e^T + e \text{diag}(A)^T - 2A \leq \text{diag}(AL)e^T + e \text{diag}(AL)^T - AL - L^T A. \tag{15}$$

Note that only $\frac{k^2-k}{2}$ of these componentwise inequalities are non redundant. From $g_* := g(x_*) = 0$ we obtain another k inequalities, i.e. $\|g_i\|^2 \leq \langle g_i, \bar{x}_i - x_* \rangle$, which translate to

$$\text{diag}(A) \leq b + \text{diag}(AL). \tag{16}$$

Defining $U := I - L$, relations (15) and (16) can be shortened slightly to

$$\text{diag}(AU)e^T + e \text{diag}(AU)^T \leq AU + U^T A$$

and

$$\text{diag}(AU) \leq b.$$

Let $e_k \in \mathbb{R}^k$ denote the k -th unit vector, $\mathbb{S}^n := \{X \in \mathbb{R}^{n \times n} \mid X = X^T\}$ denote the space of symmetric matrices and $\mathbb{S}_+^n := \{X \in \mathbb{S}^n \mid x^T X x \geq 0 \forall x \in \mathbb{R}^n\}$ the convex cone of positive semidefinite matrices. Consider the chain of inequalities

$$\begin{aligned} \|g(\bar{x}_k)\|^2 &= \underset{y_0 \in \mathbb{R}, y_1 \in \mathbb{R}^k, Y_2 \in \mathbb{S}^k}{\text{maximize}} (Y_2)_{kk} \mid \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \in \mathbb{S}_+^{k+1}, y_0 \leq R^2, \text{diag}(Y_2 U) \leq y_1 \\ &\quad \mid \text{diag}(Y_2 U)e^T + e \text{diag}(U^T Y_2)^T \leq Y_2 U + U^T Y_2 \\ &\quad \mid y_0 = R^2, y_1 = b, Y_2 = A \\ &\leq \underset{y_0 \in \mathbb{R}, y_1 \in \mathbb{R}^k, Y_2 = Y_2^T \in \mathbb{S}^k}{\text{maximize}} (Y_2)_{kk} \mid \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \in \mathbb{S}_+^{k+1}, y_0 \leq R^2, \text{diag}(Y_2 U) \leq y_1 \\ &\quad \mid \text{diag}(Y_2 U)e^T + e \text{diag}(U^T Y_2)^T \leq Y_2 U + U^T Y_2 \end{aligned} \tag{17}$$

$$\leq \underset{\xi \in \mathbb{R}_+, X \in \mathbb{S}^k \cap \mathbb{R}^{k \times k}}{\text{minimize}} R^2 \xi \mid \begin{pmatrix} \xi & -\frac{1}{2} \text{diag}(X)^T \\ -\frac{1}{2} \text{diag}(X) & U F(X) + F(X) U^T \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & e_k e_k^T \end{pmatrix} \in \mathbb{S}_+^{k+1} \tag{18}$$

for

$$F(X) := \text{Diag}(Xe) + \frac{1}{2} \text{Diag}(\text{diag}(X)) - X, \tag{19}$$

where $\text{Diag}(\cdot)$ denotes the square diagonal matrix with its vector argument on the diagonal. The first equality follows from construction, the first inequality from relaxing,

and the second inequality from weak conic duality as detailed in Sect. 5. We conclude the proof by showing feasibility of $\hat{\xi} := \frac{1}{k^2} > 0$ and

$$\hat{X} := \frac{1}{k^2} \begin{pmatrix} 0 & 1 \cdot 2 & 0 & \dots & \dots & 0 \\ 1 \cdot 2 & 0 & 2 \cdot 3 & 0 & \dots & \vdots \\ 0 & 2 \cdot 3 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & (k-2)(k-1) & 0 \\ 0 & \dots & 0 & (k-2)(k-1) & 0 & (k-1)k \\ 0 & \dots & 0 & 0 & (k-1)k & 2k \end{pmatrix} \in \mathbb{R}^{k \times k}$$

for the last optimization problem (18). First note that $\hat{X} = \hat{X}^T$ is symmetric and nonnegative. A short computation reveals, that the equality

$$UF(\hat{X}) + F(\hat{X})U^T = 2e_k e_k^T$$

holds true: Define the diagonal matrix $D := \frac{1}{k} \text{Diag}([1, \dots, k]^T) \in \mathbb{R}^{k \times k}$, together with the strict upper triangular matrix

$$P := \begin{pmatrix} 0 & 1 & \dots & 1 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} \in \mathbb{R}^{k \times k} \tag{20}$$

and the bidiagonal matrix

$$B := \begin{pmatrix} 0 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & & 1 & 0 \end{pmatrix} \in \mathbb{S}^k.$$

The matrices U , \hat{X} and $F(\hat{X})$ can now be expressed as

$$\begin{aligned} U &= I + 2DPD^{-1}, \quad \hat{X} = DBD + \frac{2}{k} e_k e_k^T \quad \text{and} \\ F(\hat{X}) &= 2D^2 - e_k e_k^T - DBD. \end{aligned} \tag{21}$$

Combining the equalities (21), $De_k = e_k$ and $D^{-1}e_k = e_k$, yields

$$\begin{aligned} UF(\hat{X}) &= 2D^2 - e_k e_k^T - DBD + 4DPD - 2DPe_k e_k^T - 2DPBD \\ &= D(2I - e_k e_k^T - B + 4P - 2Pe_k e_k^T - 2PB)D \end{aligned} \tag{22}$$

and using (22) we compute

$$\begin{aligned}
 & UF(\hat{X}) + F(\hat{X})U^T - 2e_k e_k^T \\
 &= D(4I - 2e_k e_k^T - 2B + 4P - 2Pe_k e_k^T - 2PB + 4P^T - 2e_k e_k^T P^T - 2BP^T - 2e_k e_k^T)D \\
 &= D(\underbrace{4I + 4P + 4P^T}_{=4ee^T} - 4e_k e_k^T - 2\underbrace{Pe_k e_k^T}_{e - e_k} - 2e_k \underbrace{e_k^T P^T}_{e^T - e_k^T} - 2B - 2PB - 2BP^T)D \\
 &= D(\underbrace{4ee^T - 2ee_k^T - 2e_k e^T - 2B - 2PB - 2BP^T}_{=-4ee^T})D \\
 &= 0,
 \end{aligned}$$

which implies $UF(\hat{X}) + F(\hat{X})U^T = 2e_k e_k^T$ as we claimed above. Consequently

$$\begin{pmatrix} \hat{\xi} & -\frac{1}{2}\text{diag}(\hat{X})^T \\ -\frac{1}{2}\text{diag}(\hat{X}) & UF(\hat{X}) + F(\hat{X})U^T \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & e_k e_k^T \end{pmatrix} = \begin{pmatrix} \frac{1}{k^2} & -\frac{1}{k}e_k^T \\ -\frac{1}{k}e_k & e_k e_k^T \end{pmatrix} \geq 0$$

is positive semidefinite and as a result, $\hat{\xi}$ and \hat{X} is feasible for (18). Hence

$$\|g(\bar{x}_k)\|^2 \leq R^2 \hat{\xi} = \frac{\|\bar{x}_1 - x_*\|^2}{k^2},$$

which yields the desired result after taking the square root. □

Remark 2.2 The matrix \hat{X} in the above proof carrying the weights $j(j + 1)$ used in (6) were obtained by solving (18) with YALMIP [11] in combination with the SDP solver Sedumi [13] for small values of k . In order to provide a theoretical proof that the points $\hat{\xi}$ and \hat{X} above are not only feasible but actually optimal for (18) and to prove tightness of the derived bound, we refer to Example 3.1 below, which was derived from a, numerically obtained, low-rank optimal solution of (17). More precisely, after numerically determining the optimal value of (18) a linear equation was added to (18) requiring that $(Y_2)_{kk}$ equals this value, and then the trace of $Y_{2,2}$ was minimized with the intention to find the optimal solution with minimum rank. This optimal solution was then used to derive Example 3.1 below proving the tightness of (2). In fact for any optimal solution of the SDP relaxation (17), there exists at least one Lipschitz continuous operator $\tilde{T}_k : \mathbb{R}^d \rightarrow \mathbb{R}^d$ for appropriately chosen d with some fixed point x_* such that the inequality in (18) is tight: This follows from appropriately labeling the columns of the symmetric square root of such an optimal solution and a Lipschitz extension argument. Specifically the Kirszbraun-Theorem [7] allows a Lipschitz extension of an operator that is Lipschitz on a discrete set to the entire space. For further details we refer to Section 4.2 of [12].

3 Tightness and choice of step lengths

Example 3.1 We consider the following one-dimensional real example with fixed point $x_* = 0$ and starting point $x_0 \neq 0$. Let $k \in \mathbb{N}$ be given. A nonexpansive mapping

proving tightness of (2) can then be set up as follows: Let $T : \mathbb{R} \rightarrow \mathbb{R}$ be defined via

$$T(x) := \begin{cases} x + \frac{2R}{k+1} & \text{if } x \leq -\frac{R}{k+1} \\ -x & \text{if } -\frac{R}{k+1} < x < \frac{R}{k+1} \\ x - \frac{2R}{k+1} & \text{if } \frac{R}{k+1} \leq x \end{cases} \tag{23}$$

for some fixed $k \in \mathbb{N}$ and $R := \|x_0 - x_*\|_2$ with $x_0 \in \mathbb{R}$ and $x_* := 0$. Note that T satisfies $T(x_*) = 0 = x_*$ and is 1-Lipschitz continuous, i.e. nonexpansive, because it is piece-wise linear, continuous and the derivative is bounded in norm by one ($|T'| \leq 1$) whenever it exists. We will now show that applying the Halpern-Iteration results in an equality in (2) for the k -th iterate, i.e.

$$\|\frac{1}{2}(x_k - T(x_k))\|_2 = \frac{\|x_0 - x_*\|_2}{k+1}$$

is satisfied. This means that the bound (2) can not be improved without making further assumptions, as the operator above would otherwise pose a counterexample. For the first k iterates of the Halpern-Iteration (1) we can obtain

$$x_j = x_0 \underbrace{\left(1 - \frac{j}{k+1}\right)}_{\geq \frac{1}{k+1}} \quad \text{for } j \in 0, \dots, k$$

inductively: for $x_0 = 0 = x_*$ there is nothing to prove. The case $0 < x_0 = \|x_0 - x_*\|_2 = R$ follows by using the definition of T and considering for $j \in \{0, \dots, k-1\}$ the iterates

$$\begin{aligned} x_{j+1} &= \frac{x_0}{j+2} + \left(1 - \frac{1}{j+2}\right) T\left(\underbrace{x_j}_{\geq \frac{x_0}{k+1} = \frac{R}{k+1}}\right) \\ &= \frac{x_0}{j+2} + \left(1 - \frac{1}{j+2}\right) \left(x_j - \frac{2R}{k+1}\right) \\ &= \frac{x_0}{j+2} + \left(1 - \frac{1}{j+2}\right) \left(x_0 \left(1 - \frac{j}{k+1}\right) - \frac{2x_0}{k+1}\right) \\ &= x_0 \left(\frac{1}{j+2} + \left(1 - \frac{1}{j+2}\right) \left(1 - \frac{j}{k+1} - \frac{2}{k+1}\right)\right) \\ &= x_0 \left(\frac{1}{j+2} + 1 - \frac{j+2}{k+1} - \frac{1}{j+2} + \frac{1}{k+1}\right) \\ &= x_0 \underbrace{\left(1 - \frac{j+1}{k+1}\right)}_{\geq \frac{1}{k+1}}, \end{aligned}$$

which imply, that

$$\|\frac{1}{2}(x_k - T(x_k))\|_2 = \|\frac{1}{2}\left(x_0 \left(1 - \frac{k}{k+1}\right) - \left(x_0 \left(1 - \frac{k}{k+1}\right) - \frac{2R}{k+1}\right)\right)\|_2$$

$$= \frac{R}{k+1} = \frac{\|x_0 - x_*\|_2}{k+1}$$

holds true. The case $x_0 < 0$ follows from the operators point symmetry, i.e. $T(-x) = -T(x)$. This completes the proof of tightness. \square

While Example 3.1 shows that the bound (2) is best possible for the Halpern-Iteration with $\lambda_k = 1/(k + 2)$, the rate of convergence could be improved for this example, if the values λ_k were chosen appropriately less than $1/(k + 2)$. Let us illustrate next that this is not true in general, i.e. that choosing smaller values of λ_k does not always provide faster convergence. Let H be a Hilbert space with a countable orthonormal Schauder basis $\{e_j\}_{j \in \mathbb{N}}$ and T be the linear operator defined by $T(e_j) = e_{j+1}$ for $j \in \mathbb{N}$. Hence the unique fixed point is $x_* = 0$. When choosing $x_0 = e_1$, then for any choice of step lengths $\lambda_j \in [0, 1]$, the k -th iterate always lies in the convex hull of e_1, \dots, e_{k+1} , and the choice of λ_j minimizing the error $\|x_k - x_*\|$ is precisely the step length $\lambda_j = 1/(j + 2)$ for all $1 \leq j \leq k$. This step length leads to a slightly faster rate of convergence than (2) for this example, namely $\frac{1}{2}\|x_k - T(x_k)\| \leq \frac{\|x_0 - x_*\|}{\sqrt{2(k+1)}}$. While this step length does not minimize the residual $\frac{1}{2}\|x_k - T(x_k)\|$ it shows that smaller values of λ_j such as $\lambda_j := \rho/(j + 2)$ for all j with $\rho \in [0, 1)$ lead to larger residuals. On the other hand larger values $\lambda_j > 1/(j + 2)$ for all j lead to larger residuals for Example 3.1.

4 Conclusions

We have derived a new and tight complexity bound for the Halpern-Iteration with coefficients chosen as $\lambda_k = \frac{1}{k+2}$. The proof based on semidefinite programming can in principle be adapted for other choices of parameters and fixed point iterations, again leading to tight complexity bounds. For example, for the Krasnoselski–Mann (KM) iteration (see e.g. [1])

$$x_{k+1} := (1 - t)x_k + t T(x_k)$$

with some constant stepsize $t \in [\frac{1}{2}, 1]$ a proof can be found in [12] (Theorem 4.9), where

$$L := -2 \begin{pmatrix} 0 & t & t & \dots & t \\ 0 & 0 & t & \dots & t \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & t \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{k \times k}$$

is used to to define inequalities of the form (15). However, while in practice the KM-Iteration with constant stepsize may often perform much better than the Halpern-Iteration, its worst-case complexity is an order of magnitude worse—and the convergence analysis based on semidefinite programming is considerably longer.

Acknowledgements Open Access funding provided by Projekt DEAL. I would like to thank Florian Jarre for careful proofreading of this paper. I would also like to sincerely thank the anonymous referees for their constructive criticism that helped to improve the presentation of the paper and that led to the direct proof based on a weighted sum.

Compliance with ethical standards

Conflict of interest The author declares that he has no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

5 Appendix, a technical duality result

While the duality used in (17) and (18) is the well known weak conic duality, the format of the problems (17) and (18) is quite intricate. Here, the explicit derivation of the dual problem therefore is derived in detail: Define the Euclidean space $\mathbb{E} := \mathbb{R} \times \mathbb{S}^k \times \mathbb{S}^{k+1}$ and the closed convex cone $\mathcal{K} := \mathbb{R}_+ \times (\mathbb{S}^k \cap \mathbb{R}_+^{k \times k}) \times \mathbb{S}_+^{k+1}$. We denote the trace inner product $A \cdot B := \text{trace}(AB)$ for all symmetric matrices A, B . Then \mathcal{K} is self-dual with respect to the canonical inner product $\langle X, Y \rangle_{\mathbb{E}} := x_1^T y_1 + X_2 \cdot Y_2 + X_3 \cdot Y_3$, which we define for any

$$\begin{pmatrix} x_1 \\ X_2 \\ X_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \in \mathbb{E}.$$

We proceed by restating problems (17) and (18) in standard form and exploit (weak) conic duality. In fact, by adding slack variables $S \in \mathcal{K}$, we can write (17) equivalently as a conic optimization problem in dual standard form

$$\underset{Y \in \mathbb{S}^{n+1}, S \in \mathcal{K}}{\text{maximize}} \{ \tilde{B} \cdot Y \mid \mathcal{A}^*(Y) + S = \tilde{C}, S \in \mathcal{K} \} \quad (24)$$

for

$$\tilde{C} := \begin{pmatrix} R^2 \\ 0 \\ 0 \end{pmatrix} \in \mathbb{E},$$

$$\tilde{B} = \begin{pmatrix} 0 & 0 \\ 0 & e_k e_k^T \end{pmatrix} \in \mathbb{S}^{k+1} \quad (25)$$

and the linear operator $\mathcal{A}^* : \mathbb{S}^{n+1} \rightarrow \mathbb{E}$

$$\mathcal{A}^* \left(\begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \right) := \begin{pmatrix} \text{diag}(Y_2 U) e^T + e \text{diag}(U^T Y_2)^T - (Y_2 U + U^T Y_2) + \text{Diag}(\text{diag}(Y_2 U) - y_1) & y_0 \\ - \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} & \end{pmatrix}. \tag{26}$$

The constraint $\mathcal{A}^*(Y) + S = \tilde{C}$, $S \in \mathcal{K}$ is a direct translation of the constraints in (17) except for the diagonal entries. Here it is used that the term $\text{diag}(Y_2 U) e^T + e \text{diag}(U^T Y_2)^T - (Y_2 U + U^T Y_2)$ has an all zero diagonal, which is why we can “place” the constraints $\text{diag}(Y_2 U) - y_1 \leq 0$ on the diagonal.

Recall that the optimal value of the conic optimization problem (24) in dual standard form is upper bounded by the optimal value of the conic optimization in primal standard form

$$\underset{X \in \mathbb{E}}{\text{minimize}} \{ \langle \tilde{C}, \tilde{X} \rangle_{\mathbb{E}} \mid \mathcal{A}(\tilde{X}) = \tilde{B}, \tilde{X} \in \mathbb{E}, \tilde{X} \in \mathcal{K} \} \tag{27}$$

where the linear operators $\mathcal{A}^* : \mathbb{S}^{n+1} \rightarrow \mathbb{E}$ and $\mathcal{A} : \mathbb{E} \rightarrow \mathbb{S}^{n+1}$ are adjoint, i.e. \mathcal{A} is the unique operator that satisfies

$$\langle \mathcal{A}^*(Y), \tilde{X} \rangle_{\mathbb{E}} = Y \cdot \mathcal{A}(\tilde{X}) \tag{28}$$

for all $Y \in \mathbb{S}^{n+1}$, $\tilde{X} \in \mathbb{E}$. The “ Y_2 -part” of the adjoint of the second component of \mathcal{A}^* can be obtained from

$$\begin{aligned} & (\text{diag}(Y_2 U) e^T + e \text{diag}(U^T Y_2)^T - (Y_2 U + U^T Y_2) + \text{Diag}(\text{diag}(Y_2 U))) \cdot X_1 \\ &= Y_2 \cdot \underbrace{(U(\text{Diag}(X_1 e) + \frac{1}{2} \text{Diag}(\text{diag}(X_1)) - X_1))}_{=F(X_1)} \\ &+ \underbrace{(\text{Diag}(X_1 e) + \frac{1}{2} \text{Diag}(\text{diag}(X_1)) - X_1) U^T}_{=F(X_1)} \end{aligned} \tag{29}$$

leading to the full description of \mathcal{A}

$$\mathcal{A} \left(\begin{pmatrix} \xi \\ X_1 \\ X_2 \end{pmatrix} \right) = \begin{pmatrix} \xi & -\frac{1}{2} \text{diag}(X_1)^T \\ -\frac{1}{2} \text{diag}(X_1) & U F(X_1) + F(X_1) U^T \end{pmatrix} - X_2 \tag{30}$$

where, as before $F(X) = \text{Diag}(X e) + \frac{1}{2} \text{Diag}(\text{diag}(X)) - X$. This shows that we can rewrite (27) in the claimed form (18) by eliminating the (semidefinite) variable X_2 . This establishes the inequality in (17) and (18). \square

References

1. Cominetti, R., Soto, J.A., Vaisman, J.: On the rate of convergence of Krasnoselskii–Mann iterations and their connection with sums of Bernoullis. *Isr. J. Math.* **199**(2), 757–772 (2014)
2. Gu, G., Yang, J.: Optimal nonergodic sublinear convergence rate of proximal point algorithm for maximal monotone inclusion problems (2019). arXiv preprint [arXiv:1904.05495](https://arxiv.org/abs/1904.05495)
3. Gu, G., Yang, J.: On the optimal linear convergence factor of the relaxed proximal point algorithm for monotone inclusion problems (2019). arXiv preprint [arXiv:1905.04537](https://arxiv.org/abs/1905.04537)
4. Gu, G., Yang, J.: On the optimal ergodic sublinear convergence rate of the relaxed proximal point algorithm for variational inequalities (2019). arXiv preprint [arXiv:1905.06030](https://arxiv.org/abs/1905.06030)
5. Halpern, B.: Fixed points of nonexpanding maps. *Bull. Am. Math. Soc.* **73**(6), 957–961 (1967)
6. Kim, D.: Accelerated proximal point method for maximally monotone operators (2019). arXiv preprint [arXiv:1905.05149](https://arxiv.org/abs/1905.05149)
7. Kirszbraun, M.D.: Über die zusammenziehende und Lipschitzsche Transformationen. *Fund. Math.* **22**, 77–108 (1934)
8. Kohlenbach, U.: On quantitative versions of theorems due to FE Browder and R Wittmann. *Adv. Math.* **226**(3), 2764–2795 (2011)
9. Kohlenbach, U., Leuştean, L.: Effective metastability of Halpern iterates in CAT (0) spaces. *Adv. Math.* **231**(5), 2526–2556 (2012)
10. Leuştean, L.: Rates of asymptotic regularity for halpern iterations of nonexpansive mappings. *J. UCS* **13**(11), 1680–1691 (2007)
11. Lofberg, J.: YALMIP: a toolbox for modeling and optimization in MATLAB. In: 2004 IEEE International Symposium on Computer Aided Control Systems Design, IEEE, pp. 284–289 (2004)
12. Lieder, F.: Projection based methods for conic linear programming—optimal first order complexities and norm constrained quasi newton methods, (Doctoral dissertation, Universitäts- und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf) (2018). <https://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-49971/Dissertation.pdf>
13. Sturm, J.F.: Implementation of interior point methods for mixed semidefinite and second order cone optimization problems. *Optim. Methods Softw.* **17**(6), 1105–1154 (2002)
14. Sabach, S., Shtern, S.: A first order method for solving convex bilevel optimization problems. *SIAM J. Optim.* **27**(2), 640–660 (2017)
15. Taylor, A.B., Hendrickx, J.M., Glineur, F.: Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Math. Program.* **161**(1–2), 307–345 (2017)
16. Ryu, E.K., Taylor, A.B., Bergeling, C., Giselsson, P.: Operator splitting performance estimation: tight contraction factors and optimal parameter selection (2018). arXiv preprint [arXiv:1812.00146](https://arxiv.org/abs/1812.00146)
17. Wittmann, R.: Approximation of fixed points of nonexpansive mappings. *Arch. der Math.* **58**(5), 486–491 (1992)
18. Xu, H.K.: Iterative algorithms for nonlinear operators. *J. Lond. Math. Soc.* **66**(1), 240–256 (2002)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.