



Proximal primal–dual best approximation algorithm with memory

E. M. Bednarczuk^{1,2} · A. Jezierska^{1,3} · K. E. Rutkowski²

Received: 16 May 2017 / Published online: 27 August 2018
© The Author(s) 2018

Abstract

We propose a new modified primal–dual proximal best approximation method for solving convex not necessarily differentiable optimization problems. The novelty of the method relies on introducing memory by taking into account iterates computed in previous steps in the formulas defining current iterate. To this end we consider projections onto intersections of halfspaces generated on the basis of the current as well as the previous iterates. To calculate these projections we are using recently obtained closed-form expressions for projectors onto polyhedral sets. The resulting algorithm with memory inherits strong convergence properties of the original best approximation proximal primal–dual algorithm. Additionally, we compare our algorithm with the original (non-inertial) one with the help of the so called *attraction property* defined below. Extensive numerical experimental results on image reconstruction problems illustrate the advantages of including memory into the original algorithm.

Keywords Proximal algorithm with memory · Primal–dual algorithm · Best approximation of the Kuhn–Tucker set · Inclusions with maximally monotone operators · Attraction property · Image reconstruction

✉ K. E. Rutkowski
k.rutkowski@mini.pw.edu.pl

E. M. Bednarczuk
Ewa.Bednarczuk@ibspan.waw.pl

A. Jezierska
Anna.Jezierska@ibspan.waw.pl

¹ Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

² Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

³ Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Gdańsk, Poland

1 Introduction

Motivated by problems arising in the field of inverse problems, signal processing, computer vision and machine learning, there has been an increasing interest in primal–dual methods [12,32,33]. Over the last years, substantial progress has been made. Among others, the recent advances concern block algorithms [18,24], asynchronous methods [21,43], generalizations of projection algorithms [22,28] and introduction of memory effect.

While versions with memory of several proximal primal–dual algorithms already exist [16,17,34,41,46], in this paper we propose a new way of introducing memory effect in projection algorithms by studying algorithm [2]. We consider the following convex optimization problem

$$\min_{p \in H} f(p) + g(Lp), \quad (1)$$

where H and G are two real Hilbert spaces, $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$, $g : G \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper convex lower semi-continuous functions and $L : H \rightarrow G$ is a bounded linear operator. Under suitable regularity conditions problem (1) is equivalent to the problem of finding $p \in H$ such that

$$0 \in \partial f(p) + L^* \partial g(Lp), \quad (2)$$

where $\partial(\cdot)$ denotes the subdifferential set-valued operator. Problem (2) is of the form

$$0 \in A(p) + L^* B(Lp), \quad (P)$$

where $A : H \rightrightarrows H$ and $B : G \rightrightarrows G$ are maximally monotone set-valued operators.

Different approaches to solve (P) have been proposed e.g. in [5,26,48]. In particular, primal–dual approaches to solve (1) may lead to formulations which can be represented as in (P), see e.g. [1,2,9,10,13,15,23] and the references therein. Recently, the primal–dual approach has been applied in [54] to a more general form of (P) involving the sum of two maximally monotone operators and a monotone operator. The case when A is maximal monotone and B is strongly monotone was considered in [48]. The overview of primal–dual approaches to solve (P) has been recently proposed in [32].

Some algorithms to solve (1) which rely on including x_{n-1} into the definition of x_{n+1} were proposed in [3,4,9,15,30,31,35–39,41,42]. They are mostly based on discretizations of the second order differential system related to the problem (2). This system, called *heavy ball with friction*, is exploited in order to accelerate convergence. Indeed, the introduction of the inertial term was shown to improve the speed of convergence significantly [30,31].

In [46] Pesquet and Pustelnik proposed a primal method to solve (1) with inertial effect introduced through inertia parameters. The method explores information from more than one previous steps and allows finding zeros of the sum of an arbitrary finite number of maximally monotone operators (see also [26]).

For monotone inclusion problems (P) inertial proximal algorithms and fixed-points iterations have been proposed in [3,4,8,9,11,34,39,40,48].

In the present paper we propose a new projection algorithm with memory. We introduce a memory effect into projection algorithms by relying on successive projections onto polyhedral sets constructed with the help of halfspaces originating from current and previous iterates. To the best of our knowledge this way of introducing memory has not been considered yet.

By applying to problem (P) the generalized Fenchel–Rockafellar duality framework [44, Corollary 2.12] (see also Corollary 2.4 of [45]) we obtain the dual inclusion problem which amounts to finding $v^* \in G$ such that

$$0 \in -LA^{-1}(-Lv^*) + B^{-1}v^*. \tag{D}$$

By [44, Corollary 2.12], a point $p \in H$ solves (P) if and only if $v^* \in G$ solves (D) and $(p, v^*) \in Z$, where

$$Z := \{(p, v^*) \in H \times G \mid -L^*v^* \in Ap \text{ and } Lp \in B^{-1}v^*\}. \tag{3}$$

In the case when $L = Id$ and $H = G$, the set Z reduces to the *extended solution set* $S_e(A, B)$ as defined in [25]. The set Z is a closed convex subset of $H \times G$ (see e.g. [7, Proposition 23.39]).

The Fenchel–Rockafellar dual problem of (1) takes the form (see [45])

$$\min_{v^* \in G} f^*(-L^*v^*) + g^*(v^*), \tag{4}$$

where f^* denotes the conjugate function [47]. In this case set Z is of the form

$$Z = \{(p, v^*) \in H \times G \mid -Lv^* \in \partial f(x) \text{ and } v^* \in \partial g(Lx)\}. \tag{5}$$

1.1 Projection methods

The idea of finding a point in Z is based on the fact that

$$Z \subset \{(p, v^*) \in H \times G \mid \varphi(p, v^*) \leq 0\} := H_\varphi,$$

where $\varphi(p, v^*) := \langle p - a \mid a^* + L^*v^* \rangle + \langle b^* - v^* \mid Lp - b \rangle$, $(a, a^*) \in \text{graph}A$, $(b, b^*) \in \text{graph}B$. This suggests the following iterative scheme for finding a point in Z based on projections onto H_φ : for any $(p_0, v_0^*) \in H \times G$ and relaxation parameters $\lambda_n \in (0, 2)$, $n \in \mathbb{N}$ let

$$(p_{n+1}, v_{n+1}^*) := (p_n, v_n) + \lambda_n (P_{H_n}(p_n, v_n^*) - (p_n, v_n^*)), \tag{6}$$

where $H_n := \{(p_n, v_n^*) \in H \times G \mid \varphi(p_n, v_n^*) \leq 0\}$ with φ_n defined for suitably chosen $(a_n, a_n^*) \in \text{graph}A$, $(b_n, b_n^*) \in \text{graph}B$ ([1, Proposition 2.3], see also [25, Lemma 3]) and $P_D(x)$ denoting the projection of x onto the set D . For $L = Id$ this iteration scheme has been proposed by Eckstein and Svaiter [25] and its fundamental convergence properties has been investigated in [25, Proposition 1, Proposition 2].

Further convergence properties of (6) have been investigated in [1] and [6, Theorem 2]. The sequence generated by (6) is Fejér monotone with respect to set Z and, in general, only its weak convergence is guaranteed.

Modifications of (6) to force strong convergence have been proposed in [2,51,52,54,56]. Recently, asynchronous block-iterative methods are proposed in [21,24].

1.2 The aim

In the present paper we propose a primal–dual projection algorithm with memory to solve (P) which relies on finding a point in the set Z defined by (3). The origin of our idea goes back to the algorithm of Haugazeau [7, Corollary 29.8], who proposed an algorithm for finding the projection of $x_0 \in H$ onto the intersection of a finite number of closed convex sets by using projections of x_0 onto intersections of two halfspaces. These halfspaces are defined on the basis of the current iterate x_n (see also [51–53]).

In our approach we take into account projections of x_0 onto intersections of three halfspaces which are defined on the basis of not only x_n but also x_{n-1} .

The contribution of the paper is as follows.

- We apply formulas for projections onto intersections of three half-spaces in Hilbert spaces derived in [50]. We show that in the considered cases (Proposition 4) the complete enumeration is not required (Proposition 5).
- We propose a number of iterative schemes with memory for solving primal–dual problems defined by (P) and (D).
- We apply our iterative schemes to propose a proximal algorithm with memory to solve minimization problem defined by a finite sum of convex functions.
- We provide convergence comparison of the proposed algorithm with its non-memory version in terms of *attraction property* (Proposition 7).
- We perform an experimental study aiming at comparing the best approximation algorithm proposed in [2] and our algorithm.

The organization of the paper is as follows. In Sect. 2 we propose the underlying iterative schemes with memory and we formulate basic convergence results. In Sect. 3 we provide several versions of the iterative scheme with memory. One of the main ingredients is a closed-form formula for projectors onto polyhedral sets introduced in [50]. In Sect. 4 we perform the convergence comparison of the proposed iterative schemes. In Sect. 5 we cast our general idea so as to be able to solve optimization problem of minimization of the sum of two convex, not necessarily differentiable functions. In Sect. 6 we present the results of the numerical experiment.

2 The proposed approach

In Sect. 2.1 we recall generic Fejér Approximation Scheme for finding an element from the set Z defined by (3) and its basic properties. In Sect. 2.2 we propose refinements of Fejér Approximation Scheme which are based on the idea proposed by Haugazeau [27], see also [7, Corollary]. The crucial issue of the proposed refinements is to improve convergence properties.

In the sequel, for any $x \in H \times G$ we write $x = (p, v^*)$, where $p \in H$ and $v^* \in G$.

2.1 Successive Fejér approximations iterative scheme

Let H, G be real Hilbert spaces and let Z be defined by (3). Let $\{H_n\}_{n \in \mathbb{N}} \subset H \times G$, be a sequence of convex closed sets such that $Z \subset H_n, n \in \mathbb{N}$. The projections of any $x \in H$ onto H_n are uniquely defined.

Iterative Scheme 1 Generic Fejér Approximation Iterative Scheme

```

Choose an initial point  $x_0 \in H \times G$ 
Choose a sequence of parameters  $\{\lambda_n\}_{n \geq 0} \in (0, 2)$ 
for  $n = 0, 1 \dots$  do
     $x_{n+1} = x_n + \lambda_n(P_{H_n}(x_n) - x_n)$ 
end for
return
    
```

Theorem 1 ([1, Proposition 3.1], see also [20]) *For any sequence generated by Iterative Scheme 1 the following hold:*

1. $\{x_n\}_{n \in \mathbb{N}} \subset H \times G$ is Fejér monotone with respect to the set Z , i.e.

$$\forall n \in \mathbb{N} \forall z \in Z \|x_{n+1} - z\| \leq \|x_n - z\|,$$

2. $\sum_{n=0}^{+\infty} \lambda_n(2 - \lambda_n) \|P_{H_n}(x_n) - x_n\|^2 < +\infty$,
3. if

$$\forall x \in H \times G \forall \{k_n\}_{n \in \mathbb{N}} \subset \mathbb{N} \ x_{k_n} \rightarrow x \implies x \in Z,$$

then $\{x_n\}_{n \in \mathbb{N}}$ converges weakly to a point in Z .

In [1] the sets H_n appearing in Iterative Scheme 1 are defined as closed halfspaces H_{a_n, b_n^*} ,

$$\begin{aligned}
 H_{a_n, b_n^*} &:= \left\{ x \in H \times G \mid \langle x \mid s_{a_n, b_n^*}^* \rangle \leq \eta_{a_n, b_n^*} \right\}, \\
 s_{a_n, b_n^*}^* &:= (a_n^* + L^* b_n^*, b_n - L a_n), \\
 \eta_{a_n, b_n^*} &:= \langle a_n \mid a_n^* \rangle + \langle b_n \mid b_n^* \rangle,
 \end{aligned}
 \tag{7}$$

with

$$\begin{aligned}
 a_n &:= J_{\gamma_n A} (p_n - \gamma_n L^* v_n^*), \quad b_n := J_{\mu_n B} (L p_n + \mu_n v_n^*), \\
 a_n^* &:= \gamma_n^{-1} (p_n - a_n) - L^* v_n^*, \quad b_n^* := \mu_n^{-1} (L p_n - b_n) + v_n^*,
 \end{aligned}$$

where for any maximally monotone operator D and constant $\xi > 0, J_{\xi D}(x) = (Id + \xi D)^{-1}(x)$. Parameters $\mu_n, \gamma_n > 0$ are suitable defined. It easy to see $H_{\varphi_n} = H_{a_n, b_n^*}$, where $\varphi_n = \varphi(a_n, b_n^*)$.

For $H_n = H_{a_n, b_n^*}$ Theorem 1 can be strengthened in following way.

Theorem 2 [1, Proposition 3.5] *For any sequence generated by Iterative Scheme 1 with H_n defined by (7) the following hold:*

1. $\{x_n\}_{n \in \mathbb{N}} = \{(p_n, v_n^*)\}_{n \in \mathbb{N}}$ is Fejér monotone with respect to the set Z ,
2. $\sum_{n=0}^{+\infty} \|a_n^* + L^* b_n^*\|^2 < +\infty$ and $\sum_{n=0}^{+\infty} \|La_n - b_n\|^2 < +\infty$,
3. $\sum_{n=0}^{+\infty} \|p_{n+1} - p_n\|^2 < +\infty$ and $\sum_{n=0}^{+\infty} \|v_{n+1}^* - v_n^*\|^2 < +\infty$,
4. $\sum_{n=0}^{+\infty} \|p_n - a_n\|^2 < +\infty$ and $\sum_{n=0}^{+\infty} \|v_n^* - b_n^*\|^2 < +\infty$,
5. $\{x_n\}_{n \in \mathbb{N}}$ converges weakly to a point in Z .

2.2 Best approximation iterative schemes

Here we study iterative best approximation schemes in the form of Iterative Scheme 2. For any $x, y \in H \times G$ we define

$$H(x, y) := \{h \in H \times G \mid \langle h - y \mid x - y \rangle \leq 0\}.$$

As previously, let $\{H_n\}_{n \in \mathbb{N}} \subset H \times G$ be a sequence of closed convex sets, $Z \subset H_n$ for $n \in \mathbb{N}$.

Iterative Scheme 2 Generic primal–dual best approximation iterative scheme

Choose an initial point $x_0 = (p_0, v_0^*) \in H \times G$

Choose a sequence of parameters $\{\lambda_n\}_{n \geq 0} \in (0, 1]$

for $n = 0, 1 \dots$ **do**

Fejérian step

$$x_{n+1/2} = x_n + \lambda_n (P_{H_n}(x_n) - x_n)$$

 Let C_n be a closed convex set such that $Z \subset C_n \subset H(x_n, x_{n+1/2})$.

Haugazeau step

$$x_{n+1} = P_{H(x_0, x_n) \cap C_n}(x_0)$$

end for

return

The choice of $C_n = H(x_n, x_{n+1/2})$ has been already investigated in [2]. There it has been shown that this choice allows to achieve strong convergence of the constructed sequence $\{x_n\}_{n \in \mathbb{N}}$ under relatively mild conditions.

Our aim is to propose and investigate other choices of C_n defined with the help of not only $x_n, x_{n+1/2}$ but also x_{n-1} and/or $x_{n-1+1/2}$. For such choices of C_n with memory the Iterative scheme 2 becomes an iterative scheme with memory, i.e. in the construction of the next iterate x_{n+1} not only current iterate x_n but also x_{n-1} is taken into account. In the sequel we refer to the Iterative Scheme 2 with $C_n = H(x_n, x_{n+1/2})$ as a scheme without memory and we compare it with Iterative Scheme 2, where C_n are with memory (see Proposition 4 below).

The Fejérian step in Iterative Scheme 2 coincides with what has been defined in Iterative Scheme 1 and was previously discussed in [1,20].



Convergence properties of sequences $\{x_n\}_{n \in \mathbb{N}}$ generated by Iterative Scheme 2 are summarized in Proposition 1 based on Proposition 2.1. of [2] which, in turn, is based on Proposition 3.1. of [19].

Proposition 1 *Let Z be a nonempty closed convex subset of $H \times G$ and let $x_0 = (p_0, v_0^*) \in H \times G$. Let $\{C_n\}_{n \in \mathbb{N}}$ be any sequence satisfying $Z \subset C_n \subset H(x_n, x_{n+1/2}), n \in \mathbb{N}$. For the sequence $\{x_n\}_{n \in \mathbb{N}}$ generated by Iterative Scheme 2 the following hold:*

1. $Z \subset H(x_0, x_n) \cap C_n$ for $n \in \mathbb{N}$,
2. $\|x_{n+1} - x_0\| \geq \|x_n - x_0\|$ for $n \in \mathbb{N}$,
3. $\sum_{n=0}^{+\infty} \|x_{n+1} - x_n\|^2 < +\infty$,
4. $\sum_{n=0}^{+\infty} \|x_{n+1/2} - x_n\|^2 < +\infty$.
5. If

$$\forall x \in H \times G \forall \{k_n\}_{n \in \mathbb{N}} \subset \mathbb{N} \quad x_{k_n} \rightarrow x \implies x \in Z,$$

then $x_n \rightarrow P_Z(x_0)$.

Proof The proof follows the lines of the proof of Proposition 2.1 of [2]. The proof of assertion 3 and 5 coincide with the respective parts of the proof of Proposition 2.1 of [2] and is omitted here. We provide the proofs of assertions 1, 2, 4 for completeness.

1. First we show that $Z \subset H(x_0, x_n)$. For $n = 0, x_1 = P_{H(x_0, x_0) \cap C_0}(x_0)$, so $Z \subset H(x_0, x_1)$. Furthermore, $H(x_0, x_n) \cap C_n \subset H(x_0, P_{H(x_0, x_n) \cap C_n}(x_0))$ and

$$\begin{aligned} Z \subset H(x_0, x_n) &\implies Z \subset H(x_0, x_n) \cap C_n \\ &\implies Z \subset H(x_0, P_{H(x_0, x_n) \cap C_n}(x_0)) \\ &\Leftrightarrow Z \subset H(x_0, x_{n+1}) \end{aligned}$$

2. By construction, for $n \in \mathbb{N}, x_{n+1} = P_{H(x_0, x_n) \cap C_n}(x_0)$ and $x_{n+1} \in H(x_0, x_n) \cap C_n$, so $x_{n+1} \in H(x_0, x_n)$. This implies $\|x_n - x_0\| \leq \|x_{n+1} - x_0\|$.
4. $C_n \subset H(x_n, P_{C_n}(x_n)) \subset H(x_n, x_n + \lambda_n(P_{C_n}(x_n) - x_n)) = H(x_n, x_{n+1/2})$. Since $x_{n+1} \in C_n \subset H_n \subset H(x_n, x_{n+1/2})$, we deduce that

$$\begin{aligned} \|x_{n+1/2} - x_n\|^2 &\leq \|x_{n+1} - x_{n+1/2}\|^2 + \|x_{n+1/2} - x_n\|^2 \\ &\leq \|x_{n+1} - x_{n+1/2}\|^2 + 2\langle x_{n+1} - x_{n+1/2} \mid x_{n+1/2} - x_n \rangle + \|x_{n+1/2} - x_n\|^2 \\ &\leq \|x_{n+1} - x_n\|^2. \end{aligned}$$

By item 3, $\sum_{n=0}^{+\infty} \|x_{n+1} - x_n\|^2 < +\infty$, hence $\sum_{n=0}^{+\infty} \|x_{n+1/2} - x_n\|^2 < +\infty$. □

Remark 1 Note that for $C_n = H(x_n, x_{n+1/2})$ and $H_n = H_{a_n, b_n^*}$ we obtain the primal–dual best approximation algorithm introduced by Alotaibi et al. in [2], involving projections onto the intersections of two halfspaces $H(x_0, x_n) \cap H(x_n, x_{n+1/2})$ studied in [7, Section 28.3]. Condition $Z \subset C_n \subset H(x_n, x_{n+1/2}), n \in \mathbb{N}$ allows one to consider choices of C_n other than $C_n = H(x_n, x_{n+1/2})$.

When $H_n := H_{a_n, b_n^*} \in \mathbb{N}$, where H_{a_n, b_n^*} are defined by (7), Proposition 1 takes the following form.

Proposition 2 *Let Z be a nonempty closed convex subset of $H \times G$ and let $x_0 = (p_0, v_0^*) \in H \times G$. Let $\{C_n\}_{n \in \mathbb{N}}$ be a sequence of closed convex sets satisfying the condition $Z \subset C_n \subset H(x_n, x_{n+1/2})$ and $H_n := H_{a_n, b_n^*} \in \mathbb{N}$. For any sequence $\{x_n\}_{n \in \mathbb{N}}$ generated by Iterative Scheme 2 the following hold:*

1. $\|x_{n+1} - x_0\| \geq \|x_n - x_0\|$ for all $n \in \mathbb{N}$,
2. $\sum_{n=0}^{+\infty} \|p_{n+1} - p_n\|^2 < +\infty$ and $\sum_{n=0}^{+\infty} \|v_{n+1}^* - v_n^*\|^2 < +\infty$,
3. $\sum_{n=0}^{+\infty} \|p_n - a_n\|^2 < +\infty$ and $\sum_{n=0}^{+\infty} \|Lp_n - b_n\|^2 < +\infty$,
4. $p_n \rightarrow \bar{x}$, $v_n^* \rightarrow \bar{v}^*$ and $(\bar{p}, \bar{v}^*) \in Z$.

Proof 1. The statement follows directly from item 2 of Proposition 1.

2. By Proposition 1,

$$\sum_{n=0}^{+\infty} \|p_{n+1} - p_n\|^2 + \sum_{n=0}^{+\infty} \|v_{n+1}^* - v_n^*\|^2 = \sum_{n=0}^{+\infty} \|x_{n+1} - x_n\|^2 < +\infty.$$

3. and 4. The proof is similar to the proof of [1, Proposition 3.5]. □

Remark 2 Proposition 2 shows the importance of the condition $Z \subset C_n \subset H(x_n, x_{n+1/2})$ in proving the strong convergence of Iterative Scheme 2.

3 The choice of C_n

One of the main contributions of the paper is to consider C_n which use the information from the previous step. In this way Iterative Scheme 2 becomes a scheme with memory in the sense that the construction of x_{n+1} depends not only on $x_{n+1/2}$, x_n , but also on $x_{n-1+1/2}$, x_{n-1} .

We start with the following propositions.

Proposition 3 *Let $x, u, v \in H$. Then $H(x, u) \cap H(x, v) \subset H(x, \tau u + (1 - \tau)v)$ for all $\tau \in [0, 1]$.*

Proof Let $h \in H(x, u) \cap H(x, v)$, i.e.

$$\langle h - u \mid x - u \rangle \leq 0 \quad \text{and} \quad \langle h - v \mid x - v \rangle \leq 0.$$

For any $\tau \in [0, 1]$ we have

$$\begin{aligned}
 & \langle h - \tau v - (1 - \tau)w \mid x - \tau v - (1 - \tau)w \rangle \\
 &= \langle h \mid x \rangle - \tau \langle h \mid v \rangle - \tau \langle v \mid x \rangle - (1 - \tau) \langle h \mid w \rangle - (1 - \tau) \langle w \mid x \rangle \\
 &\quad + \tau^2 \langle v \mid v \rangle + (1 - \tau)^2 \langle w \mid w \rangle + 2\tau(1 - \tau) \langle v \mid w \rangle \\
 &= \tau \langle h \mid x \rangle - \tau \langle h \mid v \rangle - \tau \langle v \mid x \rangle + \tau \langle v \mid v \rangle \\
 &\quad + (1 - \tau) \langle h \mid x \rangle - (1 - \tau) \langle h \mid w \rangle - (1 - \tau) \langle w \mid x \rangle + (1 - \tau) \langle w \mid w \rangle \\
 &\quad + \tau^2 \langle v \mid v \rangle + (1 - \tau)^2 \langle w \mid w \rangle + 2\tau(1 - \tau) \langle v \mid w \rangle - \tau \langle v \mid v \rangle - (1 - \tau) \langle w \mid w \rangle \\
 &\leq \tau^2 \langle v \mid v \rangle + (1 - \tau)^2 \langle w \mid w \rangle + 2\tau(1 - \tau) \langle v \mid w \rangle - \tau \langle v \mid v \rangle - (1 - \tau) \langle w \mid w \rangle \\
 &= \tau(\tau - 1) \langle v \mid v \rangle + (1 - \tau)(-\tau) \langle w \mid w \rangle + 2\tau(1 - \tau) \langle v \mid w \rangle \\
 &\leq \tau(\tau - 1) \|v\|^2 + (1 - \tau)(-\tau) \|w\|^2 + 2\tau(1 - \tau) \|v\| \|w\| \\
 &= \tau(\tau - 1) (\|v\|^2 - 2\|v\| \|w\| + \|w\|^2) = -\tau(1 - \tau) (\|v\| + \|w\|)^2 \leq 0.
 \end{aligned}$$

Thus $h \in H(x, \tau u + (1 - \tau)v)$. □

The following proposition provides examples of sets C_n with memory satisfying requirements of Proposition 2 (see Remark 2).

Proposition 4 For C_n defined as

$$C_n := H(x_n, x_{n+\frac{1}{2}}) \cap H(x_{n-1}, x_{n-\frac{1}{2}}) \text{ for } n \geq 1 \text{ and } C_0 = H(x_0, x_{1/2}), \tag{8}$$

$$C_n := H(x_n, x_{n+\frac{1}{2}}) \cap H(x_0, x_{n-1}) \text{ for } n \geq 1 \text{ and } C_0 = H(x_0, x_{1/2}), \tag{9}$$

$$\begin{aligned}
 C_n &:= H(x_n, x_{n+\frac{1}{2}}) \cap H(x_0, \tau_n x_n + (1 - \tau_n)x_{n-1}) \text{ for } \tau_n \in (0, 1), n \geq 1 \\
 &\text{and } C_0 = H(x_0, x_{1/2}) \tag{10}
 \end{aligned}$$

the assertions 1–5 of Proposition 1 holds.

Proof To apply Proposition 1 we need only to show that C_n are closed and convex and $Z \subset C_n \subset H_n$. The sets C_n are closed and convex as intersections of finitely many closed halfspaces. By construction of $x_{n+1/2}$ we have $Z \subset H(x_n, x_{n+1/2})$ for all $n \in \mathbb{N}$.

1. For C_n given by (8) we have $Z \subset H(x_n, x_{n+1/2}) \cap H(x_{n-1}, x_{n-1+1/2})$ since $Z \subset H(x_n, x_{n+1/2})$.
2. Let C_n be given by (9). By construction, $Z \subset H(x_0, x_{1/2}) = H(x_0, x_1) = C_0$. Let $n \in \mathbb{N}$ and suppose $Z \subset C_k = H(x_k, x_{k+1/2}) \cap H(x_0, x_{k-1})$ for all $1 \leq k \leq n$. We have

$$\begin{aligned}
 & Z \subset H(x_0, x_{n-1}) \cap H(x_{n-1}, x_{n-1+1/2}) \cap H(x_0, x_{n-2}) = C_n \cap H(x_0, x_{n-2}) \\
 & \implies Z \subset H(x_0, P_{C_n \cap H(x_0, x_{n-2})}(x_0)) \\
 & \Leftrightarrow Z \subset H(x_0, x_n) \Leftrightarrow Z \subset H(x_0, x_n) \cap H(x_{n+1}, x_{n+1+1/2}) = C_{n+1}.
 \end{aligned}$$

By induction, $Z \subset C_n$ for all $n \geq 0$.

3. Let C_n be given by (10). By construction, $Z \subset H(x_0, x_{1/2}) = H(x_0, x_1) = C_0$. By Proposition 3, we have

$$\begin{aligned} Z &\subset H(x_0, x_n) \cap H(x_0, x_{n-1}) \\ \implies Z &\subset H(x_0, \tau_n x_n + (1 - \tau_n)x_{n-1}) \text{ for all } \tau_n \in (0, 1). \end{aligned}$$

Let $n \in \mathbb{N}$ and suppose $Z \subset C_k = H(x_k, x_{k+1/2}) \cap H(x_0, x_{k-1})$ and $Z \subset H(x_0, x_k)$ for all $1 \leq k \leq n$. Then

$$\begin{aligned} Z \subset C_n &= H(x_n, x_{n+1/2}) \cap H(x_0, \tau_n x_n + (1 - \tau_n)x_{n-1}) \\ \implies Z &\subset H(x_n, x_{n+1/2}) \cap H(x_0, \tau_n x_n + (1 - \tau_n)x_{n-1}) \cap H(x_0, x_n) \\ \implies Z &\subset H(x_0, P_{H(x_0, H(x_n, x_{n+1/2})) \cap H(x_0, \tau_n x_n + (1 - \tau_n)x_{n-1}) \cap H(x_0, x_n)}(x_0)) \\ \Leftrightarrow Z &\subset H(x_0, P_{C_n \cap H(x_0, x_n)}(x_0)) = H(x_0, x_{n+1}) \\ \Leftrightarrow Z &\subset H(x_0, x_n) \cap H(x_{n+1}) \implies Z \subset H(x_0, \tau_{n+1} x_{n+1} + (1 - \tau_{n+1})x_n). \end{aligned}$$

Thus $Z \subset C_n$ for all $n \in \mathbb{N}$. □

3.1 Closed-form expressions for projectors onto intersection of three halfspaces

In this subsection we recall the closed-form formulas for projectors onto polyhedral sets as given in [50]. These halfspaces are given in a form

$$A_i = \{h \in H \times G \mid \langle h \mid u_i \rangle \leq \eta_i\}, \quad i = 1, \dots, m \tag{11}$$

where $u_i \neq 0, \eta_i \in \mathbb{R}, i = 1, \dots, m, m \in \mathbb{N}$.

Let $w_i := \langle x \mid u_i \rangle - \eta_i, i \in M := \{1, \dots, m\}$ and let $G := [\langle u_i \mid u_j \rangle]_{i,j \in M}$. For any sets $I \subset M, J \subset M, I, J \neq \emptyset$ the symbol $G_{I,J}$ denote the submatrix of G composed by rows indexed by I and columns indexed by J only. Let $s_I(a) := \#\{b \in I \mid b \leq a\}$. We define

$$B_I^a := \begin{cases} (-1)^{|s_I(a)|} & \text{if } a \in I, \\ (-1)^{|I|+1} & \text{if } a \notin I. \end{cases}$$

Theorem 3 [50, Theorem 2] *Let $m \in \mathbb{N}, m \neq 0$ and let $M = \{1, \dots, m\}$. Let $A = \bigcap_{i=1}^m A_i \neq \emptyset, x \notin A$. Let $\text{rank } G = k$. Let $\emptyset \neq I \subset M, |I| \leq k$ be such that $\det G_{I,I} \neq 0$. Let*

$$v_i := \begin{cases} \sum_{j \in I} w_j B_I^j B_I^i \det G_{I \setminus j, I \setminus i} & \text{if } |I| > 1, \\ w_i & \text{if } |I| = 1 \end{cases} \quad \text{for all } i \in I \tag{12}$$

and, whenever $I' := M \setminus I$ is nonempty, let

$$v_{i'} := \sum_{j \in I \cup \{i'\}} w_j B_I^j B_{I'}^{i'} \det G_{I, (I \cup I') \setminus j} \text{ for all } i' \in I'. \tag{13}$$

If $v_i > 0$ for $i \in I$ and $v_{i'} \leq 0$ for all $i' \in I'$, then

$$P_A(x) = x - \sum_{i \in I} \frac{v_i}{\det G_{I,I}} u_i. \tag{14}$$

Moreover, among all the elements of the set Δ of all subsets $I \subset M$ there exists at least one $I \in \Delta$ for which: (1) $\det G_{I,I} \neq 0$, (2) the coefficients $v_i, i \in I$ given by (12) are positive, (3) the coefficients $v_{i'}, i' \in I'$ given by (13) are nonpositive.

To obtain the closed-form expression formula for projection of a point on intersection of three halfspaces we propose the following finite algorithm for finding v_i as given in formula (14).

Iterative Scheme 3 Algorithm for finding $v = [v_i]_{i \in \{1,2,3\}}$

```

Let  $\mathcal{K}$  be a set of all nonempty subsets of  $K = \{1,2,3\}$ 
while  $\mathcal{K} \neq \emptyset$  do
  Choose randomly  $I \in \mathcal{K}$ 
  if  $\det G_{I,I} \neq 0$  then
    Find  $v = [v_i]_{i \in I}$  such that  $G_{I,I} v = [\langle x | u_i \rangle - \eta_i]_{i \in I}$ 
    if  $v > 0$  then
      if for all  $i \in K \setminus I, \langle x - \sum_{k \in I} v_k u_k | u_i \rangle - \eta_i \leq 0$  then
        Terminate, put  $v_i = 0$  for  $i \in K \setminus I$ 
      end if
    end if
  end if
   $\mathcal{K} := \mathcal{K} \setminus I$ 
end while
    
```

Note that Iterative Scheme 3 can be easily parallelized. For three halfspaces (i.e. $m = 3$ in (11)) at most 7 subsets $I \in \mathcal{K}$ need to be checked to calculate coefficients $v_i, i = 1, 2, 3$ of formula (14). Note that the above defined Iterative Scheme 3 can be useful for several algorithms, i.e. for computation of next iterate in [55].

On the other hand, when considering Iterative scheme 2 with halfspaces generated as

$$H(x_0, x_n) \cap C_n,$$

where C_n are as in Proposition 4 the number of iterations can be reduced to 4. This is the content of the following Proposition.

Let $W := H(x_0, x_n) \cap H(x_n, x_{n+1/2}) \cap A_3$, where A_3 is given by one the following

$$\begin{cases} H(x_n, x_{n+1/2}), \\ H(x_0, x_{n-1}), \\ H(x_0, \tau_n x_n + (1 - \tau_n)x_{n-1}), \tau_n \in (0, 1). \end{cases} \quad (15)$$

For simplicity, in Proposition 5 we use $H(a, b) = A_3$.

Proposition 5 For finding projection of x_0 onto W with the help of Iterative Scheme 3 at most 4 subsets $I \in \mathcal{K}$ need to be checked.

Proof We show that the projection of x_0 onto W does not require the cases $I = \{1\}$, $I = \{3\}$, $I = \{1, 3\}$ to be checked.

1. Suppose $I = \{1\}$. Then $v_1 = \frac{\langle x_0 - x_n \mid x_0 - x_n \rangle}{\|x_0 - x_n\|^2} = 1$, $\eta_2 = \langle x_{n+1/2} \mid x_n - x_{n+1/2} \rangle$ and for $2 \in K \setminus I$ we have

$$\langle x_0 - v_1(x_0 - x_n) \mid x_n - x_{n+1/2} \rangle - \eta_2 > 0.$$

2. Suppose $I = \{3\}$. Then

$$v_3 = \frac{\langle x_0 - b \mid a - b \rangle}{\|a - b\|^2}$$

and for $1 \in K \setminus I$ we have

$$\begin{aligned} & \langle x_0 - v_3(a - b) \mid x_0 - x_n \rangle - \eta_1 \\ &= \langle P_{H(a,b)}(x_0) - x_n \mid x_0 - x_n \rangle \\ &= \langle P_{H(a,b)}(x_0) - x_n \mid x_0 - P_{H(a,b)}(x_0) \rangle \\ &+ \langle P_{H(a,b)}(x_0) - x_n \mid P_{H(a,b)}(x_0) - x_n \rangle \geq 0. \end{aligned} \quad (16)$$

If equality in (16) holds then $P_{H(a,b)}(x_0) = x_n$. Then for $2 \in K \setminus I$ we have

$$\begin{aligned} & \langle x_0 - v_3(a - b) \mid x_n - x_{n+1/2} \rangle - \eta_2 \\ &= \langle P_{H(a,b)}(a, b) - x_{n+1/2} \mid x_n - x_{n+1/2} \rangle \\ &= \langle x_n - x_{n+1/2} \mid x_n - x_{n+1/2} \rangle > 0. \end{aligned}$$

3. Suppose, $I = \{1, 3\}$. Then

$$\begin{aligned} v_3 &= -\|x_0 - x_n\|^2 \langle a - b \mid x_0 - x_n \rangle + \langle x_0 - b \mid a - b \rangle \|x_0 - x_n\|^2 \\ &= \|x_0 - x_n\|^2 \langle x_n - b \mid a - b \rangle \leq 0 \end{aligned}$$

because $x_n \in H(a, b)$.

This shows that the choices $I = \{1\}$, $I = \{3\}$, $I = \{1, 3\}$ do not lead to suitable projection weights $v_i \geq 0$, $i = 1, 2, 3$. \square



4 Convergence analysis

In this section we analyse convergence properties of Iterative Scheme 2. To this aim we introduce *attraction property* (Proposition 7). The proposed results provide:

- new measure of quality of the solution generated by Iterative Scheme 2. Note that it was shown in [2,19] that with every iteration, x_n is further from x_0 . However, there was no results relating x_n and the solution $P_Z(x_0)$. By *attraction property*, the distance from x_n to the solution $P_Z(x_0)$ need not be decreasing, however, x_n remain in a ball centred at $P_Z(x_0)$ with radius which is a nonincreasing function of n (by (ii) of the Proposition 6);
- new evaluation criteria allowing to compare algorithms (we use them to compare experimentally algorithms with different choices of C_n) (Proposition 7).

We start with the following technical lemma.

Lemma 1 *Let U be a real Hilbert space and let $u_1, u_2, u_3 \in U, u_3 \in H(u_1, u_2), w = \frac{1}{2}(u_1 + u_3), r := \|w - u_1\|$. Then*

- (i) $\|w - u_2\| \leq \frac{1}{2}\|u_1 - u_3\|,$
- (ii) $\|u_2 - u_3\|^2 \leq b(u_2),$ where $b(\cdot) := 4r^2 - \|\cdot - u_1\|^2.$
- (iii) *Moreover, if $u_4 \in H$ and $u_2 \in H(u_1, u_4),$ then $b(u_2) \leq b(u_4).$*

Proof (i) We have

$$\begin{aligned} r &= \|w - u_1\| = \left\| \frac{1}{2}u_1 + \frac{1}{2}u_3 - u_1 \right\| = \frac{1}{2}\|u_3 - u_1\| \\ &= \left\| \frac{1}{2}u_1 + \frac{1}{2}u_3 - u_3 \right\| = \|w - u_3\|. \end{aligned}$$

By contradiction, suppose $\|w - u_2\| > \frac{1}{2}\|u_1 - u_3\|$. Since $u_3 \in H(u_1, u_2)$ we have

$$\begin{aligned} &\|u_3 - u_2\|^2 + \|u_1 - u_2\|^2 \\ &= \|u_3 - u_2\|^2 + \|u_1 - u_2\|^2 + 2\langle u_3 - u_2 \mid u_2 - u_1 \rangle + 2\langle u_3 - u_2 \mid u_1 - u_2 \rangle \\ &= \|u_3 - u_1\|^2 + 2\langle u_3 - u_2 \mid u_1 - u_2 \rangle \leq \|u_3 - u_1\|^2 = 4r^2. \end{aligned}$$

On the other hand

$$\begin{aligned} 4r^2 &= \|u_3 - u_1\|^2 \geq \|u_3 - u_2\|^2 + \|u_1 - u_2\|^2 \\ &= \|u_3 - w\|^2 - 2\langle u_3 - w \mid u_2 - w \rangle + \|u_2 - w\|^2 \\ &\quad + \|u_1 - w\|^2 - 2\langle u_1 - w \mid u_2 - w \rangle + \|u_2 - w\|^2 \\ &= 2r^2 + 2\|u_2 - w\|^2 - 2\langle u_3 + u_1 - 2w \mid u_2 - w \rangle > 4r^2, \end{aligned}$$

a contradiction.



(ii) We have

$$\begin{aligned}
 \|u_3 - u_2\|^2 &= \langle u_3 - u_2 \mid u_3 - u_2 \rangle = \langle u_3 - u_2 \mid u_1 - u_1 + u_3 - u_2 \rangle \\
 &= \langle u_3 - u_2 \mid u_1 - u_2 \rangle + \langle u_3 - u_2 \mid u_3 - u_1 \rangle \\
 &\leq \langle u_3 - u_2 \mid u_3 - u_1 \rangle = \langle u_3 - u_2 - u_1 + u_1 \mid u_3 - u_1 \rangle \\
 &= \langle u_3 - u_1 \mid u_3 - u_1 \rangle + \langle u_1 - u_2 \mid u_3 - u_1 \rangle \\
 &= 4r^2 + \langle u_1 - u_2 \mid u_2 - u_2 + u_3 - u_1 \rangle \\
 &= 4r^2 + \langle u_1 - u_2 \mid u_2 - u_1 \rangle + \langle u_1 - u_2 \mid u_3 - u_2 \rangle \\
 &\leq 4r^2 - \|u_1 - u_2\|^2.
 \end{aligned}$$

(iii) The assertion (iii) stems from the fact that $u_2 \in H(u_1, u_4)$ implies $\|u_1 - u_2\|^2 \geq \|u_1 - u_4\|^2$ and

$$\|u_3 - u_2\|^2 \leq \|u_1 - u_3\|^2 - \|u_1 - x_n\|^2 \leq \|u_1 - u_3\|^2 - \|u_1 - u_4\|^2.$$

□

We show that all the points $x_n, n \in \mathbb{N}$, generated by Iterative Scheme 2 are contained in the ball centred at $w := \frac{1}{2}(x_0 + \bar{x})$ with radius $r := \|w - x_0\| = \frac{1}{2}\text{dist}(x_0, Z)$ and the distance from x_n to the solution \bar{x} is bounded from above by a nonincreasing sequence.

Proposition 6 *Let $x_0 \in H \times G$. Any sequence $\{x_n\}_{n \in \mathbb{N}}$ generated by Iterative Scheme 2 satisfies the following:*

- (i) $\|w - x_n\| \leq \frac{1}{2}\|x_0 - \bar{x}\|, n \in \mathbb{N}$.
- (ii) $\|x_n - \bar{x}\|^2 \leq b_n$, where $b_n := 4r^2 - \|x_n - x_0\|^2 \geq 0, n \in \mathbb{N}$.
- (iii) *Moreover, if $x_n \in H(x_0, x_{n-1})$ for all $n \geq 1$, the sequence $\{b_n\}_{n \in \mathbb{N}}$ is nonincreasing. If for some $n \geq 1$ we have $x_{n-1} \neq x_n$, then*

$$\begin{aligned}
 \|\bar{x} - x_n\|^2 &< \|x_0 - \bar{x}\|^2 - \|x_0 - x_{n-1}\|^2, \\
 b_n &< b_{n-1}.
 \end{aligned} \tag{17}$$

Proof Let $n \in \mathbb{N}$. We have $\bar{x} \in H(x_0, x_n)$. We obtain (i) and (ii) by applying Lemma 1 with $u_1 = x_0, u_2 = x_n$ and $u_3 = \bar{x}$.

The assertion (iii) follows from (iii) of Lemma 1 with $u_1 = x_0, u_2 = x_n, u_3 = \bar{x}$ and $u_4 = x_{n-1}$. Moreover, if for some $n \geq 1$ we have $x_{n-1} \neq x_n$, then $\|x_0 - x_n\|^2 > \|x_0 - x_{n-1}\|^2$, which follows from 2 of Proposition 1. □

Let us note that Iterative Scheme 2 is sufficiently general to encompass algorithm 2.1 of [2] as well as any algorithm with memory introduced by C_n satisfying the requirements of Proposition 2. In consequence, Proposition 6 provides properties of sequences $\{x_n\}_{n \in \mathbb{N}}$ constructed in these algorithms.

Corollary 1 For any $x_0 \in H \times G$ and $x_n, x_{n+1/2}, n \geq 1$ generated by Iterative Scheme 2 we have

$$\|x_{n+1/2} - \bar{x}\|^2 \leq \|x_0 - \bar{x}\|^2 - \|x_n - x_0\|^2 - \|x_{n+1/2} - x_n\|^2$$

for any $\bar{x} \in H(x_0, x_n) \cap H(x_n, x_{n+1/2})$.

Proof Let $n \in \mathbb{N}$. Applying (ii) of Lemma 1 to $u_1 = x_0, u_2 = x_n, u_3 = \bar{x}$

$$\|x_n - \bar{x}\|^2 \leq \|x_0 - \bar{x}\|^2 - \|x_n - x_0\|^2. \tag{18}$$

Applying again (ii) of Lemma 1 to $u_1 = x_n, u_2 = x_{n+1/2}, u_3 = \bar{x}$ we obtain

$$\|x_{n+1/2} - \bar{x}\|^2 \leq \|x_n - \bar{x}\|^2 - \|x_{n+1/2} - x_n\|^2. \tag{19}$$

In consequence, we have $\|x_{n+1/2} - \bar{x}\| \leq \|x_n - \bar{x}\|$. Combining (18) and (19) we obtain

$$\begin{aligned} \|x_{n+1/2} - \bar{x}\|^2 &\leq \|x_n - \bar{x}\|^2 - \|x_{n+1/2} - x_n\|^2 \\ &\leq \|x_0 - \bar{x}\|^2 - \|x_n - x_0\|^2 - \|x_{n+1/2} - x_n\|^2. \end{aligned}$$

□

To prove Proposition 7, which is our main result in this section we need the following Lemma.

Lemma 2 Let U be a real Hilbert space and let $D \subset U$ be a nonempty subset of U . Let $u_1, u_2, u_3, u_4 \in U$ and $u_3 \in H(u_1, u_2) \cap H(u_2, u_4) \cap D, w = \frac{1}{2}(u_1 + u_3), r := \|w - u_1\|$.

Let $\bar{q} \in H(u_1, u_2) \cap H(u_2, u_4) \cap D$. Then $\bar{q} \in H(u_1, q)$, where $q = Q(u_1, u_2, u_4) := P_{H(u_1, u_2) \cap H(u_2, u_4)}(u_1)$ and

$$\begin{aligned} \|u_1 - \bar{q}\|^2 &\geq \|u_1 - q\|^2 + \|\bar{q} - q\|^2, \\ \|u_3 - \bar{q}\|^2 &\leq 4r^2 - \|u_1 - \bar{q}\|^2 \leq 4r^2 - \|u_1 - q\|^2 - \|\bar{q} - q\|^2. \end{aligned}$$

Proof It is immediate that $\bar{q} \in H(u_1, q)$. Thus

$$\|u_1 - \bar{q}\|^2 = \|u_1 - q\|^2 + 2\langle u_1 - q | q - \bar{q} \rangle + \|\bar{q} - q\|^2 \geq \|u_1 - q\|^2 + \|\bar{q} - q\|^2.$$

By Lemma 1, since $u_3 \in H(u_1, \bar{q})$, we have $\|u_3 - \bar{q}\| \leq 4r^2 - \|u_1 - \bar{q}\|^2$ which completes the proof. □

To compare best approximation algorithms as defined in [2] with the Iterative Scheme 2 with memory we concentrate on single step gains. To this end let us denote $q_n := P_{D(n)}(x_0), x_n := P_{D(n-1, n)}(x_0)$, where $D(n) = H(x_0, x_{n-1}) \cap H(x_{n-1}, x_{n-1+1/2})$ as e.g. in [2] and $D(n-1, n) = H(x_0, x_{n-1}) \cap H(x_{n-1}, x_{n-1+1/2}) \cap C_{n-1}$ with C_{n-1} as in Proposition 4.

Proposition 7 (Attraction property) *Sequences $\{x_n\}_{n \in \mathbb{N}}$, $\{x_{n+1/2}\}_{n \in \mathbb{N}}$ generated by Iterative Scheme 2 satisfy the following:*

- (i) $\|x_0 - x_n\|^2 \geq \|x_0 - q_n\|^2 + \|x_n - q_n\|^2$,
- (ii) $\|\bar{x} - x_n\|^2 \leq \|x_0 - \bar{x}\|^2 - \|x_0 - x_n\|^2 \leq \|x_0 - \bar{x}\|^2 - \|x_0 - q_n\|^2 - \|x_n - q_n\|^2$,

where $q_n := P_{H(x_0, x_{n-1}) \cap H(x_{n-1}, x_{n-1+1/2})}(x_0)$ and $\bar{x} = P_Z(x_0)$.

Proof The proof follows directly from Lemma 2 with $u_1 = x_0$, $u_2 = x_{n-1}$, $u_3 = \bar{x}$, $u_4 = x_{n-1+1/2}$, $\bar{q} = P_{H(x_0, x_{n-1}) \cap C_{n-1}}(x_0)$ and $D = C_{n-1}$. \square

Let us note that, in the case when $x_n = q_n$, by (ii), we have $\|\bar{x} - q_n\|^2 \leq \|x_0 - \bar{x}\|^2 - \|x_0 - q_n\|^2$. Hence, in the Iterative Scheme 2 we are interested in choices of C_n which make the difference $x_n - q_n$ large. Note that in case of $C_{n-1} = H(x_{n-1}, x_{n-1+1/2})$, $\|x_n - q_n\|^2 = 0$. For other choices of C_{n-1} the worst case leads to $\|x_n - q_n\|^2 = 0$, however, we can expect some improvement. Consequently, the proposed *attraction property* may serve as an evaluation criterion for comparing various versions of Iterative Scheme 2.

5 Proximal algorithms

Let H and G be real Hilbert spaces, let $f : H \rightarrow (-\infty, +\infty]$ and $g : G \rightarrow (-\infty, +\infty]$ be proper lower semicontinuous convex functions and let $L : H \rightarrow G$ be a bounded linear operator. Iterative Scheme 4 defined below is an application of Iterative Scheme 2 to optimization problem (1)–(4), i.e. we consider the pair of problems,

$$\min_{p \in H} F_P(p) := f(p) + g(Lp) \quad (20)$$

and the dual problem to (20),

$$\min_{v^* \in G} F_D(v^*) := f^*(-L^*v^*) + g^*(v^*). \quad (21)$$

If (20) has a solution $\bar{p} \in H$ and the regularity condition holds, e.g.

$$0 \in \text{sqli}(\text{dom } g - L(\text{dom } f)),$$

where dom denotes the effective domain of a function and for any convex closed set S

$$\text{sqli}S := \left\{ x \in S \mid \bigcup_{\lambda > 0} \lambda(S - x) \text{ is a closed linear subspace of } H \right\},$$

there exists $\bar{v}^* \in G$ solving (21) and

$$(\bar{p}, \bar{v}^*) \in Z = \{(p, v^*) \in H \times G \mid -Lv^* \in \partial f(x) \text{ and } v^* \in \partial g(Lx)\}. \quad (22)$$



Conversely, if $(\bar{p}, \bar{v}^*) \in Z$, then \bar{p} solves (20) and \bar{v}^* solves (21). The set Z defined by (22) is of the form (3), when $A = \partial f$ and $B = \partial g$.

Recall that for any $x \in H$ and any proper convex and lower semi-continuous function $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ the *proximity operator* $\text{Prox}_f(x)$ is defined as the unique solution to the optimization problem

$$\min_{y \in H} \left(f(y) + \frac{1}{2} \|x - y\|^2 \right).$$

Theorem 4 [7, Example 23.3] *Let $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper convex lower semi-continuous function, $x \in H$ and $\gamma > 0$. Then*

$$J_{\gamma \partial f}(x) = \text{Prox}_{\gamma f}(x).$$

Iterative Scheme 4 Proximal primal–dual best approximation iterative scheme

Choose an initial point $x_0 = (p_0, v_0^*) \in H \times G$ and $\varepsilon > 0$

Choose sequences of parameters $\{\lambda_n\}_{n \geq 0} \in (0, 1]$ and $\{\gamma_n\}_{n \geq 0}, \{\mu_n\}_{n \geq 0} \in [\varepsilon, 1/\varepsilon]$

for $n = 0, 1 \dots$ **do**

$$a_n = \text{Prox}_{\gamma_n f}(p_n - \gamma_n L^* v_n^*)$$

$$b_n = \text{Prox}_{\mu_n g}(L p_n + \mu_n v_n^*)$$

$$a_n^* = \gamma_n^{-1} (p_n - a_n) - L^* v_n^*$$

$$b_n^* = \mu_n^{-1} (L p_n - b_n) + v_n^*$$

$$s_n^* = (a_n^* + L^* b_n^*, b_n - L a_n)$$

$$\eta_n = (a_n \mid a_n^*) + (b_n \mid b_n^*)$$

$$H_n = \left\{ x \in H \times G \mid (x \mid s_{a_n, b_n}^*) \leq \eta_{a_n, b_n} \right\}$$

if $\|s_n^*\| = 0$ **then**

$$\bar{x} = x_n, \bar{v}^* = v_n^*$$

Terminate

else

Fejérian step

$$x_{n+1/2} = x_n + \lambda_n (P_{H_n}(x_n) - x_n)$$

Haugazeau step

Choose C_n closed convex such that $Z \subset C_n \subset H(x_n, x_{n+1/2})$

$$x_{n+1} = P_{H(x_0, x_n) \cap C_n}(x_0)$$

end if

end for

return

Convergence properties of Iterative Scheme 4 are summarized in Proposition 2.



5.1 Generalization to finite number of functions

Let M and K be natural numbers. Let $E = \bigoplus_{i=1}^M H_i \times \bigoplus_{k=1}^K G_k$, where H_i, G_k are real Hilbert spaces, $i = 1, \dots, M, k = 1, \dots, K$. Let $f_i : H_i \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g_k : G_k \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper lower semicontinuous convex functions and $L_{ik} : H_i \rightarrow G_k$ be bounded linear operators, $i = 1, \dots, M, k = 1, \dots, K$. Consider the primal problem

$$\min_{p_1 \in H_1, \dots, p_M \in H_M} \sum_{i=1}^M f_i(p_i) + \sum_{k=1}^K g_k \left(\sum_{i=1}^M L_{ik} p_i \right). \quad (23)$$

Problem formulation (23) is general enough to cover problem arising in diverse applications including signal and image reconstruction, compressed sensing and machine learning [29]. The dual problem to (23) is

$$\min_{v_1^* \in G_1, \dots, v_K^* \in G_K} \sum_{i=1}^M f_i^* \left(- \sum_{k=1}^K L_{ki}^* v_k^* \right) + \sum_{k=1}^K g_k^*(v_k^*). \quad (24)$$

Assume that

$$(\forall i \in \{1, \dots, M\}) \quad 0 \in \text{ran} \left(\partial f_i + \sum_{k=1}^K L_{ki}^* \circ \partial g_k \circ L_{ik} \right),$$

where $\text{ran} D$ denotes the range of an operator D .

Then the set

$$Z := \left\{ (p_1, \dots, p_M, v_1^*, \dots, v_K^*) \in E \mid - \sum_{k=1}^K L_{ki}^* v_k^* \in \partial f_i(p_i), \right. \\ \left. \sum_{i=1}^M L_{ik} p_i \in \partial g_k^*(v_k^*), i = 1, \dots, M, k = 1, \dots, K \right\} \quad (25)$$

is nonempty and if $(\bar{p}_1, \dots, \bar{p}_M, \bar{v}_1^*, \dots, \bar{v}_K^*) \in Z$ then $(\bar{p}_1, \dots, \bar{p}_M)$ solves (23) and $(\bar{v}_1^*, \dots, \bar{v}_K^*)$ solves (24). To find an element of set Z defined by (25) we propose the Iterative Scheme 5.

Iterative Scheme 5 Proximal primal–dual best approximation iterative scheme for finite number of functions

Choose an initial point $x_0 = (p_0, v_0^*) \in \bigoplus_{i=1}^M H_i \times \bigoplus_{k=1}^K G_k$ and $\varepsilon > 0$,
 $p_0 = (p_{1,0}, \dots, p_{M,0}), v_0^* = (v_{1,0}^*, \dots, v_{K,0}^*)$
 Choose sequences of parameters $\{\lambda_n\}_{n \geq 0} \in (0, 1]$ and $\{\gamma_n\}_{n \geq 0}, \{\mu_n\}_{n \geq 0} \in [\varepsilon, 1/\varepsilon]$
for $n = 0, 1 \dots$ **do**
 Fejerian step
 for $i = 1, \dots, M$ **do**
 $a_{i,n} = \text{Prox}_{\gamma_n f_i}(p_{i,n} - \gamma_n \sum_{k=1}^K L_{ki}^* v_{k,n}^*)$
 $a_{i,n}^* = \gamma_n^{-1}(p_{i,n} - a_{i,n}) - \sum_{k=1}^K L_{ki}^* v_{k,n}^*$
 end for
 for $k = 1, \dots, K$ **do**
 $b_{k,n} = \text{Prox}_{\mu_n g_k}(\sum_{i=1}^M L_{ik} p_{i,n} + \mu_n v_{k,n}^*)$
 $b_{k,n}^* = \mu_n^{-1}(\sum_{i=1}^M L_{ik} p_{i,n} - b_{k,n}) + v_{k,n}^*$
 $s_{M+k,n}^* = b_{k,n} - \sum_{i=1}^M L_{ik} a_{i,n}$
 end for
 for $i = 1, \dots, M$ **do**
 $s_{i,n}^* = a_{i,n}^* + \sum_{k=1}^K L_{ki}^* b_{k,n}^*$
 end for
 $s_n^* = (s_{1,n}^*, \dots, s_{M,n}^*, s_{M+1,n}^*, \dots, s_{M+K,n}^*)$
 $\eta_n = \sum_{i=1}^M \langle a_{i,n} | a_{i,n}^* \rangle + \sum_{k=1}^K \langle b_{k,n} | b_{k,n}^* \rangle$
 $H_n = \{h \in E \mid \langle h | s_n^* \rangle \leq \eta_n\}$
 if $\|s_n^*\| = 0$ **then**
 $\bar{p} = p_n, \bar{v}^* = v_n^*$
 Terminate
 else
 $x_{n+1/2} = x_n + \lambda_n(P_{H_n}(x_n) - x_n)$
 Haugazeau step
 Choose C_n closed convex such that $Z \subset C_n \subset H(x_n, x_{n+1/2})$
 $x_{n+1} = P_{H(x_0, x_n) \cap C_n}(x_0)$
 end if
end for
return

Let us note that Proposition 2 can be easily generalized to cover also the case of the set Z defined by (25).

6 Experimental results

The goal of this section is to illustrate and analyze the performance of the proposed Iterative Scheme 5 in solving problem (23), i.e. we aim at illustrating the main contribution of our work: (a) to show experimentally the influence of the choice of set C_n on the convergence of the algorithm and (b) to show experimentally that the proposed attraction property provide an additional measure of the distance of the current iterate to the solution. We provide numerical results related to simple convex image inpainting problem. The considered problem can be rewritten as an instance of (23) by setting $M = 1, K = 2, H = \mathbb{R}^{3D}$ and finding $\min_{p \in H} f_1(p) + \sum_{k=1}^2 g_k(L_k p)$, where functions f_1, g_1 and g_2 correspond to positivity constraint, data fidelity term and total variation (TV) based regularization [49], respectively. We focus on the analysis of

influence of the choice of C_n on the convergence. To this end we report the number of iterations of the algorithm with different C_n settings performed to reach a tolerance $\|p_{n+1} - p_n\| / (1 + \|p_n\|)$ less than ϵ in two successive iterations. The considered algorithms are denoted hereafter by PDBA-C0, PDBA-C1, PDBA-C2, PDBA-C3, for $C_n = H(x_n, x_{n+1/2})$ and C_n defined by (8)–(10), respectively. In the case of PDBA-C3 τ_n is set to 0.5. Numerically, the convergence rate improvement is measured by ItR defined as a ratio of the numbers of iterations consumed by PDBA-Ci (where i takes value 0, 1, 2, 3) and those consumed by PDBA-C0. The algorithms performance is illustrated by the following curves: (a) signal to noise ratio (SNR) and (b) the bounds given by Proposition 6 as a functions of iteration number.

The evaluation experiments concern the image inpainting problem which corresponds to the recovery of an image $\bar{p} \in \mathbb{R}^{3D}$ from lossy observations $y = L_1 \bar{p}$, where $L_1 \in \mathbb{R}^{3D \times 3D}$ is a diagonal matrix such that for $i = 1, \dots, D$ we have

Table 1 Reconstruction results from incomplete data with $\epsilon = 10^{-2}$, $\lambda_n = 1$, $\gamma_n = 0.005$, $\mu_n = 0.005$

	PDBA-C0	PDBA-C1	PDBA-C2	PDBA-C3
$\kappa = 20\%$				
ItR	1	0.40	1.02	1.02
SNR	24.19	24.25	24.18	24.19
TV	40.66	40.37	40.68	40.67
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^{3D}}$	0.004	0.004	0.004	0.004
$\kappa = 40\%$				
ItR	1	0.51	0.98	1.02
SNR	20.64	20.60	20.64	20.64
TV	36.06	36.04	36.05	36.07
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^{3D}}$	0.002	0.003	0.002	0.002
$\kappa = 60\%$				
ItR	1	0.44	1.01	0.36
SNR	18.26	18.28	18.26	17.90
TV	30.87	30.55	30.87	30.80
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^{3D}}$	0.002	0.002	0.002	0.002
$\kappa = 80\%$				
ItR	1	0.49	0.89	0.99
SNR	16.17	16.18	16.18	16.17
TV	23.79	23.50	23.74	23.79
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^{3D}}$	0.001	0.001	0.001	0.001
$\kappa = 90\%$				
ItR	1	0.51	1.08	0.99
SNR	14.71	14.62	14.70	14.71
TV	18.87	18.13	18.94	18.87
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^{3D}}$	0.001	0.001	0.001	0.001

$L_1(i, i) = L_1(2i, 2i) = L_1(3i, 3i) = 0$, if the pixel i in the observation image y is lost and $L_1(i, i) = L_1(2i, 2i) = L_1(3i, 3i) = 1$, otherwise. The considered optimization problem is of the form

$$\min_{p \in H} \iota_y(L_1 p) + \iota_S(p) + TV(p) \tag{26}$$

where ι is the indicator function defined as:

$$\iota_S(p) = \begin{cases} 0 & \text{if } p \in S \\ +\infty & \text{otherwise,} \end{cases} \tag{27}$$

Table 2 Reconstruction results from incomplete data with $\epsilon = 10^{-2}$, $\lambda_n = 1$, $\gamma_n = 0.01$, $\mu_n = 0.01$

	PDBA-C0	PDBA-C1	PDBA-C2	PDBA-C3
$\kappa = 20\%$				
ItR	1	0.44	1	1
SNR	24.02	24.13	24.02	24.02
TV	40.06	39.66	40.06	40.06
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^3 D}$	0.005	0.005	0.005	0.005
$\kappa = 40\%$				
ItR	1	0.54	1.02	1.03
SNR	20.52	20.56	20.51	20.53
TV	35.62	35.34	35.67	35.61
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^3 D}$	0.004	0.004	0.004	0.004
$\kappa = 60\%$				
ItR	1	0.52	1.04	0.91
SNR	18.24	18.25	18.22	18.23
TV	30.19	29.97	30.37	30.17
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^3 D}$	0.003	0.003	0.003	0.003
$\kappa = 80\%$				
ItR	1	0.59	1.21	1.02
SNR	16.15	16.16	16.15	16.15
TV	23.29	22.92	23.48	23.3202
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^3 D}$	0.002	0.002	0.002	0.002
$\kappa = 90\%$				
ItR	1	0.71	1.52	0.59
SNR	14.67	14.65	14.65	14.33
TV	18.11	17.87	18.62	16.01
$\frac{\ y - L_1 p\ _1}{(1-\bar{\kappa})^3 D}$	0.001	0.001	0.001	0.002

$TV : \mathbb{R}^{3D} \mapsto \mathbb{R}$ is a discrete isotropic total variation functional [49], i.e. for every $p \in \mathbb{R}^{3D}$, $TV(p) = g(L_2 p) := \omega \left(\sum_{d=1}^D ([\Delta^h p]_d)^2 + ([\Delta^v p]_d)^2 \right)^{1/2}$ with $L_2 \in \mathbb{R}^{6D \times 3D}$, $L_2 := [(\Delta^h)^\top (\Delta^v)^\top]^\top$, where $\Delta^h \in \mathbb{R}^{3D \times 3D}$ (resp. $\Delta^v \in \mathbb{R}^{3D \times 3D}$) corresponds to a horizontal (resp. vertical) gradient operator,

$$[\Delta^h p]_d := [(\Delta^h p)_d, (\Delta^h p)_{2d}, (\Delta^h p)_{3d}] \in \mathbb{R}^3,$$

$$[\Delta^v p]_d := [(\Delta^v p)_d, (\Delta^v p)_{2d}, (\Delta^v p)_{3d}] \in \mathbb{R}^3$$

and ω denotes regularization parameter. The function $\iota_{\mathcal{S}}(p)$ is imposing the solution to belong to the set $\mathcal{S} = [0, 1]^{3D}$. The dual problem to (26) is the following optimization problem [14, Example 3.24, 3.26, 3.27]:

Table 3 Reconstruction results from incomplete data with $\epsilon = 10^{-2}$, $\lambda_n = 1$, $\gamma_n = 1.5$, $\mu_n = 1.5$.

	PDBA-C0	PDBA-C1	PDBA-C2	PDBA-C3
$\kappa = 20\%$				
ItR	1	0.75	1	1
SNR	23.00	23.03	23.00	23.00
TV	34.91	34.99	34.91	34.91
$\frac{\ y - L_1 p\ _1}{(1-\kappa)^{3D}}$	0.011	0.011	0.011	0.011
$\kappa = 40\%$				
ItR	1	0.70	1.02	1.03
SNR	20.01	20.00	20.01	20.01
TV	31.19	31.17	31.19	31.19
$\frac{\ y - L_1 p\ _1}{(1-\kappa)^{3D}}$	0.008	0.009	0.008	0.008
$\kappa = 60\%$				
ItR	1	0.69	0.99	1.00
SNR	17.90	17.89	17.89	17.89
TV	26.60	26.63	26.59	26.59
$\frac{\ y - L_1 p\ _1}{(1-\kappa)^{3D}}$	0.006	0.006	0.006	0.006
$\kappa = 80\%$				
ItR	1	0.73	1	1.02
SNR	15.87	15.87	15.87	15.87
TV	20.46	20.52	20.48	20.49
$\frac{\ y - L_1 p\ _1}{(1-\kappa)^{3D}}$	0.004	0.004	0.004	0.004
$\kappa = 90\%$				
ItR	1	0.74	0.95	1.02
SNR	14.33	14.35	14.31	14.33
TV	15.99	16.24	15.79	16.01
$\frac{\ y - L_1 p\ _1}{(1-\kappa)^{3D}}$	0.002	0.002	0.002	0.002

$$\min_{v_1 \in G_1, v_2 \in G_2} (y | v_1) + \sup_{r \in S} (r | -L_1^* v_1 - L_2^* v_2) + TV^*(v_2) \tag{28}$$

where $TV^*(v_2) = \omega l_B(\frac{v_2}{\omega})$, convex set $B = \{v \in \mathbb{R}^{6D} : \|v\|_2 \leq 1\}$, $G_1 = \mathbb{R}^{3D}$, $G_2 = \mathbb{R}^{6D}$. In the following experiments, we consider the cases of lossy observations with κ randomly chosen pixels which are unknown.

In the following we examine the cases of κ set to 20%, 40%, 60%, 80%, 90% (hereafter $\tilde{\kappa}$ denotes a fraction of missing pixels). For all the algorithms we used the initialization $x_0 = [y, L_1 y, L_2 y]^T$. The test were performed on image `fruits` from public domain (source: www.hlevkin.com/TestImages) of size $D = 240 \times 256$.

In our first experiment, we study the influence of the choice of C_n for different settings of γ_n, μ_n and λ_n , which play a significant role in convergence analysis. The results summarized in Tables 1, 2 and 3, correspond to the choice of $\gamma_n = \mu_n$ equal to 0.005, 0.01 and 1.5, respectively. These results show that independently of the choice of parameters γ_n, μ_n algorithm PDBA-C1 leads to the best performance, while the results obtained with PDBA-C2 and PDBA-C3 are comparable to PDBA-C0. Specifically, within our setting the numbers of iterations consumed by PDBA-C1 range from 40 to 75% of those consumed by PDBA-C0, while the SNR, values of TV and inpainting residues are negligible. By inspecting Tables 1, 2 and 3, one can observe that the obtained results depend strongly upon to the choice of γ_n and μ_n .

Table 4 Reconstruction results from incomplete data with $\epsilon = 10^{-2}, \kappa = 20\%$

γ_n, μ_n	0.003	0.005	0.01	1.5	100
<i>PDBA-C0</i>					
SNR	7.28	24.19	24.02	23.00	7.64
TV	130.29	40.66	40.06	34.91	120.99
$\frac{\ y-L_1 p\ _1}{(1-\tilde{\kappa})3D}$	0.003	0.004	0.005	0.011	0.008
It no.	2	7543	6382	2883	5032
<i>PDBA-C1</i>					
SNR	7.28	24.25	24.13	23.03	7.55
TV	130.29	40.37	39.66	34.99	122.89
$\frac{\ y-L_1 p\ _1}{(1-\tilde{\kappa})3D}$	0.003	0.004	0.005	0.011	0.007
ItR	1	0.40	0.44	0.75	0.83

Table 5 Reconstruction results from incomplete coefficients with $\epsilon = 10^{-2}, \kappa = 20\%, \gamma_n = 0.005, \mu_n = 0.005$

λ_n	PDBA-C0				PDBA-C1			
	SNR0	TV	$\frac{\ y-L_1 p\ _1}{(1-\tilde{\kappa})3D}$	It0	SNR1	TV	$\frac{\ y-L_1 p\ _1}{(1-\tilde{\kappa})3D}$	It1 / It0
1	24.19	40.66	0.004	7543	24.25	40.37	0.004	0.40
0.95	24.32	40.36	0.003	5149	24.25	40.33	0.004	0.56
0.9	24.24	40.37	0.004	4584	24.26	40.26	0.004	0.57
0.8	7.30	129.24	0.005	2	7.30	129.24	0.005	1

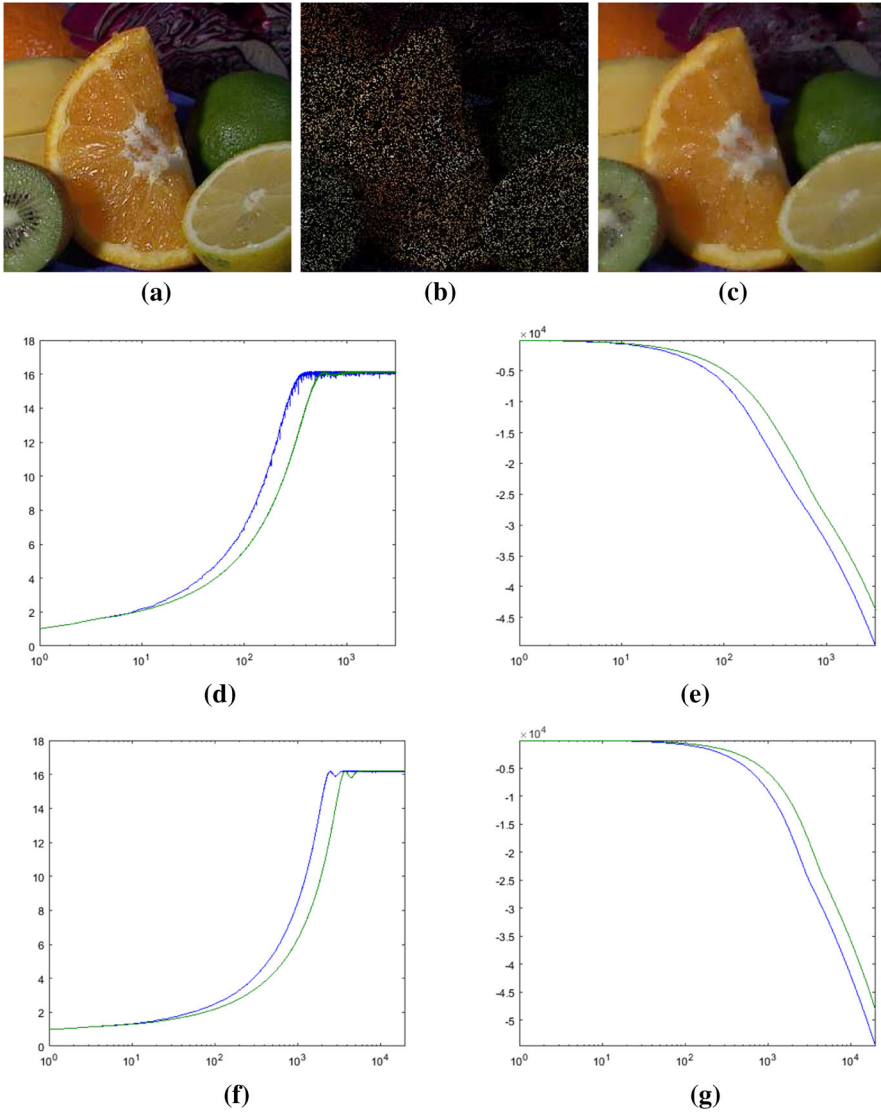


Fig. 1 **a** The 240×256 clean fruits image, **b** the same image for which 80% randomly chosen missing pixels, and **c** the solution generated by Algorithm 5 PDBA-C1 after 3000 iterations, **(d–e)** and **(f–g)** show SNR and *attraction property* values (i.e. $\|(p_0, v_0^*) - (p_n, v_n^*)\|$) versus iterations, respectively. Algorithm PDBA-C0 and PDBA-C1 are denoted in green and blue, respectively. **a** Original. **b** Degraded. **c** Ours result. **d** SNR ($\gamma_n = 0.01, \mu_n = 0.01$). **e** Bounds ($\gamma_n = 0.01, \mu_n = 0.01$). **f** SNR ($\gamma_n = 0.003, \mu_n = 0.003$). **g** Bounds ($\gamma_n = 0.003, \mu_n = 0.003$)

We would like to emphasize that ideally the termination tolerance should be a function of parameters γ_n , μ_n and λ_n . The results presented in Table 4 shows that in the case of γ_n and μ_n equal to 0.003 or 100 the tolerance should be smaller to prevent premature termination. In these cases the iteration number is very low, however the values of TV and SNR are significantly different than for the other choices. The premature termination is due to flat slope of the convergence curve. Similar effect can be observed when $\lambda_n = 0.8$ (see Table 5).

In the second experiment, we compare PDBA-C0 and PDBA-C1 (best over Algorithms with memory according to the first experiment). We present reconstruction results (see Table 4) as well as supplying convergence curves (see Fig. 1), i.e. SNR and bound as a function of iterations. Hereafter we call bounds as $-\|(p_0, v_0^*) - (p_n, v_n^*)\|^2 = -\|x_0 - x_n\|^2$ (see (ii) of Proposition 7). One can observe that PDBA-C1 leads to a faster convergence and the bounds are more tight (in the sense of Proposition 7 (ii)). The difference is the most important in the early stage of the iterations. Both algorithms slow down afterwards. For $\gamma_n = 0.01$ (resp. $\mu_n = 0.01$) both versions of the algorithm lead to some numerical oscillations in convergence, which are no more visible for settings $\gamma_n = 0.003$ (resp. $\mu_n = 0.003$).

7 Conclusions

In this paper we concentrate on a design of the novel scheme by incorporating memory into projection algorithm. We propose a new way of introducing memory effect into projection algorithms through incorporating into the algorithm projections onto polyhedral sets built as intersections of halfspaces constructed with the help of current and previous iterates. To this end we provide the closed-form expressions and the algorithm for finding projections onto intersections of three halfspaces. Moreover, we adapt the general scheme proposed in [50] to particular halfspaces which may arise in the course of our Iterative Scheme. This allows us to limit the number of steps for finding projections. Building upon these results, we propose a new primal–dual splitting algorithm with memory for solving convex optimization problems via general class of monotone inclusion problems involving parallel sum of maximally monotone operators composed with linear operators. To analyse convergence we prove the *attraction property*. The *attraction property* provides us with an evaluation criterion allowing to compare projection algorithms with and without memory. Our experimental results related to preliminary implementation of the algorithms have shown that the proposed algorithm with memory generally needs smaller number of iterations than the corresponding original one [2]. Although only three strategies of introducing memory effect are analysed in this work, the generality of the presented theoretical results allow us to address versatility of the approach by constructing various forms of the algorithm which use information from former steps.

Acknowledgements The authors would like to thank the anonymous reviewers for their careful and thoughtful comments which improved the quality of the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alotaibi, A., Combettes, P.L., Shahzad, N.: Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn–Tucker set. *SIAM J. Optim.* **24**(4), 2076–2095 (2014)
- Alotaibi, A., Combettes, P.L., Shahzad, N.: Best approximation from the Kuhn–Tucker set of composite monotone inclusions. *Numer. Funct. Anal. Optim.* **36**(12), 1513–1532 (2015)
- Alvarez, F.: Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space. *SIAM J. Optim.* **14**(3), 773–782 (2004) (**electronic**)
- Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.* **9**(1), 3–11 (2001)
- Attouch, H., Briceno-Arias, L.M., Combettes, P.L.: A parallel splitting method for coupled monotone inclusions. *SIAM J. Control Optim.* **48**(5), 3246–3270 (2010)
- Bauschke, H.H.: A note on the paper by Eckstein and Svaiter on “general projective splitting methods for sums of maximal monotone operators”. *SIAM J. Control Optim.* **48**(4), 2513–2515 (2009)
- Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* (CMS Books in Mathematics). Springer, Berlin (2011)
- Boğ, R.I., Csetnek, E.R.: An inertial alternating direction method of multipliers. *Minimax Theory Appl.* **1**(1), 29–49 (2016)
- Boğ, R.I., Csetnek, E.R.: An inertial forward–backward–forward primal–dual splitting algorithm for solving monotone inclusion problems. *Numer. Algorithms* **71**(3), 519–540 (2016)
- Boğ, R.I., Csetnek, E.R., Heinrich, A., Hendrich, C.: On the convergence rate improvement of a primal–dual splitting algorithm for solving monotone inclusion problems. *Math. Program.* **150**(2), 251–279 (2015)
- Boğ, R.I., Csetnek, E.R., Hendrich, C.: Inertial Douglas–Rachford splitting for monotone inclusion problems. *Appl. Math. Comput.* **256**, 472–487 (2015)
- Boğ, R.I., Hendrich, C.: A double smoothing technique for solving unconstrained nondifferentiable convex optimization problems. *Comput. Optim. Appl.* **54**(2), 239–262 (2013)
- Boğ, R.I., Hendrich, C.: A Douglas–Rachford type primal–dual method for solving inclusions with mixtures of composite and parallel–sum type monotone operators. *SIAM J. Optim.* **23**(4), 2541–2565 (2013)
- Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, New York (2004)
- Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal–dual algorithm. *Math. Program.* **159**, 253–287 (2015)
- Chen, C., Chan, R.H., Ma, S., Yang, J.: Inertial proximal ADMM for linearly constrained separable convex optimization. *SIAM J. Imaging Sci.* **8**(4), 2239–2267 (2015)
- Chen, C., Ma, S., Yang, J.: A general inertial proximal point algorithm for mixed variational inequality problem. *SIAM J. Optim.* **25**(4), 2120–2142 (2015)
- Chouzenoux, E., Pesquet, J.-C., Repetti, A.: A block coordinate variable metric forward–backward algorithm. *J. Glob. Optim.* **66**(3), 457–485 (2016)
- Combettes, P.L.: Strong convergence of block-iterative outer approximation methods for convex optimization. *SIAM J. Control Optim.* **38**(2), 538–565 (2000)
- Combettes, P.L.: *Encyclopedia of Optimization*. In: Fejér monotonicity in convex optimization. Springer, Boston, pp. 1016–1024 (2009)
- Combettes, P.L., Eckstein, J.: Asynchronous block-iterative primal–dual decomposition methods for monotone inclusions. *Math. Program.* **168**, 645–672 (2016)
- Combettes, P.L., Nguyen, Q.V.: Solving composite monotone inclusions in reflexive Banach spaces by constructing best Bregman approximations from their Kuhn–Tucker set. *arXiv* (2015)

23. Combettes, P.L., Pesquet, J.-C.: Primal–dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.* **20**(2), 307–330 (2012)
24. Eckstein, J.: A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers. *J. Optim. Theory Appl.* **173**(1), 155–182 (2017)
25. Eckstein, J., Svaiter, B.F.: A family of projective splitting methods for the sum of two maximal monotone operators. *Math. Progr.* **111**(1), 173–199 (2008)
26. Eckstein, J., Svaiter, B.F.: General projective splitting methods for sums of maximal monotone operators. *SIAM J. Control Optim.* **48**(2), 787–811 (2009)
27. Haugazeau, Y.: *Sur les Inequations Variationnelles et la Minimisation de Fonctionnelles Convexes*. Ph.D. thesis, Université de Paris (1968)
28. He, N., Juditsky, A., Nemirovski, A.: Mirror prox algorithm for multi-term composite minimization and semi-separable problems. *Comput. Optim. Appl.* **61**(2), 275–319 (2015)
29. Hendrich, C.: *Proximal splitting methods in nonsmooth convex optimization*. Ph.D. thesis, technische Universität Chemnitz (2014)
30. Johnstone, P.R., Moulin, P.: Local and global convergence of a general inertial proximal splitting scheme for minimizing composite functions. *Comput. Optim. Appl.* **67**(2), 259–292 (2017)
31. Jules, F., Maingé, P.E.: Numerical approach to a stationary solution of a second order dissipative dynamical system. *Optimization* **51**(2), 235–255 (2002)
32. Komodakis, N., Pesquet, J.-C.: Playing with duality: an overview of recent primal–dual approaches for solving large-scale optimization problems. *IEEE Signal Process. Mag.* **32**(6), 31–54 (2015)
33. Li, J., Chen, G., Dong, Z., Wu, Z.: A fast dual proximal-gradient method for separable convex optimization with linear coupled constraints. *Comput. Optim. Appl.* **64**(3), 671–697 (2016)
34. Lorenz, D.A., Pock, T.: An inertial forward–backward algorithm for monotone inclusions. *J. Math. Imaging Vis.* **51**(2), 311–325 (2015)
35. Maingé, P.-E.: Inertial iterative process for fixed points of certain quasi-nonexpansive mappings. *Set-Valued Anal.* **15**(1), 67–79 (2007)
36. Maingé, P.-E.: Convergence theorems for inertial KM-type algorithms. *J. Comput. Appl. Math.* **219**(1), 223–236 (2008)
37. Maingé, P.-E.: Regularized and inertial algorithms for common fixed points of nonlinear operators. *J. Math. Anal. Appl.* **344**(2), 876–887 (2008)
38. Moudafi, A.: A hybrid inertial projection-proximal method for variational inequalities. *JIPAM. J. Inequal. Pure Appl. Math.* **5**(3):Article 63, 5 (2004)
39. Moudafi, A., Elisabeth, E.: An approximate inertial proximal method using the enlargement of a maximal monotone operator. *Int. J. Pure Appl. Math.* **5**(3), 283–299 (2003)
40. Moudafi, A., Oliny, M.: Convergence of a splitting inertial proximal method for monotone operators. *J. Comput. Appl. Math.* **155**(2), 447–454 (2003)
41. Ochs, P., Brox, T., Pock, T.: iPiasco: inertial proximal algorithm for strongly convex optimization. *J. Math. Imaging Vis.* **53**(2), 171–181 (2015)
42. Ochs, P., Chen, Y., Brox, T., Pock, T.: iPiano: inertial proximal algorithm for nonconvex optimization. *SIAM J. Imaging Sci.* **7**(2), 1388–1419 (2014)
43. Peng, Z., Xu, Y., Yan, M., Yin, W.: Arock: an algorithmic framework for asynchronous parallel coordinate updates. *SIAM J. Sci. Comput.* **38**(5), A2851–A2879 (2016)
44. Pennanen, T.: *Dualization of monotone generalized equations*. Ph.D. thesis, University of Washington (1999)
45. Pennanen, T.: Dualization of generalized equations of maximal monotone type. *SIAM J. Optim.* **10**(3), 809–835 (2000)
46. Pesquet, J.-C., Pustelnik, N.: A parallel inertial proximal optimization method. *Pac. J. Optim.* **8**(2), 273–306 (2012)
47. Rockafellar, R.T.: *Convex analysis*, vol. 28. Princeton University Press, Princeton (1970)
48. Rosasco, L., Villa, S., Vū, B.C.: A stochastic inertial forward–backward splitting algorithm for multivariate monotone inclusions. *Optimization* **65**(6), 1293–1314 (2016)
49. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **60**(1–4), 259–268 (1992)
50. Rutkowski, K.E.: Closed-form expressions for projectors onto polyhedral sets in hilbert spaces. *SIAM J. Optim.* **27**(3), 1758–1771 (2017)

51. Solodov, M.V., Svaiter, B.F.: Forcing strong convergence of proximal point iterations in a Hilbert space. *Math. Progr.* **87**(1), 189–202 (2000)
52. Tang, G., Huang, N.: Strong convergence of a splitting proximal projection method for the sum of two maximal monotone operators. *Oper. Res. Lett.* **40**(5), 332–336 (2012)
53. Tang, G., Xia, F.: Strong convergence of a splitting projection method for the sum of maximal monotone operators. *Optim. Lett.* **8**(4), 1313–1324 (2014)
54. Tran-Dinh, Q., Cong Vu, B.: A new splitting method for solving composite monotone inclusions involving parallel-sum operators. *ArXiv e-prints* (2015)
55. Van Hieu, D., Anh, P.K., Muu, L.D.: Modified hybrid projection methods for finding common solutions to variational inequality problems. *Comput. Optim. Appl.* **66**(1), 75–96 (2017)
56. Wang, Y.J., Xiu, N.H., Zhang, J.Z.: Modified extragradient method for variational inequalities and verification of solution existence. *J. Optim. Theory Appl.* **119**(1), 167–183 (2003)