Set-Valued and Variational Analysis (2024) 32:10
https://doi.org/10.1007/s11228-024-00713-7

RECENT FRANCO-CHILEAN COLLABORATIONS IN VARIATIONAL ANALYSIS

Inertial Krasnoselskii-Mann Iterations

Juan José Maulén[1,2] · Ignacio Fierro[3] · Juan Peypouquet[1]

Received: 7 June 2023 / Accepted: 20 March 2024

Abstract

We establish the weak convergence of inertial Krasnoselskii-Mann iterations towards a common fixed point of a family of quasi-nonexpansive operators, along with estimates for the non-asymptotic rate at which the residuals vanish. Strong and linear convergence are obtained in the quasi-contractive setting. In both cases, we highlight the relationship with the non-inertial case, and show that passing from one regime to the other is a continuous process in terms of the hypotheses on the parameters. Numerical illustrations are provided for an inertial primal-dual method and an inertial three-operator splitting algorithm, whose performance is superior to that of their non-inertial counterparts.

Keywords Krasnoselskii-Mann iterations · Fixed points · Nonexpansive operators · Monotone inclusions · Convex optimization · Inertial methods · Acceleration

Mathematics Subject Classification 47H05 · 47H10 · 65K05 · 90C25

1 Introduction

Krasnoselskii-Mann (KM) iterations [35, 40] are at the core of numerical methods used in optimization, fixed point theory and variational analysis, since they include many fundamental splitting algorithms whose convergence can be analyzed in a unified manner. These include the *forward-backward* [37, 46] to approximate a zero of the sum of two maximally monotone operators, and its various particular instances: on the one hand, we have the *gradient projection* algorithm [31, 36], the gradient method [14] and the proximal point algorithm [11, 32, 41, 50], to cite some abstract methods, as well as the *Iterative Shrinkage-Thresholding Algorithm* (ISTA) [20, 23], to speak more concretely. KM iterations also encompass other splitting methods like *Douglas-Rachford* [29], primal-dual methods [6, 16, 18, 21, 53] and the three-operator splitting [24].

In convex optimization, first order methods can be enhanced by adding an inertial substep, motivated by physical considerations [1, 44, 48]. To our knowledge, the first extensions

✉ J. Peypouquet

1 Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Groningen, The Netherlands

2 Center for Mathematical Modelling, Mathematical Engineering Department, University of Chile, Santiago, Chile

3 BIOCORE team, Centre INRIA de l'Université de la Côte d'Azur, Sophia Antipolis, France

beyond the optimization setting was developed in [2], followed by [38, 39, 43] some years later. The main drawback of the previous results is that they require an implicit hypothesis on the sequence generated by the algorithm (the summability of a certain series) to ensure its convergence. In [2], however, this difficulty is overcome, in some special cases and for small values of the inertial parameters. These ideas were also used in [10], and then improved in [25], by adapting the inertial factors to the relaxation ones (see below). A similar principle had been used in [3], whose analysis was based on [5]. Nonasymptotic convergence rates for the residuals have been given in [34, 51]. Strong and linear convergence can be found in [52], for strictly contractive *forward-projection* operators. Other extensions have been considered in [19, 26, 27, 42]. Inexact versions are discussed in [19, 22]. See also [28] for a more thorough account of KM iterations, with and without inertia. Interest in this type of methods increased remarkably in the past decade in view of theoretical advances in the convergence theory for the *Fast Iterative Shrinkage-Thresholding Algorithm* (FISTA) [9], obtained in [4, 7, 15].

The purpose of this work is to develop further insight into the convergence properties of *inertial Krasnoselskii-Mann iterations* in their general form

$$\begin{cases} y_k &= x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} &= (1-\lambda_k)y_k + \lambda_k T_k y_k, \end{cases} \tag{1}$$

where (T_k) is a family of operators defined on a real Hilbert space $\mathcal{H}$, and the positive sequences (α_k) and (λ_k) are the *inertial* and *relaxation* (or *averaging*) parameters, respectively.

Remark 1 To fix the ideas, suppose $\inf_{k\geq 1}\lambda_k > 0$, (α_k) is bounded, and $T_k \equiv T$, where T is continuous. If x_k happens to converge to a point $\bar{x}$, then the *residual* $\|Tx_k - x_k\|$ goes to zero, and $\bar{x}$ is a fixed point of T.

Our general aim is to provide conditions on the parameter sequences and the family of operators to ensure that the sequences generated by (1) converge (weakly or strongly) to a common fixed point of the T_k's, provided there are any. More specifically, we mean to establish a setting, which is as general as possible, but such that (1) the hypotheses are interpretable and verifiable; (2) the proofs are transparent and mostly elementary; and (3) the convergence results are quantifiable in terms of appropriate sequences. We shall also see that adding the inertial term does not always make algorithms faster (this is reflected in the worst-case convergence rates), but may boost their convergence in some relevant instances. Another interesting line of research consists in identifying the combination of parameters for which the algorithm has its best numerical performance. Although we consider this highly relevant, we shall not pursue that direction here.

The paper is organized as follows: in Sect. 2 we establish the weak convergence of the iterations towards a common fixed point of the family of operators in the quasi-nonexpansive case, along with a non-asymptotic rate at which the residuals vanish. Section 3 is devoted to the strong and linear convergence in the quasi-contractive setting. In both cases, we highlight the relationship with the non-inertial case, and show that passing from one regime to the other is a continuous process in terms of parameter hypotheses and convergence rates. In Sect. 4, we discuss several instances of KM iterations, which are relevant to the numerical illustrations provided in Sect. 5, concerning an inertial primal-dual method and an inertial three-operator splitting algorithm.

2 Vanishing Residuals and Weak Convergence

An operator $T : \mathcal{H} \to \mathcal{H}$ is *quasi-nonexpansive* if $\mathrm{Fix}(T) \neq \emptyset$ and $\|Ty - p\| \leq \|y - p\|$ for all $y \in \mathcal{H}$ and $p \in \mathrm{Fix}(T)$. This implies, in particular, that

$$2\langle y - p, Ty - y\rangle \leq -\|Ty - y\|^2 \tag{2}$$

for all $y \in \mathcal{H}$ and $p \in \mathrm{Fix}(T)$.

In this section, we consider a family (T_k) of quasi-nonexpansive operators on $\mathcal{H}$, with $F := \bigcap_{k\geq 1} \mathrm{Fix}(T_k) \neq \emptyset$, along with a sequence (x_k, y_k) satisfying (1), where (α_k) is a nondecreasing sequence[1] in $[0, 1)$, and (λ_k) is a sequence in $(0, 1)$ such that $\inf_{k\geq 1} \lambda_k > 0$.

To simplify the notation, given $p \in F$, we set

$$\begin{cases} \nu_k &= \left(\lambda_k^{-1} - 1\right) \\ \delta_k &= \nu_{k-1}(1-\alpha_{k-1})\|x_k - x_{k-1}\|^2, \\ \Delta_k(p) &= \|x_k - p\|^2 - \|x_{k-1} - p\|^2, \quad \Delta_1(p) = 0 \\ C_k(p) &= \|x_k - p\|^2 - \alpha_{k-1}\|x_{k-1} - p\|^2 + \delta_k, \quad C_1(p) = \|x_1 - p\|^2. \end{cases} \tag{3}$$

At different points, and in order to simplify the computations, we shall make use of a basic property of the norm in $\mathcal{H}$: for every $x, y \in \mathcal{H}$ and $\alpha \in \mathbb{R}$, we have

$$\|\alpha x + (1-\alpha)y\|^2 = \alpha\,\|x\|^2 + (1-\alpha)\,\|y\|^2 - \alpha(1-\alpha)\,\|x - y\|^2. \tag{4}$$

The following auxiliary result will be useful in the sequel:

Lemma 2 *Let (T_k) be a family of quasi-nonexpansive operators on $\mathcal{H}$, with $F := \bigcap_{k\geq 1} \mathrm{Fix}(T_k) \neq \emptyset$, and let (x_k, y_k) satisfy (1). For each $k \geq 1$ and $p \in F$, we have*

$$\begin{aligned} &\Delta_{k+1}(p) + \delta_{k+1} + \nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 \\ &\quad \leq \alpha_k\Delta_k(p) + \left[\alpha_k(1+\alpha_k) + \nu_k\alpha_k(1-\alpha_k)\right]\|x_k - x_{k-1}\|^2. \end{aligned} \tag{5}$$

Proof Take $p \in F$. From (1), it follows that

$$\begin{aligned} \|x_{k+1} - p\|^2 &= \|y_k - p\|^2 + \lambda_k^2\|y_k - T_k y_k\|^2 + 2\lambda_k\langle y_k - p, T_k y_k - y_k\rangle \\ &\leq \|y_k - p\|^2 - \lambda_k(1-\lambda_k)\|y_k - T_k y_k\|^2, \end{aligned} \tag{6}$$

where the inequality is given by (2). Notice that

$$\|y_k - p\|^2 = \|(1+\alpha_k)(x_k - p) - \alpha_k(x_{k-1} - p)\|^2,$$

and using (4) we get

$$\|y_k - p\|^2 = (1+\alpha_k)\|x_k - p\|^2 + \alpha_k(1+\alpha_k)\|x_k - x_{k-1}\|^2 - \alpha_k\|x_{k-1} - p\|^2. \tag{7}$$

By combining expressions (6) and (7), we obtain

$$\begin{aligned} \|x_{k+1} - p\|^2 \leq{}& (1+\alpha_k)\|x_k - p\|^2 + \alpha_k(1+\alpha_k)\|x_k - x_{k-1}\|^2 - \alpha_k\|x_{k-1} - p\|^2 \\ &- \lambda_k(1-\lambda_k)\|y_k - T_k y_k\|^2. \end{aligned}$$

[1] This is just to simplify the proof and is sufficiently general for practical purposes.

Recalling from (3) that $\Delta_k(p) = \|x_k - p\|^2 - \|x_{k-1} - p\|^2$, we rewrite the latter as

$$\Delta_{k+1}(p) \le \alpha_k \Delta_k(p) + \alpha_k(1+\alpha_k)\|x_k - x_{k-1}\|^2 - \lambda_k(1-\lambda_k)\|y_k - T_k y_k\|^2. \tag{8}$$

Notice that

$$\begin{aligned}\lambda_k^2\|y_k - T_k y_k\|^2 &= \|x_{k+1} - x_k - \alpha_k(x_k - x_{k-1})\|^2 \\ &= \|(1-\alpha_k)(x_{k+1} - x_k) + \alpha_k(x_{k+1} - 2x_k + x_{k-1})\|^2, \end{aligned} \tag{9}$$

and using (4) gives

$$\begin{aligned}\lambda_k^2\|y_k - T_k y_k\|^2 &= (1-\alpha_k)\|x_{k+1} - x_k\|^2 - \alpha_k(1-\alpha_k)\|x_k - x_{k-1}\|^2 \\ &\quad + \alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2. \end{aligned} \tag{10}$$

By multiplying the latter by $\nu_k = (1-\lambda_k)/\lambda_k$, and using the definition of δ_k in (3), we rewrite this as

$$\begin{aligned}&\delta_{k+1} + \nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 \\ &\quad = \nu_k\alpha_k(1-\alpha_k)\|x_k - x_{k-1}\|^2 + \lambda_k(1-\lambda_k)\|y_k - T_k y_k\|^2. \end{aligned} \tag{11}$$

Summing (8) and (11), we obtain (5). □

We are now in a position to show that the sequence (x_n) remains anchored to the set F, while both the residuals $\|y_k - T_k y_k\|$ and the speed $\|x_k - x_{k-1}\|$ tend to 0. We shall make some assumptions on the parameter sequences (α_k) and (λ_k).

Hypothesis A There is k_0 such that $\alpha_k(1+\alpha_k) + (\lambda_k^{-1} - 1)\alpha_k(1-\alpha_k) - (\lambda_{k-1}^{-1} - 1)(1 - \alpha_{k-1}) \le 0$ for all $k \ge k_0$.

A reinforced version with strict inequality is given by:

Hypothesis B $\limsup_{k\to\infty}\left[\alpha_k(1+\alpha_k) + (\lambda_k^{-1} - 1)\alpha_k(1-\alpha_k) - (\lambda_{k-1}^{-1} - 1)(1-\alpha_{k-1})\right] < 0.$

Remark 3 With Hypothesis A or B, there exist $\varepsilon \ge 0$ and $k_0 \ge 1$ such that

$$\alpha_k(1+\alpha_k) + (\lambda_k^{-1} - 1)\alpha_k(1-\alpha_k) \le (\lambda_{k-1}^{-1} - 1)(1-\alpha_{k-1}) - \varepsilon \tag{12}$$

for all $k \ge k_0$ (if Hypothesis B holds, then $\varepsilon > 0$; otherwise, $\varepsilon = 0$). Also, under Hypothesis B, $\alpha := \sup_{k\ge1}\alpha_k < 1$ and $\lambda := \inf_{k\ge1}\lambda_k > 0$.

Theorem 4 *Let (T_k) be a family of quasi-nonexpansive operators on $\mathcal{H}$, and let (x_k, y_k) satisfy (1). Suppose that the set $F = \bigcap_{k\ge1}\mathrm{Fix}(T_k)$ is nonempty.*

i) *If Hypothesis A holds, for every $p \in F$, the sequence $\big(C_k(p)\big)_{k\ge k_0}$ is nonincreasing and nonnegative, thus $\lim_{k\to\infty} C_k(p)$ exists.*
ii) *If Hypothesis B holds, the series $\sum_{k\ge1}\|x_{k+1} - 2x_k + x_{k-1}\|^2$, $\sum_{k\ge1}\|x_k - x_{k-1}\|^2$, $\sum_{k\ge1}\delta_k$ and $\sum_{k\ge1}\|y_k - T_k y_k\|^2$ are convergent, and there is a constant $M > 0$, depending only on (α_k) and (λ_k), such that*

$$\min_{1\le k\le n}\|y_k - T_k y_k\|^2 \le \frac{M\,\mathrm{dist}(x_1, F)^2}{n}. \tag{13}$$

Moreover, for each $p \in F$, $\lim_{k\to\infty}\|x_k - p\|$ exists.

Proof Without any loss of generality, we may assume that (12) holds with $k_0 = 1$. Take any $p \in F$, and combine (12) with (5), to obtain

$$\begin{aligned}\Delta_{k+1}(p) + \delta_{k+1} + \nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 \\ \le \alpha_k\Delta_k(p) + \big[\nu_{k-1}(1-\alpha_{k-1}) - \varepsilon\big]\|x_k - x_{k-1}\|^2 \\ = \alpha_k\Delta_k(p) + \delta_k - \varepsilon\|x_k - x_{k-1}\|^2. \end{aligned} \tag{14}$$

On the one hand, (14) immediately gives

$$\Delta_{k+1}(p) \le \alpha_k\Delta_k(p) + \delta_k. \tag{15}$$

On the other, since (α_k) is nondecreasing, we have

$$\begin{aligned}C_{k+1}(p) - C_k(p) &= \Delta_{k+1}(p) - \big(\alpha_k\|x_k - p\|^2 - \alpha_{k-1}\|x_{k-1} - p\|^2\big) + \delta_{k+1} - \delta_k \\ &\le \Delta_{k+1}(p) + \delta_{k+1} - \alpha_k\Delta_k(p) - \delta_k.\end{aligned}$$

Therefore, (14) implies

$$C_{k+1}(p) + \nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 + \varepsilon\|x_k - x_{k-1}\|^2 \le C_k(p). \tag{16}$$

It ensues that $\big(C_k(p)\big)$ is nonincreasing. To show that it is nonnegative, suppose that $C_{k_1}(p) < 0$ for some $k_1 \ge 1$. Since $\big(C_k(p)\big)$ is nonincreasing,

$$\|x_k - p\|^2 - \alpha_{k-1}\|x_{k-1} - p\|^2 \le C_k(p) \le C_{k_1}(p) < 0$$

for all $k \ge k_1$. If follows that $\|x_k - p\|^2 \le \|x_{k-1} - p\|^2 + C_{k_1}(p)$, and so

$$0 \le \|x_k - p\|^2 \le \|x_{k-1} - p\|^2 + C_{k_1}(p) \le \cdots \le \|x_{k_1} - p\|^2 + (k - k_1)C_{k_1}(p)$$

for all $k \ge k_1$, which is impossible. As a consequence $\big(C_k(p)\big)$ is nonnegative, and $\lim_{k\to\infty} C_k(p)$ exists.

For ii), Inequality (12) holds with $\varepsilon > 0$. The summability of the first two series follows from (16). In particular,

$$\varepsilon\sum_{k\ge 1}\|x_k - x_{k-1}\|^2 \le C_1(p) = \|x_1 - p\|^2. \tag{17}$$

The third one is a consequence of the second one, since $\lambda := \inf_{k\ge 1}\lambda_k > 0$. For the last one, use (10) to write

$$\lambda_k^2\|y_k - T_k y_k\|^2 \le (1+\alpha)\|x_{k+1} - x_k\|^2 + \alpha(1+\alpha)\|x_k - x_{k-1}\|^2.$$

In view of (17), this gives the summability of the fourth series, with

$$n\min_{1\le k\le n}\|y_k - T_k y_k\|^2 \le \sum_{k\ge 1}\|y_k - T_k y_k\|^2 \le \frac{(1+\alpha)^2}{\varepsilon\lambda^2}\|x_1 - p\|^2.$$

Since this holds for each $p \in F$, we obtain (13) with $M = \frac{(1+\alpha)^2}{\varepsilon\lambda^2}$. Now, denoting the positive part of $d \in \mathbb{R}$ by $[d]_+$, we obtain from (15) that

$$(1-\alpha)\big[\Delta_{k+1}(p)\big]_+ + \alpha\big[\Delta_{k+1}(p)\big]_+ \le \alpha\big[\Delta_k(p)\big]_+ + \delta_k.$$

Summing for $k \geq 1$, we obtain

$$(1-\alpha)\sum_{k\geq 1}\left[\Delta_{k+1}(p)\right]_+ \leq \alpha\left[\Delta_1(p)\right]_+ + \sum_{k\geq 1}\delta_k = \sum_{k\geq 1}\delta_k < \infty.$$

By writing $h_k = \|x_k - p\|^2 - \sum_{j=1}^{k}\left[\Delta_j(p)\right]_+$, we get $h_{k+1} - h_k = \Delta_{k+1}(p) - \left[\Delta_{k+1}(p)\right]_+ \leq 0$, from which we conclude that $\lim_{k\to\infty}\|x_k - p\| = \lim_{k\to\infty} h_k$ exists. □

Remark 5 Inequality (13) implies that $\lim_{n\to\infty}\left[n\,\min_{1\leq k\leq n}\|y_k - T_k y_k\|^2\right] = 0$.

Remark 6 Hypotheses A and B are closely related, but different, from the hypotheses used in [3] for forward-backward iterations. In the non-inertial case $\alpha = 0$, Hypothesis A is just $\limsup_{k\to\infty}\lambda_k < 1$. On the other hand, since (α_k) is nondecreasing and bounded, we have $\alpha_k \to \alpha \in [0, 1]$. If $\lambda_k \to \lambda$, then Hypothesis B is reduced to

$$\lambda(1-\alpha+2\alpha^2) < (1-\alpha)^2. \tag{18}$$

For each $\alpha \in [0, 1)$, there is $\lambda_\alpha > 0$ such that (18) holds for all $\lambda < \lambda_\alpha$.

In order to prove the weak convergence of the sequences generated by Algorithm (1), we shall use the following nonautonomous extension of the concept of demiclosedness.

The family of operators $(I - T_k)$ is *asymptotically demiclosed at* 0 if for every sequence (z_k) such that $z_k \rightharpoonup z$ and $z_k - T_k z_k \to 0$, we must have $z \in F = \bigcap_{k\geq 1}\mathrm{Fix}(T_k)$.

Of course, if $T : \mathcal{H} \to \mathcal{H}$ is nonexpansive and $T_k \equiv T$, then $I - T_k$ is asymptotically demiclosed at 0. We shall discuss other examples in the next section.

Theorem 7 *Let (T_k) be a family of quasi-nonexpansive operators on $\mathcal{H}$, with $F = \bigcap_{k\geq 1}\mathrm{Fix}(T_k) \neq \emptyset$. Let (x_k, y_k) satisfy (1), and assume Hypotheses B holds. If $(I - T_k)$ is asymptotically demiclosed at* 0*, then both x_k and y_k converge weakly, as $k \to \infty$, to a point in F.*

Proof Recall that $\lim_{k\to\infty}\|y_k - T_k y_k\| = \lim_{k\to\infty}\|x_k - x_{k-1}\| = 0$, by part ii) of Theorem 4. From (1), we deduce that (y_k) and (x_k) have the same (weak and strong) limit points. Suppose $x_{n_k} \rightharpoonup x$. Then, $y_{n_k} \rightharpoonup x$ as well. Since $y_{n_k} - T_k y_{n_k} \to 0$, the asymptotic demiclosedness implies $x \in F$. Opial's Lemma [45] (see, for instance, [47, Lemma 5.2]) yields the conclusion. □

3 Strong and Linear Convergence

We now focus on the strong convergence of the sequences generated by (1), and their convergence rate. As before, we assume that (α_k) is nondecreasing but we do not assume, in principle, that $\inf_{k\geq 1}\lambda_k > 0$.

Given $q \in (0, 1)$, an operator $T : \mathcal{H} \to \mathcal{H}$ is *q-quasi-contractive* if $\mathrm{Fix}(T) \neq \emptyset$ and $\|Ty - p\| \leq q\|y - p\|$ for all $y \in \mathcal{H}$ and $p \in \mathrm{Fix}(T)$. If T is q-quasi-contractive, then $\mathrm{Fix}(T) = \{p^*\}$.

Given $\lambda, q \in (0, 1)$ and $\xi \in [0, 1]$, we define

$$\begin{aligned} Q(\lambda, q, \xi) &:= \xi\left(1-\lambda+\lambda q^2\right) + (1-\xi)(1-\lambda+\lambda q)^2 \\ &= (1-\lambda+\lambda q)^2 + \xi\lambda(1-\lambda)(1-q)^2. \end{aligned} \tag{19}$$

Notice that $Q(\lambda, q, \xi) \in (0, 1)$, and that it decreases as λ increases, or as either q or ξ decreases. The quantity $Q(\lambda, q, \xi)$ will play a crucial role in the linear convergence rate of the sequences satisfying (1). The inclusion of the auxiliary parameter ξ will also allow us to establish convergence rates, with and without inertia, in a unified manner (see the discussion in Sect. 3.2).

The following result establishes a bound on the distance to a solution after performing a standard KM step:

Lemma 8 *Let $T : \mathcal{H} \to \mathcal{H}$ be q-quasi-contractive with fixed point p^*, and let $x, y \in \mathcal{H}$ and $\lambda > 0$ be such that $x = (1-\lambda)y + \lambda Ty$. Then, for each $\xi \in [0, 1]$, we have*

$$\|x - p^*\|^2 \leq Q(\lambda, q, \xi)\|y - p^*\|^2 - \xi\lambda(1-\lambda)\|Ty - y\|^2. \tag{20}$$

Proof Notice that

$$\left\|x - p^*\right\| = \left\|(1-\lambda)(y - p^*) + \lambda(Ty - p^*)\right\|.$$

Then, using (4), we get

$$\begin{aligned}\|x - p^*\|^2 &= (1-\lambda)\|y - p^*\|^2 + \lambda\|Ty - p^*\|^2 - \lambda(1-\lambda)\|Ty - y\|^2 \\ &\leq \left(1 - \lambda + \lambda q^2\right)\|y - p^*\|^2 - \lambda(1-\lambda)\|Ty - y\|^2. \end{aligned} \tag{21}$$

On the other hand, we have

$$\|x - p^*\| \leq (1-\lambda)\|y - p^*\| + \lambda\|Ty - p^*\| \leq (1 - \lambda + \lambda q)\|y - p^*\|. \tag{22}$$

Then, inequality (20) is just a convex combination of (21) and the square of (22). □

3.1 Convergence Analysis

We now turn to the convergence of the sequences verifying (1). To simplify the notation, for each $k \in \mathbb{N}$, we set

$$\tilde{C}_k(p) = \|x_k - p^*\|^2 - \alpha_{k-1}\|x_{k-1} - p^*\|^2 + \xi\delta_k \quad \text{with} \quad \tilde{C}_1(p^*) = \|x_1 - p^*\|^2.$$

We have the following:

Proposition 9 *Let (T_k) be a sequence of operators on $\mathcal{H}$, such that $\mathrm{Fix}(T_k) \equiv \{p^*\}$ and T_k is q_k-quasi-contractive for each $k \in \mathbb{N}$. Let (x_k, y_k) satisfy (1), and let $\xi \in [0, 1]$. Write $Q_k = Q(\lambda_k, q_k, \xi)$, where Q is defined in (19). For each $k \in \mathbb{N}$, we have*

$$\begin{aligned}\|x_{k+1} - p^*\|^2 + \xi\delta_{k+1} \leq\ & Q_k\left[(1+\alpha_k)\|x_k - p^*\|^2 - \alpha_k\|x_{k-1} - p^*\|^2\right] \\ &+ \left[Q_k\alpha_k(1+\alpha_k) + \xi\nu_k\alpha_k(1-\alpha_k)\right]\|x_k - x_{k-1}\|^2. \end{aligned} \tag{23}$$

If, moreover,

$$Q_k\alpha_k(1+\alpha_k) + \xi\nu_k\alpha_k(1-\alpha_k) - \xi Q_k\nu_{k-1}(1-\alpha_{k-1}) \leq 0 \tag{24}$$

for all $k \in \mathbb{N}$, *then*

$$\tilde{C}_{k+1}(p^*) \leq \left[\prod_{j=1}^{k} Q_j\right] \|x_1 - p^*\|^2 \tag{25}$$

and

$$\|x_{k+1} - p^*\|^2 \leq \left[\alpha^k + \sum_{j=1}^{k} \alpha^{k-j} \left[\prod_{i=1}^{j} Q_i\right]\right] \|x_1 - p^*\|^2. \tag{26}$$

Proof We use (1) and (20) to obtain

$$\|x_{k+1} - p^*\|^2 \leq Q_k \|y_k - p^*\|^2 - \xi \lambda_k (1 - \lambda_k) \|y_k - T_k y_k\|^2.$$

Now, by (7), we deduce that

$$\begin{aligned}\|x_{k+1} - p^*\|^2 \leq{}& Q_k \left[(1 + \alpha_k)\|x_k - p^*\|^2 + \alpha_k(1 + \alpha_k)\|x_k - x_{k-1}\|^2 - \alpha_k \|x_{k-1} - p^*\|^2\right] \\ & - \xi \lambda_k (1 - \lambda_k) \|y_k - T_k y_k\|^2.\end{aligned}$$

On the other hand, from (11), we get

$$\xi \delta_{k+1} \leq \xi \nu_k \alpha_k (1 - \alpha_k) \|x_k - x_{k-1}\|^2 + \xi \lambda_k (1 - \lambda_k) \|y_k - T_k y_k\|^2,$$

and the last two inequalities together imply (23). For the second part, inequalities (23) and (24) together give

$$\|x_{k+1} - p^*\|^2 + \xi \delta_{k+1} \leq Q_k \left[(1 + \alpha_k)\|x_k - p^*\|^2 - \alpha_k \|x_{k-1} - p^*\|^2\right] + \xi Q_k \delta_k.$$

Subtracting $\alpha_k \|x_k - p^*\|^2$, we are left with

$$\begin{aligned}\tilde{C}_{k+1}(p^*) &\leq \big(Q_k(1 + \alpha_k) - \alpha_k\big)\|x_k - p^*\|^2 - \alpha_k Q_k \|x_{k-1} - p^*\|^2 + \xi Q_k \delta_k \\ &\leq Q_k \|x_k - p^*\|^2 - Q_k \alpha_{k-1} \|x_{k-1} - p^*\|^2 + \xi Q_k \delta_k \\ &= Q_k \tilde{C}_k(p^*),\end{aligned}$$

where the second inequality comes from α_k being nondecreasing and $Q_k \leq 1$. This gives (25), recalling that $\tilde{C}_1(p^*) = \|x_1 - p^*\|^2$. Now, since $\|x_{k+1} - p^*\|^2 - \alpha_k \|x_k - p^*\|^2 \leq \tilde{C}_{k+1}(p^*)$, we have

$$\begin{aligned}\|x_{k+1} - p^*\|^2 &\leq \alpha_k \|x_k - p^*\|^2 + \left[\prod_{j=1}^{k} Q_j\right] \|x_1 - p^*\|^2 \\ &\leq \alpha \|x_k - p^*\|^2 + \left[\prod_{j=1}^{k} Q_j\right] \|x_1 - p^*\|^2,\end{aligned}$$

which we then iterate to obtain (26). □

The preceding estimations allow us to establish the main result of this section, namely:

Theorem 10 *Let (T_k) be a sequence of operators on $\mathcal{H}$, such that $\operatorname{Fix}(T_k) \equiv \{p^*\}$ and T_k is q_k-quasi-contractive for each $k \in \mathbb{N}$. Let (x_k, y_k) satisfy (1), and let $\xi \in [0, 1]$. Write $Q_k = Q(\lambda_k, q_k, \xi)$, and assume that (24) holds for all $k \in \mathbb{N}$. We have the following:*

i) *If $\sum_{k=1}^{\infty} \lambda_k(1-q_k^2) = \infty$, then x_k converges strongly to p^*, as $k \to \infty$.*
ii) *If $\lambda_k \geq \lambda > 0$ and $q_k \leq q < 1$ for all $k \in \mathbb{N}$, then x_k converges linearly to p^*, as $k \to \infty$. More precisely,*

$$\|x_k - p^*\|^2 \leq \left[\frac{Q(\lambda, q, \xi)^{k+1} - \alpha^{k+1}}{Q(\lambda, q, \xi) - \alpha}\right] \|x_1 - p^*\|^2 = \mathcal{O}\left(Q(\lambda, q, \xi)^k\right). \tag{27}$$

Proof For part i), write $p_k = \lambda_k(1-q_k^2)$, and observe that $Q_k \leq 1 - p_k$, because Q increases with ξ. It ensues that

$$\prod_{k=1}^{K} Q_k \leq \prod_{k=1}^{K}(1-p_k) = \exp\left[\sum_{k=1}^{K} \ln(1-p_k)\right] \leq \exp\left[-\sum_{k=1}^{K} p_k\right]$$

since $\ln(1-z) \leq -z$. If $\sum_{k=1}^{\infty} \lambda_k(1-q_k^2) = \infty$, then $\prod_{k=1}^{\infty} Q_k = 0$. By (25), $\lim_{k\to\infty} \tilde{C}_k(p^*) = 0$. As in the proof of Theorem 4, we can show that the sum of the first two terms in $\tilde{C}_k(p^*)$, namely $\|x_k - p^*\|^2 - \alpha_{k-1}\|x_{k-1} - p^*\|^2$, is nonnegative. Therefore, $\lim_{k\to\infty}\left[\|x_k - p^*\|^2 - \alpha_{k-1}\|x_{k-1} - p^*\|^2\right] = 0$. If $\alpha_k \equiv 0$, the conclusion is straightforward. Otherwise, given any $\varepsilon > 0$, there is $K \in \mathbb{N}$ such that

$$\|x_k - p^*\|^2 \leq \alpha\|x_{k-1} - p^*\|^2 + \varepsilon$$

for all $k \geq K$, since α_k is nondecreasing. This implies

$$\|x_k - p^*\|^2 \leq \alpha^{k-K}\|x_K - p^*\|^2 + \varepsilon(1-\alpha)^{-1},$$

so that $\limsup_{k\to\infty}\|x_k - p^*\| \leq \varepsilon(1-\alpha)^{-1}$, and the conclusion follows.

For ii), we know that $Q(\lambda_k, q_k, \xi) \leq Q(\lambda, q, \xi)$, because Q increases either if λ decreases, and also if q increases. Gathering the common factors in the second and third terms on the left-hand side of inequality (24), we deduce that $Q \geq \alpha$ (strictly if $\alpha > 0$). Using (26), and observing that the case $Q(\lambda, q, \xi) = \alpha$ is incompatible with inequality (24), we deduce that

$$\begin{aligned}\|x_{k+1} - p^*\|^2 &\leq \alpha^k\left[\sum_{j=0}^{k}\left(\frac{Q(\lambda, q, \xi)}{\alpha}\right)^j\right]\|x_1 - p^*\|^2 \\ &= \left[\frac{\alpha^{k+1} - Q(\lambda, q, \xi)^{k+1}}{\alpha - Q(\lambda, q, \xi)}\right]\|x_1 - p^*\|^2,\end{aligned}$$

as claimed. □

3.2 Behavior with and Without Inertia

In the non-inertial case $\alpha_k \equiv 0$, (24) always holds. To simplify the explanation, suppose $q_k \equiv q \in (0, 1)$. The best convergence rate is

$$\|x_k - p^*\| = \mathcal{O}\left(q^k\right),$$

obtained from (23) in Proposition 9 with $\lambda_k \equiv 1$ and $\xi = 0$, whence we recover exactly the known convergence rate for Banach-Picard iterations. If $\alpha_k > 0$ for at least one k, the case $\xi = 0$ is ruled out, and

$$q^2 \le (1-\lambda_k+\lambda_k q)^2 = Q(\lambda_k, q, 0) \le Q(\lambda_k, q, \xi) \le Q(\lambda_k, q, 1) = 1-\lambda_k+\lambda_k q^2.$$

All inequalities are strict if $\lambda_k \in (0, 1)$. This suggests that there may be operators for which the inertial step actually deteriorate the convergence, so inertial steps should be handled with caution and this can be seen as an argument *against* the use of inertia. Actually, it is possible to find a wide variety of behaviors, *even for some of the simplest operators*, as shown by the following case study:

Example 11 Let $\lambda_k \equiv \lambda \in (0, 1)$ and $\alpha_k \equiv \alpha \in [0, 1)$. Take $q \in (0, 1]$, and consider the operator $T : \mathbb{R} \to \mathbb{R}$, defined by $Ty = -qy$, whose unique fixed point is the origin.

If $\alpha = 0$, for each $k \ge 0$, we have $x_{k+1} = Lx_k$, where we have written $L = 1-\lambda(1+q)$. Iterating from $x_0 = 1$, we obtain $|x_k| = |L|^k$. If $\lambda(1+q) = 1$, convergence occurs in one iteration.

Now, let $\alpha \in (0, 1)$, so that (1) reads

$$x_{k+1} = L\big(x_k + \alpha(x_k - x_{k-1})\big). \tag{28}$$

Here, we take $x_1 = x_0 = 1$. We can rewrite (28) in matrix form as

$$X_{k+1} = MX_k, \qquad \text{where} \qquad M = \begin{pmatrix} (1+\alpha)L & -\alpha L \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad X_k = \begin{pmatrix} x_k \\ x_{k-1} \end{pmatrix}.$$

As before, convergence occurs in one step if $L = 0$. The eigenvalues of M are

$$\mu_\pm = \frac{(1+\alpha)L \pm \sqrt{(1+\alpha)^2L^2 - 4\alpha L}}{2}.$$

Let us consider the case $L > 0$ first. If $(1+\alpha)^2L^2 < 4\alpha L$ (which is $\lambda(1+q) > (1-\alpha)^2/(1+\alpha)^2$), the eigenvalues are complex conjugates, both with modulus $|\mu_\pm| = \sqrt{\alpha L} < 1$. Now, $\sqrt{\alpha L} < L$ if, and only if, $L > \alpha$, which means that $\lambda(1+q) < 1-\alpha$. Since $|x_k| = \mathcal{O}(|\mu_\pm|)$, the inertial iterations converge strictly faster than the noninertial ones if

$$\frac{(1-\alpha)^2}{(1+\alpha)^2} < \lambda(1+q) < 1-\alpha.$$

If $L = \alpha$, the convergence rate is the same. Else, if $(1+\alpha)^2L^2 \ge 4\alpha L$, then M has two real eigenvalues (counting multiplicities), with $0 < \mu_- \le \mu_+$. But since $L \in (0, 1)$ implies $-L < -L^2$, we always have

$$\mu_+ < \frac{(1+\alpha)L + \sqrt{(1+\alpha)^2L^2 - 4\alpha L^2}}{2} = \frac{(1+\alpha)L + L\sqrt{(1-\alpha)^2}}{2} = L < 1.$$

Therefore, the inertial iterations also converge strictly faster if

$$0 < \lambda(1+q) \le \frac{(1-\alpha)^2}{(1+\alpha)^2}.$$

When $L < 0$ ($\lambda(1+q) > 1$), the matrix M will always have two real eigenvalues, one of each sign. It is easy to verify that $|\mu_+| < |\mu_-|$, which implies that $|\mu_-|$ determines the convergence (the initial condition is not an eigenvector of M, so both eigenvalues intervene). But

$$\mu_- = -\frac{(1+\alpha)|L| + \sqrt{(1+\alpha)^2L^2 + 4\alpha|L|}}{2} < -\frac{(1+\alpha)|L| + \sqrt{(1+\alpha)^2L^2}}{2} = -|L| = L.$$

In this case, the inertial algorithm performs worse than the noninertial one. Moreover, the inertial iterations do not converge if $\mu_- \le -1$, which is equivalent to

$$\lambda(1+q) \ge \frac{2(1+\alpha)}{1+2\alpha}.$$

A few comments are in order:

- For $0 < \lambda(1+q) < 1-\alpha$, the inertial iterations converge at a strictly faster linear rate than the noninertial ones, even in the noncontracting case $q = 1$.
- At the transition point $\lambda(1+q) = 1-\alpha$ the convergence rate is the same.
- In the interval $1-\alpha < \lambda(1+q) < \frac{2(1+\alpha)}{1+2\alpha}$, the inertial step is counterproductive and noninertial iterations perform better, except for the singular value $\lambda(1+q) = 1$, where both converge in one iteration. In both cases, the closer $\lambda(1+q)$ is to 1, the faster the convergence.

This behavior may be due to the *overshooting* phenomenon seen in gradient-like methods with *long* steps. More precisely, the KM part (the second subiteration) in the example is equivalent to one step of the gradient method:

$$\begin{aligned} x_{k+1} &= (1-\lambda)y_k + \lambda T y_k \\ &= (1-\lambda(1+q))y_k = y_k - \lambda\nabla\phi(y_k), \qquad \text{where} \qquad \phi(y) = \frac{1+q}{2}y^2. \end{aligned}$$

The gradient method produces zig-zagging iterations when $\lambda \in \left(\frac{1}{1+q}, \frac{2}{1+q}\right)$. The *short-step* (not zig-zagging) case $\lambda(1+q) < 1$ can be improved using inertia, whereas adding inertia to the *long-step* (zig-zagging) regime is detrimental.

- If $\lambda(1+q) \ge \frac{2(1+\alpha)}{1+2\alpha}$, the inertial iterations do not converge, while the noninertial ones do. This combination of parameters is not feasible if $q \le 1/3$. Notice that, picking λ and α satisfying (18) can be read as picking $\lambda < S(\alpha)$, with $S(\alpha) = \frac{(1-\alpha)^2}{1-\alpha+2\alpha^2}$. Calling $P(\alpha) = \frac{1+\alpha}{1+2\alpha}$, it is easy to see that

$$\lambda < S(\alpha) < P(\alpha) \le \frac{2}{1+q}P(\alpha),\ \forall q \in (0,1].$$

  Then $\lambda(1+q) < \frac{2(1+\alpha)}{1+2\alpha}$ for all $q \in (0,1]$. Therefore, this last case is incompatible with Hypotheses (A) or (B). In the limiting case $\lambda = 1$, we are led to $(1+2\alpha)q < 1$, which is precisely the condition ensuring convergence of accelerated Banach-Picard iterations discussed below.

Now, the convergence rate results given by Theorem 10 correspond to worst-case scenarios, which certainly must include cases like the one discussed in Example 11. However, this situation need not be representative of other concrete instances found in practice, in which inertia improves either the theoretical convergence rate guarantees, or the actual behavior when the algorithm is implemented. In fact, the numerical tests reported below show noticeable improvements in the performance of the selected algorithms, upon adding the inertial substep.

3.3 Accelerated Banach-Picard Iterations

When $\lambda_k \equiv 1$, (1) becomes

$$x_{k+1} = T\big(x_k + \alpha_k(x_k - x_{k-1})\big),$$

and one may wonder under which conditions these iterations converge. To simplify the exposition, we restrict our analysis to the case $\alpha_k \equiv \alpha$ and $q_k \equiv q$. Writing $e_k = \|x_k - p^*\|^2$ in (23), we deduce that

$$e_{k+1} \leq (1+\alpha)q^2 e_k - \alpha q^2 e_{k-1} + \alpha(1+\alpha)q^2 \left[(1+\zeta)e_k + \left(1+\frac{1}{\zeta}\right)e_{k-1}\right],$$

for every $\zeta > 0$. Therefore,

$$e_{k+1} - re_k \leq \big[(1+\alpha)q^2\big(1+\alpha(1+\zeta)\big) - r\big]e_k + \alpha q^2 \left[(1+\alpha)\left(1+\frac{1}{\zeta}\right) - 1\right] e_{k-1},$$

for every $r \in \mathbb{R}$. Suppose we can find $\zeta > 0$ and $r \in (0,1)$, such that

$$0 \leq \alpha q^2 \left[(1+\alpha)\left(1+\frac{1}{\zeta}\right) - 1\right] = r\big[r - (1+\alpha)q^2\big(1+\alpha(1+\zeta)\big)\big].^{2} \tag{29}$$

Then, setting $c = r - (1+\alpha)q^2\big(1+\alpha(1+\zeta)\big) \geq 0$, we will obtain

$$e_{k+1} + ce_k \leq r(e_k + ce_{k-1}).$$

We rewrite (29) as

$$\zeta r^2 - (1+\alpha)\zeta q^2(1+\alpha+\alpha\zeta)r - \alpha q^2(1+\alpha+\alpha\zeta) = 0. \tag{30}$$

Since the left-hand side is negative when $r = 0$, the equation has a solution $\hat{r} \in (0,1)$ if, and only if, there is $\zeta > 0$ such that

$$\zeta - (1+\alpha)\zeta q^2(1+\alpha+\alpha\zeta) - \alpha q^2(1+\alpha+\alpha\zeta) > 0,$$

a condition equivalent to

$$\alpha(1+\alpha)q^2\zeta^2 + \big[(1+\alpha)^2q^2 + \alpha^2q^2 - 1\big]\zeta + \alpha(1+\alpha)q^2 < 0,$$

which we rewrite as

$$\zeta^2 - m\zeta + 1 < 0 \qquad \text{with} \qquad m = \frac{1-(1+\alpha)^2q^2 - \alpha^2q^2}{\alpha(1+\alpha)q^2}$$

[2] Actually, an inequality would be enough, but it would give a worse rate.

for $\zeta > 0$. This quadratic inequality has a positive solution if, and only if, $m > 0$ and $m^2 > 4$. In other words, if $m > 2$, which means that

$$1 - (1+\alpha)^2 q^2 - \alpha^2 q^2 > 2\alpha(1+\alpha)q^2,$$

and is finally reduced to

$$(1+2\alpha)q < 1.^{3}$$

Then, although not necessarily optimal, we may set $\zeta = m/2$, and then compute r from (30). We have proved the following:

Theorem 12 *Let* (T_k) *be a sequence of operators on* $\mathcal{H}$, *such that* $\mathrm{Fix}(T_k) \equiv \{p^*\}$ *and* T_k *is* q_k*-quasi-contractive for each* $k \in \mathbb{N}$. *Let* (x_k, y_k) *satisfy* (1), *with* $\lambda_k \equiv 1$ *and* $\alpha_k \equiv \alpha \in [0, 1)$. *Assume, moreover, that* $\sup_k (1+2\alpha)q < 1$. *Then,* x_k *converges linearly to* p^*, *as* $k \to \infty$.

3.4 Some Insights into Inequality (24)

To fix the ideas, we comment on some special cases of inequality (24), especially with constant parameters:

1. In the limiting case $q_k \equiv 1$, we have $Q_k \equiv 1$. With constant parameters $\lambda_k \equiv \lambda$, $\alpha_k \equiv \alpha$, (24) becomes

$$\lambda\alpha(1+\alpha) - \xi(1-\lambda)(1-\alpha)^2 \le 0.$$

   If

$$\frac{\alpha\lambda(1+\alpha)}{(1-\lambda)(1-\alpha)^2} \le 1, \tag{31}$$

   then, there is $\xi_{\alpha,\lambda,1} \in (0, 1)$ such that (24) holds for all $\xi \in [\xi_{\alpha,\lambda,1}, 1]$. If $\xi = 1$, it is precisely the constant case in Hypothesis A (see (18) for a more direct comparison).
2. Keeping $\lambda_k \equiv \lambda \in (0, 1)$, $\alpha_k \equiv \alpha \in (0, 1)$, and fixing $\xi = 1$, let us take $q_k \equiv q \in (0, 1)$. In this case, condition (24) is equivalent to

$$\Psi(\lambda) := (1+\alpha^2)(1-q^2)\lambda^2 - \big(2\alpha^2 + (1-\alpha)(2-q^2)\big)\lambda + (1-\alpha)^2 \ge 0. \tag{32}$$

   Observe that $\Psi(0) = (1-\alpha)^2 > 0$, while $\Psi(1) = -\alpha q^2(1+\alpha) < 0$. Since Ψ is quadratic, the equation $\Psi(\lambda) = 0$ has exactly one root in $(0, 1)$, which we denote by $\lambda_{\alpha,q}$. It follows that, for each $(\alpha, q) \in [0, 1) \times (0, 1)$, inequality (32) holds for all $\lambda \le \lambda_{\alpha,q}$. The values of $\lambda_{\alpha,q}$ on $[0, 1) \times (0, 1)$ are depicted in Fig. 1. Once a value for the inertial parameter α has been selected, the best theoretical convergence rate is

$$Q(\lambda_{\alpha,q}, q, 1) = 1 - \lambda_{\alpha,q}(1-q^2).$$

   On the other hand, using the formula for the roots of a quadratic equation and some algebraic manipulations, we deduce that

$$\left[\frac{2\alpha^2 + (1-\alpha)}{2\alpha^2 + (1-\alpha)(2-q^2)}\right]\lambda_{\alpha,1} \le \lambda_{\alpha,q} \le \lambda_{\alpha,1}$$

[3] This is consistent with the behavior observed in Example 11.

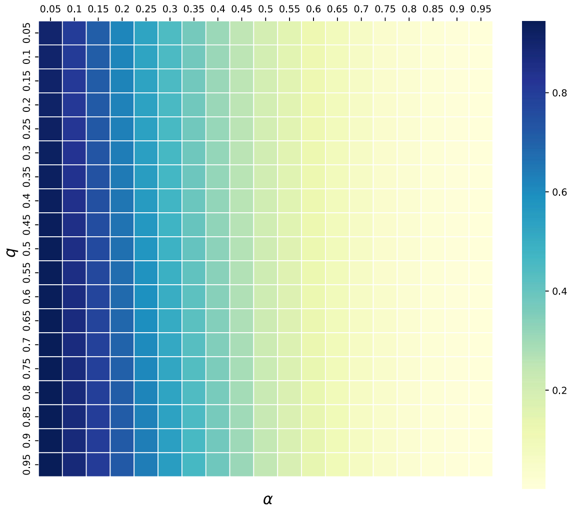


Fig. 1 Values of $\lambda_{\alpha,q}$

for every $(\alpha, q) \in [0, 1) \times (0, 1)$. Therefore, $\lambda_{\alpha,q} \to \lambda_{\alpha,1}$ as $q \to 1$, and there is no discontinuity as the contractive character is lost.

The case $\xi \in (0, 1)$ is more involved. Lower values of ξ make the constant Q smaller, but may also restrict the possible values for α and λ, in view of inequality (24). In the fully general case, if α, λ and q satisfy

$$\left[\frac{\alpha\lambda(1+\alpha)}{(1-\alpha)(1-\lambda)}\right]\left[\frac{1-\lambda+\lambda q^2}{1-\lambda+\lambda q^2-\alpha}\right] < 1$$

(this is implied by (31) if $q \neq 1$), there is $\xi_{\alpha,\lambda,q} \in (0, 1)$ such that (24) holds for all $\xi \in [\xi_{\alpha,\lambda,q}, 1]$. As $q \to 1$, we recover (31) as a limit case.

4 A Few Relevant Particular Instances

In this section, we describe very briefly some instances of Krasnoselskii-Mann algorithms, which are well-known in the optimization literature.

4.1 Averaged Operators and the Forward-Backward Method

An operator $T : \mathcal{H} \to \mathcal{H}$ is γ-averaged if there is a nonexpansive operator $R : \mathcal{H} \to \mathcal{H}$ such that $T = (1-\gamma)I + \gamma R$. In this case, $\mathrm{Fix}(T) = \mathrm{Fix}(R)$.

Let $R : \mathcal{H} \to \mathcal{H}$ be nonexpansive and let (γ_k) be a sequence in $(0, 1)$. Setting $T_k = (1 - \gamma_k)I + \gamma_k R$, (1) can be rewritten as

$$\begin{cases} y_k &= x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} &= (1-\gamma_k\lambda_k)y_k + \gamma_k\lambda_k R(y_k). \end{cases} \tag{33}$$

If $\gamma_k\lambda_k \to \eta > 0$, Hypothesis B becomes

$$\eta(1-\alpha+2\alpha^2) < (1-\alpha)^2. \tag{34}$$

It is not necessary to implement the algorithm using the operator R explicitly. However, the interval for the relaxation parameters is enlarged, and it may be convenient to over-relax. We shall come back to this point in the numerical illustrations.

Let $A:\mathcal{H}\to 2^{\mathcal{H}}$ be maximally monotone, let $B:\mathcal{H}\to\mathcal{H}$ be cocoercive with parameter β,[4] and let (ρ_k) be a sequence in $(0,2\beta)$. For each $k\geq 1$, set

$$T_k=(I+\rho_k A)^{-1}(I-\rho_k B).$$

Then, T_k is γ_k-averaged with $\gamma_k=2\beta(4\beta-\rho_k)^{-1}$, and the family $(I-T_k)$ is asymptotically demiclosed at 0 if $\inf_{k\geq 1}\rho_k>0$. If $\rho_k\to\rho$ and $\lambda_k\to\lambda$, then Hypothesis B is equivalent to

$$\lambda(1-\alpha+2\alpha^2)<\left(2-\frac{\rho}{2\beta}\right)(1-\alpha)^2.$$

If A is μ-strongly monotone, then T_k is q_k-(quasi-)contractive, with $q_k=(1+\mu\rho_k)^{-1}$. If B is μ-strongly monotone, then T_k is q_k'-(quasi-)contractive, with $q_k'=(1-\mu\rho_k)$.

4.2 Douglas-Rachford and Primal-Dual Splitting

Let $A,B:\mathcal{H}\to 2^{\mathcal{H}}$ be maximally monotone, and let (r_k) be a positive sequence. The *Douglas-Rachford* splitting method consists in iterating $z_{k+1}=T_{r_k}z_k$, for $k\geq 1$, where

$$T_r=J_{rA}\circ\big(2J_{rB}-I\big)+\big(I-J_{rB}\big)=\frac{1}{2}\big(I+(2J_{rA}-I)\circ(2J_{rB}-I)\big). \tag{35}$$

The second expression shows that T_r is averaged. Using the weak-strong closedness of the graphs of A and B, and a little algebra, one proves that the family $\big(I-T_{r_k}\big)$ is asymptotically demiclosed if $\inf_{k\geq 0}r_k>0$. Finally, observe that $\operatorname{Zer}(A+B)=J_{rB}\operatorname{Fix}(T_r)$. In this generality, there is no direct relationship between the strong monotonicity of $A+B$ and the contractivity of T_k, but the weak convergence results still hold.

More generally, let X and Y be Hilbert spaces, and consider the *primal problem*, which is to find $\hat{x}\in X$ such that

$$0\in A\hat{x}+L^*BL\hat{x},$$

where $A:X\to 2^X$ and $B:Y\to 2^Y$ are maximally monotone operators, and $L:X\to Y$ is linear and bounded. The *dual problem* is to find $\hat{y}\in Y$ such that

$$0\in B^{-1}\hat{y}-LA^{-1}(-L^*\hat{y}).$$

The primal and dual solutions, namely $\hat{x}$ and $\hat{y}$, are linked by the inclusions

$$-L^*\hat{y}\in A\hat{x}\qquad\text{and}\qquad L\hat{x}\in B^{-1}\hat{y}.$$

Remark 13 Let $f:X\to\mathbb{R}\cup\{+\infty\}$ and $g:Y\to\mathbb{R}\cup\{+\infty\}$ be closed and convex, and set $A=\partial f$ and $B=\partial g$. The inclusions above are the optimality conditions for the primal and dual (in the sense of Fenchel-Rockafellar) optimization problems

$$\min_{x\in X}\{f(x)+g(Lx)\}\qquad\text{and}\qquad\min_{y\in Y}\{g^*(y)+f^*(-L^*y)\}, \tag{36}$$

[4] An operator B is *β-cocoercive* with $\beta>0$ if $\langle Bx-By,x-y\rangle\geq\beta\,\|Bx-By\|^2$ for all $x,y\in\mathcal{H}$. The gradient of a L-smooth function is $\frac{1}{L}$-cocoercive.

respectively. Douglas-Rachford splitting applied to $A = \partial g^*$ and $B = \partial\big(f^* \circ (-L^*)\big)$ yields the *alternating direction method of multipliers* (see [30]).

In order to find a primal-dual pair, the *primal-dual* splitting algorithm (see [16]) iterates:

$$\begin{cases} x_{k+1} &= J_{\tau A}\big(x_k - \tau L^* y_k\big) \\ y_{k+1} &= J_{\sigma B^{-1}}\big(y_k + \sigma L(2x_{k+1} - x_k)\big), \end{cases} \tag{37}$$

with $\tau\sigma \|L\|^2 \leq 1$. The algorithm can be expressed as $(x_{k+1}, y_{k+1}) = T(x_k, y_k)$, where $T : X \times Y \to X \times Y$ is a 1/2-averaged operator (see [8, Remark 4.34]). As before, there is no direct relationship between the strong monotonicity of $A + L^* \circ B \circ L$ and the contractivity of T_k, but the weak convergence results still hold.

An inertial version of the primal-dual iterations is given by

$$\begin{cases} (y_k, v_k) = (x_k, u_k) + \alpha_k\,[(x_k, u_k) - (x_{k-1}, u_{k-1})] \\ p_{k+1} = J_{\tau A}(y_k - \tau L^* v_k) \\ q_{k+1} = J_{\sigma B^{-1}}(v_k + \sigma L(2p_{k+1} - y_k)) \\ (x_{k+1}, u_{k+1}) = (1 - \lambda_k)(y_k, v_k) + \lambda_k(p_{k+1}, q_{k+1}), \end{cases} \tag{38}$$

with appropriate sequences α_k and λ_k.

4.3 Three Operator Splitting

Given three maximally monotone operators A, B, C defined on the Hilbert space H, we wish to find $\hat{x} \in H$ such that

$$0 \in A\hat{x} + B\hat{x} + C\hat{x}. \tag{39}$$

If C is β-cocoercive, the *three-operator* splitting method [24] generates a sequence (z_k) by

$$\begin{cases} x_k^B = J_{\rho B}(z_k) \\ x_k^A = J_{\rho A}(2x_k^B - z_k - \rho C x_k^B) \\ z_{k+1} = z_k + \lambda_k(x_k^A - x_k^B) \end{cases} \tag{40}$$

starting from a point $z_0 \in H$. Here $\rho \in (0, 2\beta)$, $\lambda_k \in (0, 1/\gamma)$ and

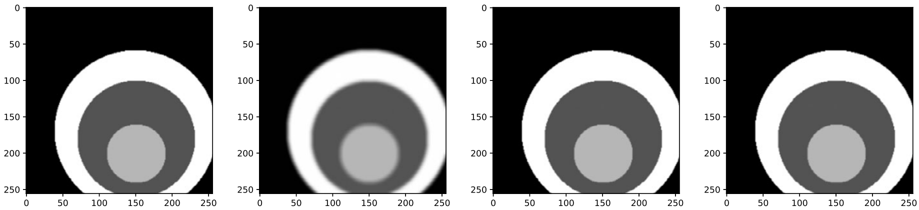
$$\gamma = \frac{2\beta}{4\beta - \rho}. \tag{41}$$

This recurrence is generated by iterating the γ-averaged operator

$$T = I - J_{\rho B} + J_{\rho A} \circ (2J_{\rho B} - I - \rho C \circ J_{\rho B}),$$

and we have $\mathrm{Zer}(A + B + C) = J_{\rho B}(\mathrm{Fix}\, T)$. Also, it gives the forward-backward method if $B = 0$ and the Douglas-Rachford method if $C = 0$. An inertial version is given by

$$\begin{cases} u_k = z_k + \alpha_k(z_k - z_{k-1}) \\ x_k^B = J_{\rho B}(u_k) \\ x_k^A = J_{\rho A}(2x_k^B - u_k - \rho C x_k^B) \\ z_{k+1} = u_k + \lambda_k(x_k^A - x_k^B), \end{cases} \tag{42}$$



(a) Original Image (b) Blurred Image (c) Recovered without inertia (d) Recovered with inertia

Fig. 2 Original, blurred and recovered images. Lowest recovered value $F^{TV}(x) = 0.1301$ (case 14 from Table 1, both methods). Tolerance $\varepsilon = 10^{-5}$, $\omega = 10^{-4}$, standard deviation $s = 10^{-3}$. $\lambda_k \equiv 1$ and α_k as in (47)

for appropriate choices of α_k, λ_k. One particular instance is given by the optimization problem

$$\min f(x) + g(x) + h(Lx), \tag{43}$$

where f, g, h are closed and convex, h has a $(1/\beta)$-Lipschitz-continuous gradient, and L is a bounded linear mapping.

5 Two Experiments in Image Processing

In this section, we test the performance of the algorithm given by iterations (1) in two of the settings described in Sect. 4. More precisely, we apply an inertial primal-dual splitting method to solve a TV-based denoising problem, and an inertial three-operator splitting algorithm to in-paint a corrupted image.

5.1 Primal-Dual Splitting and TV-Based Denoising

The algorithm will be tested in an image processing framework. Consider the problem

$$\min_{x \in \mathbb{R}^{N_1 \times N_2}} F^{TV}(x) := \frac{1}{2} \|Rx - b\|^2 + \omega \|\nabla x\|_1 , \tag{44}$$

where $x \in \mathbb{R}^{N_1 \times N_2}$ is an image to recover from a noisy observation $b \in \mathbb{R}^{M_1 \times M_2}$, $R : \mathbb{R}^{N_1 \times N_2} \to \mathbb{R}^{M_1 \times M_2}$ is a blur operator, ω is a positive parameter, and $\nabla : x \mapsto \nabla x = (D_1 x, D_2 x)$ is the classical discrete gradient, whose adjoint ∇^* is the discrete divergence. A formulation for the gradient and divergence operators can be seen on [17]. In these experiments, R will be a Gaussian blur of size 9×9, standard deviation 4 and relative boundary conditions (see [33] for details on the construction of the operator), and $\omega = 10^{-4}$. Considering the original image $\bar{x}$ in Fig. 2a composed by 256×256 pixels, the observation b is generated as $b = R\bar{x} + e$, where e is an additive zero-mean white Gaussian noise with standard deviation $s = 10^{-3}$ (Fig. 2b).

Setting $f = 0$, $g : (u, v^1, v^2) \mapsto \frac{1}{2} \|u - b\|^2 + \omega \left\|v^1\right\|_1 + \omega \left\|v^2\right\|_1$ and $L : x \mapsto (Rx, D_1 x, D_2 x)$, the problem (44) can be formulated as (36), and solved via (38). Since

$$\text{prox}_{\sigma g^*} : (u, v^1, v^2) \mapsto \left(\frac{u - \sigma b}{\sigma + 1}, v^1 - \sigma \, \text{prox}_{\frac{\omega}{\sigma}\|\cdot\|_1} \left(\frac{v^1}{\sigma} \right), v^2 - \sigma \, \text{prox}_{\frac{\omega}{\sigma}\|\cdot\|_1} \left(\frac{v^2}{\sigma} \right) \right) \tag{45}$$

we are lead to Algorithm 1.

Algorithm 1: Inertial Primal-Dual Splitting

Choose $x_0, x_1 \in \mathbb{R}^{N_1 \times N_2}$, $u_0, u_1 \in \mathbb{R}^{m_1 \times m_2}$, $v_0^1, v_1^1, v_0^2, v_1^2 \in \mathbb{R}^{N_1 \times N_2}$, $(\lambda_k)_{k\in\mathbb{N}}$ and $(\alpha_k)_{k\in\mathbb{N}}$ such that hypotheses of Theorem 4 are fulfilled, τ and σ such that $\tau\sigma \|L\|^2 \leq 1$, $\varepsilon > 0$ and $r_0 > \varepsilon$;

while $r_k > \varepsilon$ **do**

$(\bar{x}_k, \bar{u}_k, \bar{v}_k^1, \bar{v}_k^2) = (x_k, u_k, v_k^1, v_k^2) + \alpha_k[(x_k, u_k, v_k^1, v_k^2) - (x_{k-1}, u_{k-1}, v_{k-1}^1, v_{k-1}^2)]$;

$p_{k+1} = \bar{x}_k - \tau R^* \bar{u}_k - \tau D_1^* \bar{v}_k^1 - \tau D_2^* \bar{v}_k^2$;

$q_{k+1} = (\bar{u}_k + \sigma R(2p_{k+1} - \bar{x}_k) - \sigma b)/(\sigma + 1)$;

$w_{k+1}^1 = \bar{v}_k^1 + \sigma D_1(2p_{k+1} - \bar{x}_k) - \sigma \operatorname{prox}_{\omega\|\cdot\|_1/\sigma}(\bar{v}_k^1/\sigma + D_1(2p_{k+1} - \bar{x}_k))$;

$w_{k+1}^2 = \bar{v}_k^2 + \sigma D_2(2p_{k+1} - \bar{x}_k) - \sigma \operatorname{prox}_{\omega\|\cdot\|_1/\sigma}(\bar{v}_k^2/\sigma + D_2(2p_{k+1} - \bar{x}_k))$;

$(x_{k+1}, u_{k+1}, v_{k+1}^1, v_{k+1}^2) = (1 - \lambda_k)(\bar{x}_k, \bar{u}_k, \bar{v}_k^1, \bar{v}_k^2) + \lambda_k(p_{k+1}, q_{k+1}, w_{k+1}^1, w_{k+1}^2)$;

$r_k = \mathcal{R}((x_{k+1}, u_{k+1}, v_{k+1}^1, v_{k+1}^2), (x_k, u_k, v_k^1, v_k^2))$

end

return $(x_{k+1}, u_{k+1}, v_{k+1}^1, v_{k+1}^2)$

Remark 14 Since the blur operator R is a convolution, it admits a diagonal representation in the phase space. More precisely, by considering the Discrete Fourier Transform $\mathcal{F}$, we have $R = \mathcal{F}^* \Sigma \mathcal{F}$, and defining $\theta = \mathcal{F}x$, we may rewrite the problem as

$$\min_{\theta \in \mathbb{R}^{N_1 \times N_2}} \left\{ \frac{1}{2} \left\| \theta - \hat{\theta} \right\|_{\Sigma}^2 + \omega \|\mathcal{D}\theta\|_1 \right\}.$$

This reinterpretation may reduce the computational time, but does not have an impact in the relationship between inertia and relaxation that we discuss in this work.

For a stopping criterion, we consider the relative error

$$\mathcal{R}(x_{k+1}, x_k) \mapsto \frac{\|x_{k+1} - x_k\|}{\|x_k\|}. \tag{46}$$

Since the involved operator is 1/2-averaged (see [12]), we may set $\lambda_k \equiv \lambda \in (0, 2)$, as explained in Sect. 4.1.

The algorithm is tested for 17 combinations of τ, σ satisfying the critical condition $\tau\sigma \|L\|^2 = 1$ (according to [13], this tends to yield the best performance). The number $\|L\|$ is computed using an adaptation of [49, Algorithm 12]. The recovered images are collected in Figs. 2c and 2d.

Comparison in terms of the parameters τ and σ In a first stage, we compare the performance of the primal-dual splitting algorithm given by (37) (that is, Algorithm 1 with $\alpha_k \equiv 0$), and its inertial counterpart (38), with $\lambda_k \equiv 1$. The sequence $(\alpha_k)_{k\in\mathbb{N}}$ is

$$\alpha_k = \alpha \left(1 - \frac{1}{k^2}\right), \tag{47}$$

with $\alpha = 1/(3 + 0.0001)$ (condition (34) with $\eta = \lambda/2$ gives the constraint $\alpha < 1/3$ for $\lambda = 1$). Table 1 shows the execution time, number of iterations, and the objective value reached, using a tolerance $\varepsilon = 10^{-5}$. These results are depicted graphically, along with the percentage of reduction, in Fig. 3.

Table 1 TV-based denoising problem. Execution time, number of iterations and final function value for the original primal-dual algorithm and the inertial version, tolerance $\varepsilon = 10^{-5}$, $\omega = 10^{-4}$, standard deviation $s = 10^{-3}$. $\lambda_k \equiv 1$ and α_k as in (47)

Case	τ	σ	Original algorithm			Inertial algorithm		
			Time	Iterations	$F^{TV}(x)$	Time	Iterations	$F^{TV}(x)$
1	0.0004	282.8427	72.59	1565	7.30	55.11	1095	7.13
2	0.0010	122.6475	115.66	2437	2.84	86.97	1741	2.66
3	0.0024	53.183	110.16	2330	1.35	83.98	1672	1.27
4	0.0054	23.0614	98.28	2077	0.7566	72.33	1446	0.7341
5	0.0125	10	94.80	2015	0.4624	69.59	1394	0.4537
6	0.0288	4.3362	105.19	2253	0.2975	77.83	1562	0.2928
7	0.0665	1.8803	122.23	2593	0.2107	89.83	1773	0.2091
8	0.1533	0.8153	156.34	3248	0.1592	112.09	2184	0.1589
9	0.3536	0.3536	140.91	2922	0.1428	101.69	1956	0.1427
10	0.8153	0.1533	139.50	2856	0.1350	98.97	1908	0.1350
11	1.8803	0.0665	151.08	3123	0.1312	107.72	2084	0.1312
12	4.3362	0.0288	108.08	2249	0.1303	78.03	1503	0.1303
13	10	0.0125	60.28	1238	0.1301	42.78	833	0.1301
14	23.0614	0.0054	47.61	983	0.1302	35.70	693	0.1302
15	53.1830	0.0024	70.78	1466	0.1302	54.61	1065	0.1302
16	122.6475	0.0010	119.22	2471	0.1302	89.91	1762	0.1302
17	282.8427	0.0004	179.22	3767	0.1302	150.52	2999	0.1302

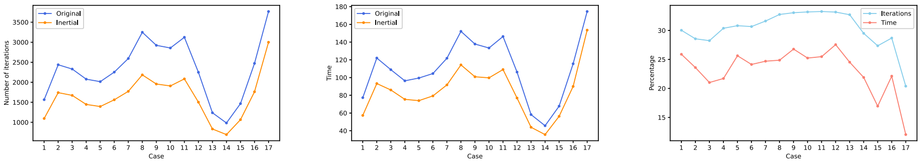


Fig. 3 Number of iterations (left), execution time (center), and percentage of reduction (right), from Table 1

Figure 4 shows the evolution of the function values, the distance to the limit and the residuals, all in logarithmic scale, for case 14. The figure also includes the plot of $k\,\|z_k - Tz_k\|^2$.

The previous experiment is repeated adding more noise to the blurred image by setting the standard deviation to 0.05.Figure 5 shows the comparison between the original image, the noisy observation and the recovered one. The parameter ω was modified with respect to the previous experiment, for an enhanced regularization. The image in Fig. 5b is noticeably more damaged, and the algorithms need more iterations and time to achieve a solution. Setting a tolerance of $\varepsilon = 10^{-4}$ and $\omega = 10^{-2}$, we consider the two best cases of Table 1 (13 and 14) and compare both algorithms using $\lambda_k \equiv 1$ and the inertial sequence as in (47). Execution time, number of iterations and objective value achieved by both methods are depicted in Table 2.

Comparison in terms of the relaxation parameter λ In principle, the parameter λ must be in $(0, 1)$. The first rows of Table 3–which refers to case 14–show that inertia provides a considerable improvement in that range. On the other hand, as discussed in Sect. 4.1,

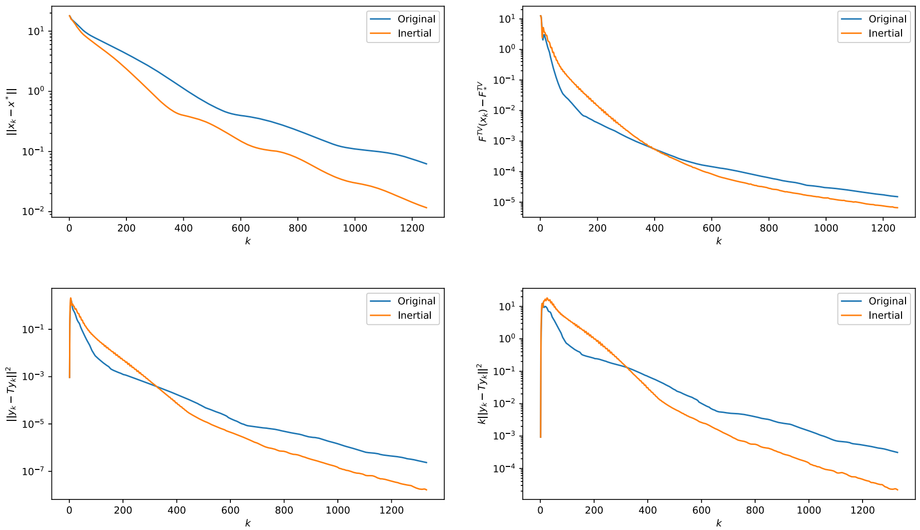
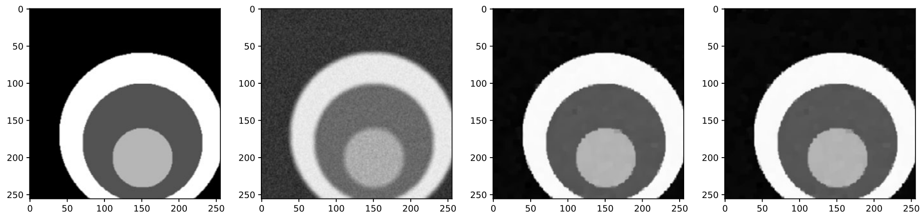


Fig. 4 TV-based denoising problem. Evolution to the distance to the computed solution (top left), objective function values (top right), residuals $\|z_k - Tz_k\|^2$ (bottom left) and $k\,\|z_k - Tz_k\|^2$ (bottom right), for case 14 in Table 1



(a) Original Image (b) Blurred Image (c) Recovered without inertia (d) Recovered with inertia

Fig. 5 Original, blurred and recovered images for the TV-based denoising problem. Lowest recovered value $F^{TV}(x) = 58.14$. $\tau = 10$, $\sigma = 0.0054$, standard deviation 0.05, tolerance $\varepsilon = 10^{-4}$, $\omega = 10^{-2}$

Table 2 Execution time, number of iterations and final function value for the original primal-dual algorithm and the inertial version, standard deviation 0.05, tolerance $\varepsilon = 10^{-4}$, $\omega = 10^{-2}$

τ	σ	Original algorithm			Inertial algorithm		
		Time	Iterations	$F^{TV}(x)$	Time	Iterations	$F^{TV}(x)$
10	0.0125	92.33	2372	58.21	63.69	1584	58.21
23.0614	0.0054	74.94	1871	58.14	52.36	1255	58.14

the KM iterations can be over-relaxed if one can quantify *a priori* the "averagedness" of the operator, which is the case here. We have therefore assessed the performance of the inertial algorithm with different values for $\lambda_k \equiv \lambda \in (0, 2)$, and the corresponding inertial parameters fulfilling condition (34).The results are shown in Table 3, along with the value of α used in (47). A graphic depiction is shown as heatmaps in Fig. 6. Larger values of the relaxation parameter λ resulted in an improvement in the perfor-

Table 3 TV-based denoising problem. Execution time, number of iterations, final function value and reduction percentage for the original primal-dual algorithm and the inertial version (case 14 for τ and σ), with tolerance $\varepsilon = 10^{-5}$, $\omega = 10^{-4}$, standard deviation $s = 10^{-3}$

λ	α	Original algorithm			Inertial algorithm			% Iterations reduction	% Time reduction
		Time	Iterations	$F^{TV}(x)$	Time	Iterations	$F^{TV}(x)$		
0.2	0.6534	119.16	2592	0.1303	49.23	992	0.1304	61.73	58.69
0.4	0.5425	74.44	1589	0.1302	40.45	799	0.1303	49.72	45.66
0.6	0.4619	62.28	1341	0.1302	39.06	773	0.1302	42.36	37.28
0.8	0.3943	54.05	1146	0.1302	33.94	730	0.1302	36.30	37.21
1.0	0.3333	46.12	983	0.1302	34.47	693	0.1302	29.50	25.26
1.2	0.2748	41.16	861	0.1301	35.17	684	0.1302	20.56	14.55
1.4	0.1352	38.22	771	0.1301	34.45	675	0.1301	12.45	9.86
1.6	0.0967	33.89	718	0.1301	33.59	655	0.1301	8.77	0.89
1.8	0.0535	32.28	679	0.1301	32.62	657	0.1301	3.24	−1.05

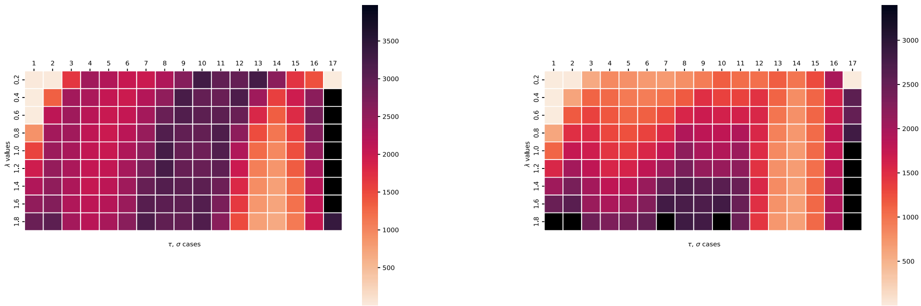


Fig. 6 Average number of iterations performed by the original (left) and inertial (right) algorithms, with tolerance $\varepsilon = 10^{-5}$, for each value of λ, and each case of τ and σ, from Table 3

mance of both algorithms, but limit the impact of inertia, as it reduces the feasible range for the limit α. Nevertheless, it is important to point out that values close to the upper and lower boundaries are usually avoided to reduce the risk of numerical instabilities.

5.2 Three-Operator Splitting and Image in-Painting

Suppose that Z is a color image represented as a 3-D tensor where $Z(:,:,1)$, $Z(:,:,2)$, $Z(:,:,3)$ are the red, green and blue channels, respectively. Consider a damaged image Y, with randomly erased pixels, represented by the white color. The positions of the erased pixels are known. Denote $\mathcal{A}$ the linear operator that selects the set of correct entries of Z (and so $\mathcal{A}^*$ is the *zero upsampling* operator). The objective is to recover the image, by filling the erased pixels.

Following [24] we consider the following formulation of the in-panting problem:

$$\min_{Z \in \mathcal{H}} F(Z) := \frac{1}{2} \|\mathcal{A}(Z - Y)\|^2 + w \left\| Z_{(1)} \right\|_* + w \left\| Z_{(2)} \right\|_*, \tag{48}$$

Algorithm 2: Inertial Three-Operator Splitting

Choose $Z_0, Z_1 \in \mathbb{R}^{m\times n}$, $(\lambda_k)_{k\in\mathbb{N}}$ and $(\alpha_k)_{k\in\mathbb{N}}$ such that hypotheses of Theorem 4 are fulfilled, $\rho \in (0,2)$, $\varepsilon > 0$ and $r_0 > \varepsilon$;

while $r_k > \varepsilon$ **do**

$U_k = Z_k + \alpha_k(Z_k - Z_{k-1})$;

$X_k^g = \mathrm{prox}_{\rho g}(U_k)$;

$Z_{k+\frac{1}{2}} = 2X_k^g - U_k - \rho\mathcal{A}^*\nabla h(\mathcal{A}X_k^g)$;

$Z_{k+1} = U_k + \lambda_k(\mathrm{prox}_{\rho f}(Z_{k+\frac{1}{2}}) - X_k^g)$;

$r_{k+1} = \mathcal{R}(Z_{k+1}, Z_k)$

end

Return Z_{n+1}, X_n^g;

(a) Original image (b) Corrupted image (c) Recovered without inertia (d) Recovered with inertia

Fig. 7 Original image (a), corrupted image with 250,000 randomly erased pixels (b), images recovered without inertia (c), and with inertia (d)

where $\mathcal{H}$ is the set of 3-D tensors, $Z_{(1)}$ is the matrix $[Z(:,:,1)\, Z(:,:,2)\, Z(:,:,3)]$, $Z_{(2)}$ is the matrix $[Z(:,:,1)^T\, Z(:,:,2)^T\, Z(:,:,3)^T]^T$, $\|\cdot\|_*$ denotes the matrix nuclear norm and w is a penalty parameter, which we take equal to 1 here, for simplicity. This problem fits in the context of (43), with $f(Z) = g(Z) = \|Z\|_*$ and $h(Z) = \frac{1}{2}\|Z - Y\|_2^2$. In this case, the operator $\nabla(h \circ \mathcal{A})$ is cocoercive with constant 1. With the error function $\mathcal{R}$ defined in (46), the iterations defined by (42) lead to Algorithm 2.

As in the previous section, Algorithm 2 will be tested in the case $\alpha_k \equiv 0$ (the algorithm studied in [24]) and, for the inertial version,

$$\alpha_k = \left(1 - \frac{1}{k}\right)\alpha, \tag{49}$$

where α satisfies the condition (34). The corresponding algorithms will be referred to as original and inertial, respectively. Algorithm (2) returns both the value of Z_k and X_k^g, since the latter represents the image solution of the problem. Throughout this section, the initial points are both set to zero.

Comparison in terms of the number of erased pixels Between 10,000 and 250,000 pixels are randomly erased from the image in Fig. 7a to obtain the one in Fig. 7b (Fig. 7c and Fig. 7d show the recovered images, as described below). We compare the number of iterations and execution time needed by both methods with step size $\rho = 1$ and $\lambda_k \equiv 1$, for a tolerance of 10^{-3}. The results are shown in Fig. 8. The reduction stands between 12% and 22% in most cases, and the improvement seems to increase with the number of erased pixels.

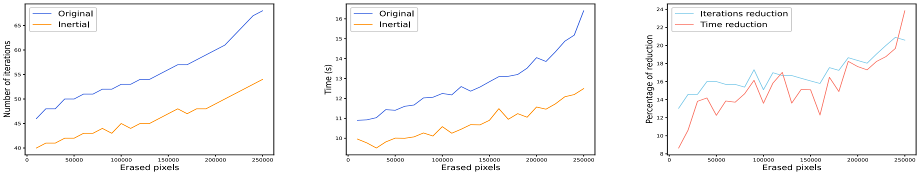


Fig. 8 In-painting problem. Number of iterations (left), execution time (center) and percentage of reduction (right) of Algorithm 2 in terms of the number of erased pixels, with step size $\rho = 1$ and relaxation parameter $\lambda_k \equiv 1$, for a tolerance of 10^{-3}

Table 4 In-painting problem. Execution time and number of iterations in terms of the step size ρ, for fixed $\lambda_k \equiv 1$ and 250,000 erased pixels

ρ	Original algorithm		Inertial algorithm	
	Time (s)	Iterations	Time (s)	Iterations
0.1	119.80	524	70.04	301
0.2	64.25	281	39.28	169
0.3	44.61	195	28.55	122
0.4	34.88	150	22.67	98
0.5	28.20	123	19.80	83
0.6	23.90	104	17.17	73
0.7	21.13	91	15.46	66
0.8	18.46	81	14.08	61
0.9	16.74	74	13.68	58
1.0	15.81	69	13.25	56
1.1	14.87	65	12.94	56
1.2	14.60	64	13.24	56
1.3	14.34	63	13.23	57
1.4	14.67	64	13.39	58
1.5	14.55	64	13.90	60

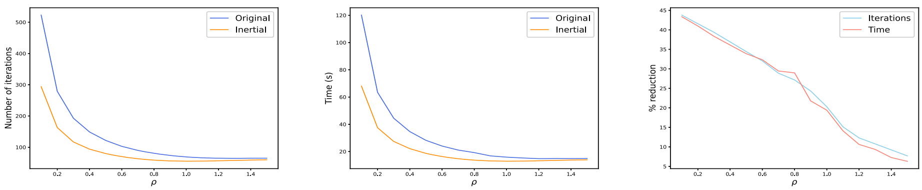


Fig. 9 In-painting problem. Number of iterations (left), execution time (center) and percentage of reduction (right) in terms of the step size ρ for a fixed $\lambda_k \equiv 1$ and 250,000 erased pixels

Comparison in terms of the step size Both algorithms are tested for the same image with 250,000 randomly erased pixels for $\lambda_k \equiv 1$ and different values of the step size ρ. For the inertial version, the constant α in (49) is adapted accordingly. The results are reported in Table 4 and depicted graphically in Fig. 9. The percentage of reduction is noticeably higher

Table 5 In-painting problem. Execution time and number of iterations for different values of λ, for a fixed value $\rho = 1$ and 250,000 erased pixels

λ	Original algorithm		Inertial algorithm	
	Time (s)	Iterations	Time (s)	Iterations
0.6	24.47	108	13.57	56
0.7	21.28	94	11.65	51
0.8	18.67	83	12.64	55
0.9	16.94	75	12.76	56
1.0	15.52	69	12.76	56
1.1	14.28	63	12.51	55
1.2	13.35	59	12.53	54
1.3	12.52	55	11.90	52
1.4	12.04	52	11.71	51

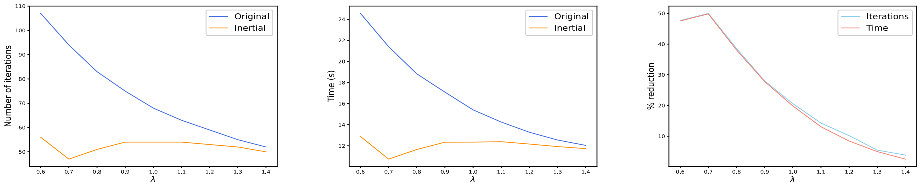


Fig. 10 In-painting problem. Number of iterations (left), execution time (center) and percentage of reduction (right) in terms of the relaxation parameter λ for a fixed $\rho = 1$ and 250,000 erased pixels

for lower values of ρ (always above 20% when $\rho \leq 1$). This is to be expected, since larger values of ρ require lower values of α, which limits the effect of inertia.

Comparison in terms of the relaxation parameter Finally, we fix the value $\rho = 1$, and compare the performance of the two methods for different values of the relaxation parameter λ, which, as before, limit the possible range for the inertial parameter α in view of condition (34). The results are presented in Table 5, and shown graphically in Fig. 10. As with the step size, the reduction is greater for lower values of λ, which is consistent with the loss of the inertial character imposed by condition (34). Nevertheless, observe that over-relaxing with $\lambda = 1.2$ or $\lambda = 1.4$ gives better results (both in number of iterations and execution time) than keeping λ in a neighborhood of 1.

The evolution of the function values, the distance to the limit and the residuals are shown (in logarithmic scale) in Fig. 11 for 250,000 erased pixels, using $\rho = 1$ and $\lambda_k \equiv 1$. As in the previous example, the sequence $k\,\|z_k - Tz_k\|^2$ tends to zero, allowing us to conjecture again an asymptotic rate of $o(1/k)$. For the comparison between the original, corrupted (with 250,000 erased pixels) and recovered images, we refer the reader back to Fig. 7.[5]

[5] For the sake of a fair visual comparison, we follow the implementation used in [24], as described in https://damek.github.io/ThreeOperators.html, which differs slightly from the description given in Sect. 4.3 in that it contains a Bregman update.

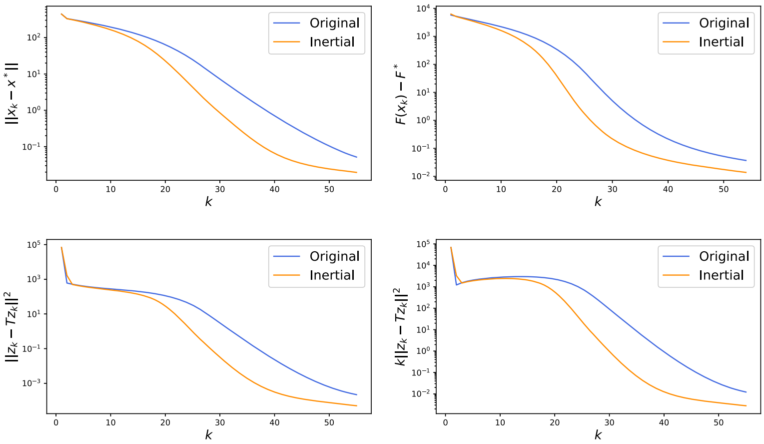


Fig. 11 In-painting problem. Evolution to the distance to the computed solution (top left), objective function values (top right), residuals $\|z_k - Tz_k\|^2$ (bottom left) and $k\,\|z_k - Tz_k\|^2$ (bottom right), for 250,000 erased pixels using $\rho = 1$ and $\lambda_k \equiv 1$

Declarations

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alvarez, F.: On the minimizing property of a second order dissipative system in Hilbert spaces. SIAM J. Control Optim. **38**(4), 1102–1119 (2000)
2. 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)
3. Attouch, H., Cabot, A.: Convergence of a relaxed inertial forward–backward algorithm for structured monotone inclusions. Appl. Math. Optim. **80**(3), 547–598 (2019)
4. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. SIAM J. Optim. **26**(3), 1824–1834 (2016)
5. Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. Math. Program. **174**, 391–432 (2019)
6. 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)
7. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. **168**(1–2), 123–175 (2018)
8. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017). With a foreword by Hédy Attouch
9. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)

10. Boţ, R.I., Csetnek, E.R., Hendrich, C.: Inertial Douglas–Rachford splitting for monotone inclusion problems. Appl. Math. Comput. **256**, 472–487 (2015)
11. Brézis, H., Lions, P.L.: Produits infinis de résolvantes. Isr. J. Math. **29**(4), 329–345 (1978)
12. Briceno-Arias, L.M.: Forward-Douglas–Rachford splitting and forward-partial inverse method for solving monotone inclusions. Optimization **64**(5), 1239–1261 (2015)
13. Briceño-Arias, L., Roldán, F.: Primal-dual splittings as fixed point iterations in the range of linear operators. J. Glob. Optim. (2022)
14. Cauchy, A., et al.: Méthode générale pour la résolution des systemes d'équations simultanées. C. R. Sci. Paris **25**(1847), 536–538 (1847)
15. Chambolle, A., Dossal, C.: On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". J. Optim. Theory Appl. **166**(3), 968–982 (2015)
16. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011)
17. Chambolle, A., Caselles, V., Cremers, D., Novaga, M., Pock, T.: An introduction to total variation for image analysis. Theor. Found. Numer. Methods Sparse Recovery **9**, 263–340 (2010)
18. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**(5–6), 475–504 (2004)
19. Combettes, P.L., Glaudin, L.E.: Quasi-nonexpansive iterations on the affine hull of orbits: from Mann's mean value algorithm to inertial methods. SIAM J. Optim. **27**(4), 2356–2380 (2017)
20. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Model. Simul. **4**(4), 1168–1200 (2005)
21. Condat, L.: A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**(2), 460–479 (2013)
22. Cortild, D., Peypouquet, J.: Krasnoselskii-Mann iterations: Inertia, perturbations and approximation (2024). arXiv:2401.16870
23. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. **57**(11), 1413–1457 (2004)
24. Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. Set-Valued Var. Anal. **25**(4), 829–858 (2017)
25. Dong, Y.: New inertial factors of the Krasnosel'skiı-Mann iteration. Set-Valued Var. Anal. **29**, 145–161 (2021)
26. Dong, Q.-L., Yuan, H.-b.: Accelerated Mann and cq algorithms for finding a fixed point of a nonexpansive mapping. Fixed Point Theory Appl. **2015**(1), 1 (2015)
27. Dong, Q.-L., Cho, Y.J., Rassias, T.M.: General inertial Mann algorithms and their convergence analysis for nonexpansive mappings. Appl. Nonlinear Anal., 175–191 (2018)
28. Dong, Q.-L., Cho, Y.J., He, S., Pardalos, P.M., Rassias, T.M.: The Krasnosel'skii-Mann Iterative Method: Recent Progress and Applications. Springer, Berlin (2022)
29. Douglas, J. Jr., Rachford, H.H. Jr.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Am. Math. Soc. **82**, 421–439 (1956)
30. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
31. Goldstein, A.A.: Convex programming in Hilbert space. Bull. Am. Math. Soc. **70**(5), 709–710 (1964)
32. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**(2), 403–419 (1991)
33. Hansen, P.C., Nagy, J.G., O'Leary, D.P.: Deblurring Images. Fundamentals of Algorithms, vol. 3. SIAM, Philadelphia (2006). Matrices, spectra, and filtering
34. Iyiola, O.S., Shehu, Y.: New convergence results for inertial Krasnoselskii–Mann iterations in Hilbert spaces with applications. Results Math. **76**(2), 1–25 (2021)
35. Krasnosel'skii, M.A.: Two comments on the method of successive approximations. Usp. Math. Nauk **10**, 123–127 (1955)
36. Levitin, E.S., Polyak, B.T.: Constrained minimization methods. USSR Comput. Math. Math. Phys. **6**(5), 1–50 (1966)
37. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)
38. Lorenz, D.A., Pock, T.: An inertial forward-backward algorithm for monotone inclusions. J. Math. Imaging Vis. **51**(2), 311–325 (2014)
39. Maingé, P.-E.: Convergence theorems for inertial KM-type algorithms. J. Comput. Appl. Math. **219**(1), 223–236 (2008)
40. Mann, W.R.: Mean value methods in iteration. Proc. Am. Math. Soc. **4**(3), 506–510 (1953)
41. Martinet, B.: Regularisation, d'inéquations variationelles par approximations succesives. Revue Française d'informatique et de Recherche operationelle (1970)

42. Moudafi, A.: A reflected inertial Krasnoselskii-type algorithm for Lipschitz pseudo-contractive mappings. Bull. Iran. Math. Soc. **44**, 1109–1115 (2018)
43. Moudafi, A., Oliny, M.: Convergence of a splitting inertial proximal method for monotone operators. J. Comput. Appl. Math. **155**(2), 447–454 (2003)
44. Nesterov, Y.: A method for solving the convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. Proc. Acad. Sci. USSR **269**, 543–547 (1983)
45. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Am. Math. Soc. **73**(4), 591–597 (1967)
46. Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J. Math. Anal. Appl. **72**(2), 383–390 (1979)
47. Peypouquet, J.: Convex Optimization in Normed Spaces. SpringerBriefs in Optimization. Springer, Cham (2015)
48. Polyak, B.: Some methods of speeding up the convergence of iteration methods. USSR Comput. Math. Math. Phys. **4**, 1–17 (1964)
49. Pustelnik, N.: Méthodes proximales pour la résolution de problèmes inverses: application à la tomographie par émission de positrons. PhD thesis, Université Paris-Est (2010)
50. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**(5), 877–898 (1976)
51. Shehu, Y.: Convergence rate analysis of inertial Krasnoselskii–Mann type iteration with applications. Numer. Funct. Anal. Optim. **39**(10), 1077–1091 (2018)
52. Shehu, Y., Gibali, A., Sagratella, S.: Inertial projection-type methods for solving quasi-variational inequalities in real Hilbert spaces. J. Optim. Theory Appl. **184**(3), 877–894 (2020)
53. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv. Comput. Math. **38**(3), 667–681 (2013)

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