Inertial Krasnoselskii-Mann Iterations

Ignacio Fierro, Juan José Maulén, Juan Peypouquet

October 2022

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 worst case estimates for the 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 parameter hypotheses and convergence rates. Numerical illustrations 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 (2020) 47H05 · 47H10 · 65K05 · 90C25

1 Introduction

Krasnoselskii-Mann (KM) iterations [28, 32] 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* [30, 37] 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 [24, 29], the gradient method [13] and the proximal point algorithm [33, 41, 10, 25], to cite some abstract methods, as well as the *Iterative Shrinkage-Thresholding Algorithm* (ISTA) [20, 18], to speak more concretely. KM iterations also encompass other splitting methods like *Douglas-Rachford* [22], primal-dual methods [16, 3, 17, 43, 19] and the three-operator splitting [21].

Convex optimization methods can be enhanced, by adding an inertial substep, motivated by physical considerations [39, 34, 1]. Extensions beyond the optimization setting have been developed in [2] and, more recently, in [31, 9, 4, 42, 27], to mention a few. 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) [8], obtained in [15, 5, 6].

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

$$\left\{ \begin{array}{rcl} y_k & = & x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} & = & (1-\lambda_k)y_k + \lambda_k T_k(y_k), \end{array} \right. \tag{1}$$

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.

The paper is organized as follows: in Section 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 worst case estimates for the 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 Section 4, we discuss several instances of KM iterations, which are relevant to the numerical illustrations provided in Section 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

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

The following auxiliary result will be useful in the sequel:

Lemma 1. *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*

$$\Delta_{k+1}(p) + \delta_{k+1} + \nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 \leq \alpha_k\Delta_k(p) + \left[\alpha_k(1 + \alpha_k) + \nu_k\alpha_k(1 - \alpha_k)\right]\|x_k - x_{k-1}\|^2. \tag{4}$$

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

$$\|x_{k+1} - p\|^2 = \|y_k - p\|^2 + \lambda_k^2\|y_k - T_k y_k\|^2 + 2\lambda_k\langle y_k - p, T_k y_k - y_k\rangle \leq \|y_k - p\|^2 - \lambda_k(1 - \lambda_k)\|y_k - T_k y_k\|^2, \tag{5}$$

where the inequality is given by (2). From

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

and

$$2\alpha_k\langle x_k - p, x_k - x_{k-1}\rangle = \alpha_k\|x_k - p\|^2 + \alpha_k\|x_k - x_{k-1}\|^2 - \alpha_k\|x_{k-1} - p\|^2,$$

we deduce that

$$\|y_k - p\|^2 = (1 + \alpha_k)\|x_k - p\|^2 + \alpha_k(1 + \alpha_k)\|x_k - x_{k-1}\|^2 - \alpha_k\|x_{k-1} - p\|^2. \tag{6}$$

By combining expressions (5) and (6), we obtain

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

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

$$\Delta_{k+1}(p) \leq \alpha_k\Delta_k(p) + \alpha_k(1 + \alpha_k)\|x_k - x_{k-1}\|^2 - \lambda_k(1 - \lambda_k)\|y_k - T_k y_k\|^2. \tag{7}$$

In turn,

$$\lambda_k^2\|y_k - T_k y_k\|^2 = \|x_{k+1} - x_k\|^2 + \alpha_k^2\|x_k - x_{k-1}\|^2 - 2\alpha_k\langle x_{k+1} - x_k, x_k - x_{k-1}\rangle, \tag{8}$$

and

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

together give

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

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

$$\delta_{k+1} + \nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 = \nu_k\alpha_k(1 - \alpha_k)\|x_k - x_{k-1}\|^2 + \lambda_k(1 - \lambda_k)\|y_k - T_k y_k\|^2. \tag{9}$$

Summing (7) and (9), we obtain (4). □

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

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 faster than $1/\sqrt{k}$, as occurs with standard KM iterations.

Proposition 2. *Let (T_k) be a family of quasi-nonexpansive operators on $\mathcal{H}$, and let (x_k, y_k) satisfy (1). Take $p \in F = \bigcap_{k\geq 1} \mathrm{Fix}(T_k)$.*

i) Assume that there is $k_0 \geq 1$ such that

$$\alpha_k(1+\alpha_k) + (\lambda_k^{-1} - 1)\alpha_k(1-\alpha_k) - (\lambda_{k-1}^{-1} - 1)(1-\alpha_{k-1}) \leq 0 \tag{10}$$

for all $k \geq k_0$. Then, the sequence $\big(C_k(p)\big)_{k\geq k_0}$ is nonincreasing and nonnegative, thus $\lim_{k\to\infty} C_k(p)$ exists.

ii) Assume further that

$$\limsup_{k\to\infty} \big[\alpha_k(1+\alpha_k) + (\lambda_k^{-1} - 1)\alpha_k(1-\alpha_k) - (\lambda_{k-1}^{-1} - 1)(1-\alpha_{k-1})\big] < 0. \tag{11}$$

Then, the series $\sum_{k\geq 1} \alpha_k \|x_{k+1} - 2x_k + x_{k-1}\|^2$, $\sum_{k\geq 1} \|x_k - x_{k-1}\|^2$, $\sum_{k\geq 1} \delta_k$ and $\sum_{k\geq 1} \|y_k - T_k y_k\|^2$ are convergent. Moreover, $\lim_{k\to\infty} \|x_k - p\|$ exists, and

$$\lim_{k\to\infty} k\|y_k - T_k y_k\|^2 = \lim_{k\to\infty} k\|x_k - x_{k-1}\|^2 = 0.$$

Proof. With either (10) or (11), there exist $\varepsilon \geq 0$ and $k_0 \geq 1$ such that

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

for all $k \geq k_0$ (if (11) holds, then $\varepsilon > 0$; otherwise, $\varepsilon = 0$). Without any loss of generality, we may assume that $k_0 = 1$. Combining this with (4), we obtain

$$\begin{aligned}\Delta_{k+1} + \delta_{k+1} + \nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 &\leq \alpha_k\Delta_k + \big[\nu_{k-1}(1-\alpha_{k-1}) - \varepsilon\big]\|x_k - x_{k-1}\|^2 \\ &= \alpha_k\Delta_k + \delta_k - \varepsilon\|x_k - x_{k-1}\|^2. \end{aligned}\tag{13}$$

On the one hand, (13) immediately gives

$$\Delta_{k+1} \leq \alpha_k\Delta_k + \delta_k. \tag{14}$$

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

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

Therefore, (13) implies

$$C_{k+1}(p) + \nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 + \varepsilon\|x_k - x_{k-1}\|^2 \leq C_k(p). \tag{15}$$

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 $k_1 \geq 1$. Since $\big(C_k(p)\big)$ is nonincreasing,

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

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

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

for all $k \geq k_1$, 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 at once from (15). The third one is a consequence of the second one, since $\inf_{k\geq 1}\lambda_k > 0$. For the last one, we use (8).
Now, denoting the positive part of $d \in \mathbb{R}$ by $[d]_+$, we obtain from (14) that

$$[\Delta_{k+1}]_+ \leq \alpha[\Delta_k]_+ + \delta_k,$$

where $\alpha := \sup_{k\geq 1}\alpha_k < 1$ by virtue of (11). Iterating this inequality, we deduce that

$$[\Delta_{j+1}]_+ \leq \alpha^j [\Delta_1]_+ + \sum_{k=1}^{j} \alpha^{j-k}\delta_k,$$

and so

$$\sum_{j=1}^{\infty} [\Delta_{j+1}]_+ \leq \frac{1}{1-\alpha}\left[[\Delta_1]_+ + \sum_{j=1}^{\infty}\delta_j \right] < \infty.$$

Now, write $h_k = \|x_k - p\|^2 - \sum_{j=1}^{k} [\Delta_j]_+$. We have

$$h_{k+1} - h_k = \Delta_{k+1} - [\Delta_{k+1}]_+ \leq 0,$$

from which we conclude that $\lim_{k\to\infty} \|x_k - p\|$ exists. Finally, from (13), we deduce that

$$[\Delta_{k+1}]_+ + \delta_{k+1} \leq [\Delta_k]_+ + \delta_k.$$

We have already shown that

$$\sum_{k=1}^{\infty} \big([\Delta_k]_+ + \delta_k\big) < \infty.$$

An elementary property of real sequences (see Lemma 5 below) implies that $\lim_{k\to\infty} k\,[\Delta_k]_+ = \lim_{k\to\infty} k\,\delta_k = 0$, which allows us to conclude. □

Remark 3. Inequalities (10) and (11) are closely related, but different, from the hypotheses used in [4] for forward-backward iterations. In the non-inertial case $\alpha = 0$, (11) 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 (11) is reduced to

$$\lambda(1-\alpha+2\alpha^2) < (1-\alpha)^2. \tag{16}$$

For each $\alpha \in [0,1)$, there is $\lambda_\alpha > 0$ such that (16) 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. We shall discuss other examples in the next section.

Theorem 4. *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 (11) 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 Proposition 2. 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 [36] (see, for instance, [38, Lemma 5.2]) yields the conclusion. □

In the proof of Proposition 2, we used the following result concerning real sequences:

Lemma 5. *Let (ζ_k) be a nonincreasing sequence of positive numbers such that $\sum_{k\geq 1}\zeta_k < \infty$. Then $\lim_{k\to\infty}[k\,\zeta_k] = 0$.*

Proof. On the one hand, since (ζ_k) is nonincreasing, for each $k \geq 1$, we have

$$k\,\zeta_k - (k-1)\,\zeta_{k-1} = \zeta_k + (k-1)(\zeta_k - \zeta_{k-1}) \leq \zeta_k.$$

But $k\,\zeta_k \geq 0$ and the right-hand side is summable, so $\lim_{k\to\infty}[k\,\zeta_k]$ exists. On the other hand, since

$$\sum_{k\geq 1}\frac{1}{k}[k\,\zeta_k] = \sum_{k\geq 1}\zeta_k < \infty \qquad \text{and} \qquad \sum_{k\geq 1}\frac{1}{k} = \infty,$$

we must have $\lim_{k\to\infty}[k\,\zeta_k] = 0$. □

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

$$Q(\lambda, q, \xi) := \xi\big(1 - \lambda + \lambda q^2\big) + (1-\xi)(1 - \lambda + \lambda q)^2 = (1 - \lambda + \lambda q)^2 + \xi\lambda(1-\lambda)(1-q)^2. \tag{17}$$

Notice that $Q(\lambda, q, \xi)$ 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 Subsection 3.2).

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

Lemma 6. *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*

$$\|x - p^*\|^2 \leq Q(\lambda, q, \xi)\|y - p^*\|^2 - \xi\lambda(1-\lambda)\|Ty - y\|^2. \tag{18}$$

Proof. On the one hand, the equalities

$$\begin{array}{rcccl} \|x - p^*\|^2 &=& \|y - p^* + \lambda(Ty - y)\|^2 &=& \|y - p^*\|^2 + \lambda^2\|Ty - y\|^2 + 2\lambda\langle y - p^*, Ty - y\rangle \\ -\lambda\|Ty - p^*\|^2 &=& -\lambda\|Ty - y + y - p^*\|^2 &=& -\lambda\|Ty - y\|^2 - \lambda\|y - p^*\|^2 - 2\lambda\langle y - p^*, Ty - y\rangle, \end{array}$$

together give

$$\begin{aligned} \|x - p^*\|^2 &= (1-\lambda)\|y - p^*\|^2 + \lambda\|Ty - p^*\|^2 - \lambda(1-\lambda)\|Ty - y\|^2 \\ &\leq \big(1 - \lambda + \lambda q^2\big)\|y - p^*\|^2 - \lambda(1-\lambda)\|Ty - y\|^2. \end{aligned} \tag{19}$$

On the other hand, we have

$$\|x - p^*\| = \|(1-\lambda)(y - p^*) + \lambda(Ty - p^*)\| \leq (1-\lambda)\|y - p^*\| + \lambda\|Ty - p^*\| \leq (1 - \lambda + \lambda q)\|y - p^*\|. \tag{20}$$

Then, inequality (18) is just a convex combination of (19) and the square of (20). □

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 7. *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 (17). For each $k \in \mathbb{N}$, we have*

$$\begin{aligned} \|x_{k+1} - p^*\|^2 + \xi\delta_{k+1} \leq\ & Q_k\left[(1+\alpha_k)\|x_k - p^*\|^2 - \alpha_k\|x_{k-1} - p^*\|^2\right] \\ & + \big[Q_k\alpha_k(1+\alpha_k) + \xi\nu_k\alpha_k(1-\alpha_k)\big]\|x_k - x_{k-1}\|^2. \end{aligned} \tag{21}$$

If, moreover,

$$Q_k\alpha_k(1+\alpha_k) + \xi\nu_k\alpha_k(1-\alpha_k) - \xi Q_k\nu_{k-1}(1-\alpha_{k-1}) \leq 0 \tag{22}$$

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

$$\tilde{C}_{k+1}(p^*) \leq \left[\prod_{j=1}^{k} Q_j\right]\|x_1 - p^*\|^2 \tag{23}$$

and

$$\|x_{k+1} - p^*\|^2 \leq \left[\alpha^k + \sum_{j=1}^{k}\alpha^{k-j}\left[\prod_{i=1}^{j} Q_i\right]\right]\|x_1 - p^*\|^2. \tag{24}$$

Proof. We use (1) and (18) to obtain

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

Now, by (6), we deduce that

$$\|x_{k+1}-p^*\|^2 \le Q_k\left[(1+\alpha_k)\|x_k-p^*\|^2+\alpha_k(1+\alpha_k)\|x_k-x_{k-1}\|^2-\alpha_k\|x_{k-1}-p^*\|^2\right]-\xi\lambda_k(1-\lambda_k)\|y_k-T_ky_k\|^2$$

On the other hand, from (9), we get

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

and the last two inequalities together imply (21). For the second part, inequalities (21) and (22) together give

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

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

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

This gives (23), recalling that $\tilde{C}_1(p^*)=\|x_1-p^*\|^2$. Now, since $\|x_{k+1}-p^*\|^2-\alpha_k\|x_k-p^*\|^2\le\tilde{C}_{k+1}(p^*)$, we have

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

which we then iterate to obtain (24). □

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

Theorem 8. *Let (T_k) be a sequence of operators on 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 $Q_k=Q(\lambda_k,q_k,\xi)$, and assume that (22) 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 $\lambda_k\ge\lambda>0$ and $q_k\le q<1$ for all $k\in\mathbb{N}$, then x_k converges linearly to p^*, as $k\to\infty$. More precisely,

$$\|x_k-p^*\|^2\le\left[\frac{\alpha^{k+1}-Q(\lambda,q,\xi)^{k+1}}{\alpha-Q(\lambda,q,\xi)}\right]\|x_1-p^*\|^2=\mathcal{O}\left(Q(\lambda,q,\xi)^k\right). \tag{25}$$

Proof. For part i), observe that if $\sum_{k=1}^{\infty}\lambda_k(1-q_k^2)=\infty$, then $\prod_{k=1}^{\infty}Q_k=0$. By (23), $\lim_{k\to\infty}\tilde{C}_k(p^*)=0$. As in the proof of Proposition 2, 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 $\alpha_k\equiv0$, the conclusion is straightforward. Otherwise, given any $\varepsilon>0$, there is $K\in\mathbb{N}$ such that

$$\|x_k-p^*\|^2\le\alpha_1\|x_{k-1}-p^*\|^2+\varepsilon$$

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

$$\|x_k-p^*\|^2\le\alpha_1^{k-K}\|x_K-p^*\|^2+\varepsilon(1-\alpha_1)^{-1},$$

so that $\limsup_{k\to\infty}\|x_k-p^*\|\le\varepsilon(1-\alpha_1)^{-1}$, and the conclusion follows.
For ii), we know that $Q(\lambda_k,q_k,\xi)\le Q(\lambda,q,\xi)$, 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 (22), we deduce that $Q\ge\alpha$ (strictly if $\alpha>0$). Using (24), and observing that the case $Q(\lambda,q,\xi)=\alpha$ is incompatible with inequality (22), we deduce that

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

as claimed. □

3.2 Some insights into inequality (22) and the convergence rate

To fix the ideas, we comment on some special cases of inequality (22), especially with constant parameters:

1. In the non-inertial case $\alpha_k \equiv 0$, (22) holds if either $\xi = 0$ or $\lambda_k \leq 1$ for all k, as in (10). This is less restrictive than (11) (see Remark 3). If we take $\xi = 0$, the best convergence rate is

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

which is obtained from Theorem 8 with $\lambda_k \equiv 1$. If $\alpha_k > 0$ for at least one k, the case $\xi = 0$ is ruled out. From a theoretical perspective, this seems to be an argument *against* the use of inertia, since

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

However, the results above correspond to worst-case scenarios, and need not be representative of concrete instances found in practice, in which inertia does improve the theoretical convergence rate guarantees (see Subsection 4.2 below, and the commented references). Moreover, the numerical tests reported below show noticeable improvements in the performance of the selected algorithms.

2. In the limiting case $q_k \equiv 1$, we have $Q_k \equiv 1$. With constant parameters $\lambda_k \equiv \lambda$, $\alpha_k \equiv \alpha$, (22) becomes

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

If

$$\frac{\alpha\lambda(1+\alpha)}{(1-\lambda)(1-\alpha)^2} \leq 1, \tag{26}$$

then, there is $\xi_{\alpha,\lambda,1} \in (0,1)$ such that (22) holds for all $\xi \in [\xi_{\alpha,\lambda,1}, 1]$. If $\xi = 1$, it is precisely the constant case in (10) (see (16) for a more direct comparison).

3. 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 (22) is equivalent to

$$\Psi(\lambda) := (1+\alpha^2)(1-q^2)\lambda^2 - \big(2\alpha^2 + (1-\alpha)(2-q^2)\big)\lambda + (1-\alpha)^2 \geq 0. \tag{27}$$

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 (27) holds for all $\lambda \leq \lambda_{\alpha,q}$. The values of $\lambda_{\alpha,q}$ on $[0,1) \times (0,1)$ are depicted in Figure 1. Once a value for the

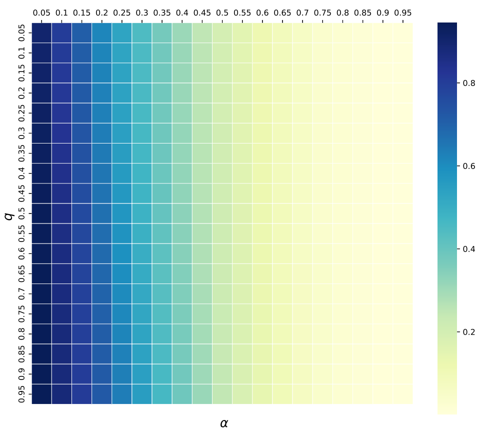


Figure 1: Values of $\lambda_{\alpha,q}$.

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

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

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 (22). In the fully general case, if α, λ and q satisfy

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

then, there is $\xi_{\alpha,\lambda,q}\in(0,1)$ such that (22) holds for all $\xi\in[\xi_{\alpha,\lambda,q},1]$. As $q\to 1$, we recover (26) as a limit case.

4 Examples

4.1 Averaged Operators

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

$$\left\{\begin{array}{rcl} y_k & = & x_k+\alpha_k(x_k-x_{k-1})\\ x_{k+1} & = & (1-\gamma_k\lambda_k)y_k+\gamma_k\lambda_k R(y_k),\end{array}\right. \tag{28}$$

and (11) becomes

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

If $\gamma_k\lambda_k\to\eta>0$, this is

$$\eta(1-\alpha+2\alpha^2)<(1-\alpha)^2. \tag{29}$$

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.

4.2 Euler Iterations and Gradient Descent

An operator B is β-*cocoercive* with $\beta>0$ if $\langle Bx-By,x-y\rangle\geq\beta\left\|Bx-By\right\|^2$ for all $x,y\in\mathcal{H}$.

Let $B:\mathcal{H}\to\mathcal{H}$ be cocoercive with constant β, and let (ρ_k) be a sequence in $(0,2\beta)$. For each $k\geq 1$, set

$$T_k=I-\rho_k B.$$

Then, T_k is nonexpansive (thus quasi-nonexpansive) and $(\rho_k/2\beta)$-averaged. If $\rho_-:=\inf_{k\geq 1}\rho_k>0$, the family $(I-T_k)$ is asymptotically demiclosed. If $\lambda_k\rho_k\to\sigma$, (11) becomes

$$\sigma(1-\alpha+2\alpha^2)<2\beta(1-\alpha)^2.$$

Now, let $f:\mathcal{H}\to\mathcal{H}$ be convex and differentiable, and assume ∇f is Lipschitz-continuous with constant L. Then, $B=\nabla f$ is cocoercive with constant $\beta=1/L$. If, moreover, f is strongly convex with parameter μ and $\rho_k\leq 2/(L+\mu)$, then T_k is q_k-quasi-contractive with

$$q_k=1-\frac{2\mu L\rho_k}{L+\mu}\leq 1-\frac{2\mu L\rho_-}{L+\mu}=:q.$$

Therefore, (T_k) is q-quasi-contractive. Considering the non-inertial case ($\alpha_k \equiv 0$), $\lambda_k \equiv 1$ and the fixed-sted choice $\rho_k = 2/(\mu + L)$, the algorithm exhibits a rate of convergence

$$f(x_k) - f^* \leq \frac{L}{2}\left(\frac{Q-1}{Q+1}\right)^{2k} \|x_0 - x^*\|^2,$$

where $Q = L/\mu$ is the *condition number* ([35, Theorem 2.1.15]. Introducing the inertial term, and using

$$\rho_k = 1/L \qquad \text{and} \qquad \alpha_k \equiv \left(\frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}\right),$$

it turns into *constant step scheme, III* [35], which has a rate of convergence of

$$f(x_k) - f^* \leq \min\left\{\left(1-\sqrt{\frac{\mu}{L}}\right)^k, \frac{4L}{(2\sqrt{L}+k\sqrt{\mu})^2}\right\}\left(f(x_0) - f^* + \frac{\mu}{2}\|x_0 - x^*\|^2\right)$$

Here, (11) can be written as

$$\lambda < \frac{2Q}{1-\sqrt{Q}+2Q},$$

which gives the condition for the convergence of Nesterov's constant step scheme with constant relaxation λ.

4.3 Proximal and Forward-Backward Methods

Let $M : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let (ρ_k) be a positive sequence. The *proximal method* consists in iterating

$$z_{k+1} = (I + \rho_k M)^{-1} z_k, \tag{30}$$

for $k \geq 1$. The operator $T_k = J_{\rho_k M} := (I + \rho_k M)^{-1}$ is nonexpansive, $\frac{1}{2}$-averaged, and $Z = \bigcap_{k\geq 1} \mathrm{Fix}(T_k) = M^{-1}0$. If $\lambda_k \to \lambda$, inequality (11) is reduced to

$$\lambda(1 - \alpha + 2\alpha^2) < 2(1-\alpha)^2.$$

As before, the family $(I - T_k)$ is asymptotically demiclosed at 0 if $\inf_{k\geq 1} \rho_k > 0$. To see this, let (z_k) be a sequence in $\mathcal{H}$ such that $z_k \rightharpoonup z$ and $z_k - T_k z_k \to 0$. We must show that $0 \in Mz$. By the definition of T_k, we have

$$\frac{1}{\rho_k}(z_k - T_k z_k) \in M(T_k z_k).$$

The left-hand side converges strongly to zero, while $T_k z_k \rightharpoonup z$. We conclude by the weak-strong closedness of the graph of M.

Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $B : \mathcal{H} \to \mathcal{H}$ be cocoercive with parameter β, 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 $\gamma_k = 2\beta(4\beta - \rho_k)^{-1}$. If $\rho_k \to \rho$ and $\lambda_k \to \lambda$, then (11) is equivalent to

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

As in the proximal case, the family $(I - T_k)$ is asymptotically demiclosed at 0 if $\inf_{k\geq 1} \rho_k > 0$.

4.4 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

$$T_r = J_{rA} \circ \big(2J_{rB} - I\big) + \big(I - J_{rB}\big) = \frac{1}{2}\big(I + (2J_{rA} - I) \circ (2J_{rB} - I)\big). \tag{31}$$

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 $\text{Zer}(A+B) = J_{rB}\,\text{Fix}(T_r)$.

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 9. 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

$$\min_{x\in X}\{f(x) + g(Lx)\} \qquad \text{and} \qquad \min_{y\in Y}\{g^*(y) + f^*(-L^*y)\}, \tag{32}$$

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 [23]).

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

$$\begin{cases} x_{k+1} &= J_{\tau A}\big(x_k - \tau L^* y_k\big) \\ y_{k+1} &= J_{\sigma B^{-1}}\big(y_k + \sigma L(2x_{k+1} - x_k)\big), \end{cases} \tag{33}$$

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 [7, Remark 4.24]).

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

$$\begin{cases} (y_k, v_k) = (x_k, u_k) + \alpha_k\,[(x_k, u_k) - (x_{k-1}, u_{k-1})] \\ p_{k+1} = J_{\tau A}(y_k - \tau L^* v_k) \\ q_{k+1} = J_{\sigma B^{-1}}(v_k + \sigma L(2p_{k+1} - y_k)) \\ (x_{k+1}, u_{k+1}) = (1 - \lambda_k)(y_k, v_k) + \lambda_k(p_{k+1}, q_{k+1}), \end{cases} \tag{34}$$

with appropriate sequences α_k and λ_k.

In [12], the authors propose the *Split Douglas-Rachford* algorithm

$$\begin{cases} v_k &= \Sigma\big(I - J_{\Sigma^{-1}B}\big)\big(Lx_k + \Sigma^{-1}y_k\big) \\ x_{k+1} &= J_{\Upsilon A}\big(x_k - \Upsilon L^* v_k\big) \\ y_{k+1} &= \Sigma L(x_{k+1} - x_k) + v_k, \end{cases} \tag{35}$$

where Υ and Σ are elliptic linear operators that induce an *ad-hoc* metric and account for preconditioning.

4.5 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

$$0 \in A\hat{x} + B\hat{x} + C\hat{x}. \tag{36}$$

If C is β-cocoercive, the *three-operator* splitting method [21] generates a sequence (z_k) by

$$\begin{cases} x_k^B = J_{\rho B}(z_k) \\ x_k^A = J_{\rho A}(2x_k^B - z_k - \rho C x_k^B) \\ z_{k+1} = z_k + \lambda_k(x_k^A - x_k^B) \end{cases} \tag{37}$$

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

$$\gamma = \frac{2\beta}{4\beta - \rho}. \tag{38}$$

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

$$\begin{cases} u_k = z_k + \alpha_k(z_k - z_{k-1}) \\ x_k^B = J_{\rho B}(u_k) \\ x_k^A = J_{\rho A}(2x_k^B - u_k - \rho C x_k^B) \\ z_{k+1} = u_k + \lambda_k(x_k^A - x_k^B), \end{cases} \tag{39}$$

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

$$\min f(x) + g(x) + h(Lx), \tag{40}$$

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

5 Numerical Illustrations

In this section, we test the performance of the algorithm given by iterations (1) in two of the settings described in section 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

$$\min_{x \in \mathbb{R}^{N_1 \times N_2}} F^{TV}(x) := \frac{1}{2} \|Rx - b\|^2 + w \|\nabla x\|_1, \tag{41}$$

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, w 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 [14]. In these experiments, R will be a Gaussian blur of size 9×9, standard deviation 4 and relative boundary conditions (see [26] for details on the construction of the operator), and $w = 10^{-4}$. Considering the original image $\bar{x}$ in Figure 3a 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 10^{-3} (Figure 3b).

Setting $f = 0$, $g : (u, v^1, v^2) \mapsto \frac{1}{2} \|u - b\|^2 + w \|v^1\|_1 + w \|v^2\|_1$ and $L : x \mapsto (Rx, D_1 x, D_2 x)$, the problem (41) can be formulated as (32), and solved via (34). Since

$$\mathrm{prox}_{\sigma g^*} : (u, v^1, v^2) \mapsto \left(\frac{u - \sigma b}{\sigma + 1}, v^1 - \sigma\, \mathrm{prox}_{\frac{w}{\sigma}\|\cdot\|_1}\left(\frac{v^1}{\sigma}\right), v^2 - \sigma\, \mathrm{prox}_{\frac{w}{\sigma}\|\cdot\|_1}\left(\frac{v^2}{\sigma}\right) \right) \tag{42}$$

we are lead to Algorithm 1.
For a stopping criterion, we consider the relative error

$$\mathcal{R}(x_{k+1}, x_k) \mapsto \frac{\|x_{k+1} - x_k\|}{\|x_k\|}. \tag{43}$$

Since the involved operator is 1/2-averaged (see [11]), we may set $\lambda_k \equiv \lambda \in (0, 2)$, as explained in Section 4.1.

The algorithm is tested for 17 combinations of τ, σ satisfying the critical condition $\tau\sigma \|L\|^2 = 1$ (according to [12], this tends to yield the best performance). The number $\|L\|$ is computed using an adaptation of [40, Algorithm 12].

Algorithm 1:

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 Proposition 2 are fulfilled, τ and σ such that $\tau\sigma \|L\|^2 \leq 1$, $\varepsilon > 0$ and $r_0 > \varepsilon$;

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

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

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

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

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

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

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

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

end

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

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 (33) (that is, Algorithm 1 with $\alpha_k \equiv 0$), and its inertial counterpart (34), with $\lambda_k \equiv 1$. The sequence $(\alpha_k)_{k\in\mathbb{N}}$ is

$$\alpha_k = \alpha \left(1 - \frac{1}{k^2}\right), \tag{44}$$

with $\alpha = 1/(3 + 0.0001)$ (condition (29) with $\eta = \lambda/2$ gives the constraint $\alpha < 1/3$). Table 1 shows the execution time, number of iterations, and the value for the objective value reached, using a tolerance $\varepsilon = 10^{-5}$. These results are depicted graphically, along with the percentage of reduction, in Figure 2. The recovered images are collected in Figures 3c and 3d.

			Original algorithm			Inertial algorithm		
Case	τ	σ	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

Table 1: Execution time, number of iterations and final function value for the original primal-dual algorithm and the inertial version, with tolerance $\varepsilon = 10^{-5}$.

Comparison in terms of the relaxation parameter λ. For both algorithms, case 14 showed the best performance in terms of iterations and execution time. We now assess the performance of the inertial algorithm with different values for $\lambda_k \equiv \lambda \in (0, 2)$, and the corresponding inertial parameters fulfilling condition (29). The results are shown in Table 2, along with the value of α used in (44). A graphic depiction is shown as heatmaps in Figure 4. Larger values of the relaxation parameter λ resulted in an improvement in the performance of both algorithms, but limit the impact of inertia, as it reduces the feasible range for the limit α. A more thorough study on the selection of these parameters is the object of a forthcoming article.

Finally, Figure 5 shows the evolution of the function values, the distance to the limit and the residuals, all in logarithmic scale, for case 14. As stated in Proposition 2, the sequence $k\,\|z_k - Tz_k\|^2$ tends to zero.

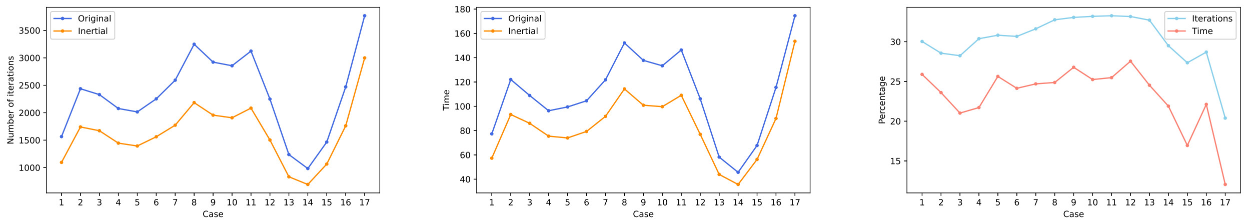
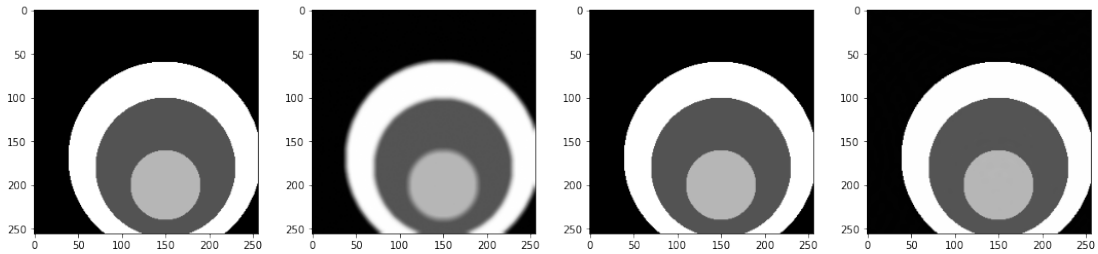


Figure 2: Number of iterations (left), execution time (center), and percentage of reduction (right), from Table 1.



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

Figure 3: Original, blurred and recovered images. Lowest recovered value $F^{TV}(x) = 0.1301$ (case 13, both methods).

		Original algorithm			Inertial algorithm			% Iterations	% Time
λ	α	Time	Iterations	$F^{TV}(x)$	Time	Iterations	$F^{TV}(x)$	reduction	reduction
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

Table 2: Execution time, number of iterations, final function value and reduction percentage for the original primal-dual algorithm and the inertial version (case 14), with tolerance $\varepsilon = 10^{-5}$.

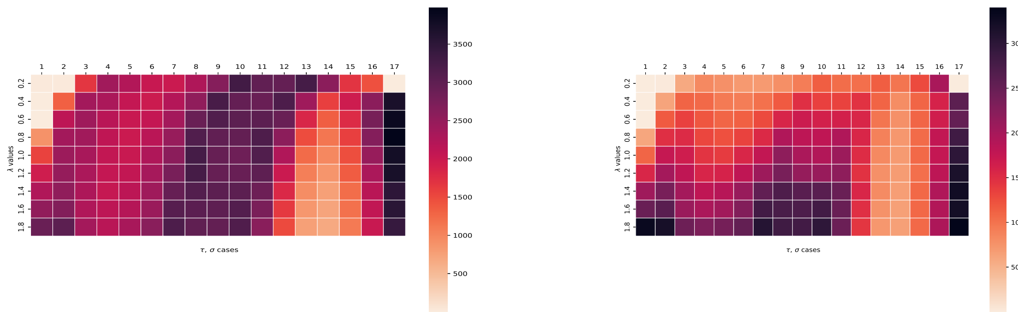


Figure 4: Average number of iterations performed by the original (left) and inertial (right) algorithms, with tolerance $\varepsilon = 10^{-5}$, for each value of λ, and each case of τ and σ, from Table 2.

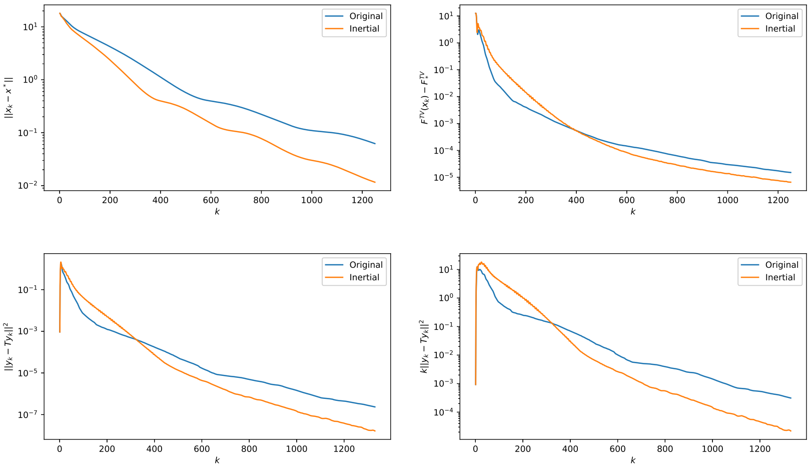


Figure 5: Evolution to the distance to the computed solution (top left), objective function values (top right), residuals $\|z_k - Tz_k\|^2$ (bottom left) and $k\|z_k - Tz_k\|^2$ (bottom right), for case 14.

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

$$\min_{Z\in\mathcal{H}} F(Z) := \frac{1}{2}\|\mathcal{A}(Z-Y)\|^2 + w\left\|Z_{(1)}\right\|_* + w\left\|Z_{(2)}\right\|_*, \tag{45}$$

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 (40), 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 (43), the iterations defined by (39) lead to Algorithm 2.

Algorithm 2:

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 Proposition 2 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;

As in the previous section, Algorithm 2 will be tested in the case $\alpha_k \equiv 0$ (the algorithm studied in [21]) and, for the inertial version,

$$\alpha_k = \left(1 - \frac{1}{k}\right)\alpha, \tag{46}$$

where α satisfies the condition (29). 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 10000 and 250000 pixels are randomly erased from the image in Figure 10a to obtain the one in Figure 10b. 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 Figure 6. The reduction stands between 12% and 22% in most cases, and the improvement seems to increase with the number of erased pixels.

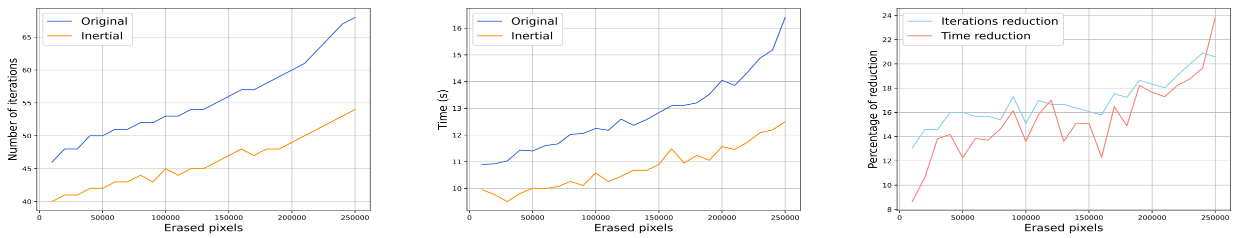


Figure 6: Number of iterations (left), execution time (center) and percentage of reduction (right) in terms of the number of erased pixels, with step size $\rho = 1$ and relaxation parameter $\lambda_k \equiv 1$, for a tolerance of 10^{-3}.

Comparison in terms of the step size. Both algorithms are tested for the same image with 250000 randomly erased pixels for $\lambda_k \equiv 1$ and different values of the step size ρ. For the inertial version, the constant α in (46) is adapted accordingly. The results are reported in Table 3 and depicted graphically in Figure 7. The percentage of reduction is noticeably higher 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.

	Original algorithm		Inertial algorithm	
ρ	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

Table 3: Execution time and number of iterations in terms of the step size ρ.

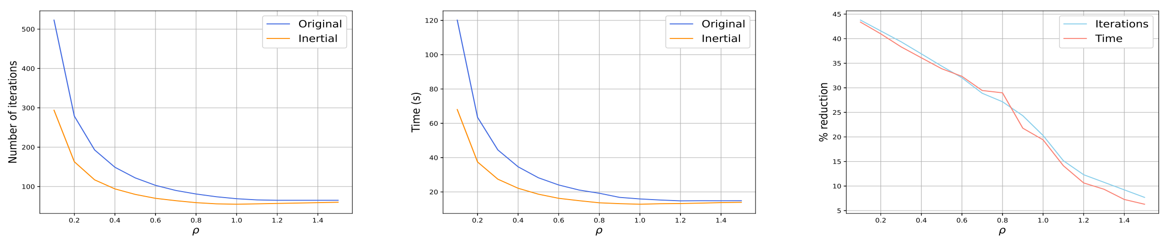


Figure 7: Number of iterations (left), execution time (center) and percentage of reduction (right) in terms of the step size ρ.

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 (29). The results are presented in Table 4, and shown graphically in Figure 8. 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 (29). 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.

	Original algorithm		Inertial algorithm	
λ	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

Table 4: Execution time and number of iterations for different values of λ.

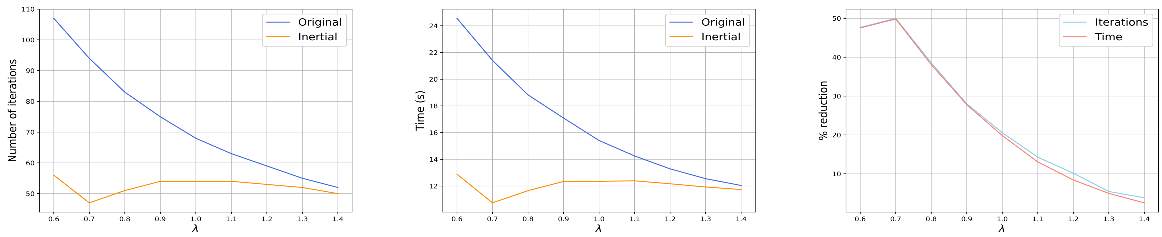


Figure 8: Number of iterations (left), execution time (center) and percentage of reduction (right) in terms of the relaxation parameter λ.

The evolution of the function values, the distance to the limit and the residuals are shown (in logarithmic scale) in Figure 9 for 250000 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, in agreement with Proposition 2. Finally, Figure 10 shows the original, corrupted (with 250000 erased pixels) and recovered images.[2]

References

[1] F. Alvarez. On the Minimizing Property of a Second Order Dissipative System in Hilbert Spaces. *SIAM Journal on Control and Optimization*, 38(4):1102–1119, 2000.

[2] F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Analysis*, 9(1):3–11, 2001.

[3] H. Attouch, L. M. Briceno-Arias, and P. L. Combettes. A parallel splitting method for coupled monotone inclusions. *SIAM Journal on Control and Optimization*, 48(5):3246–3270, 2010.

[4] H. Attouch and A. Cabot. Convergence of a relaxed inertial forward–backward algorithm for structured monotone inclusions. *Applied Mathematics & Optimization*, 80(3):547–598, 2019.

[5] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.*, 168(1-2, Ser. B):123–175, 2018.

[6] H. Attouch and J. Peypouquet. 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.

[2] For the sake of a fair visual comparison, we follow the implementation used in [21], as described in https://damek.github.io/ThreeOperators.html, which differs slightly from the description given in Section 4.5 in that it contains a Bregman update.

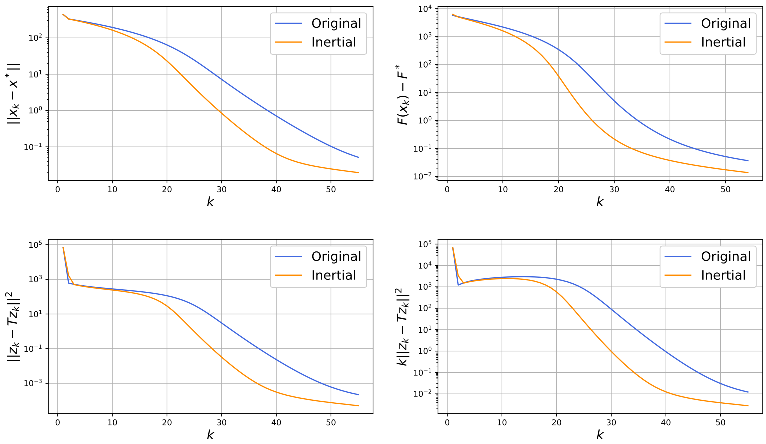


Figure 9: Evolution to the distance to the computed solution (top left), objective function values (top right), residuals $\|z_k - Tz_k\|^2$ (bottom left) and $k\,\|z_k - Tz_k\|^2$ (bottom right), for 250000 erased pixels using $\rho = 1$ and $\lambda_k \equiv 1$.

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

Figure 10: Original image (a), corrupted image with 250000 randomly erased pixels (b), images recovered without inertia (c), and with inertia (d).

[7] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham, second edition, 2017. With a foreword by Hédy Attouch.

[8] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[9] R. I. Boţ, E. R. Csetnek, and C. Hendrich. Inertial douglas–rachford splitting for monotone inclusion problems. *Applied Mathematics and Computation*, 256:472–487, 2015.

[10] H. Brézis and P. L. Lions. Produits infinis de résolvantes. *Israel Journal of Mathematics*, 29(4):329–345, 1978.

[11] L. M. Briceno-Arias. Forward-douglas–rachford splitting and forward-partial inverse method for solving monotone inclusions. *Optimization*, 64(5):1239–1261, 2015.

[12] L. Briceño-Arias and F. Roldán. Primal-dual splittings as fixed point iterations in the range of linear operators. *Journal of Global Optimization*, 2022.

[13] A. Cauchy et al. Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, 25(1847):536–538, 1847.

[14] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery*, 9(263-340):227, 2010.

[15] A. Chambolle and C. Dossal. On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". *J. Optim. Theory Appl.*, 166(3):968–982, 2015.

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

[17] P. L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5-6):475–504, 2004.

[18] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale modeling & simulation*, 4(4):1168–1200, 2005.

[19] L. Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of optimization theory and applications*, 158(2):460–479, 2013.

[20] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

[21] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set-valued and variational analysis*, 25(4):829–858, 2017.

[22] J. Douglas, Jr. and H. H. Rachford, Jr. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.

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

[24] A. A. Goldstein. Convex programming in hilbert space. *Bulletin of the American Mathematical Society*, 70(5):709–710, 1964.

[25] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM journal on control and optimization*, 29(2):403–419, 1991.

[26] P. C. Hansen, J. G. Nagy, and D. P. O'Leary. *Deblurring images*, volume 3 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006. Matrices, spectra, and filtering.

[27] O. S. Iyiola and Y. Shehu. New convergence results for inertial krasnoselskii–mann iterations in hilbert spaces with applications. *Results in Mathematics*, 76(2):1–25, 2021.

[28] M. A. Krasnosel'skii. Two comments on the method of successive approximations. *Usp. Math. Nauk*, 10:123–127, 1955.

[29] E. S. Levitin and B. T. Polyak. Constrained minimization methods. *USSR Computational mathematics and mathematical physics*, 6(5):1–50, 1966.

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

[31] D. A. Lorenz and T. Pock. An inertial forward-backward algorithm for monotone inclusions. *Journal of Mathematical Imaging and Vision*, 51(2):311–325, 2014.

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

[33] B. Martinet. Regularisation, d'inéquations variationelles par approximations succesives. *Revue Française d'informatique et de Recherche operationelle*, 1970.

[34] Y. Nesterov. A method for solving the convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. *Proceedings of the USSR Academy of Sciences*, 269:543–547, 1983.

[35] Y. Nesterov. *Lectures on convex optimization*, volume 137 of *Springer Optimization and Its Applications*. Springer, Cham, 2018. Second edition of [MR2142598].

[36] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bulletin of the American Mathematical Society*, 73(4):591–597, 1967.

[37] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in hilbert space. *Journal of Mathematical Analysis and Applications*, 72(2):383–390, 1979.

[38] J. Peypouquet. *Convex optimization in normed spaces*. SpringerBriefs in Optimization. Springer, Cham, 2015. Theory, methods and examples, With a foreword by Hedy Attouch.

[39] B. Polyak. Some methods of speeding up the convergence of iteration methods. *Ussr Computational Mathematics and Mathematical Physics*, 4:1–17, 12 1964.

[40] N. Pustelnik. *Méthodes proximales pour la résolution de problèmes inverses: application à la tomographie par émission de positrons*. PhD thesis, Université Paris-Est, 2010.

[41] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.

[42] Y. Shehu and A. Gibali. Inertial krasnoselskii–mann method in banach spaces. *Mathematics*, 8(4):638, 2020.

[43] B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38(3):667–681, 2013.