university of groningen
faculty of science and engineering
mathematics and applied mathematics

Perturbed Inertial Krasnoselskii-Mann Iterations

Bachelor's Project Mathematics

May 2023

Student: D. Cortild

First supervisor: Prof Dr J. G. Peypouquet

Second assessor: Prof Dr C. A. Bertoglio

Abstract

Fixed point problems are essential to a lot of work in engineering and applied sciences nowadays. In this paper, we look into perturbed inertial Krasnoselskii-Mann iterations, building on the known-to-converge Krasnoselskii-Mann iterations and their inertial or perturbed versions. We consider a general inertial scheme, which includes both the heavy-ball method by Polyak and the momentum approach by Nesterov. The perturbations are added on each step and show the stability of the provided algorithm. We first establish weak convergence in the quasi-nonexpansive case and strong convergence in the quasi-contractive setting. We then lay out generalisations and examples, from which the real interest in these iterations surfaces. Finally, we explore the link between Krasnoselkii-Mann iterations and the solutions to minimisation problems, namely by the three-operator splitting method. We then illustrate this splitting scheme through an application to the image inpainting problem.

Keywords Krasnoselskii-Mann iterations, Fixed point iterations, Nonexpansive operators, Inertial methods, Perturbed methods, Minimisation problems, Image inpainting

Mathematics Subject Classification (2020) 46N10, 47H09, 47H10, 47N10, 65K10

Contents

1 Introduction

The fixed point problem is a frequently encountered mathematical concept in various fields of engineering and applied sciences. In this problem, the objective is to find a point, called a *fixed point*, that does not move under the action of a self-mapping $T\colon \mathcal{H} \to \mathcal{H}$ on a real Hilbert space $\mathcal{H}$, namely a point $\hat{x} \in \mathcal{H}$ such that $T\hat{x} = \hat{x}$. Fixed point problems arise in various areas of mathematics, and iterative solution schemes are actively studied and utilised.

A first approach to the problem, established by Émile Picard in 1893 [28], consists of applying the operator recursively. This scheme is one of the simplest methods to approximate a fixed point and is called the *Picard iterative scheme*. It is given by

$$x_{k+1} = Tx_k,$$

for an initial guess $x_0 \in \mathcal{H}$. These iterations were proven to converge for contractive operators by Stefan Banach in 1922 [2].

Mark Krasnoselskii's work in 1955 showed that Picard iterations don't always converge for nonexpansive operators [16], which created a need for a new algorithm. The intuition behind this new algorithm is that the point x_k obtained from the previous iteration was likely to be a good approximation, so it could be beneficial to incorporate it into the next iteration. This led to the development of the now-known-as *Krasnoselskii-Mann iterations* [16, 21], which provide a weighted average between the Picard iterate and the previous point. The iteration can be expressed as

$$x_{k+1} = (1 - \lambda_k)x_k + \lambda_k Tx_k,$$

where (λ_k) are *relaxation* parameters. The proof for nonexpansive operators was established by Bernard Martinet in 1972 [22]. The importance and interest of studying Krasnoselskii-Mann iterations lie in how they generalise many known splitting algorithms and thus allow for a unified convergence analysis study. To cite a few, these iterations include the *Douglas-Rachford method* [11], the *forward-backward method* [33], the *primal-dual method* [6], the *proximal method* [30] and the *three-operator splitting method* [7].

Although the aforementioned iterations are known to converge theoretically, any small error can render the guarantee for convergence invalid. Such errors could arise due to rounding errors or imprecisions when evaluating the operator T at the point x_k. In some scenarios, such as for proximal operators, evaluating Tx_k exactly may not be possible, further enhancing the errors. Hence, a more realistic version of the Krasnoselskii-Mann iteration can be expressed as

$$x_{k+1} = (1 - \lambda_k)x_k + \lambda_k(Tx_k + e_k) + r_k,$$

where (e_k) represent the imprecisions in the evaluation of Tx_k, and (r_k) account for rounding errors.

To ensure convergence in the presence of these added errors, researchers have proposed studying a perturbed version of the Krasnoselskii-Mann iterations [15, 18, 35]. The *perturbed Krasnoselskii-Mann iteration* is given by

$$x_{k+1} = (1 - \lambda_k)x_k + \lambda_k Tx_k + \varepsilon_k,$$

where (ε_k) represent *perturbations*. By selecting $\varepsilon_k := \lambda_k e_k + r_k$, this perturbed version incorporates the previous imprecisions and rounding errors. It is also interesting to note that the perturbations do not need to account for errors.

The addition of momentum has proven to be a useful technique in optimisation. The concept was first introduced by Boris Teodorovich Polyak in 1964 [29], who showed that the heavy-ball method accelerates convergence in certain problems. Although originally proposed for gradient descent methods, it may be extended to Krasnoselskii-Mann iterations [9]. The *accelerated Krasnoselskii-Mann iteration* is given by

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

where (α_k) are *acceleration* parameters.

The concept was later generalised by Yurii Nesterov in 1983, also initially proposed on gradient methods for faster convergence [24]. Since then, many algorithms have been improved by the addition of this more popular acceleration step [13, 31]. It has also been proposed to extend the Krasnoselskii-Mann iterations with such an inertial step [10, 12, 32]. The *inertial Krasnoselskii-Mann iteration* is given by

$$\begin{cases} z_k &= x_k + \beta_k(x_k - x_{k-1}) \\ x_{k+1} &= (1-\lambda_k)z_k + \lambda_k T z_k, \end{cases}$$

where (β_k) are *acceleration* parameters.

The previous two acceleration schemes may be combined into a more general algorithm, incorporating both types of inertia [8]. The *general inertial Krasnoselskii-Mann iteration* is then given by

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

where (α_k) and (β_k) are *acceleration* parameters. One can observe that setting $\beta_k \equiv 0$ results in the heavy-ball method proposed by Boris Polyak, whereas $\alpha_k \equiv \beta_k$ results in the acceleration scheme proposed by Yurii Nesterov. As such, this scheme does indeed generalise both previously mentioned acceleration methods. By taking the final natural choice of parameters $\alpha_k \equiv 0$, we get another algorithm, referred to as the *reflected Krasnoselskii-Mann iterations* [8], inspired by the reflected gradient method [19, 20], obtained through a similar acceleration. The latter shall however not be covered here.

In this paper, we shall combine the addition of perturbations and the addition of inertia. We note that errors can occur in the Krasnoselskii-Mann step and both the acceleration steps of the algorithm, and hence there is a need to account for them in all phases. The *perturbed general inertial Krasnoselskii-Mann iteration* is then given by

$$\begin{cases} y_k &= x_k + \alpha_k(x_k - x_{k-1}) + \varepsilon_k \\ z_k &= x_k + \beta_k(x_k - x_{k-1}) + \rho_k \\ x_{k+1} &= (1-\lambda_k)y_k + \lambda_k T z_k + \theta_k, \end{cases}$$

where $(\varepsilon_k), (\rho_k)$ and (θ_k) are *perturbations*. For the rest of this paper, we shall drop the term"*general*" to make the name less cumbersome.

We shall moreover solve a more general problem than the earlier fixed point problem, namely we seek a shared fixed point of a family of operators $T_k\colon \mathcal{H} \to \mathcal{H}$. To this end, we use the same algorithm as before, with the slight change that we apply the operator T_k at iteration k instead of the constant operator T, inspired by the recent work of Ignacio Fierro, Juan José Maulén and Juan Peypouquet [12]. The algorithm that we study is thus given by

$$\begin{cases} y_k &= x_k + \alpha_k(x_k - x_{k-1}) + \varepsilon_k \\ z_k &= x_k + \beta_k(x_k - x_{k-1}) + \rho_k \\ x_{k+1} &= (1-\lambda_k)y_k + \lambda_k T_k z_k + \theta_k. \end{cases}$$

This algorithm is particularly useful for problems that involve multiple operators, as it allows to simultaneously converge to a common fixed point of all of them.

In addition to accounting for errors, perturbations in the algorithm may also be chosen artificially to solve a specific problem. For instance, if the operators T_k do not have a common fixed point, then we can shift all the operators towards new operators $\tilde{T}_k$ that share a fixed point, and incorporate the shift within the perturbations. This modified algorithm can then be used to prove convergence for a broader class of operators, including those that do not necessarily share a fixed point. This might be useful in scenarios where the operators T_k are successive approximations of a difficult-to-compute operator, such as a proximal or integral operator, where the approximate operators do not necessarily share a fixed point but the fixed point of interest is that of the limiting operator.

The paper is divided into five sections. In Section 2, we create a comprehensive list of the results used throughout this thesis. In Section 3, we lay out proofs of the weak and strong convergence of the aforementioned algorithm, and present some generalisations and examples. In Section 4, we link the fixed point problem to an optimisation problem and realise that the present algorithm includes a well-known splitting scheme. In Section 5, we present an example of the optimisation algorithm developed earlier, namely by applying it to the image inpainting problem. Visual results and plots shall be given to illustrate the matters. Finally, in Section 6, we conclude the paper and give some directions for further work, as well as a personal reflection on the work done.

2 Preliminaries

Some basic definitions and properties from functional and convex analysis, as well as linear algebra and subdifferential calculus, are given below. A reader well familiar with those subjects may skip to the next section, as every result presented shall be well-known and established. Although listed with the main ideas required for the proofs, the results shall not be proven, as they do not enter the main focus of the paper.

2.1 From Linear Algebra

We assume the basics of linear algebra and only mention a non-standard result by John von Neumann from 1937. An elementary proof, relying on doubly-stochastic matrices, was developed in 1973 by Leon Mirsky [23].

Theorem 2.1 (von Neumann Trace Inequality)**.** *For any two real matrices* $X, Y \in \mathbb{R}^{n\times n}$*,*

$$|tr(X^TY)| \leq \sum_{i=1}^{n} \sigma_i(X)\sigma_i(Y),$$

where $\sigma_i(X)$ *and* $\sigma_i(Y)$ *represent the ordered singular values of* X *and* Y *respectively.*

This result will prove itself useful when determining the proximal operator of the nuclear norm in a later section.

2.2 From Functional Analysis

We recall a few needed definitions and basic properties from functional analysis [27], as well as establish the notation used, and introduce some new results.

Throughout this thesis, we let $\mathcal{H}$ denote a real Hilbert space with associated inner product $\langle\cdot,\cdot\rangle$, and induced norm $\|\cdot\|$. We denote strong (or norm) convergence by $\to$ and weak convergence by $\rightharpoonup$, and define $\mathrm{Fix}(T) \coloneqq \{x \in \mathcal{H} \colon Tx = x\}$, the fixed points of T, $\mathrm{Zer}(T) \coloneqq \{x \in \mathcal{H} \colon Tx = 0\}$, the zeros of T, and $\mathrm{Graph}\ (T) \coloneqq \{(x, y) \colon x \in \mathcal{H}, y = Tx\}$, the graph of T.

We introduce the notion of *Kuratowski convergence* [5] in the context of a Hilbert space $\mathcal{H}$, which describes the limit of a sequence of sets. Namely, for a sequence (A_n) of subsets of $\mathcal{H}$, the *Kuratowski limit inferior* of (A_n) as $n \to \infty$ is defined as

$$\mathrm{Li}\ A_n \coloneqq \{x \in \mathcal{H} \mid \text{for all open neighbourhoods } U \text{ of } x,\, U \cap A_n \neq \emptyset \text{ for large enough } n\}.$$

Intuitively, we may regard $\mathrm{Li}\ A_n$ as the set of points where the sets A_n accumulate.

We will employ various adjectives to describe operators and their properties that characterise their behaviour in different contexts. An operator $T \colon \mathcal{H} \to \mathcal{H}$ is called

- *nonexpansive* if $\|Tx - Ty\| \leq \|x - y\|$ for all $x, y \in \mathcal{H}$,
- *quasi-nonexpansive* if $\mathrm{Fix}(T) \neq \emptyset$ and $\|Ty - p\| \leq \|y - p\|$ for all $y \in \mathcal{H}$ and $p \in \mathrm{Fix}(T)$,
- *q-contractive* for $q \in (0, 1)$ if $\|Tx - Ty\| \leq q\|x - y\|$ for all $x, y \in \mathcal{H}$,

- *q-quasi-contractive* for $q \in (0,1)$ if $\operatorname{Fix}(T) \neq \emptyset$ and $\|Ty - p\| \leq q\|y - p\|$ for all $y \in \mathcal{H}$ and $p \in \operatorname{Fix}(T)$,
- *firmly nonexpansive* if $\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y\rangle$ for all $x, y \in \mathcal{H}$,
- *τ-cocoercive* for $\tau > 0$ if τT is firmly nonexpansive, namely if $\tau\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y\rangle$ for all $x, y \in \mathcal{H}$,
- *γ-averaged* for $\gamma \in (0,1)$ if there exists a nonexpansive operator $R\colon \mathcal{H} \to \mathcal{H}$ such that $T = (1-\gamma)I + \gamma R$, and
- *demiclosed at $y \in \mathcal{H}$* if, for any sequence $(x_k) \subset \mathcal{H}$ with $x_k \rightharpoonup x$ and $Tx_k \to y$, it holds that $Tx = y$.

The previous definition may be extended to a family of operators (T_k), where $T_k\colon \mathcal{H} \to \mathcal{H}$. Indeed, (T_k) is called *asymptotically demiclosed at $y \in \mathcal{H}$* if, for every sequence $(x_k) \subset \mathcal{H}$ such that $x_k \rightharpoonup x$ and $T_k x_k \to y$, it follows that $T_k x = y$ for all k. Note that this definition is not standard.

Determining whether an operator is demiclosed may appear arduous at first. William Browder first introduced an important result that aids in determining this in 1966, and a simplified proof was presented by Zdzislaw Opial in 1967 [26, Lemma 2].

Theorem 2.2 (Browder's Demiclosedness Principle)**.** *Let $T\colon \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator. Then $I - T$ is demiclosed at any $y \in \mathcal{H}$.*

Lastly, we introduce a result that owes its name to Zdzislaw Opial. The proof is based on an argument by contradiction, and on simple identities in Hilbert spaces [26].

Lemma 2.3 (Opial's Lemma)**.** *Let $\mathcal{H}$ be a real Hilbert space, S a nonempty subset of $\mathcal{H}$, and (x_k) a sequence in $\mathcal{H}$. Assume*

- *the quantity $\lim_{k\to\infty} \|x_k - s\|$ exists for every $s \in S$ and*
- *every weak limit point of (x_k) lies in S.*

Then $x_k \rightharpoonup x$ for some point $x \in S$.

This result is convenient in proving the weak convergence of a sequence.

2.3 From Convex Analysis

We shall now introduce some notions from convex analysis [4, 25]. Through the rest of this paper, $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ denotes a set-valued operator, in comparison to $A\colon \mathcal{H} \to \mathcal{H}$ which denotes a value-valued operator. For a set-valued operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$, we denote the set of zeros of A by $\operatorname{Zer}(A) \coloneqq \{x \in \mathcal{H}\colon 0 \in Ax\}$ and the graph of A by $\operatorname{Graph}(A) \coloneqq \{(x,y) \in \mathcal{H}\times\mathcal{H}\colon y \in Ax\}$, similarly to the notation previously introduced. Moreover, an operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ is called *firmly nonexpansive* if $\|u - v\|^2 \leq \langle x - y, u - v\rangle$ for all $(x,u), (y,v) \in \operatorname{Graph}(A)$.

We extend the concept of monotonicity from scalar functions to operators. An operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ is called *monotone* if $\langle x - y, u - v\rangle \geq 0$ for all $(x,u), (y,v) \in \operatorname{Graph}(A)$. Moreover,

A is called *maximally monotone* if there does not exist another monotone operator $B\colon \mathcal{H} \rightrightarrows \mathcal{H}$ such that Graph $(A) \subsetneq$ Graph (B).

We introduce the following theorem by Jean-Bernard Baillon and Georges Haddad from 1977, which allows us to link Lipschitz continuity of a gradient with its cocoercivity. The proof relies on different basic inequalities on convex functions, manipulated to obtain the desired result [1, Corollary 10].

Theorem 2.4 (Baillon-Haddad Theorem)**.** *Let $f\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable, and let $\beta > 0$. Then f has $1/\tau$-Lipschitz continuous gradient if and only if ∇f is τ-cocoercive.*

Finally, for a given operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$, we define its *resolvent* as

$$J_A\colon \mathcal{H} \rightrightarrows \mathcal{H}, \quad x \mapsto (I + A)^{-1}(x).$$

It should be noted that the resolvent operator, denoted by J_A, is firmly nonexpansive and single-valued when A is maximally monotone.

2.4 From Subdifferential Calculus

Finally, we introduce some notions of subdifferential calculus [4, 27, 34], which deals with the study of subdifferentials, which are generalisations of derivatives to convex functions that are not necessarily differentiable. We also recall the notion of proximal operators and their relationship with subdifferentials.

For a proper, convex, and lower-semicontinuous function $f\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ we define its *subdifferential* at a point $x \in \mathcal{H}$ as

$$\partial f(x) \coloneqq \{u \in \mathcal{H}\colon f(x) - f(y) \leq \langle u, x - y\rangle \ \forall y \in \mathcal{H}\}.$$

Subdifferentials play a crucial role in determining the minimisers of functions. In fact, for $f\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ a proper, convex, and lower-semicontinuous function, it holds that $\hat{x}$ is a minimum of f if, and only if, $0 \in \partial f(\hat{x})$. This is known as Fermat's Rule. It is also worth recalling that if we furthermore require f to be strictly convex, then we can guarantee a unique minimiser.

The first result about subdifferentials concerns the chain rule, which is similar to the chain rule of multivariate functions. A proof presented by Juan Peypouquet is an interesting application of the Hahn-Banach Separation Theorem [27, Proposition 3.28].

Theorem 2.5 (Subgradient Chain Rule)**.** *Let $f\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper, convex, and lower-semicontinuous, and let $L\colon \mathcal{H} \to \mathcal{H}$ be a bounded linear operator. It holds that*

$$\partial(f \circ L)(x) \supset L^* \circ \partial f \circ L(x)$$

for all $x \in \mathcal{H}$. Equality holds for all $x \in \mathcal{H}$ if there exists a $x_0 \in ran(L)$ such that f is continuous at x_0.

The following theorem by Jean-Jacques Moreau and Ralph Tyrrell Rockafellar is similar in its proof, although somewhat more technical [27, Theorem 3.30].

Theorem 2.6 (Moreau-Rockafeller Theorem)**.** *Let $f, g\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper, convex, and lower-semicontinuous functions. It holds that*

$$\partial(f+g)(x) \supset \partial f(x) + \partial g(x)$$

for all $x \in \mathcal{H}$. Equality holds for all $x \in \mathcal{H}$ if there exists a $x_0 \in dom(g)$ such that f is continuous at x_0.

Finally, for a proper, convex, and lower-semicontinuous function $f\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ we define its *proximal operator* at the point $x \in \mathcal{H}$ as the point

$$\mathrm{prox}_f(x) \coloneqq \mathrm{argmin}_{u \in \mathcal{H}} \left(\frac{1}{2}\|x-u\|^2 + f(u) \right).$$

Notice that the strongly convex nature of the objective functions assures a unique minimum.

We recall that the proximal operator is tightly linked with the resolvent. Indeed, it holds that for a proper, convex, and lower-semicontinuous function $f\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, ∂f is maximally monotone, and $J_{\partial f} \equiv \mathrm{prox}_f$ is single-valued and firmly nonexpansive.

3 Perturbed Inertial Krasnoselskii-Mann Iterations

Recall we aim to solve a problem of the following form: For a family of operators (T_k) on a real Hilbert space $\mathcal{H}$,

$$\text{find } \hat{x} \in \mathcal{H} \text{ such that } \hat{x} \in \bigcap_{k\geq 1} \operatorname{Fix}(T_k). \tag{1}$$

This section explores the convergence of *perturbed inertial Krasnoselskii-Mann iterations*. The algorithm in its most general form, for an initial guess $x_0, x_1 \in \mathcal{H}$, is given by, for $k \geq 1$,

$$\begin{cases} y_k &= x_k + \alpha_k(x_k - x_{k-1}) + \varepsilon_k \\ z_k &= x_k + \beta_k(x_k - x_{k-1}) + \rho_k \\ x_{k+1} &= (1-\lambda_k)y_k + \lambda_k T_k z_k + \theta_k, \end{cases} \tag{2}$$

where $(\alpha_k) \subset (0,1)$ and $(\beta_k) \subset [0,1)$ are sequences of *inertial parameters*, $(\lambda_k) \subset (0,1)$ is a sequence of *relaxation parameters*, $(\varepsilon_k), (\rho_k), (\theta_k) \subset \mathcal{H}$ are sequences of *perturbations* and (T_k) is a family of operators on the real Hilbert space $\mathcal{H}$.

To simplify the proof of convergence, we set $\theta_k \equiv 0$. As such, the algorithm is given by, for $k \geq 1$,

$$\begin{cases} y_k &= x_k + \alpha_k(x_k - x_{k-1}) + \varepsilon_k \\ z_k &= x_k + \beta_k(x_k - x_{k-1}) + \rho_k \\ x_{k+1} &= (1-\lambda_k)y_k + \lambda_k T_k z_k. \end{cases} \tag{3}$$

The equivalence of Algorithms (2) and (3) is proven later in the section, and our simplification assumption is thus justified.

In Section 3.1, we delve into some basic properties needed for the following proofs. In Section 3.2 we prove, in the quasi-nonexpansive case and under some additional restrictions, the weak convergence of (x_k), (y_k) and (z_k) to a same fixed point of all (T_k). In Section 3.3 we establish, in the quasi-contractive case and under some different assumptions, the strong convergence of (x_k) to the, under the newly imposed conditions, unique fixed point of all (T_k). In Section 3.4, we lay out generalisations of the convergence theorems. Finally, in Section 3.5, we present some examples of the previously obtained results.

3.1 Basic Properties

This part is dedicated to some fundamental properties which will be made use of extensively over the next sections. Their proofs are brief and depend largely on elementary principles.

The first property allows to bound the norm of a sum by the weighted sum of the norms. This is particularly useful when a significant amount is known about one term, but not so much about the other.

Property 3.1. *For any $x, y \in \mathcal{H}$ and $\gamma > 0$, it holds that*

$$(1-\gamma)\|x\|^2 + \left(1 - \frac{1}{\gamma}\right)\|y\|^2 \leq \|x+y\|^2 \leq (1+\gamma)\|x\|^2 + \left(1 + \frac{1}{\gamma}\right)\|y\|^2.$$

Proof. Fix $x, y \in \mathcal{H}$ and $\gamma > 0$. First, observe that the Cauchy-Schwarz Inequality and the Arithmetic-Geometric Inequality imply that

$$|\langle x, y\rangle| = \left|\left\langle \sqrt{2\gamma}x, \frac{1}{\sqrt{2\gamma}}y \right\rangle\right| \leq \|\sqrt{2\gamma}x\| \cdot \left\|\frac{1}{\sqrt{2\gamma}}y\right\| \leq \gamma\|x\|^2 + \frac{1}{\gamma}\|y\|^2.$$

The right-most inequality follows, because

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle \leq \|x\|^2 + \|y\|^2 + \gamma\|x^2\| + \frac{1}{\gamma}\|y^2\|.$$

The left-most inequality follows in a similar fashion, by expanding $-\|x+y\|^2$. □

We also introduce a property that permits the deduction of the summability of a sequence.

Property 3.2. *Let $a \in (0,1)$, $(d_k) \subset \mathbb{R}$ be nonnegative and summable, and $(\Omega_k) \subset \mathbb{R}$ be some sequence such that, for all $k \geq 1$,*

$$\Omega_{k+1} \leq a\Omega_k + d_k.$$

Then $\sum_{k=1}^{\infty}[\Omega_k]_+$ is convergent, where $[\cdot]_+ = \max(0, \cdot)$.

Proof. Taking $[\cdot]_+$ on both sides does not interfere with the given inequality since it is an increasing operation. Hence, for all $k \geq 1$,

$$\begin{aligned}
[\Omega_{k+1}]_+ &\leq [a\Omega_k + d_k]_+ \\
&= \begin{cases} a\Omega_k + d_k & \text{if } a\Omega_k \geq -d_k \\ 0 & \text{else} \end{cases} \\
&= \begin{cases} a\Omega_k + d_k & \text{if } \Omega_k \geq 0 \\ a\Omega_k + d_k & \text{if } 0 \geq a\Omega_k \geq -d_k \\ 0 & \text{else} \end{cases} \\
&\leq \begin{cases} a\Omega_k + d_k & \text{if } \Omega_k \geq 0 \\ d_k & \text{if } 0 \geq a\Omega_k \geq -d_k \\ d_k & \text{else} \end{cases} \\
&= a[\Omega_k]_+ d_k.
\end{aligned}$$

Iterating this argument yields

$$[\Omega_{k+1}]_+ \leq a[\Omega_k]_+ + d_k \leq \cdots \leq a^k[\Omega_1]_+ + \sum_{j=1}^{k} a^{k-j} d_j.$$

Summing over all values of $k \geq 1$ gives

$$\begin{aligned}\sum_{k=0}^{\infty}[\Omega_{k+1}]_{+} &\leq \frac{1}{1-a}[\Omega_1]_{+} + \sum_{k=0}^{\infty}\sum_{j=1}^{k} a^{k-j} d_j \\ &= \frac{1}{1-a}[\Omega_1]_{+} + \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} a^{i} d_j \\ &= \frac{1}{1-a}[\Omega_1]_{+} + \frac{1}{1-a}\sum_{j=1}^{\infty} d_j.\end{aligned}$$

Since $\sum_{j=1}^{\infty} d_j$ converges, we conclude that $\sum_{k=1}^{\infty}[\Omega_k]_{+}$ is convergent, as wanted. □

Finally, we introduce a property allowing us to determine the convergence of a given sequence.

Property 3.3. *Let $(\xi_k) \subset \mathbb{R}$ be a sequence such that $\xi_{k+1} - \xi_k \leq d_k$, where $(d_k) \subset \mathbb{R}$ is summable, and such that $\xi_k \geq \eta_k - \eta_{k-1}$, where $(\eta_k) \subset \mathbb{R}$ is nonnegative. Then (ξ_k) is convergent. Moreover, if (d_k) is nonnegative, the limit of (ξ_k) is nonnegative.*

Proof. We first define $\chi_k := \xi_k - \sum_{i=1}^{k-1} d_i$. By assumption, it holds that $\chi_{k+1} - \chi_k \leq 0$, so (χ_k) is nonincreasing. Additionally, we shall prove that (χ_k) is bounded by below by $-\Lambda$ where $\Lambda := \sum_{i=1}^{\infty} d_i < +\infty$. To this extent, suppose that there exists a k_0 such that $\chi_{k_0} < -\Lambda$. Then, for all $k \geq k_0$, it would hold that

$$\chi_{k_0} \geq \chi_k \geq \xi_k - \Lambda \geq \eta_k - \eta_{k-1} - \Lambda,$$

and hence we deduce that

$$\eta_k \leq \chi_{k_0} + \Lambda + \eta_{k-1}.$$

Iterating this will then yield

$$0 \leq \eta_k \leq (k - k_0)\left(\chi_{k_0} + \Lambda\right) + \eta_{k_0}.$$

The right-hand side is a linear function in k with a negative slope, which is impossible since it is positive for arbitrarily large values of k. Hence we conclude that (χ_k) must be bounded by below by $-\Lambda$, and its nonincreasingness thus guarantees its convergence. Since we know that (d_k) is summable, this also guarantees the convergence of (ξ_k).

If (d_k) is nonnegative, it follows from the summability of (d_k) that $\sum_{i=k}^{\infty} d_i \to 0$. Since $\chi_k \geq -\Lambda$, it holds that $\xi_k \geq -\sum_{i=k}^{\infty} d_i$, proving that the limit of (ξ_k) is nonnegative. □

With these properties in mind, we are ready to tackle the proof of weak convergence of the sequence produced by Algorithm (3).

3.2 Weak Convergence

To facilitate and make possible the proof of weak convergence, we shall gradually add some restrictions to the parameters. Additionally, to facilitate future notations, for all $p \in \mathcal{H}$ and $\gamma_1, \gamma_2 > 0$, we define, for $k \geq 1$,

$$\begin{cases} \nu_k & := & \lambda_k^{-1} - 1, \\ \Delta_k(p) & := & \|x_k - p\|^2 - \|x_{k-1} - p\|^2, \\ B_k(\gamma_1, \gamma_2) & := & (1+\gamma_1)\Big[(1-\lambda_k)\alpha_k(1+\alpha_k) + \lambda_k\beta_k(1+\beta_k)\Big] + (1-\gamma_2)\nu_k\alpha_k(1-\alpha_k), \\ A_k(\gamma_1, \gamma_2) & := & \left(1+\gamma_1^{-1}\right)\Big[(1-\lambda_k)\|\varepsilon_k\|^2 + \lambda_k\|\rho_k\|^2\Big] + \nu_k\left(\gamma_2^{-1} - 1\right)\|\varepsilon_k\|^2. \end{cases}$$

Intending to deduce an inductive relationship between Δ_{k+1} and Δ_k, we introduce the following lemma. The subsequent results will aim to simplify most terms not related to Δ_{k+1} or Δ_k, and eventually prove the convergence of $\|x_k - p\|$ for any $p \in \bigcap_{k\geq 1} \mathrm{Fix}(T_k)$, which then will lead to the weak convergence of (x_k).

Lemma 3.4. *Let $(\alpha_k) \subset (0,1)$ be nondecreasing, let $(\beta_k) \subset [0,1]$, let $(\lambda_k) \subset (0,1)$ be such that $0 < \lambda \leq \lambda_k$ for all $k \geq 1$, let $(\varepsilon_k), (\rho_k) \subset \mathcal{H}$, and let $T_k \colon \mathcal{H} \to \mathcal{H}$ be quasi-nonexpansive such that $F := \bigcap_{k\geq 1} Fix(T_k) \neq \emptyset$. Also let (x_k, y_k, z_k) be generated by Algorithm (3). Then, for all $p \in F$, $\gamma_1 > 0$, $\gamma_2 \in (0,1)$, and $k \geq 1$,*

$$\begin{aligned} \Delta_{k+1}(p) \leq & (1+\gamma_1)\Big[(1-\lambda_k)\alpha_k + \lambda_k\beta_k\Big]\Delta_k(p) + \gamma_1\|x_k - p\|^2 \\ & + B_k(\gamma_1,\gamma_2)\|x_k - x_{k-1}\|^2 - (1-\gamma_2)\nu_k(1-\alpha_k)\|x_{k+1} - x_k\|^2 \\ & - (1-\gamma_2)\nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 + A_k(\gamma_1,\gamma_2). \end{aligned}$$

Proof. Fix some $p \in F$. Using the definition of x_k and the quasi-nonexpansiveness of T_k, it follows that

$$\begin{aligned} \|x_{k+1} - p\|^2 &= \|(1-\lambda_k)y_k + \lambda_k T_k z_k - p\|^2 \\ &= \|y_k - p\|^2 + \lambda_k^2\|T_k z_k - y_k\|^2 + 2\lambda_k\langle y_k - p, T_k z_k - y_k\rangle \\ &= (1-\lambda_k)\|y_k - p\|^2 - \lambda_k(1-\lambda_k)\|T_k z_k - y_k\|^2 + \lambda_k\|Tz_k - p\|^2 \\ &\leq (1-\lambda_k)\|y_k - p\|^2 - \lambda_k(1-\lambda_k)\|T_k z_k - y_k\|^2 + \lambda_k\|z_k - p\|^2. \end{aligned} \tag{4}$$

The first term of the right-hand side of Equation (4) may be rewritten, using the definition of y_k and Property 3.1, as

$$\begin{aligned} \|y_k - p\|^2 &= \|x_k - p + \alpha_k(x_k - x_{k-1}) + \varepsilon_k\|^2 \\ &\leq (1+\gamma_1)\|x_k - p + \alpha_k(x_k - x_{k-1})\|^2 + \left(1+\gamma_1^{-1}\right)\|\varepsilon_k\|^2 \\ &= (1+\gamma_1)\left(\|x_k - p\|^2 + \alpha_k^2\|x_k - x_{k-1}\|^2 + 2\alpha_k\langle x_k - p, x_k - x_{k-1}\rangle\right) \\ &\quad + \left(1+\gamma_1^{-1}\right)\|\varepsilon_k\|^2 \\ &= (1+\gamma_1)\left((1+\alpha_k)\|x_k - p\|^2 + \alpha_k(1+\alpha_k)\|x_k - x_{k-1}\|^2 - \alpha_k\|x_{k-1} - p\|^2\right) \\ &\quad + \left(1+\gamma_1^{-1}\right)\|\varepsilon_k\|^2, \end{aligned} \tag{5}$$

for any $\gamma_1 > 0$. Analogously, the last term of Equation (4) may be bounded by

$$\begin{aligned}\|z_k - p\|^2 \leq (1+\gamma_1)\left((1+\beta_k)\|x_k - p\|^2 + \beta_k(1+\beta_k)\|x_k - x_{k-1}\|^2 - \beta_k\|x_{k-1} - p\|^2\right) \\ + \left(1+\gamma_1^{-1}\right)\|\rho_k\|^2.\end{aligned} \tag{6}$$

Using once again the definition of x_k and y_k, along with Property 3.1, we may rewrite the middle term of Equation (4) as

$$\begin{aligned}-\lambda_k^2\|T_k z_k - y_k\|^2 &= -\|x_{k+1} - y_k\|^2 \\ &= -\|x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) - \varepsilon_k\|^2 \\ &\leq -(1-\gamma_2)\|x_{k+1} - x_k - \alpha_k(x_k - x_{k-1})\|^2 + \left(\gamma_2^{-1} - 1\right)\|\varepsilon_k\|^2 \\ &= -(1-\gamma_2)\|x_{k+1} - x_k\|^2 - (1-\gamma_2)\alpha_k^2\|x_k - x_{k-1}\|^2 \\ &\quad + (1-\gamma_2)\alpha_k\langle x_{k+1} - x_k, x_k - x_{k-1}\rangle + \left(\gamma_2^{-1} - 1\right)\|\varepsilon_k\|^2 \\ &= -(1-\gamma_2)(1-\alpha_k)\|x_{k+1} - x_k\|^2 + (1-\gamma_2)\alpha_k(1-\alpha_k)\|x_k - x_{k-1}\|^2 \\ &\quad - (1-\gamma_2)\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 + \left(\gamma_2^{-1} - 1\right)\|\varepsilon_k\|^2,\end{aligned}$$

for all $\gamma_2 \in (0,1)$. We multiply this equation by ν_k to rewrite it as

$$\begin{aligned}-\lambda_k(1-\lambda_k)\|y_k - T_k z_k\|^2 \leq &-(1-\gamma_2)\nu_k(1-\alpha_k)\|x_{k+1} - x_k\|^2 \\ &+ (1-\gamma_2)\nu_k\alpha_k(1-\alpha_k)\|x_k - x_{k-1}\|^2 \\ &- (1-\gamma_2)\nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 \\ &+ \left(\gamma_2^{-1} - 1\right)\nu_k\|\varepsilon_k\|^2.\end{aligned} \tag{7}$$

Combining Equations (4), (5), (6) and (7), gives, for all $\gamma_1 > 0$, $\gamma_2 \in (0,1)$,

$$\begin{aligned}\Delta_{k+1}(p) \leq &(1+\gamma_1)\Big[(1-\lambda_k)\alpha_k + \lambda_k\beta_k\Big]\Delta_k(p) + \gamma_1\|x_k - p\|^2 \\ &+ B_k(\gamma_1,\gamma_2)\|x_k - x_{k-1}\|^2 - (1-\gamma_2)\nu_k(1-\alpha_k)\|x_{k+1} - x_k\|^2 \\ &- (1-\gamma_2)\nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 + A_k(\gamma_1,\gamma_2),\end{aligned}$$

which yields the wanted inequality. □

The term $\|x_k - p\|^2$ presents certain difficulties since it is not accompanied by a corresponding $-\|x_{k-1} - p\|^2$, but the following proposition allows us to bound it in terms of Δ_{k+1} and Δ_k, which is closer to what we are trying to achieve.

Proposition 3.5. *For all $p \in F$, $\xi > 0$, $\gamma \in (0, 3\xi/4]$ and $k \geq 1$, it holds that*

$$\gamma\|x_k - p\|^2 - \xi\|x_{k+1} - 2x_k + x_{k-1}\|^2 \leq \frac{\xi}{4}\Delta_{k+1}(p) - \xi\Delta_k(p).$$

Proof. Firstly we apply Property 3.1 twice with $\gamma = \frac{1}{2}$ to obtain

$$\begin{aligned}-\|x_{k+1} - 2x_k + x_{k-1}\|^2 &= -\|(x_{k+1} - p) - 2(x_k - p) + (x_{k-1} - p)\|^2 \\ &\leq -\frac{1}{2}\|(x_{k+1} - p) - 2(x_k - p)\|^2 + \|x_{k-1} - p\|^2 \\ &\leq \frac{1}{4}\|x_{k+1} - p\|^2 - 2\|x_k - p\|^2 + \|x_{k-1} - p\|^2.\end{aligned}$$

Adding the term $\gamma\|x_k - p\|^2$ then yields that

$$\gamma\|x_k - p\|^2 - \xi\|x_{k+1} - 2x_k + x_{k-1}\|^2 \leq \frac{\xi}{4}\|x_{k+1} - p\|^2 + (\gamma - 2\xi)\|x_k - p\|^2 + \xi\|x_{k-1} - p\|^2.$$

Substituting $\gamma \leq 3\xi/4$ generates the wanted result. ☐

We may now introduce the main ingredient for the proof of weak convergence.

Proposition 3.6. *Let $(\alpha_k) \subset (0,1)$ be nondecreasing, let $(\beta_k) \subset [0,1]$, let $(\lambda_k) \subset (0,1)$ be such that $0 < \lambda \leq \lambda_k$ for all $k \geq 1$, and let $(\varepsilon_k), (\rho_k) \subset \mathcal{H}$ be such that $\sum_{k=1}^{\infty} \|\varepsilon_k\|^2, \sum_{k=1}^{\infty} \|\rho_k\|^2 < \infty$. Let $T_k \colon \mathcal{H} \to \mathcal{H}$ be quasi-nonexpansive such that $F := \bigcap_{k\geq 1} Fix(T_k) \neq \emptyset$. Furthermore, assume that*

$$\limsup_{k\to\infty} (1 - \lambda_k)\alpha_k(1 + \alpha_k) + \lambda_k\beta_k(1 + \beta_k) + \nu_k\alpha_k(1 - \alpha_k) - \nu_{k-1}(1 - \alpha_{k-1}) < 0.$$

Let (x_k, y_k, z_k) be generated by Algorithm (3). Then $\sum_{k=1}^{\infty} \|x_k - x_{k-1}\|^2$ and $\sum_{k=1}^{\infty} \|T_k z_k - y_k\|^2$ are convergent. Moreover, $\|x_k - p\|$ is convergent for all $p \in F$.

Proof. Firstly, fix $p \in F$. For the rest of this proof, we define $\Delta_k := \Delta_k(p)$.

Notice that the assumption is never verified if $\limsup \lambda_k = 1$. As we are interested in asymptotic results, we may shift the sequence and suppose that $\sup \lambda_k < 1$, and thus define

$$\nu := \min(1, \inf \nu_k) > 0.$$

Additionally, $\alpha = \sup \alpha_k = 1$ also contradicts the assumption, and thus $\alpha < 1$.

We select $\gamma_1 > 0$ and $\gamma_2 \in (0,1)$ such that $(1 + \gamma_1)((1 - \lambda_k)\alpha_k + \lambda_k\beta_k) < 1$ for all $k \geq 1$, such that $\gamma_1 \leq 3(1 - \gamma_2)\nu\alpha_0/4$, and such that

$$\begin{aligned} \limsup_{k\to\infty} (1 + \gamma_1)\Big[(1 - \lambda_k)\alpha_k(1 + \alpha_k) + \lambda_k\beta_k(1 + \beta_k)\Big] & \\ + (1 - \gamma_2)\Big[\nu_k\alpha_k(1 - \alpha_k) - \nu_{k-1}(1 - \alpha_{k-1})\Big] &< 0. \end{aligned} \tag{8}$$

For the remainder of this proof, we shall define $A_k := A_k(\gamma_1, \gamma_2)$ and $B_k := B_k(\gamma_1, \gamma_2)$.

Inequality (8) means that there exists an $\varepsilon > 0$ and a $k_0 > 0$ such that, for all $k \geq k_0$,

$$\begin{aligned} B_k - (1 - \gamma_2)\nu_{k-1}(1 - \alpha_{k-1}) &= (1 + \gamma_1)\Big[(1 - \lambda_k)\alpha_k(1 + \alpha_k) + \lambda_k\beta_k(1 + \beta_k)\Big] \\ &\quad + (1 - \gamma_2)\nu_k\alpha_k(1 - \alpha_k) - (1 - \gamma_2)\nu_{k-1}(1 - \alpha_{k-1}) \\ &\leq -\varepsilon. \end{aligned}$$

Since we are interested in asymptotic results, shifting the sequence does not affect the outcome, and we can thus assume for simplicity that $k_0 = 1$.

Combining Lemma 3.4 with the above, and using that $\nu \leq \nu_k$ and $\alpha_0 \leq \alpha_k$, we get that

$$\begin{aligned}\Delta_{k+1} \leq &(1+\gamma_1)\Big[(1-\lambda_k)\alpha_k + \lambda_k\beta_k\Big]\Delta_k + \gamma_1\|x_k - p\|^2 \\ &+ B_k\|x_k - x_{k-1}\|^2 - (1-\gamma_2)\nu_k(1-\alpha_k)\|x_{k+1} - x_k\|^2 \\ &- (1-\gamma_2)\nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 + A_k \\ \leq &(1+\gamma_1)\Big[(1-\lambda_k)\alpha_k + \lambda_k\beta_k\Big]\Delta_k + \gamma_1\|x_k - p\|^2 \\ &+ (1-\gamma_2)\nu_{k-1}(1-\alpha_{k-1})\|x_k - x_{k-1}\|^2 - (1-\gamma_2)\nu_k(1-\alpha_k)\|x_{k+1} - x_k\|^2 \\ &- \varepsilon\|x_k - x_{k-1}\|^2 - (1-\gamma_2)\nu\alpha_0\|x_{k+1} - 2x_k + x_{k-1}\|^2 + A_k.\end{aligned}$$

By Proposition 3.5 with $\xi = (1-\gamma_2)\nu\alpha_0 \in (0,1)$ using that $\gamma_1 \in (0, 3\xi/4]$, by using that $(1+\gamma_1)((1-\lambda_k)\alpha_k + \lambda_k\beta_k) \leq 1$, and by setting $\delta_k \coloneqq (1-\gamma_2)\nu_{k-1}(1-\alpha_{k-1})\|x_k - x_{k-1}\|^2$, we obtain

$$(1-\xi/4)\,\Delta_{k+1} \leq (1-\xi)\Delta_k + \delta_k - \delta_{k+1} - \varepsilon\|x_k - x_{k-1}\|^2 + A_k. \tag{9}$$

We now define C_k as

$$C_k \coloneqq (1-\xi/4)\,\|x_k - p\|^2 - (1-\xi)\,\|x_{k-1} - p\|^2 + \delta_k.$$

Combining this with the previous gives us

$$\begin{aligned}C_{k+1} - C_k &\leq (1-\xi/4)\,\Delta_{k+1} - (1-\xi)\,\Delta_k + \delta_{k+1} - \delta_k \\ &\leq -\varepsilon\|x_k - x_{k-1}\|^2 + A_k.\end{aligned} \tag{10}$$

Notice that, by construction,

$$C_k \geq (1-\xi/4)\,\|x_k - p\|^2 - (1-\xi/4)\,\|x_{k-1} - p\|^2.$$

Applying Property 3.3 with $\xi_k = C_k$, $d_k = A_k$ and $\eta_k = (1-\xi/4)\,\|x_k - p\|^2 \geq 0$ shows that (C_k) is convergent.

By Inequality (10), we obtain that

$$\lim_{k\to\infty} C_k - C_1 = \sum_{k=1}^{\infty}[C_{k+1} - C_k] \leq -\varepsilon\sum_{k=1}^{\infty}\|x_k - x_{k-1}\|^2 + \sum_{k=1}^{\infty}A_k.$$

Since $\varepsilon > 0$, and since $\sum_{k=1}^{\infty} A_k < \infty$, we can conclude that $\sum_{k=1}^{\infty}\|x_k - x_{k-1}\|^2$ is convergent.

Additionally, by Property 3.1 with $\gamma = 1$, it holds that

$$\begin{aligned}\lambda_k^2\|T_k z_k - y_k\|^2 &= \|x_{k+1} - y_k\|^2 \\ &= \|x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}) - \varepsilon_k\|^2 \\ &\leq 2\|x_{k+1} - x_k - \alpha_k(x_k - x_{k-1})\|^2 + 2\|\varepsilon_k\|^2 \\ &\leq 4\|x_{k+1} - x_k\|^2 + 4\alpha_k^2\|x_k - x_{k-1}\|^2 + 2\|\varepsilon_k\|^2.\end{aligned}$$

As such, $\sum_{k=1}^{\infty}\|T_k z_k - y_k\|^2$ is convergent.

Finally, we prove that $\|x_k - p\|$ is convergent. Recall Equation (9), which implies that

$$\Delta_{k+1} \leq a\Delta_k + d_k,$$

where

$$a := \frac{4-4\xi}{4-\xi} \quad \text{and} \quad d_k := \frac{4}{4-\xi}\delta_k + \frac{4}{4-\xi}A_k.$$

It holds that $a \in (0,1)$ since $\xi \in (0,1)$. By definition, assumption, and the previous, we know that (d_k) is nonnegative and summable. Thus, by Property 3.2, $\sum_{k=1}^{\infty}[\Delta_k]_+$ is convergent. We define $\xi_k := \|x_k - p\|^2 - \sum_{j=1}^{k}[\Delta_j]_+$, which is bounded by below by $-\sum_{k=1}^{\infty}[\Delta_k]_+$, and is nonincreasing since

$$\xi_{k+1} - \xi_k = \|x_{k+1} - p\|^2 - \|x_k - p\|^2 - [\Delta_{k+1}]_+ = \Delta_{k+1} - [\Delta_{k+1}]_+ \leq 0.$$

It must thus be convergent, which implies that $\|x_k - p\|^2$ is convergent, as wanted. □

With all these results, we may finally conclude the weak convergence of (x_k), (y_k) and (z_k) to a point in F.

Theorem 3.7. *Let $(\alpha_k) \subset (0,1)$ be nondecreasing, let $(\beta_k) \subset [0,1]$, let $(\lambda_k) \subset (0,1)$ be such that $0 < \lambda \leq \lambda_k$ for all $k \geq 1$, and let $(\varepsilon_k), (\rho_k) \subset \mathcal{H}$ such that $\sum_{k=1}^{\infty}\|\varepsilon_k\|^2, \sum_{k=1}^{\infty}\|\rho_k\|^2 < \infty$. Let $T_k\colon \mathcal{H} \to \mathcal{H}$ be quasi-nonexpansive such that $(I - T_k)$ is asymptotically demiclosed and such that $F := \bigcap_{k\geq 1} Fix(T_k) \neq \emptyset$. Furthermore, assume that*

$$\limsup_{k\to\infty}(1-\lambda_k)\alpha_k(1+\alpha_k) + \lambda_k\beta_k(1+\beta_k) + \nu_k\alpha_k(1-\alpha_k) - \nu_{k-1}(1-\alpha_{k-1}) < 0.$$

Let (x_k, y_k, z_k) be generated by Algorithm (3). Then (x_k), (y_k) and (z_k) converge weakly to a same point in F as $k \to \infty$.

Proof. Firstly notice that $(x_{k+1} - x_k)$ converges strongly to 0 by Proposition 3.6. Hence, by the definition of y_k and z_k, it follows that (x_k), (y_k) and (z_k) share the same set of weak limit points, since $\varepsilon_k, \rho_k \to 0$. Moreover,

$$y_k - z_k = (\alpha_k - \beta_k)(x_k - x_{k-1}) + \varepsilon_k - \rho_k \to 0.$$

By Proposition 3.6 again, we know that $\lim_{k\to\infty}\|x_k - p\|$ exists for all $p \in F$. Additionally, Proposition 3.6 also shows that

$$(I - T_k)z_k = (y_k - T_k z_k) - (y_k - z_k) \to 0.$$

The asymptotic demiclosedness of $(I - T_k)$ thus implies that for any weak limit point z of (z_k), thus also for any weak limit point of (x_k), we have $z \in F$.

We now apply Opial's Lemma 2.3 to deduce that (x_k) must converge weakly to some value $x \in F$, and by earlier discussion, (y_k) and (z_k) must also converge weakly to this same $x \in F$. □

3.3 Strong Convergence

We now tackle the strong convergence of the sequence. We start by again defining, for all $\gamma_1, \gamma_2 > 0$ and for $k \geq 1$,

$$\begin{cases} \nu_k & := \quad \lambda_k^{-1} - 1, \\ B_k(\gamma_1, \gamma_2) & := \quad (1+\gamma_1)\Big[(1-\lambda_k)\alpha_k(1+\alpha_k) + \lambda_k q_k^2 \beta_k(1+\beta_k)\Big] + (1-\gamma_2)\nu_k\alpha_k(1-\alpha_k), \\ A_k(\gamma_1, \gamma_2) & := \quad (\gamma_1^{-1}+1)\Big[(1-\lambda_k)\|\varepsilon_k\|^2 + \lambda_k q_k^2 \|\rho_k\|^2\Big] + \nu_k(\gamma_2^{-1}-1)\|\varepsilon_k\|^2. \end{cases}$$

The following proves the strong convergence of the given sequence.

Theorem 3.8. *Let $(\alpha_k) \subset (0,1)$ be nondecreasing, let $(\beta_k) \subset [0,1]$, let $(\lambda_k) \subset (0,1)$ be such that $0 < \lambda \leq \lambda_k$ for all $k \geq 1$, and let $(\varepsilon_k), (\rho_k) \subset \mathcal{H}$ such that $\sum_{k=1}^\infty \|\varepsilon_k\|^2, \sum_{k=1}^\infty \|\rho_k\|^2 < \infty$. Let $T_k \colon \mathcal{H} \to \mathcal{H}$ be q_k-quasi-contractive with Fix$(T_k) = \{p^*\}$ and $q_k \leq q < 1$. Furthermore, assume that*

$$\limsup_{k\to\infty} (1-\lambda_k)\alpha_k(1+\alpha_k) + \lambda_k q_k^2\beta_k(1+\beta_k) + \nu_k\alpha_k(1-\alpha_k) - \Big((1-\lambda_k)\alpha_k + \lambda_k q_k^2 \beta_k\Big)\nu_{k-1}(1-\alpha_{k-1}) < 0$$

Let (x_k, y_k, z_k) be generated by Algorithm (3). Then (x_k) converges strongly to p^* as $k \to \infty$.

Proof. Firstly, define

$$R_k := (1-\lambda_k)\alpha_k + \lambda_k q_k^2 \beta_k, \quad Q_k := 1 - \lambda_k + \lambda_k q_k^2, \quad \text{and} \quad Q := 1 - \lambda + \lambda q^2,$$

and notice that it holds that $R_k \leq Q_k \leq Q$ for all $k \geq 1$. Select $\gamma_1 > 0$ and $\gamma_2 \in (0,1)$ such that $(1+\gamma_1)Q < 1$ and such that, for some $k_0 > 0$,

$$B_k(\gamma_1, \gamma_2) - (1+\gamma_1)(1-\gamma_2)R_k\nu_{k-1}(1-\alpha_{k-1}) \leq 0 \tag{11}$$

holds for $k \geq k_0$, which is possible by the limiting inequality assumption. Since we are interested in limiting results, we may suppose that $k_0 = 1$.

Observe that by the q_k-quasi-contractivity of T_k, we may write

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

We notice that Equations (5), (6) and (7) still hold, and that they yield

$$\begin{aligned} \|x_{k+1} - p^*\|^2 \leq\ & (1+\gamma_1)\Big[(1-\lambda_k)(1+\alpha_k) + \lambda_k q_k^2(1+\beta_k)\Big]\|x_k - p^*\|^2 \\ & - (1+\gamma_1)\Big[(1-\lambda_k)\alpha_k + \lambda_k q_k^2\beta_k\Big]\|x_{k-1} - p^*\|^2 \\ & + B_k(\gamma_1, \gamma_2)\|x_k - x_{k-1}\|^2 \\ & - (1-\gamma_2)\nu_k(1-\alpha_k)\|x_{k+1} - x_k\|^2 \\ & - (1-\gamma_2)\nu_k\alpha_k\|x_{k+1} - 2x_k + x_{k-1}\|^2 + A_k(\gamma_1, \gamma_2). \end{aligned}$$

Using the definitions of R_k and Q_k, along with Equation (11) and the fact that $\|x_{k+1} - 2x_k + x_{k-1}\|^2 \geq 0$, we rewrite the previous as

$$\begin{aligned}\|x_{k+1} - p^*\|^2 \leq &(1+\gamma_1)(R_k + Q_k)\|x_k - p^*\|^2 \\ &- (1+\gamma_1)R_k\|x_{k-1} - p^*\|^2 \\ &+ (1+\gamma_1)(1-\gamma_2)R_k\nu_{k-1}(1-\alpha_{k-1})\|x_k - x_{k-1}\|^2 \\ &- (1-\gamma_2)\nu_k(1-\alpha_k)\|x_{k+1} - x_k\|^2 \\ &+ A_k(\gamma_1, \gamma_2).\end{aligned}$$

Since $(1+\gamma_1)Q_{k-1} < 1$ and $R_k \leq Q$, this may be written as

$$C_{k+1} \leq (1+\gamma_1)R_k C_k + A_k(\gamma_1, \gamma_2), \tag{12}$$

where

$$C_k \coloneqq \|x_k - p^*\|^2 - (1+\gamma_1)Q_{k-1}\|x_{k-1} - p^*\| + (1-\gamma_2)\nu_{k-1}(1-\alpha_{k-1})\|x_k - x_{k-1}\|^2.$$

We notice that all coefficients present in A_k are bounded by above, thus implying the summability of $(A_k(\gamma_1, \gamma_2))$. In specific, Property 3.3 guarantees the convergence of (C_k) to some $C \geq 0$ by setting $\xi_k = C_k$, $d_k = A_k(\gamma_1, \gamma_2) \geq 0$, and $\eta_k = (1+\gamma_1)Q\|x_k - p^*\|^2$.

As such, for any $\varepsilon > 0$, there exists a $k_0 \geq 1$ such that, for all $k \geq k_0$, it holds that $C_k \in (C - \varepsilon, C + \varepsilon)$ and $A_k(\gamma_1, \gamma_2) < \varepsilon$. The latter follows since $(A_k(\gamma_1, \gamma_2))$ is nonnegative and summable. Hence Inequality (12) implies that

$$C - \varepsilon \leq (1+\gamma_1)Q(C + \varepsilon) + \varepsilon,$$

or equivalently that

$$\varepsilon \geq \frac{(1 - (1+\gamma_1)Q)C}{2 + (1+\gamma_1)Q}.$$

As such we see that $C = 0$, since if $C > 0$, the right-hand side would be strictly positive, contradicting that $\varepsilon > 0$ was arbitrary. As such, we can conclude that $C_k \to 0$.

Note that $\|x_k - x_{k-1}\|^2$ is always positive, and hence we must see that

$$\limsup_{k\to\infty} \|x_k - p^*\|^2 - (1+\gamma_1)Q_{k-1}\|x_{k-1} - p^*\|^2 \leq 0.$$

Hence, for all $\varepsilon > 0$, there is a $k_0 \geq 1$ such that, for all $k \geq k_0$, it holds that

$$\|x_k - p^*\|^2 \leq (1+\gamma_1)Q\|x_{k-1} - p^*\|^2 + \varepsilon.$$

Iterating this then yields

$$\|x_k - p^*\|^2 \leq ((1+\gamma_1)Q)^{k-k_0}\|x_{k_0} - p^*\|^2 + \frac{\varepsilon}{1 - (1+\gamma_1)Q},$$

from which it follows that (x_k) converges strongly to p^*, since $\varepsilon > 0$ is arbitrary. □

3.4 Generalisations

In this subsection, we present two generalisation directions of Theorem 3.7, in which we allow for a stronger algorithm and a broader range of operators.

Stability in Krasnoselskii-Mann Step

The convergence of Algorithm (3) proves that the inertial steps are stable under perturbations. We recall that in Algorithm (2) we have a third sequence of perturbations $(\theta_k) \subset \mathcal{H}$, which perturb the Krasnoselskii-Mann step of the algorithm. As mentioned earlier, the convergence of Algorithm (3) also implies the convergence of Algorithm (2). Indeed, by setting $\tilde{x}_k = x_k - \theta_{k-1}$, the latter is equivalent to

$$\begin{cases} y_k &= \tilde{x}_k + \alpha_k(\tilde{x}_k - \tilde{x}_{k-1}) + \theta_k + \alpha_k(\theta_k - \theta_{k-1}) + \varepsilon_k \\ z_k &= \tilde{x}_k + \beta_k(\tilde{x}_k - \tilde{x}_{k-1}) + \theta_k + \beta_k(\theta_k - \theta_{k-1}) + \rho_k \\ \tilde{x}_{k+1} &= (1-\lambda_k)y_k + \lambda_k T_k z_k. \end{cases}$$

As such, if $(\varepsilon_k), (\rho_k)$ and (θ_k) are square-summable, then $(\theta_k + \alpha_k(\theta_k - \theta_{k-1}) + \varepsilon_k)$ and $(\theta_k + \beta_k(\theta_k - \theta_{k-1}) + \rho_k)$ are square-summable, and hence Algorithm (2) is equivalent to Algorithm (3) with different perturbation parameters and initial guess, and thus does still converge. As such, Theorem 3.7 also proves that the Krasnoselskii-Mann step of the algorithm is stable under perturbations.

Operators Not Sharing a Fixed Point

We could also relax the condition of all (T_k) to share a common fixed point. Instead, we shall suppose that $\tilde{F} \coloneqq \mathrm{Li}(\mathrm{Fix}(T_k)) \neq \emptyset$. In specific, this means there exists a sequence $(p_k) \subset \mathcal{H}$ such that $p_k \in \mathrm{Fix}(T_k)$ such that $p_k \to p^*$. We define, for each $k \geq 1$,

$$\tilde{T}_k \colon \mathcal{H} \to \mathcal{H}, \quad x \mapsto T_k(x + p_k - p^*) - p_k + p^*. \tag{13}$$

We notice that $p^* \in \mathrm{Fix}(\tilde{T}_k)$ for all $k \geq 1$, such that $p^* \in F \coloneqq \bigcap_{k\geq 1} \mathrm{Fix}(\tilde{T}_k)$. We also realise that if $q \in F$, then $q \in \mathrm{Fix}(\tilde{T}_k)$ for all $k \geq 1$, and as such

$$q = \tilde{T}_k(q) = T_k(q + p_k - p^*) - p_k + p^* \implies q + p_k - p^* \in \mathrm{Fix}(T_k),$$

which, since $\|p_k - p^*\| \to 0$, implies that $q \in \tilde{F}$. Hence we conclude that $F \subset \tilde{F}$.

Moreover, we realise that $\tilde{T}_k$ is quasi-nonexpansive since T_k is. Indeed, $\mathrm{Fix}(\tilde{T}_k)$ is nonempty as it contains p^*, and for any fixed point $\tilde{p}$ of $\tilde{T}_k$, $\tilde{p} + p_k - p^*$ is a fixed point of T_k, and thus, for any $x \in \mathcal{H}$, we see that

$$\|\tilde{T}_k x - \tilde{p}\| = \|T_k(x + p_k - p^*) - (\tilde{p} + p_k - p^*)\| \leq \|x - \tilde{p}\|,$$

which proves that $\tilde{T}_k$ is quasi-nonexpansive.

We also realise that Algorithm (2) boils down to

$$\begin{cases} \tilde{y}_k &= \tilde{x}_k + \alpha_k(\tilde{x}_k - \tilde{x}_{k-1}) + [p_{k-1} - p_k + \alpha_k(p_{k-1} - p_{k-2}) + \varepsilon_k] \\ \tilde{z}_k &= \tilde{x}_k + \beta_k(\tilde{x}_k - \tilde{x}_{k-1}) + [p_{k-1} - p_k + \beta_k(p_{k-1} - p_{k-2}) + \rho_k] \\ \tilde{x}_{k+1} &= (1-\lambda_k)\tilde{y}_k + \lambda_k \tilde{T}_k \tilde{z}_k + \theta_k, \end{cases}$$

where $\tilde{x}_k = x_k - p_{k-1} + p^*$, $\tilde{y}_k = y_k - p_k + p^*$, and $\tilde{z}_k = z_k - p_k + p^*$. Applying Algorithm (2) with the operators (T_k), initial guesses $x_0, x_1 \in \mathcal{H}$, and perturbations (ε_k), (ρ_k) and (θ_k) is thus equivalent to applying it with the operators $(\tilde{T}_k)$, initial guesses $\tilde{x}_0, \tilde{x}_1, \in H$, and perturbations $(p_{k-1} - p_k + \alpha_k(p_{k-1} - p_{k-2}) + \varepsilon_k)$, $(p_{k-1} - p_k + \beta_k(p_{k-1} - p_{k-2}) + \rho_k)$ and (θ_k). Since (ε_k), (ρ_k) and (θ_k) are already assumed square-summable, the latter are square-summable when $\|p_k - p_{k-1}\|^2$ is summable, since (α_k) and (β_k) are bounded.

We note that it is not necessary to implement $\tilde{T}_k$ or any of the auxiliary variables $\tilde{x}_k$, $\tilde{y}_k$ or $\tilde{z}_k$, which were only used to prove the convergence. Moreover, the choice of p^* and the sequence (p_k) do not affect the result, since $\mathrm{Li}(\mathrm{Fix}(T_k))$ is independent of those quantities.

The Generalisations

Both discussions above may be combined, proving that Algorithm (2) converges for a broader range of operators. As such, they give rise to the following corollary of Theorem 3.7.

Corollary 3.9. *Let $(\alpha_k) \subset (0,1)$ be nondecreasing, let $(\beta_k) \subset [0,1]$, let $(\lambda_k) \subset (0,1)$ be such that $0 < \lambda \leq \lambda_k$ for all $k \geq 1$, and let $(\varepsilon_k), (\rho_k), (\theta_k) \subset \mathcal{H}$ such that $\sum_{k=1}^{\infty} \|\varepsilon_k\|^2, \sum_{k=1}^{\infty} \|\rho_k\|^2, \sum_{k=1}^{\infty} \|\theta_k\|^2 < \infty$. Let $T_k\colon \mathcal{H} \to \mathcal{H}$ be quasi-nonexpansive such that $F := Li((Fix(T_k))) \neq \emptyset$ such that there exists a sequence $(p_k) \subset \mathcal{H}$ with $p_k \in Fix(T_k)$, $(p_k) \to p^*$, $\sum_{k=1}^{\infty} \|p_k - p_{k-1}\|^2 < \infty$, and such that $(I - \tilde{T}_k)$ is asymptotically demiclosed, where $\tilde{T}_k$ is given by Equation (13). Furthermore, assume that*

$$\limsup_{k\to\infty} (1 - \lambda_k)\alpha_k(1 + \alpha_k) + \lambda_k\beta_k(1 + \beta_k) + \nu_k\alpha_k(1 - \alpha_k) - \nu_{k-1}(1 - \alpha_{k-1}) < 0.$$

Let (x_k, y_k, z_k) be generated by Algorithm (3). Then (x_k), (y_k) and (z_k) converge weakly to a same point in F as $k \to \infty$.

Of course, the same discussion may be applied to Theorem 3.8, and thus yield the following result.

Corollary 3.10. *Let $(\alpha_k) \subset (0,1)$ be nondecreasing, let $(\beta_k) \subset [0,1)$, let $(\lambda_k) \subset (0,1)$ be such that $0 < \lambda \leq \lambda_k$ for all $k \geq 1$, and let $(\varepsilon_k), (\rho_k), (\theta_k) \subset \mathcal{H}$ such that $\sum_{k=1}^{\infty} \|\varepsilon_k\|^2, \sum_{k=1}^{\infty} \|\rho_k\|^2, \sum_{k=1}^{\infty} \|\theta_k\|^2 < \infty$. Let $T_k\colon \mathcal{H} \to \mathcal{H}$ be q_k-quasi-contractive with $Fix(T_k) = \{p_k\}$, $q_k \leq q < 1$, and $p_k \to p^*$ with $\sum_{k=1}^{\infty} \|p_k - p_{k-1}\|^2 < \infty$. Furthermore, assume that*

$$\limsup_{k\to\infty} (1-\lambda_k)\alpha_k(1+\alpha_k) + \lambda_k q_k^2\beta_k(1+\beta_k) + \nu_k\alpha_k(1-\alpha_k) - \Big((1-\lambda_k)\alpha_k + \lambda_k q_k^2\beta_k\Big)\nu_{k-1}(1-\alpha_{k-1}) < 0$$

Let (x_k, y_k, z_k) be generated by Algorithm (3). Then (x_k) converges strongly to p^* as $k \to \infty$.

3.5 Examples

This final subsection is dedicated to examples of the results established previously. These are often more friendly to apply, as the complicated conditions are relaxed.

Averaged Operators

If we suppose that $T_k \equiv T$ is nonexpansive, then Browder's Demiclosedness Principle 2.2 assures that $(I - T) \equiv (I - T_k)$ is asymptotically demiclosed.

By letting $T_k = (1-\gamma_k)I + \gamma_k R$ be γ_k-averaged, where $\gamma > 0$ and R is nonexpansive, Algorithm (2) may be rewritten as

$$\begin{cases} \delta_k &= [(1-\lambda_k)\alpha_k + (1-\gamma_k)\lambda_k\beta_k]/(1-\gamma_k\lambda_k) \\ y_k &= x_k + \delta_k(x_k - x_{k-1}) \\ z_k &= x_k + \beta_k(x_k - x_{k-1}) + \rho_k \\ x_{k+1} &= (1-\gamma_k\lambda_k)y_k + \gamma_k\lambda_k R z_k + [(1-\lambda_k)\varepsilon_k + (1-\gamma_k)\lambda_k\rho_k + \theta_k]. \end{cases}$$

Hence applying Algorithm (2) with the γ_k-averaged operator T_k, relaxation parameters (λ_k), acceleration parameters (α_k) and (β_k), and perturbations (ε_k), (ρ_k) and (θ_k) is completely equivalent to applying it with the nonexpansive operator R, relaxation parameters $(\gamma_k\lambda_k) \subset (0,1)$, acceleration parameters (δ_k) and (β_k), and different perturbations. Additionally, $I - R$ is asymptotically demiclosed, making the check for asymptotic demiclosedness redundant. Since it is easily checked that $\mathrm{Fix}(R) = \mathrm{Fix}(T_k)$, the found solution will be valid. If moreover $\gamma_k\lambda_k \to \eta \in (0,1)$, $\beta_k \to \beta$ and $\delta_k \to \delta$, the inequality conditions becomes

$$(1-\eta)\delta(1+\delta) + \eta\beta(1+\beta) - (\eta^{-1} - 1)(\delta - 1)^2 < 0.$$

As such, if the inequality is satisfied, (δ_k) is nondecreasing, and the remaining conditions of Theorem 3.7 are satisfied, the convergence is guaranteed.

The interest in this lies in the fact that it is not necessary to implement the algorithm with the operator R implicitly. We also do not need to test for asymptotic demiclosedness in Theorem 3.7 anymore, which substantially simplifies the conditions. Moreover, since we only require $(\gamma_k\lambda_k) \subset (0,1)$, it might be interesting to overrelax and set $1 < \lambda_k < \gamma_k^{-1}$.

Uniformly Convergent Continuous Operators

Instead of supposing that all the operators are averaged with respect to the same operator, which is a strong condition, we might simply suppose that the operators are nonexpansive and converge uniformly to some operator T with $\mathrm{Fix}(T) \neq \emptyset$. For a somewhat representative solution to exist, we shall suppose that $\mathrm{Fix}(T_k) = \mathrm{Fix}(T) + \{d_k\}$, where $(d_k) \subset \mathcal{H}$ such that $d_k \to 0$ and such that $\sum_{k=1}^{\infty} \|d_k - d_{k-1}\|^2 < \infty$. We also assume that T is weakly sequentially continuous, to make the following observations possible.

It holds that T is nonexpansive. Indeed, we notice that

$$\|Tx - Ty\| \leq \|Tx - T_k x\| + \|T_k x - T_k y\| + \|T_k y - Ty\|,$$

for all $k \geq 1$, and $x, y \in \mathcal{H}$. The first and last terms go to zero due to the uniform convergence of (T_k), and the second term is smaller than $\|x - y\|$ by nonexpansiveness of T_k. As such, we conclude that

$$\|Tx - Ty\| \leq \|x - y\|,$$

for all $x, y \in H$, which means that T is nonexpansive.

We then observe that Li $\text{Fix}(T_k) = \text{Fix}(T)$. Indeed, fixing an $x \in \text{Fix}(T)$ means that $x + d_k \in \text{Fix}(T_k)$, and since $d_k \to 0$, $x \in \text{Li Fix}(T_k)$. On the other hand, if $x \in \text{Li Fix}(T_k)$, then there must exist a sequence $(\delta_k) \subset \mathcal{H}$ such that $\delta_k \to 0$ with $x + \delta_k \in \text{Fix}(T_k)$. As such, we can write $x + \delta_k = x_k + d_k$, where $x_k \in \text{Fix}(T)$, which is equivalent to $x_k = x - d_k + \delta_k$. As $d_k - \delta_k \to 0$, we obtain $x_k \to x$. Finally, using that $\text{Fix}(T)$ is closed since T is assumed weakly sequentially continuous, we get that $x \in \text{Fix}(T)$.

Now, since $\text{Fix}(T) \neq \emptyset$, we may select $p^* \in \text{Fix}(T)$, and define $p_k := p^* + d_k \in \text{Fix}(T_k)$ such that $p_k \to p^*$. Moreover, since $\sum_{k=1}^{\infty} \|d_k - d_{k-1}\|^2 < \infty$, we obtain that $\sum_{k=1}^{\infty} \|p_k - p_{k-1}\|^2 < \infty$. Defining the operators $\tilde{T}_k$ as in Equation (13) produces operators $\tilde{T}_k$ such that $\text{Fix}(\tilde{T}_k) = \text{Fix}(T)$ by construction. Moreover, it holds that $\tilde{T}_k \to T$ uniformly too, since, for all $x \in \mathcal{H}$,

$$
\begin{aligned}
\frac{\|\tilde{T}_k x - Tx\|}{\|x\|} &\leq \frac{\|T_k(x + d_k) - T(x + d_k)\| + \|T(x + d_k) - Tx\| + \|d_k\|}{\|x\|} \\
&\leq \|T_k - T\| \cdot \frac{\|x + d_k\|}{\|x\|} + \frac{2\|d_k\|}{\|x\|} \\
&\to 0,
\end{aligned}
$$

since T is nonexpansive, $T_k \to T$ uniformly and $d_k \to 0$.

Additionally, if we have a sequence $(x_n) \subset \mathcal{H}$ such that $x_n \rightharpoonup x$, we observe that

$$
|\langle y, \tilde{T}_k x_k - Tx \rangle| \leq |\langle y, \tilde{T}_k x_k - Tx_k \rangle| + |\langle y, Tx_k - Tx \rangle|
$$

for all $y \in \mathcal{H}$. The first term goes to 0 because (x_k) is bounded since weakly convergent, and $\|\tilde{T}_k - T\| \to 0$ by uniform convergence of $(\tilde{T}_k)$. The second term goes to zero since T is assumed weakly sequentially continuous. As such, $\tilde{T}_k x_k \rightharpoonup Tx$. If we also have $(I - \tilde{T}_k)x_k \to 0$, we firstly get that $\tilde{T}_k x_k \rightharpoonup x$, and hence by unicity of the weak limit, $Tx = x$, or equivalently that $x \in \text{Fix}(T) = \text{Fix}(\tilde{T}_k)$ for all $k \geq 1$. As such, the family of operators $(I - \tilde{T}_k)$ is asymptotically demiclosed.

Finally, by Corollary 3.9, under the right choice of parameters, the sequences generated by Algorithm (2) converge weakly to a point in Li $\text{Fix}(T_k) = \text{Fix}(T)$.

Although the condition $\text{Fix}(T_k) = \text{Fix}(T) + \{d_k\}$ seems rather restrictive, it is interesting to notice it is automatically verified if each operator has a unique fixed point.

4 Application to Optimisation

Optimisation problems occur in various scenarios and are often represented as

$$\min_{x \in C} f(x),$$

where $C \subset \mathcal{H}$ is a subset of the Hilbert space $\mathcal{H}$, called the constraint set, and $f \colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a function, called the objective function.

In our case, we will be interested in solving minimisation problems over the entire space, where the objective function can be written as the sum of three functions. Such a scheme is called a *three-operator splitting scheme* [7]. We shall start by rewriting the minimisation problem in the form of Problem (1) depending on various operators emerging from the problem. Next, we prove that these operators verify certain conditions, allowing us to prove the convergence of Algorithm (2) to a solution of the problem. Finally, we provide the algorithm in terms of the objective function.

4.1 To a Fixed Point Problem

We aim to solve the following unconstrained minimisation problem

$$\min_{x \in \mathcal{H}} f(x) + g(x) + h(Lx), \tag{14}$$

where $f, g \colon \mathcal{H} \to \mathbb{R}$ are proper, convex, and lower-semicontinuous, $h \colon \mathcal{H} \to \mathbb{R}$ has a $1/\tau$-Lipschitz-continuous gradient, and $L \colon \mathcal{H} \to \mathcal{H}$ is a bounded linear operator. Furthermore, we assume the objective function is strictly convex, such that the problem has a unique minimiser.

We first notice that by Fermat's Rule, the Subgradient Chain Rule 2.5 and the Moreau-Rockafeller Theorem 2.6, this is equivalent to finding $\hat{x} \in \mathcal{H}$ such that

$$0 \in \partial(f + g + h \circ L)(\hat{x}) = (\partial f + \partial g + \nabla(h \circ L))(\hat{x}).$$

We know that ∂f and ∂g are maximally monotone, and that $\nabla(h \circ L)$ is $\tau/\|L\|_{\text{op}}$-cocoercive by the Baillon-Hadded Theorem 2.4.

With this in mind, we reformulate the problem into the following more general problem: For $A, B \colon \mathcal{H} \rightrightarrows \mathcal{H}$ maximally monotone operators and $C \colon \mathcal{H} \to \mathcal{H}$ a cocoercive operator,

$$\text{find } \hat{x} \in \mathcal{H} \text{ such that } \hat{x} \in \operatorname{Zer}(A + B + C). \tag{15}$$

A little manipulation is still required to write this as a fixed point problem, which the next lemma should clear up [7, Lemma 2.2].

Lemma 4.1. *Let $A, B \colon \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone and $C \colon \mathcal{H} \to \mathcal{H}$ cocoercive. Define, for a fixed $\rho > 0$,*

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

Then $Zer(A + B + C) = J_{\rho B}(Fix(T))$.

Proof. Take any element $x \in \mathrm{Zer}(A+B+C)$. Then it follows that $-\rho Cx \in \rho Ax + \rho Bx$, and hence there exists elements $\alpha \in \rho Ax$ and $\beta \in \rho Bx$ such that $\alpha + \beta = -\rho Cx$. Define $y \coloneqq \beta + x$, to obtain that there exists a $y \in \mathcal{H}$ such that

$$y - x = \beta \in \rho Bx \quad \text{and} \quad x - y - \rho Cx = \alpha \in \rho Ax.$$

From this, it follows directly that

$$y \in (I + \rho B)(x) \quad \text{and} \quad 2x - y - \rho Cx \in (I + \rho A)(x).$$

As such, using that $J_{\rho A}$ and $J_{\rho B}$ are single-valued since A and B are maximally monotone,

$$x = J_{\rho B}(y) \quad \text{and} \quad x = J_{\rho A}(2x - y - \rho Cx).$$

Combining these two equations yields that

$$J_{\rho B}(y) = J_{\rho A}(2J_{\rho B} - I - \rho C \circ J_{\rho B})(y).$$

As a consequence, $y \in \mathrm{Fix}(T)$, proving that $x \in J_{\rho B}(\mathrm{Fix}(T))$, and hence $\mathrm{Zer}(A+B+C) \subset J_{\rho B}(\mathrm{Fix}(T))$.

We notice all the steps are equivalences, hence proving that $\mathrm{Zer}(A+B+C) = J_{\rho B}(\mathrm{Fix}(T))$. □

As such, Problem (15) may be rewritten as follows: For $A, B\colon \mathcal{H} \rightrightarrows \mathcal{H}$ maximally monotone operators and $C\colon \mathcal{H} \to \mathcal{H}$ a cocoercive operator, where T is defined as in Lemma 4.1,

$$\text{find } \hat{x} \in \mathcal{H} \text{ such that } \hat{x} \in \mathrm{Fix}(T).$$

Setting $T_k \equiv T$ would thus bring this problem into the form of Problem (1). It thus gives a chance to Algorithm (2) to converge, if certain conditions are verified. In that case the generated sequence (x_k) would converge to a point $\hat{x} \in \mathrm{Fix}(T)$, and $J_{\rho B}(\hat{x})$ would be a zero of $A + B + C$, and hence the unique solution to Problem (14).

4.2 Proof of Convergence

To prove that Algorithm (2) produces a convergent sequence under the right parameters with the definition of T given in Lemma 4.1, we will prove that T is averaged and apply the discussion of Section 3.5.

However, before proving that T is averaged, we introduce an alternative characterisation of averaged operators [3, Proposition 2.2].

Property 4.2. *Let $T\colon \mathcal{H} \to \mathcal{H}$ and let $\gamma \in (0,1)$. Then T is γ-averaged if and only if, for every $x, y \in \mathcal{H}$, it holds that*

$$\|Tx - Ty\|^2 \leq \|x - y\|^2 - \frac{1-\gamma}{\gamma}\|(I-T)x - (I-T)y\|^2.$$

Proof. Let $R = (1/\gamma)T + (1 - 1/\gamma)I$. Then T is γ-averaged if and only if R is nonexpansive. Note that, for some fixed $x, y \in \mathcal{H}$, it holds that

$$2\left\langle x - y, \frac{(T-I)x}{\gamma} - \frac{(T-I)y}{\gamma}\right\rangle = \|Rx - Ry\|^2 - \|x-y\|^2 - \frac{\|(I-T)x-(I-T)y\|^2}{\gamma^2}$$

and that

$$\frac{2}{\gamma}\langle x - y, (T-I)x - (T-I)y\rangle = \frac{\|Tx - Ty\|^2}{\gamma} - \frac{\|x-y\|^2}{\gamma} - \frac{\|(I-T)x-(I-T)y\|^2}{\gamma}.$$

Combining those yields

$$\|x-y\|^2 - \|Rx - Ry\|^2 = \frac{1}{\gamma}\|Tx - Ty\|^2 - \frac{1}{\gamma}\|x-y\|^2 - \frac{1}{\gamma}\left(1 - \frac{1}{\gamma}\right)\|(I-T)x-(I-T)y\|^2.$$

If R is nonexpansive, the left-hand side is nonnegative for every $x, y \in \mathcal{H}$, implying the wanted inequation. Similarly, if the original equation holds for every $x, y \in \mathcal{H}$, then the right-hand side is nonnegative for every $x, y \in \mathcal{H}$, proving that R is nonexpansive. □

A first step towards proving that the operator T is averaged lies in the following lemma [7, Lemma 3.3], which will help reach an inequality similar to the one obtained in the alternative characterisation above.

Lemma 4.3. *Let $T \coloneqq U + T_1 \circ V$ and $W \coloneqq I - (2U + V)$, where $U, T_1 \colon \mathcal{H} \to \mathcal{H}$ are firmly nonexpansive, and $V \colon \mathcal{H} \to \mathcal{H}$. Then, for all $x, y \in \mathcal{H}$, we have*

$$\|Tx - Ty\|^2 \leq \|x-y\|^2 - \|(I-T)x - (I-T)y\|^2 - 2\langle T_1 \circ Vx - T_1 \circ Vy, Wx - Wy\rangle.$$

Proof. Fix some $x, y \in \mathcal{H}$, and denote $\tilde{T} = T_1 \circ V$ such that $T = U + \tilde{T}$. Then it follows that, since U and T_1 are firmly nonexpansive,

$$\begin{aligned}
\|Tx - Ty\|^2 &= \|Ux - Uy\|^2 + \|\tilde{T}x - \tilde{T}y\|^2 + 2\langle Ux - Uy, \tilde{T}x - \tilde{T}y\rangle \\
&\leq \langle Ux - Uy, x - y\rangle + \langle \tilde{T}x - \tilde{T}y, Vx - Vy\rangle + 2\langle Ux - Uy, \tilde{T}x - \tilde{T}y\rangle \\
&= \langle Ux - Uy, x - y\rangle + \langle \tilde{T}x - \tilde{T}y, (I-W)x - (I-W)y\rangle \\
&= \langle Tx - Ty, x - y\rangle + \langle T_1 \circ Vx - T_1 \circ Vy, Wx - Wy\rangle.
\end{aligned}$$

Next, we notice that

$$2\langle Tx - Ty, x - y\rangle = \|x-y\|^2 + \|Tx - Ty\|^2 - \|(I-T)x - (I-T)y\|^2,$$

which, combined with the previous equation, yields the wanted result. □

Now we may show that T is averaged under certain conditions [7, Proposition 3.1]. Note that the choice of ρ in the following lemma corresponds to the *step size*.

Lemma 4.4. *Let $T_1, T_2 \colon \mathcal{H} \to \mathcal{H}$ be firmly nonexpansive, and let $C \colon \mathcal{H} \to \mathcal{H}$ be τ-cocoercive. Then, for any $\rho \in (0, 2\tau)$, the operator*

$$T \coloneqq I - T_2 + T_1 \circ (2T_2 - I - \rho C \circ T_2)$$

is $\gamma \coloneqq \frac{2\tau}{4\tau - \rho}$-averaged.

Proof. Let $U := I - T_2$, $V := 2T_2 - I - \rho C \circ T_2$, and $W := \rho C \circ T_2 = I - (2U + V)$. For any $\varepsilon > 0$ and $x, y \in \mathcal{H}$,

$$\begin{aligned}
-2\langle T_1 \circ Vx - T_1 \circ Vy, Wx - Wy\rangle \\
&= 2\langle (I - T - T_2)x - (I - T - T_2)y, Wx - Wy\rangle \\
&= 2\langle (I - T)x - (I - T)y, Wx - Wy\rangle - 2\langle T_2x - T_2y, Wx - Wy\rangle \\
&\le 2\langle (I - T)x - (I - T)y, Wx - Wy\rangle - 2\rho\tau\|C \circ T_2x - C \circ T_2y\|^2 \\
&\le \varepsilon\|(I - T)x - (I - T)y\|^2 + \frac{\|Wx - Wy\|^2}{\varepsilon} - 2\rho\tau\|C \circ T_2x - C \circ T_2y\|^2 \\
&= \varepsilon\|(I - T)x - (I - T)y\|^2 + \rho\left(\frac{\rho}{\varepsilon} - 2\tau\right)\|C \circ T_2x - \rho C \circ T_2y\|^2,
\end{aligned}$$

where the first inequality follows from the cocoercivity of C and the second from the Cauchy-Schwarz Inequality and the Arithmetic-Geometric Inequality. In specific, the inequality holds for $\varepsilon = \rho/2\tau > 0$, for which it reduces to

$$-2\langle T_1 \circ Vx - T_1 \circ Vy, Wx - Wy\rangle \le \frac{\rho}{2\tau}\|(I - T)x - (I - T)y\|^2.$$

On the other hand, since T_2 is assumed firmly nonexpansive, we know U is too. Note that $T = U + T_1 \circ V$, hence Lemma 4.3 applies, giving us

$$-2\langle T_1 \circ Vx - T_1 \circ Vy, Wx - Wy\rangle \ge \|Tx - Ty\|^2 - \|x - y\|^2 + \|(I - T)x - (I - T)y\|^2.$$

Combining the two previous inequalities yields

$$\|Tx - Ty\|^2 - \|x - y\|^2 + \|(I - T)x - (I - T)y\|^2 \le \frac{\rho}{2\tau}\|(I - T)x - (I - T)y\|^2,$$

which in turn implies that

$$\|Tx - Ty\|^2 \le \|x - y\|^2 + \frac{\rho - 2\tau}{2\tau}\|(I - T)x - (I - T)y\|^2.$$

Since $\frac{2\tau - \rho}{2\tau} = \frac{1-\gamma}{\gamma}$ for $\gamma = \frac{2\tau}{4\tau - \rho}$, T is γ-averaged by Property 4.2. □

Setting $T_1 = J_{\rho A}$ and $T_2 = J_{\rho B}$, which are firmly nonexpansive as they are the resolvents of maximally monotone operators, makes the definitions of T in Lemmas 4.1 and 4.4 coincide. As such, T is averaged with parameter $\frac{2\tau}{4\tau - \rho}$.

By the discussion in Section 3.5, Algorithm (2), under the right parameters, generates a weakly convergent sequence (x_k) to a fixed point $\hat{x}$ of T, such that $J_{\rho B}(\hat{x}) = \operatorname{prox}_{\rho g}(\hat{x})$ corresponds to a solution of Problem (14).

4.3 Back to Optimisation

Recall that we are solving a problem of the following form: For $f, g\colon \mathcal{H} \to \mathbb{R}$ proper, lower-semicontinuous, and convex functions, $h\colon \mathcal{H} \to \mathbb{R}$ a function with $1/\tau$-Lipschitz continuous gradient, and $L\colon \mathcal{H} \to \mathcal{H}$ a bounded linear operator such that $f + g + h \circ L$ is strictly convex, find

$$\min_{x\in\mathcal{H}} f(x) + g(x) + h(Lx).$$

In Section 4.1, the solution set was shown to be the singleton $J_{\rho\partial g}(\mathrm{Fix}(T)) = \mathrm{prox}_{\rho g}(\mathrm{Fix}(T))$, where T is defined as

$$T \coloneqq I - \mathrm{prox}_{\rho g} + \mathrm{prox}_{\rho f} \circ (2\mathrm{prox}_{\rho g} - I - \rho L^* \circ \nabla h \circ L \circ \mathrm{prox}_{\rho g}). \tag{16}$$

This operator T was proven averaged with parameter $\gamma \coloneqq \frac{2\tau}{4\tau - \rho\|L\|_{\mathrm{op}}}$ in Section 4.2. Hence Corollary 3.9 applies as per our discussion about averaged operators in Section 3.5, which guarantees the weak convergence of (x_k) under certain conditions on the parameters. We also simplify some other conditions. We namely require $(\alpha_k) \subset (0,1)$ and $(\beta_k) \subset [0,1]$ to be nondecreasing converging to $\alpha \in (0,1)$ and $\beta \in [0,1]$, $(\lambda_k) \subset (0, 1/\gamma)$ to be converging to $\lambda > 0$ such that $\lambda \leq \lambda_k$ for all $k \geq 1$, and $(\varepsilon_k), (\rho_k), (\theta_k) \subset \mathcal{H}$ such that $\sum_{k=1}^{\infty} \|\varepsilon_k\|^2, \sum_{k=1}^{\infty} \|\rho_k\|^2, \sum_{k=1}^{\infty} \|\theta_k\|^2 < \infty$. We also require that

$$(1-\gamma\lambda)\delta(1+\delta) + \gamma\lambda\beta(1+\beta) - \left((\gamma\lambda)^1 - 1\right)(\delta - 1)^2 < 0, \tag{17}$$

where

$$\delta \coloneqq [(1-\lambda)\alpha + (1-\gamma)\lambda\beta]/(1-\gamma\lambda),$$

and that $[(1-\lambda)\alpha_k + (1-\gamma)\lambda\beta_k]/(1-\gamma\lambda)$ is nondecreasing, although the latter is guaranteed by the nondecreasingness of (α_k) and (β_k).

The shared weak limit point of (x_k), (y_k) and (z_k) will be a fixed point $\hat{x}$ of T, such that $\mathrm{prox}_{\rho g}(\hat{x})$ solves Problem (14).

Breaking down the algorithm into smaller steps avoids the unnecessary overhead of computing $\mathrm{prox}_{\rho g}$ multiple times. The algorithm thus becomes the following.

Require: $x_0, x_1 \in \mathcal{H}$, $\rho \in (0, 2\tau/\|L\|_{\mathrm{op}})$, $(\alpha_k) \subset (0,1)$ and $(\beta_k) \subset [0,1]$ nondecreasing converging to $\alpha \in (0,1)$ and $\beta \in [0,1]$, $(\lambda_k) \subset (0,1/\gamma)$ converging to $\lambda \in (0,1/\gamma)$ such that Equation (17) holds, $\varepsilon > 0$, $(\varepsilon_k), (\rho_k), (\theta_k) \subset \mathcal{H}$ such that $\sum \|\varepsilon_k\|^2, \sum \|\rho_k\|^2, \sum \|\theta_k\|^2 < \infty$, and an error function $\mathcal{R}\colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}_+$.

$k \leftarrow 1$
while $\mathcal{R}(x_k, x_{k-1}) > \varepsilon$ **do**
  $y_k \leftarrow x_k + \alpha_k \cdot (x_k - x_{k-1}) + \varepsilon_k$
  $z_k \leftarrow x_k + \beta_k \cdot (x_k - x_{k-1}) + \rho_k$
  $x_k^g \leftarrow \mathrm{prox}_{\rho g}(z_k)$
  $x_k^T \leftarrow z_k - x_k^g + \mathrm{prox}_{\rho f}(2 \cdot x_k^g - z_k - \rho L^*(\nabla h(L(x_k^g)))$
  $x_{k+1} \leftarrow (1-\lambda_k) \cdot y_k + \lambda_k \cdot x_k^T + \theta_k$
  $k \leftarrow k + 1$
end while
return $\mathrm{prox}_{\rho g}(x_k)$

---

5 Image Inpainting

Given an image and a region of erased pixels in the image, the *image inpainting problem* consists of restoring these erased pixels in order not to stand out with respect to their surroundings, recreating a complete image looking realistic to the human eye. The region of erased pixels may be a contiguous area covering an unwanted part of the image, such as a scratch or fold mark, but might also be a set of randomly lost data in the image.

Figure 1: Original image (left), image with erased pixels (centre), and restored image (right). Process not obtained through the described algorithm.

5.1 Mathematical Model

In this subsection, we develop a mathematical model of the image inpainting problem described above.

Let us represent an image X of M by N pixels by a tensor in $\mathcal{H} := [0,1]^{M\times N\times 3}$, in which each of the three layers represent the red, green and blue colour channels. We note that $\mathcal{H}$ is a Hilbert space with the Frobenius inner product and norm naturally extended to 3-dimensional tensors.

Let Ω be an element of $\{0,1\}^{M\times N}$ such that $\Omega_{ij} = 0$ indicates that the pixel at position (i,j), on all colour channels, has been damaged. Denote by $\mathcal{A}$ the linear operator that maps an image to an image whose elements in Ω have been erased. More formally,

$$\mathcal{A}\colon \mathcal{H} \to \mathcal{H}, \quad X \mapsto \tilde{X}, \quad \text{where } \tilde{X}_{ijk} = \Omega_{ij} \cdot X_{ijk}.$$

One easily observes that $\mathcal{A}$ is a projection map since Ω_{ij} takes values in $\{0,1\}$, and thus, for any $X \in \mathcal{H}$,

$$\mathcal{A}\mathcal{A}X = \sum_{i,j,k} \Omega_{ij}^2 X_{ijk} = \sum_{i,j,k} \Omega_{ij} X_{ijk} = \mathcal{A}X.$$

It is also self-adjoint as, for any $X, Y \in \mathcal{H}$,

$$\langle \mathcal{A}X, Y \rangle = \sum_{i,j,k} \Omega_{ij} X_{ijk} Y_{ijk} = \langle X, \mathcal{A}Y \rangle.$$

As a final observation, we notice that $\mathcal{A}$ is bounded since

$$\|\mathcal{A}X\|^2 = \sum_{i,j,k} \Omega_{ij}^2 X_{ijk}^2 \le \sum_{i,j,k} X_{ijk}^2 = \|X\|^2.$$

The equality takes place for $X \in \operatorname{ran}(\mathcal{A})$, proving that $\|\mathcal{A}\|_{\mathrm{op}} = 1$.

Let X be an image and call $X_{\mathrm{corrupt}} = \mathcal{A}X$ the damaged image. The objective is to recover an image from X_{corrupt} that could be a "good approximation" to X. Here, we define a "good approximation" to be an image that mainly overlaps on the points where $\Omega_{ij} = 1$, and which looks smooth. If the damaged area is not too important, this should resemble the original image.

One way to describe the smoothness of an image involves low-rank approximations [17]. Indeed, smooth images typically have a lower rank, and images with more complicated, or irregular, details, such as noise, tend to have a higher rank. This is because complex details require more basis functions to be represented, whereas simple smooth patterns often rely linearly on each other.

It is thus natural to imagine that a good smoothened version of a non-smooth image can be found via its low-rank approximation. This may be visualised by the idea in Figure 2.

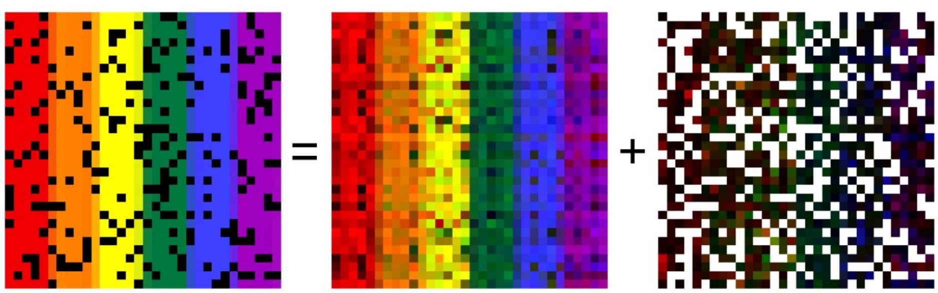

Figure 2: The original (noisy) image (left), its low-rank approximation (centre), and the remains (right).

Hence we could formulate the image inpainting problem as

$$\min_{X \in \mathcal{H}} \frac{1}{2}\|\mathcal{A}X - X_{\mathrm{corrupt}}\|^2 \quad \text{subject to } \operatorname{rank}(X) \le k,$$

for some pre-fixed integer value k.

This formulation however is cumbersome in that it requires to choose a k before the execution of the program, and hence to try multiple values of k. Additionally, the rank of a matrix is a discrete value, thus not leaving as much room for optimisation as a continuous variable would. Instead, we may replace the constraint by adding the nuclear norm of X to the objective function, which is defined by $\|X\|_* := \sum_i \sigma_i(X)$ and consists of a continuous value representing how linked the columns are [14]. A lower value represents a strong linkage, often also characterised by a low rank, although this is not necessarily the case. The nuclear norm not only measures the linear dependence of each row to each other as the rank does but also takes into consideration the difference between linearly dependent rows. Although slightly different, this quantity is still of interest and preserves the wanted properties.

Since the nuclear norm is only defined for matrices, and not for 3-dimensional tensors, we construct the two matrices

$$X_{(1)} \coloneqq [X_{\cdot\cdot 1}\ X_{\cdot\cdot 2}\ X_{\cdot\cdot 3}] \quad \text{and} \quad X_{(2)} \coloneqq \left[X_{\cdot\cdot 1}^T\ X_{\cdot\cdot 2}^T\ X_{\cdot\cdot 3}^T\right]^T.$$

We define the image inpainting problem, for $\sigma \in \mathbb{R}_{>0}$ a *regularisation parameter*, as

$$\min_{X \in \mathcal{H}} \frac{1}{2}\|\mathcal{A}X - X_{\text{corrupt}}\|^2 + \sigma\|X_{(1)}\|_* + \sigma\|X_{(2)}\|_*. \tag{18}$$

Notice this problem fits Problem (14), by writing $f(X) = \sigma\|X_{(1)}\|_*$, $g(X) = \sigma\|X_{(2)}\|_*$, $h(X) = \frac{1}{2}\|X - X_{\text{corrupt}}\|^2$, and $L = \mathcal{A}$. Since f and g are positive multiples of a norm, they are proper, lower-semicontinuous, and convex. Since h is the square of a norm, it is strongly continuous, and one can easily verify that $\nabla h(X) = X - X_{\text{corrupt}}$. Thus ∇h has a $1/\tau$-Lipschitz continuous gradient, for $\tau = 1$. Moreover, $\mathcal{A}$ is a linear bounded operator with operator norm 1, as discussed earlier. As such, the discussion in Section 4 applies and ensures the convergence of the presented algorithm.

5.2 Adapted Algorithm

The algorithm given at the end of Section 4.3 boils down to the following in the case of image inpainting. Using the self-adjointness of $\mathcal{A}$, that $\|\mathcal{A}\|_{\text{op}} = 1$, that $\mathcal{A}$ is a linear projection map, that $\mathcal{A}(X_{\text{corrupt}}) = X_{\text{corrupt}}$, that $\tau = 1$, and that $\nabla h(X) = X - X_{\text{corrupt}}$, help to simplify it substantially. We denote $\gamma \coloneqq 2/(4 - \rho)$.

Require: $X_0, X_1 \in \mathcal{H} = [0,1]^{M \times N \times 3}$, $\rho \in (0,2)$, $(\alpha_k) \subset (0,1)$ and $(\beta_k) \subset [0,1]$ nondecreasing converging to $\alpha \in (0,1)$ and $\beta \in [0,1]$, $(\lambda_k) \subset (0, 1/\gamma)$ converging to $\lambda \in (0, 1/\gamma)$ such that Equation (17) holds, $\varepsilon > 0$, $(\varepsilon_k), (\rho_k), (\theta_k) \subset \mathcal{H}$ such that $\sum \|\varepsilon_k\|^2, \sum \|\rho_k\|^2, \sum \|\theta_k\|^2 < \infty$, and an error function $\mathcal{R}\colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}_+$.

$k \leftarrow 1$

while $\mathcal{R}(X_k, X_{k-1}) > \varepsilon$ **do**

  $Y_k \leftarrow X_k + \alpha_k \cdot (X_k - X_{k-1}) + \varepsilon_k$

  $Z_k \leftarrow X_k + \beta_k \cdot (X_k - X_{k-1}) + \rho_k$

  $X_k^g \leftarrow \operatorname{prox}_{\rho g}(Z_k)$

  $X_k^T \leftarrow Z_k - X_k^g + \operatorname{prox}_{\rho f}\left(2 \cdot X_k^g - Z_k - \rho \cdot \left(\mathcal{A}(X_k^g) - X_{\text{corrupt}}\right)\right)$

  $X_{k+1} \leftarrow (1 - \lambda_k) \cdot Y_k + \lambda_k \cdot X_k^T + \theta_k$

  $k \leftarrow k + 1$

end while

return $\operatorname{prox}_{\rho g}(X_k)$

This algorithm makes use of the proximal operator of the nuclear norm, which should thus be computed. We first introduce another proximal operator, namely the one of the absolute value, also called the *Soft Threshold operator*.

Proposition 5.1. *Let $\lambda > 0$ and $f(x) = \lambda|x|$. Then it holds that*

$$prox_f(x) = sgn(x) \cdot \max(|x| - \lambda, 0).$$

Proof. We recall that

$$\text{prox}_f(x) = \text{argmin}_{y\in\mathbb{R}} \left(\frac{1}{2}(x-y)^2 + \lambda|y|\right).$$

We notice that the objective function is strongly convex, hence there is a unique minimiser $y^* \in \mathbb{R}$. If y^* is strictly positive, then it must be a root of the derivative of the objective function since it is differentiable on $\mathbb{R}_{>0}$. In specific, it must hold that

$$y^* = x - \lambda.$$

Since we assumed $y^* > 0$, we deduce that $\text{prox}_f(x) = x - \lambda$ for $x > \lambda$, and that the minimiser cannot be strictly positive for $x \leq \lambda$. Analogously, if $x < -\lambda$, we can conclude that $\text{prox}_f(x) = x + \lambda$, and that the minimiser cannot be strictly negative if $x \geq -\lambda$. Combining the two secondary remarks we conclude the minimiser must be 0 for $-\lambda \leq x \leq \lambda$. Hence we have that

$$\text{prox}_f(x) = \begin{cases} x - \lambda & \text{if } x > \lambda, \\ x + \lambda & \text{if } x < -\lambda, \\ 0 & \text{otherwise,} \end{cases}$$

which is equivalent to the Soft Threshold operator. □

Next, we introduce the proximal operator of the nuclear norm. We recall the nuclear norm of a matrix $X \in \mathbb{R}^{n\times n}$ is given by $\|X\|_* = \sum_{i=1}^n \sigma_i(X)$.

Proposition 5.2. *Let $\lambda \in \mathbb{R}_{>0}$ and $f(X) = \lambda\|X\|_*$. Then it holds that, for $X \in \mathbb{R}^{n\times n}$ with singular value decomposition $X = U\Sigma V^T$,*

$$prox_f(X) = U[\Sigma - \lambda I]_+ V^T,$$

where $[\cdot]_+ = \max(\cdot, 0)$ is evaluated component-wise.

Proof. Denote by $F(Y) = \frac{1}{2}\|X - Y\|^2 + \lambda\|Y\|_*$ and by X^* the minimiser of F, which exists and is unique since F is strongly convex.

Recall that the associated inner product to the Frobenius norm is given by $\langle A, B\rangle = \text{tr}(A^T B)$ for matrices $A, B \in \mathbb{R}^{n\times n}$, and that $\|A\|^2 = \text{tr}(A^T A) = \sum_{i=1}^n \sigma_i(A)^2$. We thus get that, for any $X, Y \in \mathbb{R}^{n\times n}$,

$$\frac{1}{2}\|X - Y\|^2 = \frac{1}{2}\|X\|^2 - \text{tr}(X^T Y) + \frac{1}{2}\|Y\|^2 = \frac{1}{2}\sum_{i=1}^n \sigma_i(X)^2 - \text{tr}(X^T Y) + \frac{1}{2}\sum_{i=1}^n \sigma_i(Y)^2.$$

Applying the von Neumann Trace Inequality 2.1 yields

$$\frac{1}{2}\|X - Y\|^2 \geq \frac{1}{2}\sum_{i=1}^n \sigma_i(X)^2 - \sum_{i=1}^n \sigma_i(X)\sigma_i(Y) + \frac{1}{2}\sum_{i=1}^n \sigma_i(Y)^2 = \frac{1}{2}\sum_{i=1}^n (\sigma_i(X) - \sigma_i(Y))^2.$$

As such, using the definition of the nuclear norm,

$$F(Y) \geq \frac{1}{2} \sum_{i=1}^{n} \left((\sigma_i(X) - \sigma_i(Y))^2 + \lambda \sigma_i(Y) \right).$$

Now it holds that

$$\begin{aligned}
F(X^*) &= \min_{Y \in \mathbb{R}^{n \times n}} F(Y) \\
&\geq \min_{Y \in \mathbb{R}^{n \times n}} \sum_{i=1}^{n} \left(\frac{1}{2} (\sigma_i(X) - \sigma_i(Y))^2 + \lambda \sigma_i(Y) \right) \\
&\geq \sum_{i=1}^{n} \min_{Y \in \mathbb{R}^{n \times n}} \left(\frac{1}{2} (\sigma_i(X) - \sigma_i(Y))^2 + \lambda \sigma_i(Y) \right) \\
&= \sum_{i=1}^{n} \min_{\sigma \in \mathbb{R}_{\geq 0}} \left(\frac{1}{2} (\sigma_i(X) - \sigma)^2 + \lambda \sigma \right) \\
&\geq \sum_{i=1}^{n} \min_{\sigma \in \mathbb{R}} \left(\frac{1}{2} |\sigma_i(X) - \sigma|^2 + \lambda |\sigma| \right).
\end{aligned}$$

We know that each minimisation problem is independent of the others and is solved by $\text{prox}_{\lambda|\cdot|}(\sigma_i(X))$, which by Proposition 5.1 is equal to $[\sigma_i(X) - \lambda]_+$ since $\sigma_i(X)$ is always positive. Hence the previous reduces to

$$F(X^*) \geq \sum_{i=1}^{n} \sigma_i(X) \cdot \min(\sigma_i(X), \lambda).$$

Now define $\tilde{X} = U[\Sigma - \lambda I]_+ V^T$. Since the right and left singular vectors of X and $\tilde{X}$ coincide by construction, it is easy to see that the inequality in the von Neumann Trace Inequality 2.1 results in an equality, and thus that

$$F(\tilde{X}) = \sum_{i=1}^{n} \left(\left(\sigma_i(X) - \sigma_i(\tilde{X}) \right)^2 + \lambda \sigma_i(\tilde{X}) \right).$$

Now using that $\sigma_i(\tilde{X}) = [\sigma_i(X) - \lambda]_+$, we deduce that

$$F(\tilde{X}) = \sum_{i=1}^{n} \sigma_i(X) \cdot \min(\sigma_i(X), \lambda).$$

Hence we conclude that it holds that $F(X^*) \geq F(\tilde{X})$, and since X^* is the unique minimiser of F, this implies that $X^* = \tilde{X}$, as wanted. □

Now we may deduce the proximal operators of the functions f and g provided above. We denote by ι_1 and ι_2 the inverse operators of $\cdot_{(1)}$ and $\cdot_{(2)}$, namely

$$\iota_1 \colon \mathbb{R}^{M \times 3N} \to \mathbb{R}^{M \times N \times 3}, \quad [A \ B \ C] \mapsto [A, B, C]$$

and

$$\iota_2 \colon \mathbb{R}^{3M\times N} \to \mathbb{R}^{M\times N\times 3}, \quad [A\ B\ C]^T \mapsto [A^T, B^T, C^T].$$

We observe that $\cdot_{(i)}$ and ι_i are linear norm-preserving bijections, and as such,

$$\begin{aligned}
\mathrm{prox}_{\rho f}(X) &= \operatorname*{argmin}_{Y\in\mathbb{R}^{M\times N\times 3}} \left(\frac{1}{2}\|X - Y\|^2 + \rho f(Y)\right) \\
&= \operatorname*{argmin}_{Y\in\mathbb{R}^{M\times N\times 3}} \left(\frac{1}{2}\|X_{(1)} - Y_{(1)}\|^2 + \rho\sigma\|Y_{(1)}\|_*\right) \\
&= \iota_1\left(\operatorname*{argmin}_{\tilde{Y}\in\mathbb{R}^{M\times 3N}} \left(\frac{1}{2}\|X_{(1)} - \tilde{Y}\|^2 + \rho\sigma\|\tilde{Y}\|_*\right)\right) \\
&= \iota_1\left(\mathrm{prox}_{\rho\sigma\|\cdot\|_*}(X_{(1)})\right).
\end{aligned}$$

Analogously, we deduce that

$$\mathrm{prox}_g(X) = \iota_2\left(\mathrm{prox}_{\rho\sigma\|\cdot\|_*}(X_{(2)})\right).$$

Everything required for implementing the algorithm has thus been provided. The next subsection will focus on the results of the given algorithm.

5.3 Numerical Results

In this part, we will study the algorithm given in the previous subsection. The image to be inpainted has dimensions 512×512. We shall pick a fixed tolerance of $\varepsilon = 10^{-3}$, and a fixed error function $\mathcal{R} \colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}_+$ defined by

$$\mathcal{R}(X_k, X_{k-1}) = \frac{\|X_k - X_{k-1}\|}{\|X_{k-1}\|}.$$

All corrupted images shall be so randomly, with a certain percentage of pixels erased. To ensure the termination of the algorithm, we enforce a maximal number of iterations of 100, after which we consider that the algorithm did not converge. We always set X_0 to be the zero matrix, and $X_1 = X_{\mathrm{corrupt}}$, such that the inertial effect is on a good trajectory at the start.

To simplify the algorithm, we set the relaxation parameters $\lambda_k \equiv \lambda \in (0, 1/\gamma)$ to be constant. We run multiple versions of the algorithm, listed below.

1. *Accelerated* version: The acceleration parameters are set to $\alpha_k \equiv (1-1/k)\alpha$ and $\beta_k = 0$, where $\alpha \in (0,1)$ is such that
$$(1-\gamma\lambda)\delta(1+\delta) - \left((\gamma\lambda)^{-1} - 1\right)(\delta - 1)^2 < 0$$
is tight, where $\delta \coloneqq (1-\lambda)\alpha/(1-\gamma\lambda)$. The perturbation parameters $\varepsilon_k, \rho_k, \theta_k \equiv 0$ are identically zero.

2. *Inertial* version: The acceleration parameters are set to $\alpha_k \equiv \beta_k \equiv (1-1/k)\alpha$, where $\alpha \in (0,1)$ is such that
$$\alpha(1+\alpha) - \left((\gamma\lambda)^{-1} - 1\right)(\alpha - 1)^2 < 0$$
is tight. The perturbation parameters $\varepsilon_k, \rho_k, \theta_k \equiv 0$ are identically zero.

To each version of the algorithm, we also introduce a *perturbed* version of the same algorithm, in which the perturbations are randomly distributed with mean $O(k^{-1})$.

All experiments are run in Python, on a single-core 2.3 GHz Intel i7-1068NG7. The code may be found on https://github.com/DanielCortild/Perturbed-Inertial-KM-Iterations.

Ratio of Erased Pixels

We first study the evolution of the number of iterations and the execution time required by our algorithms to inpaint a randomly corrupt image at a certain percentage in $(0, 1)$, whilst fixing the step size $\rho = 1$, the relaxation parameter $\lambda = 0.5$ and the regularisation parameter $\sigma = 1$. The results are shown in Figure 3.

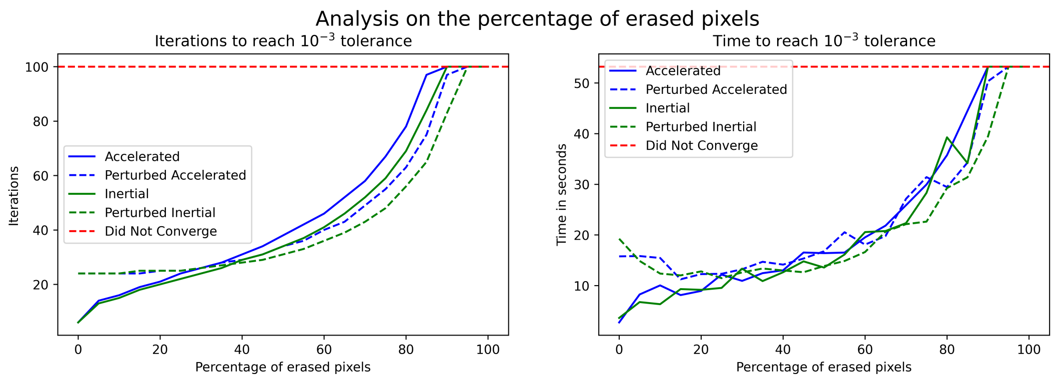

Figure 3: Variation of the ratio of erased pixels for $\rho = 1$, $\lambda = 0.5$, and $\sigma = 1$.

We observe that the larger the ratio of deleted pixels, the more iterations are required and the longer the algorithm runs before terminating, either by reaching the maximum number of iterations or by reaching the required tolerance. This was of course expected, as an image with nearly all pixels corrupted should be harder to recover than an image close to the original smooth version.

We also see that the accelerated version is slightly worse than the inertial version of the algorithm. Interestingly, we also notice that the algorithms do not seem to suffer from the addition of perturbations. The perturbed versions consistently outperform the non-perturbed versions by a few iterations. The experiments were repeated multiple times, and the same results were observed in all of them.

Step Size

Next, we fix the relaxation parameter $\lambda = 0.5$ and the regularisation parameter $\sigma = 1$, and randomly corrupt 50% of the pixels in the image. We iterate over representative values of the step size $\rho \in (0, 2)$. The results are shown in Figure 4.

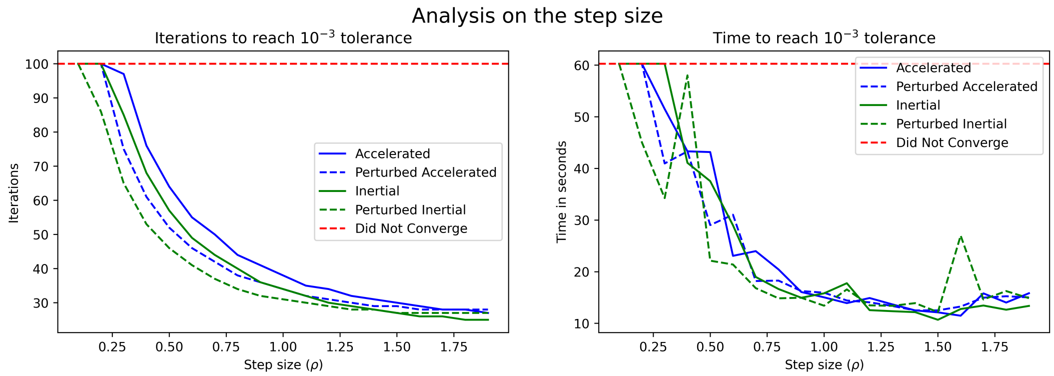

Figure 4: Variation of the step size ρ for $\lambda = 0.5$, $\sigma = 1$, and a ratio of erased pixels of 50%.

We observe that the number of iterations, along with the time spent to run the algorithm, decreases as the step size increases. As such, for both versions of the algorithms, it is beneficial to select a large step size. By selecting $\rho \approx 2$, we see that $\gamma \approx 1$, where γ is the averagedness parameter of the operator T, such that T becomes nonexpansive. Any value of ρ within the interval $(1.5, 2)$ converges in a relatively low number of iterations, and we select $\rho = 1.5$ as an ideal choice of step size, to allow for overrelaxation.

Once again, we observe that the addition of perturbations does not render slower the algorithm, and even speeds it up by a few iterations, and that the inertial version of the algorithm converges faster than the accelerated version.

Relaxation Parameter

Now we fix the step size $\rho = 1$ and the regularisation parameter $\sigma = 1$, and run the algorithms on some representative values of the relaxation parameter $\lambda \in (0, 1.5)$, on a fixed image with 50% of its pixels randomly erased. Notice that convergence is guaranteed for all such values of λ, due to the lower value of γ. The results are illustrated in Figure 5.

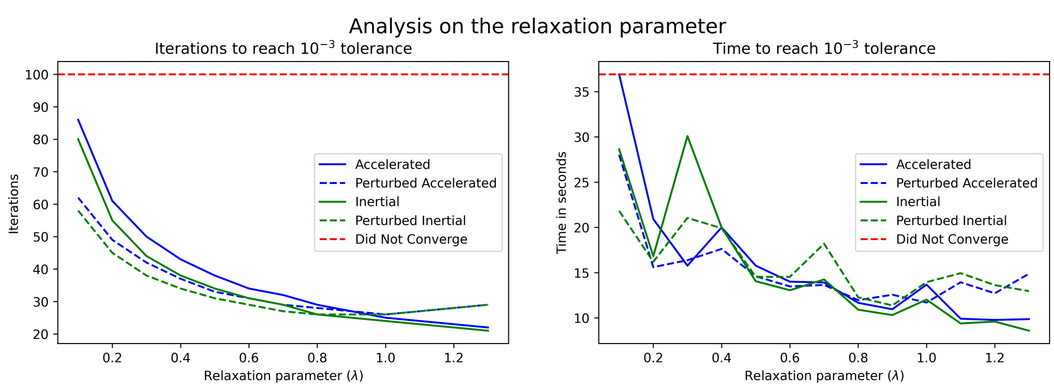

Figure 5: Variation of the relaxation parameter λ for $\rho = 1$, $\sigma = 1$, and a ratio of erased pixels of 50%.

As for the step size, we observe it to be beneficial to select a large relaxation parameter. We observe that selecting $\lambda \approx 1.3$ yields better results than keeping λ in a neighbourhood of 1.

As in the previous two experiments, we observe that the inertial version is slightly faster than the accelerated version and that the perturbations speed up the algorithm by a small factor.

Regularisation Parameter

Finally, we run a test on the regularisation parameter σ. Note that this parameter is not directly correlated with the convergence rate of the algorithm, but it is used in the definition of the problem. The purpose of the regularisation parameter is to indicate how much we prone smoothness, represented by $\|X_{(1)}\|_* + \|X_{(2)}\|_*$, over resemblance on the known pixels, represented by $\|\mathcal{A}X - X_{\text{corrupt}}\|^2$.

As such, we set the step size $\rho = 1$ and the relaxation parameter $\lambda = 0.5$, and run the algorithms on some representative values of the regularisation parameter $\sigma > 0$, on a fixed image with 50% of its pixels randomly erased. The final outputs are given in Figure 6.

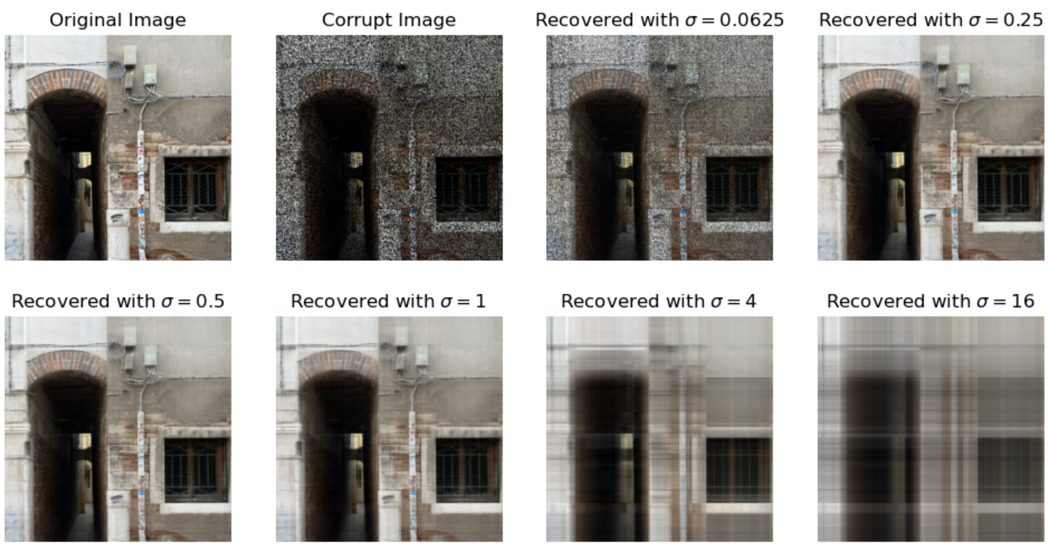

Figure 6: Results of the non-perturbed inertial version of the algorithm for different values of σ, for $\rho = 1$, $\lambda = 0.5$, and a ratio of erased pixels of 50%.

We observe that $\sigma = 0.0625$ and $\sigma \geq 1$ are unsuitable, as the first considers the smoothness too little and the latter does it too much. The results for both $\sigma = 0.25$ and $\sigma = 0.5$ seem acceptable, although the version recovered with $\sigma = 0.5$ does look better.

5.4 Visual Results

We may now group the results of the previous experiments. We shall still consider a randomly corrupted image with a fixed percentage of 50% of erased pixels. We select a regularisation parameter $\sigma = 1$, a step size $\rho = 1.8$, and a relaxation parameter $\lambda = 1.3$. We include the visual results of each version of the algorithm in Figure 7, and the convergence rates in Figure 8.

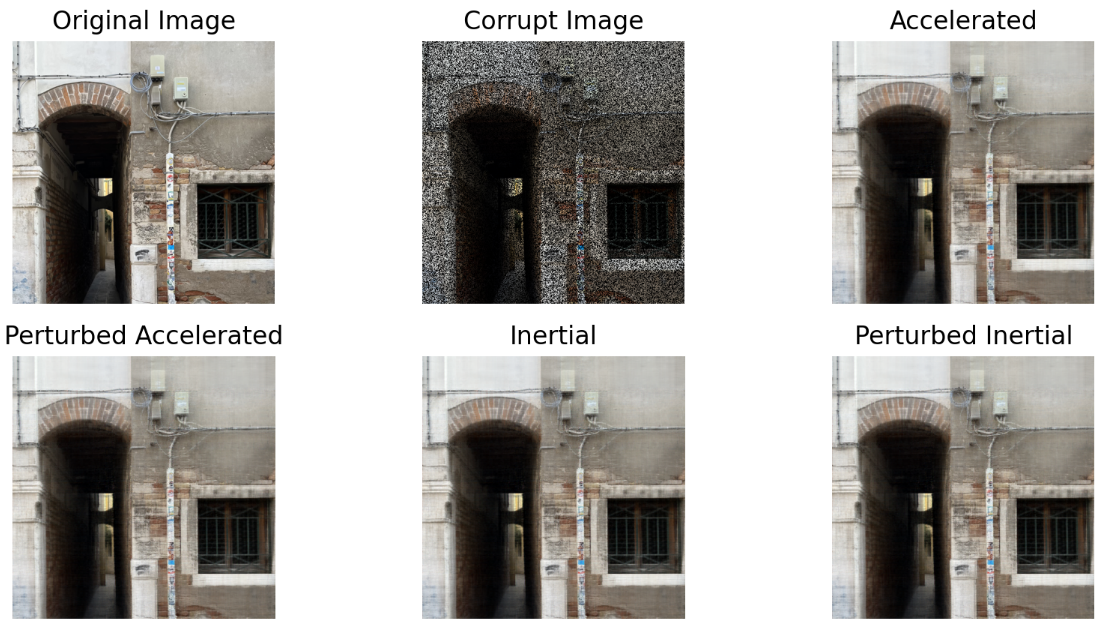

Figure 7: Inpainted image with $\lambda = 1.3$, $\rho = 1.8$, and $\sigma = 0.5$.

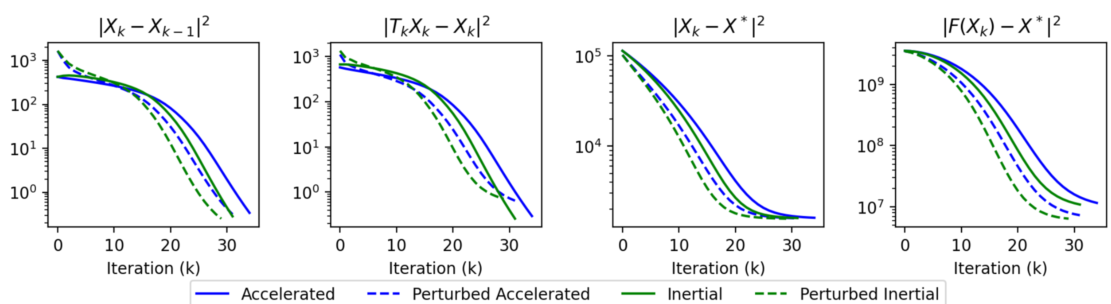

Figure 8: Convergence plots of the residuals of the iterations. F is the objective function to be minimised, and X^* and F^* represent the minimiser and minimal value of F.

We do observe, as earlier, that the perturbed versions of the algorithms outperform the non-perturbed versions slightly. We also notice that, as previously observed, the inertial version of the algorithm performs better than the accelerated version. However, all versions do converge to a smooth picture resembling the original image, as wanted.

6 Conclusion

This thesis focused on the convergence and practical applications of the *perturbed inertial Krasnoselskii-Mann iterations*. By introducing perturbations, we were able to establish the stability of both the inertial steps and the Krasnoselskii-Mann step. We also expanded the range of operators that this algorithm could be applied to, such as those not necessarily sharing a common fixed point.

Additionally, we drew a connection between fixed point problems and optimisation problems, enabling us to develop an algorithm for solving minimisation problems. This algorithm is a perturbed and inertial variant of the three-operator splitting method, and our convergence proof generalised a previously proposed approach by Damek Davis and Watao Yin [7].

Furthermore, we demonstrated the practical applicability of our method by presenting an example on the image inpainting problem. Through an extensive parameter study, we achieved visually pleasing results.

Overall, this thesis contributes to the advancement of numerical optimisation techniques and provides a framework for solving a broader range of problems in the future.

6.1 Further Research

The present paper does not represent a definitive study of perturbed inertial Krasnoselskii-Mann iterations. There remain several intriguing questions and conjectures that have yet to be explored. For instance, one could dive into the study of various rates of convergence, the convergence for a larger class of parameters, or a detailed analysis of Bregman updates, which will be elaborated on in the following passages.

Rates of Convergence

In Proposition 3.6, we proved that $\|x_k - x_{k-1}\|^2$ and $\|T_k y_k - y_k\|^2$ converge to 0. The exact rate of convergence was not determined and certainly depends on the rate of convergence of the perturbations. Following the results found in the paper by Juan José Maulén, Ignacio Fierro and Juan Peypouquet [12], I conjecture the convergence to be at least $\mathcal{O}(1/k)$, provided that $\sum k\|\varepsilon_k\|^2, \sum k\|\rho_k\|^2 < \infty$.

Moreover, in Theorem 3.8, we proved the strong convergence of (x_k) towards p^*. In the paper mentioned above, the convergence was also proven linear. This again cannot hold under the simple square-summability condition of (ε_k) and (ρ_k), but if the sequence of perturbations converges linearly to 0 the convergence might become linear.

Larger Class of Parameters

Theorems 3.7 and 3.8 prove convergence in the case $\alpha_0 > 0$. As such, this does not include the reflected algorithm, in which $\alpha_k \equiv 0$, or the static algorithm, in which $\alpha_k \equiv \beta_k \equiv 0$. For further research, the main theorems could be extended to include the case $\alpha_k \equiv 0$.

Bregman Updates

Instead of considering only Problem (14), we could consider the family of problems

$$\min_{x \in \mathcal{H}} f(x) + g(x) + h_k(Lx),$$

where $f, g \colon \mathcal{H} \to \mathbb{R}$ are proper, convex, and lower-semi-continuous, $h_k \colon \mathcal{H} \to \mathbb{R}$ have $1/\tau$-Lipschitz-continuous gradients and are strongly convex such that $\nabla h_k \to \nabla h$ uniformly for some continuously differentiable function $h \colon \mathcal{H} \to \mathbb{R}$, and $L \colon \mathcal{H} \to \mathcal{H}$ is a bounded linear operator.

The operator mentioned in Section 4.3 then becomes the following family of operators (T_k), defined by

$$T_k \coloneqq I - \operatorname{prox}_{\rho g} + \operatorname{prox}_{\rho f} \circ (2\operatorname{prox}_{\rho g} - I - \rho L^* \circ \nabla h_k \circ L \circ \operatorname{prox}_{\rho g}).$$

As such, we could find a fixed point of the limiting operator $T_k \to T$ by using the algorithm presented in Section 4.3, with non-constant h. Per Section 3.4, this will converge under certain conditions.

To speed up convergence or to produce clearer results, mostly in the case of image processing, the addition of a *Bregman update* might be beneficial. The idea is that, as above, the operator h is non-constant, and the approximate solution x_k is used to define the function h_{k+1} in a recursive manner. These updates need not occur at every iteration but could be applied based on different rulings, such as whenever a certain residual becomes small enough, or after a constant number of iterations.

Several conditions need to be verified by these updates to guarantee convergence. Most importantly, the updates need to be such that $\nabla h_k \to \nabla h$, where h_k now depends on x_{k-1}, which is no longer defined at the start. Experimentally, this does not always hold for naive implementations. As such, one could explore under what conditions these updates produce a converging sequence, and whether they are beneficial or not.

6.2 Personal Reflection

Reflecting on my thesis work, I am pleased with the overall outcome. Through proper time management and consistent effort, I was able to avoid the stress of last-minute rush and maintain steady progress throughout the entire project.

In hindsight, I realised that I could have saved myself some time by not delving too deeply into the proofs of the preliminary statements. Despite having removed them from the final version of my thesis, I found that this exercise allowed me to develop a deeper understanding of the underlying concepts and reasoning behind those results.

Overall, this experience taught me the value of patience and persistence in research, as well as the importance of balancing depth of understanding with efficient time management. I feel confident that the skills I have gained during this project will serve me well in future academic pursuits.

7 References

[1] J. B. Baillon and G. Haddad. Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones. *Israel Journal of Mathematics*, 26:137–150, June 1977.

[2] S. Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922.

[3] S. Bartz, M. N. Dao, and H. M. Phan. Conical averagedness and convergence analysis of fixed point algorithms. *Journal of Global Optimization*, 82:351–373, February 2022.

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

[5] G. Beer and J. Rodríguez-López. Topologies Associated with Kuratowski-Painlevé Convergence of Closed Sets. *Journal of Convex Analysis*, 17(3&4):805–826, January 2010.

[6] A. Chambolle and T. Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40:120–145, April 2011.

[7] D. Davis and W. Yin. A Three-Operator Splitting Scheme and its Optimization Applications. *Set-Valued and Variational Analysis*, 25:829–858, June 2017.

[8] Q.-L. Dong, Y. J. Cho, and T. M. Rassias. *Applications of Nonlinear Analysis*, chapter General Inertial Mann Algorithms and Their Convergence Analysis for Nonexpansive Mappings, pages 175–191. Springer International Publishing, Cham, June 2018.

[9] Q.-L. Dong and H. Yuan. Accelerated Mann and CQ algorithms for finding a fixed point of a nonexpansive mapping. *Fixed Point Theory and Applications*, 125, July 2015.

[10] Y. Dong. New inertial factors of the Krasnoselskii-Mann iteration. *Set-Valued and Variational Analysis*, 29:145–161, March 2021.

[11] J. Douglas and H. H. Rachford. On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables. *Transactions of the American Mathematical Society*, 82(2):421–439, December 1956.

[12] I. Fierro, J. J. Maulén, and J. Peypouquet. Inertial Krasnoselskii-Mann Iterations. October 2022.

[13] O. Güler. New Proximal Point Algorithms for Convex Minimization. *SIAM Journal on Optimization*, 2(4):649–664, January 1992.

[14] S. Huang and H. Wolkowicz. Low-rank matrix completion using nuclear norm minimization and facial reduction. *Journal of Global Optimization*, 72:5–26, September 2018.

[15] C. Kanzow and Y. Shehu. Generalized Krasnoselskii–Mann-type iterations for nonexpansive mappings in Hilbert spaces. *Computational Optimization and Applications*, 67:595–620, July 2017.

[16] M. A. Krasnosel'skii. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10(1(63)):123–127, 1955.

[17] J. Li, M. Li, and H. Fan. Image Inpainting Algorithm Based on Low-Rank Approximation and Texture Direction. *Mathematical Problems in Engineering*, 2014:1–11, December 2014.

[18] J. Liang, J. Fadili, and G. Peyré. Convergence rates with inexact non-expansive operators. *Mathematical Programming*, 159:403–434, September 2016.

[19] P. Maingé. Numerical approach to monotone variational inequalities by a one-step projected reflected gradient method with line-search procedure. *Computers & Mathematics with Applications*, 72(3):720–728, August 2016.

[20] P. Maingé and M. Gobinddass. Convergence of One-Step Projected Gradient Methods for Variational Inequalities. *Journal of Optimization Theory and Applications*, 171:146–168, July 2016.

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

[22] B. Martinet. *Algorithmes pour la résolution de problèmes d'optimisation et de minimax.* PhD thesis, Université Joseph-Fourier - Grenoble I, April 1972.

[23] L. Mirsky. A Trace Inequality of John yon Neumann. *Monatshefte für Mathematik*, 79:303–306, December 1973.

[24] 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(3):543–547, 1983.

[25] J. Nocedal and S. Wright. *Numerical Optimization.* Springer Series in Operations Research and Financial Engineering. Springer New York, NY, December 2006.

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

[27] J. Peypouquet. *Convex Optimization in Normed Spaces: Theory, Methods and Examples.* SpringerBriefs in Optimization. Springer Cham, March 2015.

[28] E. Picard. Sur l'application des méthodes d'approximations successives à l'étude de certaines équations différentielles ordinaires. *Journal de Mathématiques Pures et Appliquées*, 9:217–272, 1893.

[29] B. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, December 1964.

[30] R. T. Rockafellar. Monotone Operators and the Proximal Point Algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, August 1976.

[31] E. K. Ryu and W. Yin. *Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators.* Cambridge University Press, December 2022.

[32] Y. Shehu. Convergence Rate Analysis of Inertial Krasnoselskii–Mann Type Iteration with Applications. *Numerical Functional Analysis and Optimization*, 39:1–15, July 2018.

[33] S. Suantai, M. Noor, K. Kankam, and P. Cholamjiak. Novel forward-backward algorithms for optimization and applications to compressive sensing and image inpainting. *Advances in Difference Equations*, May 2021.

[34] S. J. Wright and B. Recht. *Optimization for Data Analysis*. Cambridge University Press, March 2022.

[35] Y. Xu. Ishikawa and Mann Iterative Processes with Errors for Nonlinear Strongly Accretive Operator Equations. *Journal of Mathematical Analysis and Applications*, 224(1):91–101, August 1998.