MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Inertial Krasnosel'skiĭ-Mann algorithms"

verfasst von / submitted by

Jacques Veloso, BSc

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien, 2017 / Vienna, 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Applied Mathematics and Scientific Computing
Betreut von / Supervisor:	Assoc. Prof. Dr. Habil. Radu Ioan Boţ

Abstract

In this paper we consider a generalized inertial version of the Krasnosel'skiĭ-Mann iteration for solving fixed-point problems. First we introduce the classic Krasnosel'skiĭ-Mann iteration and go over some results out of fixed-point theory and monotone operator theory. We then show a proof of weak convergence and present a special case of the proposed general KM-iteration, which delivers an inertial forward-backward algorithm with variable stepsize. Lastly we provide an application for solving image deblurring problems.

Contents

1 Introduction/Motivation

First of all, for the rest of this paper let $\mathcal{H}$ be a real Hilbert space with corresponding scalar product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$. Furthermore, let $\rightharpoonup$, $\rightarrow$ denote weak, respectively strong convergence.

The classical Krasnosel'skiĭ-Mann iteration is defined as

$$(\forall n \in \mathbb{N}) \quad x_{n+1} := (1-\lambda_n)x_n + \lambda_n T x_n \tag{1}$$

where $\lambda_n \in [0,1]$ are the relaxation factor, $T : D \to D$ is a self-mapping with D being a closed and convex nonempty subset of $\mathcal{H}$ and $x_0 \in D$. The Krasnosel'skiĭ-Mann iteration is a well known method in fixed-point theory, in particular for the approximation of fixed-points of nonexpansive operators. Under which conditions does it converge? It is known (see [1, Theorem 5.14]) that (1) converges weakly to a fixed-point of T, i.e

$$x_n \rightharpoonup x \in \operatorname{Fix}(T) := \{x \in D : Tx = x\}$$

if the relaxation factors $(\lambda_n)_{n\in\mathbb{N}}$ fulfill the following condition

$$\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty,$$

and if T is a nonexpansive operator. An operator $T : D \to \mathcal{H}$ is called nonexpansive if

$$\forall x, y \in D : \|Tx - Ty\|^2 \leq \|x-y\|^2.$$

Furthermore, T is called firmly nonexpansive if

$$\forall x, y \in D : \|Tx - Ty\|^2 \leq \|x-y\|^2 - \|(x-y) - (Tx - Ty)\|^2.$$

Every firmly nonexpansive opertator is obviously nonexpansive. For firmly nonexpansive T we even know (see [1, Corollary 5.16]) that (1) converges weakly to a fixed-point of T if the relaxation factors $(\lambda_n)_{n\in\mathbb{N}}$ fulfill the following condition

$$\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty, \text{ where } \lambda_n \in (0,2) \text{ for all } n \geq 0.$$

Notice that in this case in particular, we can set $\lambda_n = 1$ for all $n \in \mathbb{N}$, obtaining an iteration without relaxation factors, i.e. the Picard-iteration

$$(\forall n \in \mathbb{N}) \quad x_{n+1} := Tx_n,$$

which converges weakly to a fixed-point of the firmly nonexpansive operator T. This is not necessarily true for just nonexpansive T.

An extension of the classical Krasnosel'skiĭ-Mann iteration (1) is an inertial version of the Krasnosel'skiĭ-Mann iteration, which can provide an acceleration or a speed up of the classic iteration. For given elements x_0, x_1 of the affine set D the inertial Krasnosel'skiĭ-Mann iteration looks as follows:

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} w_n := x_n + \alpha_n(x_n - x_{n-1}), \\ x_{n+1} := (1-\lambda_n)w_n + \lambda_n T w_n, \end{array} \right. \tag{2}$$

where $\alpha_n \in [0,1]$ are the so called damping terms, and $\lambda_n \in [0,1]$ are again the relaxation factors. Here we can see that the next iterate x_{n+1} is dependent on the two previous iterates x_n and x_{n-1}. More precisely, we use (1) on a affine combination of x_n and x_{n-1}. It is shown in [3, Theorem 5] that the iterates x_n in (2) are weakly converging to a fixed-point of a nonexpansive operator T under the assumption that there exist $0 \leq \alpha_n \leq \alpha < 1$ and $\delta, \sigma, \lambda > 0$ such that

$$\delta > \frac{\alpha^2(1+\alpha)+\alpha\sigma}{1-\alpha^2} \text{ and } 0 < \lambda \leq \lambda_n \leq \frac{\delta - \alpha(\alpha(1+\alpha)+\alpha\delta+\sigma)}{\delta(1+\alpha(1+\alpha)+\alpha\delta+\sigma)}, \forall n \geq 1,$$

where the sequence $(\alpha_n)_{n\in\mathbb{N}}$ is nondecreasing with $\alpha_1 := 0$. Furthermore, in [3] they showed that

$$\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$$

which implies that $x_{n+1} - x_n \to 0$ as $n \to +\infty$.

Our main focus in this paper will be on a more general Krasnosel'skiĭ-Mann iteration. In this setting we have a sequence of nonexpansive operators $(T_n)_{n\in\mathbb{N}}$ with $T_n : D \to D$ for all $n \in \mathbb{N}$ whereas D is a nonempty subset of $\mathcal{H}$. For $x_0, x_1 \in D$ the general Krasnosel'skiĭ-Mann iteration is defined as follows

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} w_n := x_n + \alpha_n(x_n - x_{n-1}), \\ x_{n+1} := (1-\lambda_n)w_n + \lambda_n T_n w_n, \end{array} \right. \tag{3}$$

where $\alpha_n, \lambda_n \in [0,1]$ are damping terms, resp. the relaxation factors as mentioned before. In this setting we also assume that D is weak sequentially closed and affine, otherwise w_n, x_{n+1} and every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ do not necessarily have to be in D again. Furthermore, in the previous iterative methods we had a constant operator T where the associated solution set was $\operatorname{Fix}(T)$. In this case though, since $(T_n)_{n\in\mathbb{N}}$ is not necessarily constant we will show that $(x_n)_{n\in\mathbb{N}}$ converges weakly to an element of the set $S := \bigcap_{n\geq 0} \operatorname{Fix}(T_n)$, assuming it is not empty. But under which conditions does (3) converge? What kind of restrictions do we have to set on relaxation factors $(\lambda_n)_{n\in\mathbb{N}}$, the damping terms $(\alpha_n)_{n\in\mathbb{N}}$ and the operators $(T_n)_{n\in\mathbb{N}}$? We will answer these questions in Section 3, but first it is necessary to recall some preliminary results in Section 2.

2 Preliminaries

We will now list a few necessary lemmata and results for the proofs later on. The first one is a well known norm-identity, one could say it is a generalized form of the parallelogram law (set $\alpha = \frac{1}{2}$ in the following lemma).

Lemma 1. Let $\mathcal{H}$ be a real Hilbert space. For every $x, y \in \mathcal{H}$ and $\alpha \in \mathbb{R}$ it holds

$$\|\alpha x + (1-\alpha)y\|^2 + \alpha(1-\alpha)\|x-y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2.$$

Proof. For all $x, y \in \mathcal{H}$ and $\alpha \in \mathbb{R}$ we have

$$\begin{aligned}
&\|\alpha x + (1-\alpha)y\|^2 + \alpha(1-\alpha)\|x-y\|^2 \\
&= \alpha^2\|x\|^2 + 2\alpha(1-\alpha)\langle x, y\rangle + (1-\alpha)^2\|y\|^2 \\
&\quad + \alpha(1-\alpha)(\|x\|^2 - 2\langle x, y\rangle + \|y\|^2) \\
&= (\alpha^2 + \alpha(1-\alpha))\|x\|^2 + ((1-\alpha)^2 + \alpha(1-\alpha))\|y\|^2 \\
&= \alpha\|x\|^2 + (1-\alpha)\|y\|^2.
\end{aligned}$$

□

The next Lemma is a famous result from Opial (see [8]), which we need for a proof of weak convergence later on.

Lemma 2. (Opial, 1967) Let C be a nonempty subtset of $\mathcal{H}$ and $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that the following two conditions hold:

i) for every $x \in C$, $\lim_{n\to\infty}\|x_n - x\|$ exits;

ii) every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in C.

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C.

Proof. First we show that $(x_n)_{n\in\mathbb{N}}$ has at most one weak sequential cluster point. Let $x, x' \in C$ (by ii)) be two weak sequential cluster points of the sequence $(x_n)_{n\in\mathbb{N}}$ with $x_{n_k} \rightharpoonup x \in C$ and $x_{m_k} \rightharpoonup x' \in C$ as $k \to +\infty$ and define $l(y) := \lim_{n\to\infty}\|x_n - y\|$ for $y \in C$. Then it holds for all $k \in \mathbb{N}$:

$$\begin{aligned}
2\langle x_{n_k}, x - x'\rangle &= \|x_{n_k} - x'\|^2 - \|x_{n_k} - x\|^2 - \|x'\|^2 - \|x\|^2 \\
&\to 2\langle x, x - x'\rangle = l(x') - l(x) - \|x'\|^2 - \|x\|^2
\end{aligned}$$

as $k \to +\infty$ and

$$\begin{aligned}
2\langle x_{m_k}, x - x'\rangle &= \|x_{m_k} - x'\|^2 - \|x_{m_k} - x\|^2 - \|x'\|^2 - \|x\|^2 \\
&\to 2\langle x', x - x'\rangle = l(x') - l(x) - \|x'\|^2 - \|x\|^2,
\end{aligned}$$

as $k \to +\infty$, hence

$$2\left\|x - x'\right\|^2 = 2\langle x, x - x'\rangle - 2\langle x', x - x'\rangle = 0,$$

and therefore $x = x'$. Furthermore, due to i) the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded. Since $(x_n)_{n\in\mathbb{N}}$ is bounded and has at most one weak sequential cluster point, it follows that $x_n \rightharpoonup x \in \mathcal{H}$ as $n \to +\infty$. Using ii) again we get that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C, which finishes the proof. □

The next lemma is a technical result which is also crucial for the proof later on.

Lemma 3. Let $(\varphi_n)_{n\in\mathbb{N}}$, $(\delta_n)_{n\in\mathbb{N}}$ and $(\alpha_n)_{n\in\mathbb{N}}$ be sequences in $[0, +\infty)$ such that $\varphi_{n+1} \leq \varphi_n + \alpha_n(\varphi_n - \varphi_{n-1}) + \delta_n$ for all $n \geq 1$, $\sum_{n\in\mathbb{N}} \delta_n < +\infty$ and there exists a real number α with $0 \leq \alpha_n \leq \alpha < 1$ for all $n \in \mathbb{N}$. Then the following hold:

i) $\sum_{n\geq 1}[\varphi_n - \varphi_{n-1}]_+ < +\infty$, where $[t]_+ = \max\{t, 0\}$;

ii) there exists $\varphi^* \in [0, +\infty)$ such that $\lim_{n\to+\infty} \varphi_n = \varphi^*$.

Proof. Set $u_n := \varphi_n - \varphi_{n-1}$. It follows that

$$[u_{n+1}]_+ \leq \alpha_n[u_n]_+ + \delta_n \leq \alpha[u_n]_+ + \delta_n,$$

and by induction we get

$$[u_{n+1}]_+ \leq \alpha^n[u_1]_+ + \sum_{j=0}^{n-1} \alpha^j \delta_{n-j}.$$

Since $\alpha \in [0, 1)$ and the fact that $\sum_{n\in\mathbb{N}} \delta_n < +\infty$ we obtain

$$\sum_{n\geq 0}[u_{n+1}]_+ \leq \frac{1}{1-\alpha}\left([u_1]_+ + \sum_{n\geq 1} \delta_n\right) < +\infty,$$

which proves i). Furthermore, $w_n := \varphi_n - \sum_{j=1}^{n}[u_j]_+$ is bounded from below and

$$w_{n+1} := \varphi_{n+1} - [u_{n+1}]_+ - \sum_{j=1}^{n}[u_j]_+ \leq \varphi_{n+1} - \varphi_{n+1} + \varphi_n - \sum_{j=1}^{n}[u_j]_+ = w_n,$$

i.e. $(w_n)_{n\in\mathbb{N}}$ is nonincreasing and bounded from below, thus $(w_n)_{n\in\mathbb{N}}$ is convergent and so is $(\varphi_n)_{n\in\mathbb{N}}$ which finishes the proof. □

The next lemma is a useful consequence of nonexpansive operators.

Lemma 4. (Demi-closedness principle) Let $D \subseteq \mathcal{H}$ be non-empty and weak sequentially closed, $T : D \to \mathcal{H}$ nonexpansive and $(x_n)_{n\in\mathbb{N}} \subseteq D$, $x, u \in \mathcal{H}$. It holds that

$$x_n \rightharpoonup x \text{ and } x_n - Tx_n \to u \text{ as } n \to +\infty \Rightarrow x - Tx = u.$$

In particular, if we set $u = 0$ we get

$$\begin{aligned} x_n \rightharpoonup x \text{ and } x_n - Tx_n \to 0 \text{ as } n \to +\infty &\Rightarrow x = Tx \\ &\Leftrightarrow x \in \operatorname{Fix}(T). \end{aligned}$$

Proof. Let $(x_n)_{n\in\mathbb{N}} \subseteq D$ with $x_n \rightharpoonup x$ and $x_n - Tx_n \to u$ as $n \to +\infty$. Since D is weak sequentially closed, $x \in D$ and Tx is therefore well defined. Moreover, from the nonexpansiveness of T it follows for all $n \in \mathbb{N}$ that

$$\begin{aligned} \|x - Tx - u\|^2 &= \|x_n - Tx - u\|^2 - \|x_n - x\|^2 - 2\langle x_n - x, x - Tx - u\rangle \\ &= \|x_n - Tx_n - u\|^2 + 2\langle x_n - Tx_n - u, Tx_n - Tx\rangle \\ &\quad + \|Tx_n - Tx\|^2 - \|x_n - x\|^2 - 2\langle x_n - x, x - Tx - u\rangle \\ &\leq \|x_n - Tx_n - u\|^2 + 2\langle x_n - Tx_n - u, Tx_n - Tx\rangle \\ &\quad - 2\langle x_n - x, x - Tx - u\rangle \\ &\leq \|x_n - Tx_n - u\|^2 + 2\|x_n - Tx_n - u\|\|x_n - x\| \\ &\quad - 2\langle x_n - x, x - Tx - u\rangle. \end{aligned}$$

Taking the limit as $n \to +\infty$ in the last inequality and using the fact that the sequence $(\|x_n - x\|)_{n\in\mathbb{N}}$ is bounded (by the uniform boundedness principle), we obtain $x - Tx = u$. $\square$

The previous result is called the "Demi-closedness principle" since it guarantees that the graph of $Id - T$ is demi-closed; in other words it suffices to have weak convergence ($x_n \rightharpoonup x$ as $n \to +\infty$) in the domain of $Id - T$, and strong convergence ($(Id - T)x_n \to u$ as $n \to +\infty$) in the range of $(Id - T)$ to get $(Id - T)x = u$. Note that it is sufficient to assume that D is closed and convex in Lemma 4, since for every convex set $M \subseteq \mathcal{H}$ it holds (see [1, Theorem 3.32])

$$M \text{ closed} \Leftrightarrow M \text{ weak sequentially closed.}$$

Now let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-valued operator. The graph of A is defined by $\operatorname{Gr}(A) := \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$. Similarly, we can define the inverse of A, i.e. $A^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ by the equivalence: $(x, u) \in \operatorname{Gr}(A)$ if and only if $(u, x) \in \operatorname{Gr}(A^{-1})$. Furthermore, let $\operatorname{Zer}(A) := \{x \in \mathcal{H} : 0 \in Ax\} = A^{-1}(0)$ denote the set of zeros of A and $\operatorname{Ran}(A) := \bigcup_{x\in H} Ax$ its range.

Definition 5. A set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is called monotone, if

$$\forall (x,u),(y,v) \in \operatorname{Gr}(A) : \langle x-y, u-v\rangle \geq 0.$$

Furthermore, it is called maximally monotone if there is no monotone operator $B : \mathcal{H} \rightrightarrows \mathcal{H}$ such that the graph of B properly contains the graph of A on $\mathcal{H} \times \mathcal{H}$. In other words, an operator A is maximally monotone if for all $(x,u) \in \mathcal{H} \times \mathcal{H}$ it holds that

$$(x,u) \in \operatorname{Gr}(A) \Leftrightarrow \forall (y,v) \in \operatorname{Gr}(A) : \langle x-y, u-v\rangle \geq 0.$$

A popular example for maximally monotone operators is the convex subdifferential

$$\partial f(x) := \{\xi \in \mathcal{H} : f(y) - f(x) \geq \langle y-x, \xi\rangle \quad \text{for all } y \in \mathcal{H}\}$$

of a proper, convex and lower semi-continuous function f, i.e. if f is an element of the space

$$\Gamma(\mathcal{H}) := \{f : \mathcal{H} \to \overline{\mathbb{R}} : f \text{ is proper, convex and lsc}\}$$

where $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ denotes the extended real line. Another useful property of the convex subdifferential is the following

$$0 \in \partial f(x) \text{ if and only if } x \in \operatorname{argmin} f. \tag{4}$$

The next lemma is a asymptotic result about the set of zeros of the sum of two maximally monotone operators, which we again need for a proof later on (see [1, Corollary 25.5 for m=2]).

Lemma 6. Let $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone operators and the sequences $(x_n, u_n)_{n\in\mathbb{N}} \in \operatorname{Gr}(A)$, $(y_n, v_n)_{n\in\mathbb{N}} \in \operatorname{Gr}(B)$ such that

$$x_n \rightharpoonup x, \quad y_n \rightharpoonup y, \quad u_n \rightharpoonup u, \quad v_n \rightharpoonup v, \quad u_n + v_n \to 0 \text{ and } x_n - y_n \to 0$$

as $n \to +\infty$. Then $x = y \in \operatorname{Zer}(A+B)$, $(x,u) \in \operatorname{Gr}(A)$ and $(y,v) \in \operatorname{Gr}(B)$.

The resolvent of A is defined by

$$J_A = (\operatorname{Id} + A)^{-1}, \quad J_A : \operatorname{Dom}(J_A) \rightrightarrows \mathcal{H}$$

where $\operatorname{Id} : \mathcal{H} \to \mathcal{H}$ is the identity operator. The following lemma is a useful characterization of the resolvent operator of A.

Lemma 7. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$, $x, p \in \mathcal{H}$ and $\gamma > 0$. It holds:

$$p \in J_{\gamma A} x \Leftrightarrow (p, \gamma^{-1}(x-p)) \in \operatorname{Gr}(A).$$

Proof. For every $x \in \mathcal{H}$ and $\gamma > 0$ we have

$$p \in J_{\gamma A}x = (I + \gamma A)^{-1}x \Leftrightarrow x \in (I + \gamma A)p \Leftrightarrow \frac{1}{\gamma}(x - p) \in Ap$$
$$\Leftrightarrow (p, \gamma^{-1}(x - p)) \in \mathrm{Gr}(A).$$

□

With this lemma one can easily see that the fixed-point set of J_A coincides with the set of zeroes of A, i.e.

$$\mathrm{Fix}(J_A) = \mathrm{Zer}(A). \tag{5}$$

The next theorem (see [7]) is an important equivalence for maximally monotone operators , it also implies that J_A has full domain if and only if A is maximally monotone.

Theorem 8. ***(Minty, 1962)*** *An operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximally monotone if and only if*

$$\mathrm{Ran}(Id + \gamma A) = \mathcal{H} \text{ for some } \gamma > 0.$$

From this theorem and Lemma 7 we can deduce that $J_{\gamma A}$ is single valued when A is a maximal monotone operator and $\gamma > 0$, since for an arbitrary $x \in \mathcal{H}$ we get $J_{\gamma A}x \neq \emptyset$ (by Theorem 8) and for $y_1, y_2 \in J_{\gamma A}x$ with $y_1 \neq y_2$ we know that $Ay_1 = x - y_1$ and $Ay_2 = x - y_2$ by Lemma 7. Thus we obtain

$$\begin{aligned}\|y_1 - y_2\|^2 = \langle y_1 - y_2, y_1 - y_2\rangle &= \langle (y_1 - x) + (x - y_2), y_1 - y_2\rangle \\ &= \gamma\langle Ay_2 - Ay_1, y_1 - y_2\rangle \qquad \leq 0\end{aligned}$$

hence $y_1 = y_2$, and therefore $J_{\gamma A}$ is single valued. Now that we know that $J_{\gamma A}$ is single valued for maximally monotone A, we can further show that $J_{\gamma A}$ is firmly nonexpansive for maximally monotone A and $\gamma > 0$.

Corollary 9. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone and $\gamma > 0$. Then $J_{\gamma A}$ is firmly nonexpansive.

Proof. Let $x, y \in \mathcal{H}$ and $\gamma > 0$. From the argumentation above we know that $J_{\gamma A}$ is single valued, i.e. there exist $x', y' \in \mathcal{H}$ such that $x' = J_{\gamma A}x$ and $y' = J_{\gamma A}y$. Furthermore, it holds that

$$\begin{aligned}\|J_{\gamma A}x - J_{\gamma A}y\|^2 &= \left\|x' - y'\right\|^2 \\ &\leq \left\|x' - y'\right\|^2 + \gamma\langle x' - y', Ax' - Ay'\rangle \\ &= \left\|x' - y' + \gamma(Ax' - Ay')\right\|^2 - \left\|\gamma(Ax' - Ay')\right\|^2 \\ &= \|x - y\|^2 - \|(\mathrm{Id} - J_{\gamma A})x - (\mathrm{Id} - J_{\gamma A})y\|^2\end{aligned}$$

where the last equality follows from the fact that $x = (\mathrm{Id} + \gamma A)x'$ and $y = (\mathrm{Id} + \gamma A)y'$. □

Definition 10. An operator $B : \mathcal{H} \to \mathcal{H}$ is called β-cocoercive for $\beta > 0$ if for all $x, y \in \mathcal{H}$ it holds

$$\beta \|Bx - By\|^2 \leq \langle Bx - By, x - y \rangle.$$

Note that B being β-cocoercive is equivalent to B being $\frac{1}{\beta}$-Lipschitz continuous. Moreover, every cocoercive operator is in particular maximally monotone, as we will see in the next lemma.

Lemma 11. If $B : \mathcal{H} \to \mathcal{H}$ is β-cocoercive for $\beta > 0$, then B is maximally monotone.

Proof. The monotonicity of B follows immediately from the cocoercivity of B. Let $(x, u) \in \mathcal{H} \times \mathcal{H}$. It remains to show that

$$\forall y \in \mathcal{H} : \langle x - y, u - By \rangle \geq 0 \Rightarrow (x, u) \in \mathrm{Gr}(B).$$

Set $y_\alpha := x + \alpha(u - Bx)$ for $\alpha \geq 0$. We obtain for all $\alpha \geq 0$

$$\begin{aligned} -\alpha\langle u - Bx, u - By_\alpha \rangle = \langle x - y_\alpha, u - By_\alpha \rangle \geq 0 \\ \Rightarrow \langle u - Bx, u - By_\alpha \rangle \leq 0 \text{ , for all } \alpha \geq 0, \end{aligned}$$

and since B and the scalar product are continuous, it follows $\|u - Bx\|^2 \leq 0$, i.e. $u = Bx$ and $(x, u) \in \mathrm{Gr}(B)$ which finishes the proof. □

We will now introduce the well known forward-backward algorithm (see [1, Theorem 25.8]) which is a special case of the classic Krasnosel'skiĭ-Mann iteration (1).

Theorem 12. ***(Forward-Backward algorithm)*** *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $B : \mathcal{H} \to \mathcal{H}$ be β-cocoercive with $\beta > 0$, let $\gamma \in (0, 2\beta)$, and set $\delta := \min\{1, \frac{\beta}{\gamma}\} + \frac{1}{2}$. Furthermore, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$ and let $x_0 \in H$. Suppose that $\mathrm{Zer}(A + B) \neq \emptyset$ and set*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} y_n & := x_n - \gamma B x_n, \\ x_{n+1} & := (1 - \lambda_n) w_n + \lambda_n J_{\gamma A} y_n, \end{array} \right.$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathrm{Zer}(A + B)$.

If we set $T := J_{\gamma A}(\mathrm{Id} - \gamma B)$, $\lambda'_n := \frac{\lambda_n}{\delta}$ and show that T is also nonexpansive then the proof follows from the classic KM iteration in (1). We will just show the nonexpansivness of $T := J_{\gamma A}(\mathrm{Id} - \gamma B)$ under the assumptions given in Theorem 12.

Lemma 13. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $B : \mathcal{H} \to \mathcal{H}$ be β-cocoercive with $\beta > 0$ and $\gamma \in (0, 2\beta)$. Then $T := J_{\gamma A}(\mathrm{Id} - \gamma B)$ is nonexpansive.

Proof. Let $x, y \in \mathcal{H}$. From the nonexpansivness of $J_{\gamma A}$ and the β-cocoerciveness of B it follows

$$\begin{aligned}
\|T_n x - T_n y\|^2 &= \|J_{\gamma A}(Id - \gamma B)x - J_{\gamma A}(Id - \gamma B)y\|^2 \\
&\leq \|(\mathrm{Id} - \gamma B)x - (\mathrm{Id} - \gamma B)y\|^2 \\
&= \|x - y\|^2 - 2\langle x - y, \gamma(Bx - By)\rangle + \gamma^2 \|Bx - By\|^2 \\
&\leq \|x - y\|^2 - 2\langle x - y, \gamma(Bx - By)\rangle + \gamma^2 \frac{1}{\beta}\langle x - y, Bx - By\rangle \\
&= \|x - y\|^2 - \gamma \underbrace{(2 - \frac{\gamma}{\beta})}_{\geq 0} \underbrace{\langle x - y, Bx - By\rangle}_{\geq 0} \\
&\leq \|x - y\|^2
\end{aligned}$$

which finishes the proof. □

How can we use these ideas to find, for example, a minimizer of a proper, convex, lower semi-continuous f? In other words how should we choose A, B in Theorem 12 to solve

$$\operatorname*{argmin}_{x \in H} f(x), \quad \text{for } f \in \Gamma(\mathcal{H}).$$

This is where the proximal operator comes in handy. For functions $f \in \Gamma(\mathcal{H})$ we can define the proximal operator $\mathrm{Prox}_f : \mathcal{H} \to \mathcal{H}$ of f by

$$\mathrm{Prox}_f(y) := \operatorname*{argmin}_{x \in \mathcal{H}} f(x) + \frac{1}{2}\|x - y\|_2^2.$$

Prox_f is well defined for $f \in \Gamma(\mathcal{H})$ considering $J_{\partial f}$ is single valued and it holds that $\mathrm{Prox}_f = J_{\partial f}$ since

$$\begin{aligned}
p = J_{\partial f}(y) \Leftrightarrow y - p \in \partial f(p) \Leftrightarrow 0 \in \partial f(p) + \{p - y\} \\
\Leftrightarrow p = \operatorname*{argmin}_{x \in \mathcal{H}} f(x) + \frac{1}{2}\|x - y\|_2^2 = \mathrm{Prox}_f(y).
\end{aligned}$$

Furthermore, from (4) and (5) we obtain

$$\mathrm{Fix}(\mathrm{Prox}_f) = \mathrm{Fix}(J_{\partial f}) = \mathrm{Zer}(\partial f) = \mathrm{argmin}(f).$$

So if we have a function $f \in \Gamma(\mathcal{H})$ with $\mathrm{argmin}(f) \neq \emptyset$, then finding a minimizer of f is equivalent to finding a fixed-point of Prox_f which is again equivalent to finding a zero of ∂f. Furthermore, it holds that

$$\mathrm{Fix}(J_{\gamma_n A}(Id - \gamma_n B)) = \mathrm{Zer}(\gamma_n(A + B)) = \mathrm{Zer}(A + B),$$

i.e. if we substitute A in Theorem 12 with ∂f and B with ∇g, then we obtain the following algorithm (see [1, Theorem 27.9]).

Theorem 14. ***(Proximal-Gradient algorithm)*** *Let $f \in \Gamma(\mathcal{H})$, let $g : \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a $\frac{1}{\beta}$-Lipschitz continuous gradient for some $\beta > 0$, let $\gamma \in (0, 2\beta)$, and set $\delta := \min\{1, \frac{\beta}{\gamma}\} + \frac{1}{2}$. Furthermore, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$ and let $x_0 \in H$. Suppose that* $\operatorname{argmin}(f + g) \neq \emptyset$ *and set*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} y_n & := x_n - \gamma \nabla x_n, \\ x_{n+1} & := (1 - \lambda_n)x_n + \lambda_n \operatorname{Prox}_{\gamma f} y_n, \end{array} \right.$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{argmin}(f + g)$.

In the next section we prove the weak convergence of the general Krasnosel'skiĭ-Mann iteration (3), and we derive the algorithms from above as a special case of it.

3 General Krasnosel'skiĭ-Mann iteration

In the classical Krasnosel'skiĭ-Mann iteration

$$x_{n+1} = (1 - \lambda_n)x_n + \lambda_n T x_n$$

one used the demi-closedness principle (Lemma 4) for a nonexpansive operator T to prove the weak convergence of the sequence $(x_n)_{n\in\mathbb{N}}$ to an element $x \in \operatorname{Fix}(T)$. In the more general setting

$$w_n := x_n + \alpha_n(x_n - x_{n-1}),$$
$$x_{n+1} := (1 - \lambda_n)w_n + \lambda_n T_n w_n,$$

we need a similar statement for the sequence $(T_n)_{n\in\mathbb{N}}$ where $T_n : D \to \mathcal{H}$. To be more precise, in the rest of this section we assume that $(T_n)_{n\in\mathbb{N}}$ fulfills the following "demi-closedness-type" condition:

$$\begin{gathered}\text{For any subsequence } (T_{n_k})_{k\in\mathbb{N}} \text{ of } (T_n)_{n\in\mathbb{N}}, \text{ for } (x_{n_k})_{k\in\mathbb{N}} \subseteq D, x \in \mathcal{H} \\ (x_{n_k}) \rightharpoonup x \text{ and } x_{n_k} - T_{n_k}x_{n_k} \to 0 \Rightarrow x \in \bigcap_{n\geq 0} \operatorname{Fix}(T_n).\end{gathered} \tag{6}$$

We know that for the particular case where $T_n = T$ for all $n \in \mathbb{N}$ and T nonexpansive the above condition is fulfilled (if D is also weak sequentially closed), thanks to the demi-closedness principle. Unfortunately, in general it does not suffice that every operator T_n is nonexpansive, i.e. condition (6) is in general not fulfilled for nonexpansive operators $(T_n)_{n\in\mathbb{N}}$ (take, e.g. $T_n := (1 - 1/n)\operatorname{Id}$). That is a reason why we have to assume that a given sequence of operators $(T_n)_{n\in\mathbb{N}}$ has to satisfy (6).
The next theorem is the main result of this paper. It is heavily based on the work of [3] and [6].

Theorem 15. *Let D be a nonempty weak-sequentially closed affine subset of H and $T_n : D \to D$ be a sequence of nonexpansive operators such that $\bigcap_{n\geq 0} \operatorname{Fix}(T_n) \neq \emptyset$. We consider the following iterative scheme:*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} w_n & := x_n + \alpha_n(x_n - x_{n-1}), \\ x_{n+1} & := (1-\lambda_n)w_n + \lambda_n T_n w_n, \end{array} \right.$$

where x_0, x_1 are arbitrarily chosen in D, $(\alpha_n)_{n\geq 1}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_n \leq \alpha < 1$ for every $n \geq 1$ and $\lambda, \sigma, \delta > 0$ are such that

$$\delta > \frac{\alpha^2(1+\alpha)+\alpha\sigma}{1-\alpha^2} \text{ and } 0 < \lambda \leq \lambda_n \leq \frac{\delta - \alpha(\alpha(1+\alpha)+\alpha\delta+\sigma)}{\delta(1+\alpha(1+\alpha)+\alpha\delta+\sigma)} \ \forall n \geq 1.$$

Then the following statements are true:

i) $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$;

ii) if furthermore condition (6) holds, then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\bigcap_{n\geq 0} \operatorname{Fix}(T_n)$.

Proof. Let us start with the remark that, due to the choice of $\delta, \lambda_n \in (0,1)$ for every $n \geq 1$. Furthermore, we would like to notice that, since D is affine, the iterative scheme provides a well-defined sequence in D.
$i)$ Let us fix an element $y \in \bigcap_{n\geq 0} \operatorname{Fix}(T_n)$ and $n \geq 1$. It follows from Lemma 1 and the nonexpansiveness of T_n that

$$\begin{aligned} \|x_{n+1} - y\|^2 &= (1-\lambda_n)\|w_n - y\|^2 + \lambda_n \|T_n w_n - T_n y\|^2 - \lambda_n(1-\lambda_n)\|T_n w_n - w_n\|^2 \\ &\leq \|w_n - y\|^2 - \lambda_n(1-\lambda_n)\|T_n w_n - w_n\|^2 \end{aligned} \tag{7}$$

Applying Lemma 1 again, we have

$$\begin{aligned} \|w_n - y\|^2 &= \|(1+\alpha_n)(x_n - y) - \alpha_n(x_{n-1} - y)\|^2 \\ &= (1+\alpha_n)\|x_n - y\|^2 - \alpha_n\|x_{n-1} - y\|^2 + \alpha_n(1+\alpha_n)\|x_n - x_{n-1}\|^2 \end{aligned} \tag{8}$$

hence from (7) we obtain

$$\begin{aligned} &\|x_{n+1} - y\|^2 - (1+\alpha_n)\|x_n - y\|^2 + \alpha_n\|x_{n-1} - y\|^2 \\ &\leq -\lambda_n(1-\lambda_n)\|T_n w_n - w_n\|^2 + \alpha_n(1+\alpha_n)\|x_n - x_{n-1}\|^2. \end{aligned} \tag{9}$$

Furthermore, we have

$$
\begin{aligned}
\|T_n w_n - w_n\|^2 &= \left\| \frac{1}{\lambda_n}(x_{n+1} - x_n) + \frac{\alpha_n}{\lambda_n}(x_{n-1} - x_n) \right\|^2 \\
&= \frac{1}{\lambda_n^2} \|x_{n+1} - x_n\|^2 + \frac{\alpha_n^2}{\lambda_n^2} \|x_n - x_{n-1}\|^2 + 2\frac{\alpha_n}{\lambda_n^2} \langle x_{n+1} - x_n, x_{n-1} - x_n \rangle \\
&\geq \frac{1}{\lambda_n^2} \|x_{n+1} - x_n\|^2 + \frac{\alpha_n^2}{\lambda_n^2} \|x_n - x_{n-1}\|^2 \\
&\quad + \frac{\alpha_n}{\lambda_n^2}(-\rho_n \|x_{n+1} - x_n\|^2 - \frac{1}{\rho_n} \|x_n - x_{n-1}\|^2) \qquad (10)
\end{aligned}
$$

where we denote $\rho_n := \frac{1}{\alpha_n + \delta\lambda_n}$.
We derive from (9) and (10) the inequality

$$
\begin{aligned}
&\|x_{n+1} - y\|^2 - (1 + \alpha_n) \|x_n - y\|^2 + \alpha_n \|x_{n-1} - y\|^2 \\
&\leq \frac{(1 - \lambda_n)(\alpha_n \rho_n - 1)}{\lambda_n} \|x_{n+1} - x_n\|^2 + \gamma_n \|x_n - x_{n-1}\|^2, \qquad (11)
\end{aligned}
$$

where

$$\gamma_n := \alpha_n(1 + \alpha_n) + \alpha_n(1 - \lambda_n)\frac{1 - \rho_n\alpha_n}{\rho_n\lambda_n} \geq 0, \qquad (12)$$

since $\rho_n\alpha_n < 1$ and $\lambda_n \in (0, 1)$.
Again, taking into account the choice of ρ_n we have

$$\delta = \frac{1 - \rho_n\alpha_n}{\rho_n\lambda_n}$$

and from (12), it follows

$$\gamma_n = \alpha_n(1 + \alpha_n) + \alpha_n(1 - \lambda_n)\delta \leq \alpha(1 + \alpha) + \alpha\delta \quad \forall n \geq 1. \qquad (13)$$

We define the sequences $\varphi_n := \|x_n - y\|^2$ for all $n \in \mathbb{N}$ and $\mu_n := \varphi_n - \alpha_n\varphi_{n-1} + \gamma_n \|x_n - x_{n-1}\|^2$ for all $n \geq 1$. Using the monotonicity of $(\alpha_n)_{n\geq1}$ and the fact that $\varphi_n \geq 0$ for all $n \in \mathbb{N}$, we get

$$\mu_{n+1} - \mu_n \leq \varphi_{n+1} - (1+\alpha_n)\varphi_n + \alpha_n\varphi_{n-1} + \gamma_{n+1} \|x_{n+1} - x_n\|^2 - \gamma_n \|x_n - x_{n-1}\|^2.$$

Employing (11), we have

$$\mu_{n+1} - \mu_n \leq \left(\frac{(1 - \lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n} + \gamma_{n+1} \right) \|x_{n+1} - x_n\|^2 \quad \forall n \geq 1. \quad (14)$$

We claim that

$$\frac{(1 - \lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n} + \gamma_{n+1} \leq -\sigma \quad \forall n \geq 1. \qquad (15)$$

Let be $n \geq 1$. Indeed, by the choice of ρ_n, we get

$$\begin{aligned}
&\frac{(1-\lambda_n)(\alpha_n\rho_n - 1)}{\lambda_n} + \gamma_{n+1} \leq -\sigma \\
&\Leftrightarrow \lambda_n(\gamma_{n+1}+\sigma) + (\alpha_n\rho_n - 1)(1-\lambda_n) \leq 0 \\
&\Leftrightarrow \lambda_n(\gamma_{n+1}+\sigma) - \frac{\delta\lambda_n(1-\lambda_n)}{\alpha_n + \delta\lambda_n} \leq 0 \\
&\Leftrightarrow (\alpha_n + \delta\lambda_n)(\gamma_{n+1}+\sigma) + \delta\lambda_n \leq \delta.
\end{aligned}$$

By using (13), we have

$$(\alpha_n + \delta\lambda_n)(\gamma_{n+1}+\sigma) + \delta\lambda_n \leq (\alpha + \delta\lambda_n)(\alpha(1+\alpha) + \alpha\delta + \sigma) + \delta\lambda_n \leq \delta$$

where the last inequality follows by using the upper bound for $(\lambda_n)_{n\geq 1}$. Hence the claim in (15) is true. We obtain from (14) and (15) that

$$\mu_{n+1} - \mu_n \leq -\sigma \|x_{n+1} - x_n\|^2 \quad \forall n \geq 1. \tag{16}$$

The sequence $(\mu_n)_{n\geq 1}$ is nonincreasing and the bound for $(\alpha_n)_{n\geq 1}$ delivers

$$-\alpha\varphi_{n-1} \leq \varphi_n - \alpha\varphi_{n-1} \leq \mu_n \leq \mu_1 \quad \forall n \geq 1. \tag{17}$$

We obtain

$$\varphi_n \leq \alpha^n\varphi_0 + \mu_1 \sum_{k=0}^{n-1} \alpha^k \leq \alpha^n\varphi_0 + \frac{\mu_1}{1-\alpha} \quad \forall n \geq 1$$

where we notice that $\mu_1 = \varphi_1 \geq 0$ (due to the relation $\alpha_1 = 0$). Combining (16) and (17), we get for all $n \geq 1$

$$\sigma \sum_{k=1}^{n} \|x_{k+1} - x_k\|^2 \leq \mu_1 - \mu_{n+1} \leq \mu_1 + \alpha\varphi_n \leq \alpha^{n+1}\varphi_0 + \frac{\mu_1}{1-\alpha}$$

which proves i).

ii) We prove this statement by using the result of Opial in Lemma 2. We have proven that for an arbitrary $y \in \bigcap_{n\geq 0} \operatorname{Fix}(T_n)$ inequality (11) is true. On the one hand, by part i), (13) and Lemma 3 we derive that $\lim_{n\to+\infty} \|x_n - y\|$ exists (we also take into consideration that in (11) $\alpha_n\rho_n < 1$ for all $n \geq 1$). On the other hand, let x be a weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, that is, the latter has a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ fulfilling $x_{n_k} \rightharpoonup x$ as $k \to +\infty$. By part i), the definition of w_n and the upper bound for $(\alpha_n)_{n\geq 1}$, we get $w_{n_k} \rightharpoonup x$ as $k \to +\infty$. Furthermore, we have

$$\begin{aligned}
\|T_n w_n - w_n\| = \frac{1}{\lambda_n}\|x_{n+1} - w_n\| &\leq \frac{1}{\lambda}\|x_{n+1} - w_n\| \\
&\leq \frac{1}{\lambda}(\|x_{n+1} - x_n\| + \alpha\|x_n - x_{n-1}\|)
\end{aligned}$$

thus by i), we obtain that $T_{n_k}w_{n_k} - w_{n_k} \to 0$ as $k \to +\infty$. Applying now (6) for the sequence $(w_{n_k})_{k\in\mathbb{N}}$, we conclude that $x \in \bigcap_{n\geq 0} \operatorname{Fix}(T_n)$. Since the two assumptions of Lemma 2 (Opial) are verified, it follows that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\bigcap_{n\geq 0} \operatorname{Fix}(T_n)$. □

Remark 16. We can simplify the constraints for λ_n in Theorem 15 if σ tends towards to 0. By assuming that σ is very small, i.e. close to zero, we get the following approximation

$$\frac{\delta - \alpha(\alpha(1+\alpha) + \alpha\delta)}{\delta(1 + \alpha(1+\alpha) + \alpha\delta)} \approx \frac{\delta - \alpha(\alpha(1+\alpha) + \alpha\delta + \sigma)}{\delta(1 + \alpha(1+\alpha) + \alpha\delta + \sigma)}.$$

Now we will maximize the upper bound for λ_n in Theorem 15 for a given α, in other words we define

$$\lambda_\infty(\alpha) := \max_{\delta > \frac{\alpha^2}{1-\alpha}} \frac{\delta - \alpha(\alpha(1+\alpha) + \alpha\delta)}{\delta(1 + \alpha(1+\alpha) + \alpha\delta)}$$

and by solving this constrained optimization problem for δ we obtain

$$\lambda_\infty(\alpha) = \frac{\delta^* - \alpha(\alpha(1+\alpha) + \alpha\delta^*)}{\delta^*(1 + \alpha(1+\alpha) + \alpha\delta^*)} \text{ for } \delta^* = \frac{\alpha^2 + \sqrt{\alpha^4 + \frac{1-\alpha}{1+\alpha}(\alpha^3 + \alpha^2 + \alpha)}}{1-\alpha}.$$

One can see that there is a trade-off between choosing α and choosing λ_∞, more precisely, the bigger the value of α is, the smaller the value of λ_∞ has to be and vice versa (see Figure 1).

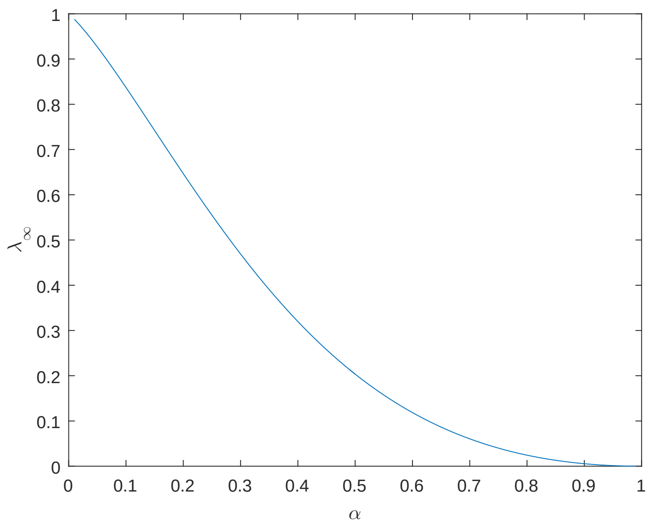

Figure 1: All the positive points below the graph of λ_∞ are feasible values for λ_n and α_n.

This more general setting allows us to make inertial methods with variable stepsize. For example, if we set $T_n := J_{\gamma_n A}(Id - \gamma_n B)$ and we could show that this particular sequence $(T_n)_{n\in\mathbb{N}}$ fulfills condition (6) and that T_n is nonexpansive for all $n \in \mathbb{N}$, we would then obtain an inertial forward-backward algorithm with variable stepsize. We already know that T_n is nonexpansive for all $n \in \mathbb{N}$ due to Lemma 13. Moreover, it is in fact true that this particular sequence $(T_n)_{n\in\mathbb{N}}$ fulfills condition (6), as we will see in the next corollary.

Corollary 17. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $B : \mathcal{H} \to \mathcal{H}$ be β-cocoercive continuous and $\inf_{n\in\mathbb{N}} \gamma_n > 0$. Suppose that $\mathrm{Zer}(A+B) \neq \emptyset$ and set $T_n := J_{\gamma_n A}(Id - \gamma_n B)$. Then $(T_n)_{n\in\mathbb{N}}$ fulfills condition (6).

Proof. First of, from Lemma 7 it follows immediately that

$$\mathrm{Fix}(J_{\gamma_n A}(Id - \gamma_n B)) = \mathrm{Zer}(\gamma_n(A + B)) = \mathrm{Zer}(A + B),$$

hence $\bigcap\limits_{n\geq 0} \mathrm{Fix}\,(T_n) = \mathrm{Zer}(A + B)$.
Now, let $x_{n_k} \rightharpoonup x$ and $x_{n_k} - T_{n_k} x_{n_k} \to 0$. We will use Lemma 6 to show that $x \in \mathrm{Zer}(A + B)$. Define $y_k := T_{n_k} x_{n_k}$ for all k. It holds $x_{n_k} - y_k \to 0$, therefore we get that $y_k \rightharpoonup x$. Since

$$x_{n_k} - y_k \in (Id + \gamma_{n_k} A)y_k + \gamma_{n_k} B x_{n_k} - y_k = \gamma_{n_k}(Ay_k + Bx_{n_k}),$$

it follows that

$$\forall k \in \mathbb{N}\ \exists\ (x_{n_k}, u_k) \in \mathrm{Gr}(B),\ (y_k, v_k) \in \mathrm{Gr}(A)) : \gamma_{n_k}(v_k+u_k) = x_{n_k} - y_k \to 0,$$

thus we obtain $(v_k+u_k) \to 0$ as $k \to +\infty$ because $\inf_{n\in\mathbb{N}} \gamma_n > 0$. Now we will show that there exist convergent subsequences of $(u_k)_{k\in\mathbb{N}}, (v_k)_{k\in\mathbb{N}}$, allowing us to use Lemma 6 in order to finish the proof. Since B is β-cocoercive we know that B is in particular maximally monotone (see Lemma 11) and we get

$$\|Bx_{n_k}\| \leq \beta^{-1}\|x_{n_k}\| + \|B0\|.$$

Furthermore, from the weak convergence of $(x_{n_k})_{k\in\mathbb{N}}$ together with the uniform boundedness principle we know that $(x_{n_k})_{k\in\mathbb{N}}$ is bounded, hence $(Bx_{n_k})_{k\in\mathbb{N}}$ is also bounded. Consequently, there exists a convergent subsequence of $(Bx_{n_k})_{k\in\mathbb{N}}$, i.e. $Bx_{n_{k_l}} = u_{k_l} \rightharpoonup u$ as $l \to +\infty$, and since $(v_k + u_k) \to 0$ as $k \to +\infty$ it follows that $v_{k_l} \rightharpoonup -u$ as $l \to +\infty$.
Finally, we can use Lemma 6 for the sequences $(x_{n_{k_l}})_{l\in\mathbb{N}}, (y_{k_l})_{l\in\mathbb{N}}, (u_{k_l})_{l\in\mathbb{N}}$ and $(v_{k_l})_{l\in\mathbb{N}}$ which gives us $x = y \in \mathrm{Zer}(A + B)$. $\square$

We are now able to formulate an inertial forward-backward algorithm with variable stepsize.

Theorem 18. ***(Inertial Forward-Backward algorithm with variable stepsize)*** *Let $f \in \Gamma(\mathcal{H})$, $g : \mathcal{H} \to \mathbb{R}$ be convex with a $\frac{1}{\beta}$-Lipschitz continuous gradient. Furthermore, let $\sup_{n\in\mathbb{N}} \gamma_n \leq 2\beta$, $\inf_{n\in\mathbb{N}} \gamma_n > 0$ and assume that* $\operatorname{argmin}(f+g) \neq \emptyset$. *We consider the following iteration scheme:*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} w_n & := x_n + \alpha_n(x_n - x_{n-1}), \\ y_n & := w_n - \gamma_n \nabla g(w_n) \\ x_{n+1} & := (1-\lambda_n) w_n + \lambda_n \operatorname{Prox}_{\gamma_n f} y_n, \end{array} \right.$$

where x_0, x_1 are arbitrarily chosen in $\mathcal{H}$, $(\alpha_n)_{n\geq 1}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_n \leq \alpha < 1$ for every $n \geq 1$ and $\lambda, \sigma, \delta > 0$ are such that

$$\delta > \frac{\alpha^2(1+\alpha)+\alpha\sigma}{1-\alpha^2} \text{ and } 0 < \lambda \leq \lambda_n \leq \frac{\delta - \alpha(\alpha(1+\alpha)+\alpha\delta+\sigma)}{\delta(1+\alpha(1+\alpha)+\alpha\delta+\sigma)} \; \forall n \geq 1.$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{argmin}(f+g)$.

Proof. Set $A = \partial f, B = \nabla g$ and define $T_n := J_{\gamma_n A}(Id - \gamma_n B) : \mathcal{H} \to \mathcal{H}$, for all $n \in \mathbb{N}$. We will show that $(T_n)_{n\in\mathbb{N}}$ fulfills condition (6) and that T_n is nonexpansive so we can use Theorem 15 which finishes the proof. From Lemma 13 we already know that T_n is nonexpansive. It remains to show that $(T_n)_{n\in\mathbb{N}}$ fulfills condition (6). Since $f \in \Gamma(\mathcal{H})$ we know that ∂f is maximally monotone and ∇g is β-cocoercive seeing that it is $\frac{1}{\beta}$-Lipschitz continuous. Furthermore, it holds that

$$\bigcap_{n\in\mathbb{N}} \operatorname{Fix}(J_{\gamma_n A}(Id - \gamma_n B)) = \operatorname{Zer}(A+B) = \operatorname{argmin}(f+g),$$

hence by applying Corollary 17 it follows that $(T_n)_{n\in\mathbb{N}}$ fulfills condition (6). ☐

The next remark shows that we can reduce Theorem 18 to a similar, yet weaker statement compared to Theorem 12.

Remark 19. If we set $T_n := J_{\gamma\partial f}(Id - \nabla g)$ with $f \in \Gamma(\mathcal{H})$, $g : \mathcal{H} \to \mathbb{R}$ convex with a $\frac{1}{\beta}$-Lipschitz continuous gradient, $\sup_{n\in N} \gamma_n \leq 2\beta$, $\alpha = 0$ and assuming $\operatorname{argmin}(f+g) \neq \emptyset$ then we get the same implications as in Theorem 12 except with the stronger assumption on the relaxation factors λ_n, since $\inf_{n\in\mathbb{N}} \lambda_n > 0$ and $\sup_{n\in\mathbb{N}} \gamma_n \leq 1$ implies $\sum_{n\in\mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

By setting $B = 0$ in Corollary 17 we immediately get the following analogous Corollary and Theorem.

Corollary 20. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator and $\inf_{n\in\mathbb{N}} \gamma_n > 0$. Suppose that $\operatorname{Zer}(A) \neq \emptyset$ and set $T_n := J_{\gamma_n A} : \mathcal{H} \to \mathcal{H}$. Then T_n fulfills condition (6).

The last corollary allows us to formulate an inertial Proximal-Point method. Note that in contrary to the classic Forward-Backward algorithm in Theorem 12, the Proximal-Point method (see [1, Theorem 23.41]) is already defined by a variable stepsize.

Theorem 21. ***(Proximal-Point method)*** *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone such that $\operatorname{Zer}(A) \neq \emptyset$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $(0, +\infty)$ such that $\sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty$, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = J_{\gamma_n A} x_n.$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Zer} A$.

By setting $A := \partial f$ in Corollary 20 one can obtain the next algorithm, with a more relaxed condition on the stepsizes γ_n (see [1, Theorem 27.1]).

Theorem 22. ***(Proximal-Point algorithm)*** *Let $f \in \Gamma(\mathcal{H})$ be such that $\operatorname{argmin}(f) \neq \emptyset$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $(0, +\infty)$ such that $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{Prox}_{\gamma_n f} x_n.$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{argmin}(f)$.

As mentioned before, we will now formulate an inertial proximal-point algorithm.

Theorem 23. ***(Inertial Proximal-Point algorithm)*** *Let $f \in \Gamma(\mathcal{H})$, $\inf_{n\in\mathbb{N}} \gamma_n > 0$ and assume that $\operatorname{argmin}(f) \neq \emptyset$. Furthermore we define the following iteration:*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{ll} w_n & := x_n + \alpha_n(x_n - x_{n-1}), \\ x_{n+1} & := (1 - \lambda_n) w_n + \lambda_n \operatorname{Prox}_{\gamma_n f} w_n, \end{array} \right.$$

where x_0, x_1 are arbitrarily chosen in $\mathcal{H}$, $(\alpha_n)_{n\geq 1}$ is nondecreasing with $\alpha_1 = 0$ and $0 \leq \alpha_n \leq \alpha < 1$ for every $n \geq 1$ and $\lambda, \sigma, \delta > 0$ are such that

$$\delta > \frac{\alpha^2(1+\alpha) + \alpha\sigma}{1 - \alpha^2} \text{ and } 0 < \lambda \leq \lambda_n \leq \frac{\delta - \alpha(\alpha(1+\alpha) + \alpha\delta + \sigma)}{\delta(1 + \alpha(1+\alpha) + \alpha\delta + \sigma)} \; \forall n \geq 1.$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{argmin}(f)$.

Proof. Use Theorem 18 with $g = 0$. □

Similarly to Remark 19, the next remark shows that we can reduce Theorem 23 to a similar, yet weaker statement compared to Theorem 21.

Remark 24. If we set $T_n := J_{\gamma_n \partial f}$ with $f \in \Gamma(\mathcal{H})$, $\inf_{n\in\mathbb{N}} \gamma_n > 0$,$\alpha = 0$ and assuming $\operatorname{argmin}(f) \neq \emptyset$ then we get the same statement as in Theorem 22 except with the stronger assumption on the stepsizes γ_n, since

$$\inf_{n\in\mathbb{N}} \gamma_n > 0 \Rightarrow \sum_{n\in\mathbb{N}} \gamma_n = +\infty.$$

4 Applications

In the following we will use the inertial forward-backward algorithm with variable stepsize (Theorem 18) to solve a deblurring problem of the following kind. We are given an observed noisy grayscale image $B \in [0,1]^{256\times 256}$ with

$$B = G * X + \eta,$$

where $*$ denotes the convolution between matrices with respect to Neumann (mirror) boundary conditions, G is a Gauss filter of size 9×9 with standard deviation 4, $X \in [0,1]^{256\times 256}$ is the original grayscale image and η is zero-mean white Gaussian noise with standard deviation 10^{-3}. The original image X and the observed noisy image B can be seen in Figure 2.

Figure 2: Original image (left) and the observed noisy image.

We will solve the following optimization problem:

$$\operatorname*{argmin}_{x\in\mathbb{R}^{256^2}} \frac{1}{2}\,\|RWx - b\|_2^2 + \beta\|x\|_1 \tag{18}$$

where $b = \operatorname{vec}(B) \in [0,1]^{256^2}$ is the vectorization of B (formed by stacking the columns of B into a single column vector b), $\beta > 0$ is a regularization parameter, R is a matrix representing the blur operator and W is a matrix representing the inverse of a three stage Haar wavelet transform. To be more precise, one can write (w.r.t Neumann boundary conditions)

$$G * X = MXM$$

as a product of matrices, with M being a sum of a Hankel and Toeplitz matrix (for further detail see [5]). Furthermore, it holds that

$$\operatorname{vec}(MXM) = (M^T \otimes M)\operatorname{vec}(X) =: R\operatorname{vec}(X)$$

where $\otimes$ denotes the Kronecker product, defined by

$$U \otimes V = \begin{bmatrix} u_{11}V & \dots & u_{1n}V \\ \vdots & \ddots & \vdots \\ u_{m1}V & \dots & u_{mn}V \end{bmatrix} \in \mathbb{R}^{ms \times nt}$$

for matrices $U \in \mathbb{R}^{m \times n}, V \in \mathbb{R}^{s \times t}$. Similarly, we can derive the inverse Haar wavelet transform operator W. In our case, $R, W \in \mathbb{R}^{256^2 \times 256^2}$ are fortunately sparse matrices, thus making the computational effort to compute RWx not too big. It is common to use the l_1-regularization in (18), since we are minimizing over the wavelet domain and most images have a sparse wavelet representation in the wavelet domain.

Now we apply the forward-backward algorithm for $g(x) := \frac{1}{2}\|RWx - b\|_2^2$ and $f(x) := \beta\|x\|_1$ with $\beta \in \{1\text{e-}2, 1\text{e-}3, 1\text{e-}4, 1\text{e-}5\}$ consecutively (see Figure 3). The gradient of g is $\nabla g(x) = W^T R^T RWx - W^T R^T b$ and Lipschitz continuous with Lipschitz constant 1. The proximal operator of f is the shrinkage thresholding operator, i.e.

$$\text{Prox}_{\beta\gamma_n\|\cdot\|_1}(x) = \begin{cases} x - \beta\gamma_n & \text{for } x \geq \beta\gamma_n, \\ 0 & \text{for } -\beta\gamma_n \leq x \leq \beta\gamma_n, \\ x + \beta\gamma_n & \text{for } x \leq -\beta\gamma_n. \end{cases}$$

Furthermore, let F_n denote the objective function value of the n-th iteration. We can see in Figure 3. that the lower the value of β is, the better the image quality is. Moreover, we can see in Figure 4 that we get a better convergence rate if we choose bigger values for the damping terms α_n. In this example, the algorithm will converge very slowly after 50 iterations, which can be seen in Figure 5. [2]

(a) $F_{100} = 42.81$, $\beta =$ 1e-2

(b) $F_{200} = 42.8$, $\beta =$ 1e-2

(c) $F_{100} = 4.8$, $\beta =$ 1e-3

(d) $F_{200} = 4.7$, $\beta =$ 1e-3

(e) $F_{100} = 0.62$, $\beta =$ 1e-4

(f) $F_{200} = 0.58$, $\beta =$ 1e-4

(g) $F_{100} = 0.15$, $\beta =$ 1e-5

(h) $F_{200} = 0.1085$, $\beta =$ 1e-5

Figure 3: Objective function values after the 100-th and 200-th iteration for $\beta \in \{1\text{e-}2, 1\text{e-}3, 1\text{e-}4, 1\text{e-}5\}$ and for the parameters $\lambda_n = 0.92$ and $\alpha_n = 0.05$, $\gamma_n = 2$.

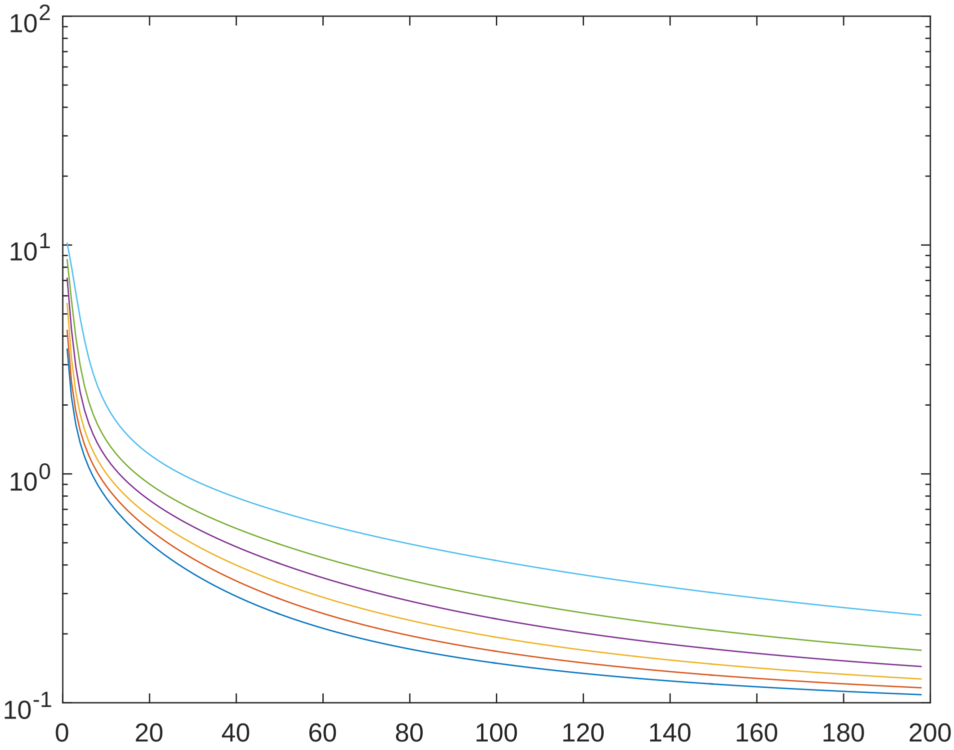

Figure 4: From bottom to top: We do a semilogy plot of the objective function values F_n up to 200 iterations for the parameters $\gamma_n = 2$ and $(\alpha_n = 0.05, \lambda_n = 0.92)$, $(\alpha_n = 0.2, \lambda_n = 0.64)$, $(\alpha_n = 0.3, \lambda_n = 0.46)$, $(\alpha_n = 0.4, \lambda_n = 0.31)$, $(\alpha_n = 0.5, \lambda_n = 0.2)$, $(\alpha_n = 0.6, \lambda_n = 0.1)$ respectively.

Figure 5: From left to right: The observed blurred noisy image B, the solution after 50 iterations, the original image X and the values of the objective function $F_n - F_{10000}$ from the iterations 1 to 200. Here we used the same parameters as in Figure 3 (h).

5 Appendix

The following is a summary of this paper written in german.

Zusammenfassung

In dieser Arbeit betrachten wir eine verallgemeinerte Version von dem Krasnosel'skiĭ-Mann Algorithmus, der bekannt für das Lösen von Fixpunkt Problemen ist. Als erstes stellen wir den klassischen Krasnosel'skiĭ-Mann Algorithmus vor und führen ein paar notwendige Resultate aus der Fixpunkt - und Operator Theorie vor. Des Weiteren beweisen wir die schwache Konvergenz und betrachten einen Spezialfall vom verallgemeinerten KM Algorithmus, der insbesondere ein inertialer Forward-Backward Algorithmus mit variabler Schrittweite ist. Schlussendlich zeigen wir eine Anwendung zum lösen von "image deblurring" Problemen.

References

[1] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer Publishing Company, Incorporated, 1st edition, 2011.

[2] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Img. Sci.*, 2(1):183–202, March 2009.

[3] Radu Ioan Boţ, Ernö Robert Csetnek, and Christopher Hendrich. Inertial douglas–rachford splitting for monotone inclusion problems. *Applied Mathematics and Computation*, 256:472 – 487, 2015.

[4] Antonin Chambolle and Charles Dossal. How to make sure the iterates of FISTA converge. working paper or preprint, September 2014.

[5] Per Christian Hansen, James G. Nagy, and Dianne P. O'Leary. *Deblurring Images: Matrices, Spectra, and Filtering (Fundamentals of Algorithms 3) (Fundamentals of Algorithms)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2006.

[6] Paul-Emile Maingé. Convergence theorems for inertial km-type algorithms. *Journal of Computational and Applied Mathematics*, 219(1):223 – 236, 2008.

[7] George J. Minty. Monotone (nonlinear) operators in hilbert space. *Duke Math. J.*, 29(3):341–346, 09 1962.

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