MASTERARBEIT / MASTER THESIS

Titel der Masterarbeit / Title of the Master Thesis

„Deriving primal-dual splitting methods
for nonsmooth optimization problems: theory and applications“

verfasst von / submitted by

Guillermo D’Esposito, BSc

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

Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von / Supervisor:	Dr. Ernö Robert Csetnek, Privatdoz. MSc

Abstract

In this thesis we extend some theoretical results about the forward-reflected-backward splitting method and two of its variants, namely, the relaxed-inertial-forward-reflected-backward splitting and the three operator splitting presented in [12]. These methods are intended for solving monotone inclusions problems requiring only Lipschitz continuity of the single-valued operator.
After an introduction to monotone operator theory and convex analysis, we show the linear convergence of the forward-reflected-backward splitting method with variable step-size, the linear convergence of the relaxed-intertial-forward-reflected-backward splitting method as well as the linear convergence of the three operator splitting method.
We then derive methods to composite inclusion problems using a well known product space technique and show an application to a general structured non-smooth convex minimization problem. Lastly we provide numerical experiments comparing the above methods to a variant of the forward-backward method proposed in [14] and the error-free version of the forward-backward-forward method proposed in [10]. The numerical tests were made on a system with Intel i5-7400 (4) 3.5 GHz and the python code can be found on the attached USB flash drive.

Zusammenfassung

Diese Arbeit beinhaltet verschiedene erweiterte Konvergenzresultate des “forward-reflected-backward splitting” Verfahrens und zwei seiner Varianten, nämlich das “relaxed-intertial-forward-reflected-backward splitting” und das “three operator splitting”, welche in [12] präsentiert wurden.
Diese Verfahren sind für die Lösung von “monotone inclusion problems” gedacht und erfordern nur Lipschitzstetigkeit des einwertigen Operators.
Nach einer Einführung in die Theorie monotoner Operatoren und konvexer Analysis, wird die lineare Konvergenz mit variabler Schrittweite des “forward-reflected-backward” Verfahrens, die lineare Konvergenz des “relaxed-intertial-forward-reflected-backward” Verfahrens sowie die lineare Konvergenz des “three operator” Verfahrens gezeigt. Danach werden Verfahren für “composite inclusion problems” abgeleitet und eine Anwendung zu einem allgemeinen konvexen Optimierungsproblem gezeigt.
Abschließend werden die o.a. Verfahren mit einer Variante des “forward-backward” Verfahrens aus [14] und mit der error-free Version des “forward-backward-forward” Ver-

fahrens aus [10] in numerichen Experimenten verglichen, welche mit einem Desktop-Computer mit Intel i5-7400 (4) 3.5 GHz gemacht wurden und die gebundenen Exemplare enthalten einen USB-Stick mit dem Python-Code.

Contents

1 Introduction

Throughout this master's thesis we consider *real Hilbert spaces* denoted by $\mathcal{H}, \mathcal{K}, \mathcal{G}, \mathcal{G}_i$. For a Hilbert space $\mathcal{H}$, its *inner product* is denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and its *associated norm* by $\|\cdot\|_{\mathcal{H}} = \sqrt{\langle \cdot, \cdot \rangle_{\mathcal{H}}}$. We will avoid subscripts when there is no risk of confusion.

For $\mathcal{G}_i$, $i = 1, \ldots, m$, real Hilbert spaces, we denote by

$$\mathcal{G} = \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_m,$$

their Hilbert direct sum endowed with inner product and associated norm defined as follows:

$$\langle v, w \rangle_{\mathcal{G}} := \sum_{i=1}^{m} \langle v_i, w_i \rangle_{\mathcal{G}_i} \qquad \text{and} \qquad \|v\|_{\mathcal{G}} := \sqrt{\sum_{i=1}^{m} \|v_i\|_{\mathcal{G}_i}^2},$$

where $v = (v_1, \ldots, v_m), w = (w_1, \ldots, w_m) \in \mathcal{G}$.

Several times we will use the *parallelogram law*:

$$2 \langle a_1 - a_2, b_1 - b_2 \rangle = \|a_1 - b_2\|^2 - \|a_1 - b_1\|^2 + \|a_2 - b_1\|^2 - \|a_2 - b_2\|^2, \tag{1.1}$$

as well as the *Cauchy-Schwarz* inequality (CS) $|\langle x, y \rangle| \leqslant \|x\|\|y\|$ in combination with the *Cauchy inequality*:

$$(\forall\, a, b \in \mathbb{R}),\ (\forall \delta > 0) \qquad 2ab \leqslant \delta a^2 + \frac{b^2}{\delta}. \tag{1.2}$$

By $\mathscr{B}(\mathcal{H}, \mathcal{G})$ we denote the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$ and by $\mathcal{H}^*$ the dual space of $\mathcal{H}$. If $L \in \mathscr{B}(\mathcal{H}, \mathcal{G})$, we denote by L^* its adjoint, i.e., the unique operator $L^* \in \mathscr{B}(\mathcal{G}, \mathcal{H})$, that satisfies

$$(\forall x \in \mathcal{H}),\ (\forall v \in \mathcal{G}) \quad \langle Lx, v \rangle_{\mathcal{G}} = \langle x, L^* v \rangle_{\mathcal{H}}.$$

By $\mathbb{N}$ we denote the set of *natural numbers* $\{0, 1, 2, \dots\}$ and by $\mathbb{N}_1 = \mathbb{N}\backslash\{0\}$ the set of natural numbers without 0. By $\mathbb{R}$ we denote the set of *real numbers*, by $\mathbb{R}_+$ the set of nonnegative real numbers and by $\mathbb{R}_{++}$ the set of strictly positive real numbers. We also denote the *extended real line* by $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$.

For two nonempty subsets C and D in $\mathcal{H}$, their *Minkowski sum* is defined as

$$C + D := \{c + d \,|\, c \in C \text{ and } d \in D\},$$

and for every $\lambda \in \mathbb{R}$ the λ-scaled set λC is defined as

$$\lambda C := \{\lambda c \,|\, c \in C\}.$$

If Λ is a nonempty subset of $\mathbb{R}$ then

$$\Lambda C := \bigcup_{\lambda \in \Lambda} \lambda C.$$

A subset C of $\mathcal{H}$ is called:

- ***convex*** if $(\forall x, y \in C)$ and $\lambda \in [0,1]$

$$\lambda x + (1-\lambda) y \in C.$$

- a ***cone*** if $C = \mathbb{R}_{++} C$.

Be aware that in the literature a cone is sometimes defined as $C = R_+ C$. Anyway, in the context of this master's thesis, there is no difference, since we only use the notion of cone in definition 1.5, where 0 belongs to the set.

Definition 1.1. Let C be a subset of $\mathcal{H}$.

- The smallest linear subspace of $\mathcal{H}$ containing C is called the ***span*** of C

$$\operatorname{span} C := \bigcap \{D \subseteq \mathcal{H} \,|\, C \subseteq D \text{ and } D \text{ is a linear subspace}\}.$$

- The smallest cone in $\mathcal{H}$ containing C is called the ***conical hull*** of C

$$\operatorname{cone} C := \bigcap \{D \subseteq \mathcal{H} \,|\, C \subseteq D \text{ and } D \text{ is a cone}\}.$$

For $x \in \mathcal{H}$ and $\lambda \in \mathbb{R}_{++}$, the ***open ball*** centered at x with radius λ is denoted by

$$B_{\mathcal{H}}(x; \lambda) := \{y \in \mathcal{H} \,|\, \|y - x\| < \lambda\}.$$

For a subset C of $\mathcal{H}$, its ***interior*** is denoted by

$$\operatorname{int} C := \{x \in C \,|\, \exists \lambda \in \mathbb{R}_{++} \text{ such that } B_{\mathcal{H}}(x; \lambda) \subset C\},$$

and its ***topological closure*** by

$$\overline{C} := \bigcap \{F \subseteq \mathcal{H} \,|\, C \subseteq F \text{ and } F \text{ is a closed set}\}.$$

Definition 1.2. A sequence $\{x_n\}_{n \in \mathbb{N}}$ in $\mathcal{H}$ is said to converges to x^*

- ***weakly***:

$$x_n \rightharpoonup x^* :\Leftrightarrow (\forall y \in \mathcal{H}) \lim_{n \to \infty} \langle x_n - x^*, y \rangle = 0.$$

- ***strongly:***
$$x_n \to x^* :\Leftrightarrow \lim_{n\to\infty} \|x_n - x^*\| = 0.$$

- **Q-linearly:** if there is a constant $r \in (0,1)$ such that
$$\lim_{n\to\infty} \frac{\|x_{n+1} - x^*\|}{\|x_n - x^*\|} \leqslant r.$$

- **R-linearly:** if there is a sequece $\{\varepsilon_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}$ that converges Q-linearly to 0 such that:
$$\|x_n - x^*\| \leqslant \varepsilon_n, \qquad \forall n \in \mathbb{N}.$$

We need two results regarding convergence, the first one is often referred as ***Opial's Lemma***:

Lemma 1.3. Let $\{x_n\}_{n\in\mathbb{N}}$ be a bounded sequence in $\mathcal{H}$. Then there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ that converges weakly.

Proof. See [1],Lemma 2.45. □

Lemma 1.4. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and let Z be a nonempty subset of $\mathcal{H}$. Suppose that for every $x \in Z, \{\|x_n - x\|\}_{n\in\mathbb{N}}$ converges and that every weak sequential cluster point of $\{x_n\}_{n\in\mathbb{N}}$ belongs to Z. Then $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Proof. See [1], Lemma 2.47. □

A nice result involving weak and strong convergence is given in ([1], Corollary 2.52). Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and let $\bar{x} \in \mathcal{H}$. Then

$$x_n \to \bar{x} \Leftrightarrow x_n \rightharpoonup \bar{x} \text{ and } \|x_n\| \to \|\bar{x}\|.$$

The following definitions are weaker notions of interiority:

Definition 1.5. Let $C \subseteq \mathcal{H}$, be convex.

- The ***core*** of C is
$$\operatorname{core} C := \{x \in C \mid \operatorname{cone}(C - x) = \mathcal{H}\}.$$

- The ***strong relative interior*** of C is
$$\operatorname{sri} C := \{x \in C \mid \operatorname{cone}(C - x) = \overline{\operatorname{span}}(C - x)\}.$$

- The ***relative interior*** of C is
$$\operatorname{ri} C := \{x \in C \mid \operatorname{cone}(C - x) = \operatorname{span}(C - x)\}.$$

Since $\operatorname{cone} C \subseteq \operatorname{span} C \subseteq \overline{\operatorname{span}}\, C \subseteq \mathcal{H}$ for every convex set C in $\mathcal{H}$, it follows that

$$\operatorname{int} C \subseteq \operatorname{core} C \subseteq \operatorname{sri} C \subseteq \operatorname{ri} C \subseteq C.$$

The above inclusions can be strict and examples can be found in ([1], Chapter 6).

Proposition 1.6. *Let $\mathcal{H}$ be a real Hilbert space, let $\mathcal{G}$ and $\mathcal{K}$ be real finite-dimensinal Hilbert spaces, let $L_1 \in \mathscr{B}(\mathcal{H}, \mathcal{G})$, let $L_2 \in \mathscr{B}(\mathcal{G}, \mathcal{K})$ and let C, D and E be nonempty convex subsets of $\mathcal{H}$, $\mathcal{G}$ and $\mathcal{K}$ respectively. Then the following hold:*

(i) $\operatorname{sri} E = \operatorname{ri} E$.

(ii) $\operatorname{ri} L_2(D) = L_2(\operatorname{ri} D)$.

(iii) *If* $\operatorname{ri} D \cap \operatorname{ri} L_1(C) \neq \varnothing$

$$\textit{then } 0 \in \operatorname{sri}\,(D - L_1(C)). \tag{1.3}$$

Proof. See [1], Corollary 6.15 and Proposition 6.19. □

1.1 Monotone Operator Theory

Let $\mathcal{X}$ and $\mathcal{Y}$ be two nonempty sets. We denote by $2^{\mathcal{X}}$ the power set of $\mathcal{X}$ and by $A : \mathcal{X} \to 2^{\mathcal{Y}}$ a set-valued operator A from $\mathcal{X}$ to $\mathcal{Y}$. The ***domain*** of A is denoted by

$$\operatorname{dom} A := \{x \in \mathcal{X} \mid Ax \neq \varnothing\},$$

its ***range*** by

$$\operatorname{ran} A := \{Ax \subseteq \mathcal{Y} \mid x \in \operatorname{dom} A\},$$

its ***graph*** by

$$\operatorname{gra} A := \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid y \in Ax\},$$

and the **inverse** of A, denoted by A^{-1}, is defined through its graph:

$$\operatorname{gra} A^{-1} = \{(y, x) \in \mathcal{Y} \times \mathcal{X} \mid (x, y) \in \operatorname{gra} A\}.$$

If $\mathcal{Y}$ is a vector space, we denote the ***set of zeros*** of A by

$$\operatorname{zer} A = A^{-1}(0) = \{x \in \mathcal{X} \mid 0 \in Ax\}.$$

Definition 1.7. Let $\mathcal{H}$ be a real Hilbert space. An operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is said to be:

- ***Monotone*** if
$$(\forall\, (x, u), (y, v) \in \operatorname{gra} A)\ \ \langle x - y, u - v \rangle_{\mathcal{H}} \geqslant 0. \tag{1.4}$$

- ***Maximally monotone*** if A is monotone and $\forall (x, u) \in \mathcal{H} \times \mathcal{H}$
$$(x, u) \in \operatorname{gra} A \Leftrightarrow (\forall\, (y, v) \in \operatorname{gra} A)\ \ \langle x - y, u - v \rangle_{\mathcal{H}} \geqslant 0. \tag{1.5}$$

- ***Strongly monotone with constant*** $\rho \in \mathbb{R}_{++}$ if

$$(\forall\,(x,u),(y,v) \in \operatorname{gra} A)\ \langle\, x-y, u-v\,\rangle_{\mathcal{H}} \geqslant \rho\|x-y\|^2. \tag{1.6}$$

Remark 1.8. A single-valued operator T from $\mathcal{H}$ to $\mathcal{G}$, with $\operatorname{dom} T = D \subseteq \mathcal{H}$ will be denoted by $T : D \to \mathcal{G}$. Note that $T : \mathcal{H} \to \mathcal{G}$, means that T has full domain.

From the definition follows that the inverse A^{-1} of a maximally monotone operator A, is maximally monotone as well, but the sum of maximally monotone operators is not necessarily maximally monotone and we need extra assumptions:

Proposition 1.9. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $B : \mathcal{H} \to \mathcal{H}$ be monotone and continuous. Then $A + B$ is maximallity monotone.*

Proof. By ([1], Corollary 20.28) follows that B is maximally monotone, and by ([1], Corollary 25.5) that $A + B$ is maximally monotone since $\operatorname{dom} B = \mathcal{H}$. □

Proposition 1.10. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $z, u \in \mathcal{H}$ and let $\lambda \in \mathbb{R}_{++}$. Then*

$$x \mapsto u + \lambda A(x + z)$$

is maximally monotone.

Proof. See [1], Proposition 20.22. □

Proposition 1.11. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\{(x_n, u_n)\}_{n\in\mathbb{N}}$ be sequence in $\operatorname{gra} A$ and let $(x, u) \in \mathcal{H} \times \mathcal{H}$. If one of the following conditions holds:*

(i) $x_n \to x$ *and* $u_n \rightharpoonup u$,

(ii) $x_n \rightharpoonup x$ *and* $u_n \to u$.

Then $(x, u) \in \operatorname{gra} A$.

Proof. See [1], Proposition 20.38. □

Definition 1.12. Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be an operator and let $\gamma \in \mathbb{R}_{++}$. The ***resolvent*** of γA, denoted by $\mathrm{J}_{\gamma A}$, is defined as:

$$\mathrm{J}_{\gamma A} : \mathcal{H} \to 2^{\mathcal{H}} \quad : \quad x \mapsto (\mathrm{Id} + \gamma A)^{-1}(x).$$

From the definition of the resolvent it is clear that:

$$\operatorname{dom} \mathrm{J}_{\gamma A} = \operatorname{ran}(\mathrm{Id} + \gamma A), \qquad \text{and} \qquad \operatorname{ran} \mathrm{J}_{\gamma A} = \operatorname{dom} A.$$

In order to characterize the resolvent, we need further definitions:

Definition 1.13. Let D be a non-empty subset of $\mathcal{H}$ and let $\beta \in \mathbb{R}_{++}$. An operator $T : D \to \mathcal{H}$ is called:

- ***Lipschitz continuous with constant*** β if

$$(\forall x, y \in D) \quad \|Tx - Ty\| \leqslant \beta\|x - y\|. \tag{1.7}$$

- ***Nonexpansive*** if it is Lipschitz continuous with constant 1, i.e.,

$$(\forall x, y \in D) \quad \|Tx - Ty\| \leqslant \|x - y\|. \tag{1.8}$$

- ***Firmly nonexpansive*** if

$$(\forall x, y \in D) \quad \|Tx - Ty\|^2 \leqslant \langle\, x - y, Tx - Ty\,\rangle\,. \tag{1.9}$$

- β-***cocoercive*** if βT is firmly nonexpansive, i.e.,

$$(\forall x, y \in D) \quad \beta\|Tx - Ty\|^2 \leqslant \langle\, x - y, Tx - Ty\,\rangle\,. \tag{1.10}$$

Proposition 1.14. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be an operator with* $\operatorname{dom} A \neq \varnothing$*, let* $D = \operatorname{ran}(\mathrm{Id} + A)$*, and set $T = \mathrm{J}_A|_D$. Then the following hold:*

(i) A is monotone $\Leftrightarrow$ T is firmly nonexpansive.

(ii) A is maximally monotone $\Leftrightarrow$ $(\forall \gamma \in \mathbb{R}_{++})$ $\mathrm{J}_{\gamma A}$ is firmly nonexpansive and $D = \mathcal{H}$.

Proof. See [1], Proposition 23.10. □

Remark 1.15. Note that this proposition implies that the resolvent J_A of a monotone operator A is single-valued.

Proposition 1.16 (Resolvent Calculus)**.** *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $x, z \in \mathcal{H}$ and let $\gamma \in \mathbb{R}_{++}$. Then the following hold:*

(i) Let $B = A + \gamma\mathrm{Id}$. Then

$$\mathrm{J}_B(x) = \mathrm{J}_{(1+\gamma)^{-1}A}\left(\frac{x}{1+\gamma}\right). \tag{1.11}$$

(ii) Let $Bx = Ax + z$. Then

$$\mathrm{J}_B(x) = \mathrm{J}_A(x - z). \tag{1.12}$$

(iii) Let $Bx = A(x - z)$. Then

$$\mathrm{J}_B(x) = z + \mathrm{J}_A(x - z). \tag{1.13}$$

(iv)

$$\mathrm{Id} = J_{\gamma A} + \gamma\, \mathrm{J}_{\gamma^{-1}A^{-1}} \circ \gamma^{-1}\mathrm{Id}\,. \tag{1.14}$$

(v) Let $B(x) = \rho A(\rho x)$. Then

$$\mathrm{J}_B(x) = \frac{1}{\rho}\,\mathrm{J}_{\rho^2 A}(\rho x), \quad \forall \rho \neq 0. \tag{1.15}$$

Proof. (i), (ii), (iii) See [1],Proposition 23.17.
(iv) See [1], Proposition 23.20.
(v) See [1], Corollary 23.26. □

Proposition 1.17. *Let $A_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be maximally monotone for each $i \in \{1, \dots, m\}$, and let*

$$M : \mathcal{G} := \bigoplus_{i \in I} \mathcal{G}_i \to 2^{\mathcal{G}} \quad : \quad (x_1, \dots, x_m) \mapsto (A_1 x_1, \dots, A_m x_m). \tag{1.16}$$

Then M is maximally monotone and

$$\mathrm{J}_M : \mathcal{G} \to \mathcal{G} \quad : \quad (x_1, \dots, x_m) \mapsto (\mathrm{J}_{A_1} x_1, \dots, \mathrm{J}_{A_m} x_m). \tag{1.17}$$

Proof. See [1], Proposition 23.18. □

The next proposition will be the main tool for proving linear convergence of the algorithms in the next section:

Proposition 1.18. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be monotone, let $D = \mathrm{ran}\,(\mathrm{Id} + A)$ and let $\rho \in \mathbb{R}_{++}$. Then*

$$A \text{ is } \rho\text{-strongly monotone} \Leftrightarrow \mathrm{J}_{A|_D} \text{ is } (1+\rho)\text{-cocoervice.} \tag{1.18}$$

Proof. See [1],Proposition 23.13. □

Many problems of nonlinear analysis can be reduced to solving a monotone inclusion problem, i.e.

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Mx$$

where $M : \mathcal{H} \to 2^{\mathcal{H}}$ is a monotone operator.

When M is maximally monotone, the above problem could "theoretically" be solved by **proximal-point algorithm**, (ref. [1], Theorem 23.41) which asserts that:

$$x_{n+1} = \mathrm{J}_{\gamma_n M}(x_n) \rightharpoonup \bar{x} \in \mathrm{zer}\, M,$$

whenever $\gamma_n \in \mathbb{R}_{++}$ with $\sum_{n \in \mathbb{N}} \gamma_n^2 = +\infty$, and $\mathrm{zer}\, M \neq \varnothing$ for $x_0 \in \mathcal{H}$ arbitrary. Moreover, Proposition 1.10 implies that the trivial substitution

$$M'(x) := M(x) - z,$$

allows us to solve any problem of the form:

$$\text{find } x \in \mathcal{H} \text{ such that } z \in Mx,$$

since M' is maximally monotone and its resolvent can be computed using (1.12).

When the operator M represents the sum of two operators, the formula to compute its resolvent is not useful in practise (see [1], Corollary 25.34) and the **"splitting methods"** arise in such a situation, providing an algorithm which acts on each operator separately. Let's recall a simplified version of three of them, assuming that $\operatorname{zer}(A+B) \neq \varnothing$ and $x_0 \in \mathcal{H}$:

- **Douglas-Rachford** (ref. [1], Theorem 26.11)
Assumptions: $A, B : \mathcal{H} \to 2^{\mathcal{H}}$ are both maximally monotone.
Step: $\gamma \in \mathbb{R}_{++}$.
Update rule:

$$\left\lfloor \begin{aligned} y_n &= \mathrm{J}_{\gamma B}(x_n) \\ x_{n+1} &= \mathrm{J}_{\gamma A}(2y_n - x_n) \end{aligned} \right.$$

Assertions:

(i) $y_n \rightharpoonup \bar{x} \in \operatorname{zer}(A+B)$,

(ii) If A or B is uniformly monotone then $y_n \to \bar{x}$, where $\bar{x}$ is the unique solution.

Pros: It solves the problem when A and B are set valued operators.
Cons: The computation of the resolvent of both operators is required.

- **Forward-backward** (ref. [1], Theorem 26.14)
Assumptions: $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone, $B : \mathcal{H} \to \mathcal{H}$ is β-cocoercive with $\beta \in \mathbb{R}_{++}$.
Step: $\gamma \in (0, 2\beta)$
Update rule:

$$\left\lfloor \begin{aligned} y_n &= x_n - \gamma B x_n \\ x_{n+1} &= \mathrm{J}_{\gamma A}(y_n) \end{aligned} \right.$$

Assertion:

(i) $x_n \rightharpoonup \bar{x} \in \operatorname{zer}(A+B)$,

(ii) If A or B is uniformly monotone then $x_n \to \bar{x}$, where $\bar{x}$ is the unique solution.

Pros: It requires the computation of only one resolvent.
Cons: It is not suitable when A and B are set-valued operators.

- **Forward-backward-forward** (Tseng) (ref. [1], Theorem 26.17)
Assumptions: $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone, $B : \mathcal{H} \to \mathcal{H}$ is monotone and β-Lipschitz continuous with $\beta \in \mathbb{R}_{++}$.
Step: $\gamma \in \left(0, \frac{1}{\beta}\right)$

Update rule:

$$\left\lfloor \begin{aligned} y_n &= x_n - \gamma B x_n \\ p_n &= \mathrm{J}_{\gamma A}(y_n) \\ x_{n+1} &= p_n + \gamma(Bx_n - Bp_n) \end{aligned} \right.$$

Assertion:

(i) $x_n \rightharpoonup \bar{x} \in \operatorname{zer}(A+B)$,

(ii) If A or B is uniformly monotone at $\bar{x}$ then $x_n \to \bar{x}$.
Pros: It requires Lypschitz continuity of the operator B which is less restrictive than cocoercivity.
Cons: It is not suitable when A and B are set-valued operators.

The **forward-reflected-backward (FRB)** method, proposed by Malitsky-Tam [12], which is the motivation of this master's thesis, requires only Lipschitz continuity of the operator B as in the case of FBF.

1.2 Convex Analysis

Definition 1.19. Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be a function.

- The ***effective domain*** of f is

$$\operatorname{dom} f := \{x \in \mathcal{H} \mid f(x) < +\infty\}.$$

- f is called ***proper*** if

$$-\infty \notin f(\mathcal{H}) \text{ and } \operatorname{dom} f \neq \varnothing.$$

- f is ***convex*** if $\forall \lambda \in (0,1)$, and $(\forall x, y \in \operatorname{dom} f)$

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y).$$

- f is ***lower-semicontiunous*** (lsc) at x if

$$f(x) \leqslant \liminf_{y \to x} f(y) = \sup_{\varepsilon > 0} \inf_{y \in B(x,\varepsilon)} f(y).$$

The set of proper lower semicontinuous convex functions from $\mathcal{H}$ to $\overline{\mathbb{R}}$ is denoted by $\Gamma(\mathcal{H})$.

Definition 1.20. Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be a proper function, let $x \in \operatorname{dom} f$ and let $y \in \mathcal{H}$.

- The ***directional derivative of*** f at x in the direction y is defined as

$$f'(x; y) := \lim_{\alpha \downarrow 0} \frac{f(x+\alpha y) - f(x)}{\alpha},$$

whenever this limit exists in $\overline{\mathbb{R}}$.

- If $f'(x,\cdot)$ is linear and bounded, then f is said to be ***Gâteaux differentiable*** at x and the ***Gâteaux gradient*** of f at x is the unique vector $\nabla f(x) \in \mathcal{H}$ (its existence is guaranteed by the Riesz representation theorem) such that

$$f'(x,\cdot) = \langle\, \cdot, \nabla f(x)\,\rangle\,. \tag{1.19}$$

- f is called ***Fréchet differentiable*** at $x \in \operatorname{int}(\operatorname{dom} f)$ if there exists a bounded linear functional in $\mathcal{H}^* \simeq \mathcal{H}$ (Riesz), called the ***Fréchet gradient*** of f at x, also denoted by $\nabla f(x)$, such that

$$\lim_{y\to 0} \frac{f(x+y) - f(x) - \langle\, y, \nabla f(x)\,\rangle}{\|y\|} = 0. \tag{1.20}$$

Remark 1.21. Let $f \in \Gamma(\mathcal{H})$ and let $x \in \operatorname{dom} f$. Suppose that f is Gâteaux differentiable on $B(x;\varepsilon)$ for some $\varepsilon \in \mathbb{R}_{++}$. Then f is Fréchet differentiable at x if and only if ∇f is continuous at x. Moreover, if a convex function is Gâteaux differentiable on its domain, then its gradient is monotone.

Proof. See [1], Corollary 17.42 & Proposition 17.7. □

Definition 1.22. Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be a proper function. The ***subdifferential*** of f is the set-valued operator

$$\partial f : \mathcal{H} \to 2^{\mathcal{H}} \quad : \quad x \mapsto \{u \in \mathcal{H} \,|\, (\forall y \in \mathcal{H})\ \langle\, y - x, u\,\rangle + f(x) \leqslant f(y)\}. \tag{1.21}$$

When a function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is proper, convex and Gâteaux differentiable at $x \in \operatorname{dom} f$, then

$$\partial f(x) = \{\nabla f(x)\}, \tag{1.22}$$

(see [1], Proposition 17.31). Moreover, is not difficult to see, that for poper functions $f : \mathcal{H} \to \overline{\mathbb{R}}$, $g : \mathcal{G} \to \overline{\mathbb{R}}$, and $L \in \mathscr{B}(\mathcal{H},\mathcal{G})$, such that $\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing$, the following inclusion holds:

$$\partial f(x) + (L^* \circ (\partial g) \circ L)\,(x) \subseteq \partial(f + g \circ L)(x) \quad \forall x \in \mathcal{H}. \tag{1.23}$$

Definition 1.23. Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be a function. The ***Fenchel conjugate*** f^* of f is defined as

$$f^* : \mathcal{H} \to \overline{\mathbb{R}} \quad : \quad u \mapsto \sup_{x\in\mathcal{H}}\{\langle\, x, u\,\rangle - f(x)\}. \tag{1.24}$$

Example 1.24. Let $C \subseteq \mathcal{H}$ and $f(x) = \iota_C(x)$ be the **indicator function** of C, defined as:

$$\iota_C(x) := \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

Then

$$f^*(u) = \sup_{x\in\mathcal{H}}\{\langle\, x, u\,\rangle - \iota_C(x)\} = \sup_{x\in C}\{\langle\, x, u\,\rangle\} = \sigma_C(u),$$

is the **support function** of C, which is usually denoted by σ_C.

Example 1.25. Continuing with the last example, let $C = \overline{B}_{\mathcal{H}}(0;\lambda)$ with $\lambda \in \mathbb{R}_{++}$. Then

$$\left(\iota_{\overline{B}_{\mathcal{H}}(0;\lambda)}\right)^*(u) = \sup_{\|x\|\leqslant\lambda} \{\langle x, u \rangle_{\mathcal{H}}\} = \lambda\|u\|_{\mathcal{H}}. \tag{1.25}$$

Proposition 1.26. *Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be a proper function. Then the following hold:*

(i) *f^* is lower-semicontinuous and convex.*

(ii) *$\forall\, x, u \in \mathcal{H}$:*

$$f(x) + f^*(u) \geqslant \langle x, u \rangle, \tag{1.26}$$

with equality if and only if $u \in \partial f(x)$.

(iii) *$f^{**} \leqslant f$, and*

$$f^{**} = f \Leftrightarrow f \in \Gamma(\mathcal{H}). \tag{1.27}$$

(iv)

$$u \in \partial f(x) \Rightarrow x \in \partial f^*(u). \tag{1.28}$$

Proof. (i) See [1], Proposition 13.13.

(ii) See [1], Proposition 13.15. (***Young-Fenchel*** inequality).

(iii) See [1], Theorem 13.37. (***Fenchel-Moreau*** theorem).

(iv) It is a consequence of (ii) and (iii).

□

Corollary 1.27.

$$\text{If } f \in \Gamma(\mathcal{H}) \Rightarrow (\partial f)^{-1} = \partial f^*. \tag{1.29}$$

Example 1.28. Since $C := \overline{B}_{\mathcal{H}}(0, \lambda)$ is a closed convex set, it follows that $\iota_C \in \Gamma(\mathcal{H})$, and the Fenchel-Moreau Theorem implies:

$$(\lambda\|\cdot\|_{\mathcal{H}})^* \underset{(1.25)}{=} \left(\iota_{\overline{B}_{\mathcal{H}}(0;\lambda)}\right)^{**} \underset{(1.27)}{=} \iota_{\overline{B}_{\mathcal{H}}(0;\lambda)}. \tag{1.30}$$

Proposition 1.29. *For each $i \in \{1, \ldots, m\}$, let $y_i \in \mathcal{H}_i$ and let $f_i : \mathcal{H}_i \to \overline{\mathbb{R}}$ be a proper function. Then the following hold:*

$$\text{(i)} \quad f^* : \bigoplus_{i=1}^m \mathcal{H}_i \to \overline{\mathbb{R}} \quad : \quad (u_1, \ldots, u_m) \mapsto \sum_{i=1}^m f_i^*(u_i) + \langle y_i, u_i \rangle, \tag{1.31}$$

$$\text{(ii)} \quad \partial f : \bigoplus_{i=1}^m \mathcal{H}_i \to \bigtimes_{i=1}^m 2^{\mathcal{H}_i} \quad : \quad (x_1, \ldots, x_m) \mapsto \bigtimes_{i=1}^m \partial f_i(x_i - y_i). \tag{1.32}$$

where $f := \bigoplus_{i=1}^m f_i : \bigoplus_{i=1}^m \mathcal{H}_i \to \overline{\mathbb{R}} : \quad (x_1, \ldots, x_m) \mapsto \sum_{i=1}^m f_i(x_i - y_i)$.

Proof. (i) Combine ([1], Proposition 13.30) and ([1], Proposition 13.23).

(ii) See [1], Proposition 16.9.

□

Example 1.30. The last proposition provides a great flexibility. Let $f_i = \lambda_i \| \cdot \|_{\mathcal{H}_i}$ for $i = 1, \dots, m$ and $\lambda_i \in \mathbb{R}_{++}$. Define

$$f : \bigtimes_{i=1}^{m} \mathcal{H}_i \to \mathbb{R} \quad : \quad x \mapsto \sum_{i=1}^{m} f_i(x_i). \tag{1.33}$$

Then

$$f^* = \sum_{i=1}^{m} f_i^* = \sum_{i=1}^{m} \iota_{\overline{B}_{\mathcal{H}_i}(0,\lambda_i)} = \iota_C, \tag{1.34}$$

where

$$C = \bigtimes_{i=1}^{m} \overline{B}_{\mathcal{H}_i}(0; \lambda_i).$$

Now, let $\mathcal{H} = \mathbb{R}^{m \times n}$, with $\langle A, B \rangle_{\mathcal{H}} = \operatorname{tr}(A^t B)$. Set

$$f : \mathcal{H} \times \mathcal{H} \to \mathbb{R} \quad : \quad (A, B) \mapsto \sum_{i=1}^{m} \sum_{j=1}^{n} \|(A_{ij}, B_{ij})\|_{\mathbb{R}^2} \tag{1.35}$$

It is easy to verify that f is a norm in $\mathcal{H} \times \mathcal{H}$ and that $\mathcal{H} \times \mathcal{H} = \bigtimes_{k=1}^{mn} \mathbb{R}^2$. Therefore, calling $\|(\cdot,\cdot)\|_\times := f$ and letting $\lambda \in \mathbb{R}_{++}$, we obtain

$$(\lambda \|(\cdot,\cdot)\|_\times)^* \underset{\substack{(1.30)\\(1.31)}}{=} \iota_C(\cdot,\cdot) \tag{1.36}$$

where

$$C = \left[\, \overline{B}_{\mathbb{R}^2}(0; \lambda)\right]^{m \times n},$$

and $\| \cdot \|_{\mathbb{R}^2}$ denotes the standard euclidean norm in $\mathbb{R}^2$.

The set of minimizers of a function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is denoted by

$$\operatorname{Argmin} f := \{x \in \mathcal{H} \mid f(x) = \inf f(\mathcal{H})\}.$$

Lemma 1.31. [Fermat's rule] Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be a proper function. Then

$$\operatorname{Argmin} f = \operatorname{zer} \partial f = \{x \in \mathcal{H} \mid 0 \in \partial f(x)\}. \tag{1.37}$$

If $f \in \Gamma(\mathcal{H})$ then

$$\operatorname{Argmin} f = \partial f^*(0). \tag{1.38}$$

Proof. This follows directly by the definition of the subdifferential and Corollary 1.27. □

Definition 1.32. Let $f, g : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be functions. The ***infimal convolution*** of f and g is defined as

$$f \,\square\, g : \mathcal{H} \to \overline{\mathbb{R}} \quad : \quad x \mapsto \inf_{y \in \mathcal{H}} \{f(y) + g(x - y)\}. \tag{1.39}$$

The infimal convolution is called ***exact*** at x whenever the infimun in (1.39) attained. If the infimal convolution is exact for all $x \in \mathcal{H}$, it is denoted by $\boxdot$.

Proposition 1.33. *Let $f, g : \mathcal{H} \to \overline{\mathbb{R}}$ be proper functions. Then the following hold:*

(i) $(f \,\square\, g)^* = f^* + g^*$. (1.40)

(ii) $(f + g)^* = f^* \boxdot g^* \in \Gamma(\mathcal{H})$, (1.41)

if, for instance, $f, g \in \Gamma(\mathcal{H})$ and $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$.

Proof. (i) See [1], Proposition 13.24.

(ii) See [1], Theorem 15.3.

□

Remark 1.34. A list of conditions satisfying $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$ can be found in [1], Proposition 15.5. Moreover, for proper functions f, g, the inequality $(f + g)^* \leqslant f^* \,\square\, g^*$ is always true.

Proposition 1.35. *Let $f \in \Gamma(\mathcal{H})$, let $g \in \Gamma(\mathcal{G})$, and let $L \in \mathscr{B}(\mathcal{H}, \mathcal{G})$ such that*

$$0 \in \operatorname{sri}\left(\operatorname{dom} g - L(\operatorname{dom} f)\right).$$

Then the following hold:

(i) $\partial(f + g \circ L) = \partial f + L^* \circ \partial g \circ L$. (1.42)

(ii) $\inf_{x \in \mathcal{H}} f(x) + g(Lx) = -\min_{v \in \mathcal{G}} f^*(-L^* v) + g^*(v)$. (1.43)

Proof. (i) See [1], Theorem 16.47.

(ii) See [1], Theorem 15.23. See Remark 1.50 at the end of this section.

□

Definition 1.36. Let $f \in \Gamma(\mathcal{H})$. The ***Proximity operator*** of f is defined as

$$\operatorname{Prox}_f : \mathcal{H} \to \mathcal{H} \quad : \quad x \mapsto \operatorname*{Argmin}_{y \in \mathcal{H}} f(y) + \frac{1}{2}\|x - y\|^2. \tag{1.44}$$

Proposition 1.37. *Let $f \in \Gamma(\mathcal{H})$. Then the following hold:*

(i) ∂f *is maximally monotone.*

(ii) $\operatorname{Prox}_f = \mathrm{J}_{\partial f}$. (1.45)

(1.46)

Proof. (i) See [1], Theorem 20.25 (Moreau).

(ii) See [1], Proposition 16.44.

□

Corollary 1.38. *Combining Proposition 1.16, Proposition 1.37 and Corollary 1.27 we obtain the* ***Moreau's decomposition****:*

$$x = \operatorname{Prox}_{\gamma f}(x) + \gamma \operatorname{Prox}_{\gamma^{-1} f^*}(\gamma^{-1} x) \tag{1.47}$$

$\forall x \in \mathcal{H}, \forall \gamma \in \mathbb{R}_{++}$ *whenever* $f \in \Gamma(\mathcal{H})$.

Corollary 1.39. *Let* $f_i \in \Gamma(\mathcal{H}_i)$*, let* $x_i, z_i \in \mathcal{H}_i$ *let* $\alpha_i \neq 0$ *and set*

$$f = \bigoplus_{i=1}^{m} f_i(\alpha_i x_i - z_i). \tag{1.48}$$

Then combining Proposition 1.16, Proposition 1.17 and Proposition 1.29, we obtain the following usefull rule:

$$\operatorname{Prox}_{\gamma f}(x_1, \ldots, x_m) = \bigtimes_{i=1}^{m} \frac{1}{\alpha_i} \left(z_i + \operatorname{Prox}_{\gamma \alpha_i^2 f_i}(\alpha_i x_i - z_i) \right). \tag{1.49}$$

Next, let us compute a couple of proximity operators which will be used in our numerical expermients:

Example 1.40. [Induced Norm] It is clear from the definition that the proximity operator of the indicator function of a nonempty closed convex set is the projection operator, i.e., let $\varnothing \neq C \subseteq \mathcal{H}$ be closed and convex and let $f = \iota_C$, then

$$\operatorname{Prox}_{\gamma f}(x) = P_C(x), \quad \forall \gamma \in \mathbb{R}_{++}. \tag{1.50}$$

where P_C denotes the projection onto a nonempty closed convex set C.

Using Example 1.28, where $(\lambda \| \cdot \|_{\mathcal{H}})^* = \iota_{\overline{B}(0;\lambda)}$ we obtain:

$$\begin{aligned} \operatorname{Prox}_{\lambda \| \cdot \|_{\mathcal{H}}}(x) &\underset{(1.47)}{=} x - \operatorname{Prox}_{\iota_{\overline{B}_{\mathcal{H}}(0;\lambda)}}(x) \\ &= x - P_{\overline{B}_{\mathcal{H}}(0;\lambda)}(x) \\ &= \left(1 - \frac{\lambda}{\max\{\, \lambda, \|x\|_{\mathcal{H}} \,\}} \right) x. \end{aligned} \tag{1.51}$$

Observe that for:

- $\mathcal{H} = \mathbb{R}$, we obtain the **soft threeshold** function $\mathcal{T}_\lambda$ on $[-\lambda, \lambda]$:

$$\begin{aligned}\mathcal{T}_\lambda(x) &= \operatorname{sign}(x)\max\{|x| - \lambda, 0\}\\ &= \begin{cases} x-\lambda, & x>\lambda,\\ 0, & |x|<\lambda,\\ x+\lambda, & x<-\lambda.\end{cases}\end{aligned} \tag{1.52}$$

Example 1.41 (l_1 norm)**.** Let $\mathcal{H} = \mathbb{R}^n$ and let $f(x) = \lambda\|x - z\|_1 = \lambda\sum_{i=1}^n |x_i - z_i|$. Then

$$\begin{aligned}\operatorname{Prox}_{\gamma f}(x) &= \bigtimes_{i=1}^{n} z_i + \mathcal{T}_{\gamma\lambda}(x_i - z_i)\\ &= \bigtimes_{i=1}^{n} z_i + \operatorname{sign}(x_i - z_i)\max\{|x_i - z_i| - \gamma\lambda, 0\}\end{aligned} \tag{1.53}$$

$$\operatorname{Prox}_{\gamma f^*}(x) = P_{[-\lambda,\lambda]^n}(x - \gamma z). \tag{1.54}$$

Example 1.42. [Pointwise l_2 norm] Let $\mathcal{G} = \mathcal{H} \oplus \mathcal{H}$ with $\mathcal{H} = \mathbb{R}^{m\times n}$ and for $\lambda \in \mathbb{R}_{++}$ set

$$f : \mathcal{G} \to \mathbb{R} \quad : \quad (A, B) \mapsto \lambda\|(A, B)\|_\times$$

Then

$$\operatorname{Prox}_{\gamma f^*}(A, B) = P_C(A, B), \tag{1.55}$$

where

$$C = \left[\ \overline{B}_{\mathbb{R}^2(0;\lambda)}\right]^{m\times n},$$

and $\|(\cdot,\cdot)\|_\times$ is the norm defined in (1.35).

Definition 1.43. Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be proper and let $\beta \in \mathbb{R}_{++}$. We say that f is ***β-strongly convex*** if $\forall x, y \in \operatorname{dom} f$ and $\forall\lambda \in (0, 1)$

$$f(\lambda x + (1-\lambda)y) + \lambda(1-\lambda)\frac{\beta}{2}\|x - y\|^2 \leqslant \lambda f(x) + (1-\lambda)f(y) \tag{1.56}$$

It is easy to verify (see [1], Example 22.4) that

$$\text{If } f \text{ is } \beta\text{-strongly convex } \Rightarrow \partial f \text{ is } \beta\text{-strongly monotone.} \tag{1.57}$$

Proposition 1.44. *Let $f \in \Gamma(\mathcal{H})$. Then the following hold:*

(i) If f is β-strongly convex $\Rightarrow f$ is supercoercive, i.e., $\displaystyle\lim_{\|x\|\to+\infty} \frac{f(x)}{\|x\|} = +\infty.$ (1.58)

(ii) f is β-strongly convex $\Leftrightarrow f^*$ is Fréchet differentiable on $\mathcal{H}$ and ∇f^* is β-cocoervice. (1.59)

Proof. (i) See [1], Corollary 11.17.

(ii) See [1], Theorem 18.15.

□

Remark 1.45. Let $f : \mathcal{H} \to \mathbb{R}$ be a convex Fréchet differentiable function. The Baillon-Haddad Theorem asserts that

$$\nabla f \text{ is } \beta\text{-Lipschitzian} \Leftrightarrow \nabla f \text{ is } \frac{1}{\beta}\text{-cocoercive.} \tag{1.60}$$

In particular, if the gradient of a convex continuous differentiable function is nonexpansive, then it is actually firmly nonexpansive.

1.2.1 Fenchel-Rockafellar Duality

Definition 1.46. Let $F : \mathcal{H} \to \overline{\mathbb{R}}$ and $G : \mathcal{G} \to \overline{\mathbb{R}}$, be proper functions, and let $L \in \mathscr{B}(\mathcal{H}, \mathcal{G})$.

The ***primal problem*** associated with the composite function $F + G \circ L$ is

$$\underset{x \in \mathcal{H}}{\text{minimize}}\, F(x) + G(Lx), \tag{1.61}$$

its ***dual problem*** is

$$\underset{v \in \mathcal{G}}{\text{minimize}}\, F^*(-L^*v) + G^*(v), \tag{1.62}$$

the ***optimal primal value*** is

$$\mu = \inf_{x \in \mathcal{H}} F(x) + G(Lx),$$

the ***dual optimal value*** is

$$\mu^* = \inf_{v \in \mathcal{G}} F^*(-L^*v) + G^*(v),$$

and the ***the duality gap*** is

$$\Delta(F, G, L) = \begin{cases} 0 & \text{if } \mu = -\mu^* \in \{-\infty, +\infty\} \\ \mu + \mu^* & \text{otherwise} \end{cases} \tag{1.63}$$

Remark 1.47. It is always true that $-\mu^* \leqslant \mu$, this result is known as ***weak-duality*** and can be shown using the Fenchel-Young inequality. When $\mu = -\mu^*$ we say that ***strong-duality*** holds. A comprehensive analysis of assumptions that guarantee strong-duality is beyond the scope of this master's thesis and my knowledge. We keep it as simple as posible where our main argument is given in Proposition 1.35.

We close this section with two results regarding existence of solutions to the above Problem. The first one ensures the existence of a primal solution, and the second one states a relation between primal and dual solutions.

Proposition 1.48. *Let $f \in \Gamma(\mathcal{H})$, let $g \in \Gamma(\mathcal{G})$, and let $L \in \mathscr{B}(\mathcal{H},\mathcal{G})$ such that*

$$\operatorname{dom} g \cap L(\operatorname{dom} f) \neq \varnothing.$$

Suppose that one of the following holds:

(a) f is supercoercive.

(b) f is coercive (i.e., $\lim_{\|x\|\to+\infty} f(x) = +\infty$) and g is bounded from below.

Then $f + g \circ L$ is coercive and it has a minimizer over $\mathcal{H}$.

Proof. See [1], Corollary 11.16 with $g \circ L \in \Gamma(\mathcal{H})$ in place of g. □

Proposition 1.49. *Let $f \in \Gamma(\mathcal{H}), g \in \Gamma(\mathcal{G})$ and let $L \in \mathscr{B}(\mathcal{H},\mathcal{G})$ be such*

$$0 \in \operatorname{sri}(\operatorname{dom} g - L(\operatorname{dom} f)). \tag{1.64}$$

Then there exists $v \in \mathcal{G}$ solution to the dual problem (1.62)*, the duality gap* (1.63) *is zero, i.e., stong duality holds and*

$$\operatorname{Argmin}(f + g \circ L) = \partial f^*(-L^* v) \cap L^{-1}(\partial g^*(v)). \tag{1.65}$$

Proof. See [1], Theorem 19.1, Corollary 19.2. □

Remark 1.50. The set in (1.65) could be empty and a list of conditions satisfying (1.64) can be found in ([1], Proposition 15.24).

2 Malitsky-Tam Algorithms

In this section we briefly recall some results presented in [12], from which we will derive methods to solve general monotone inclusion problems. We also include their proofs, where we made three small contributions, namely, the relaxation of a fixed λ in ([12], Theorem 2.9) obtaining linear convergence under the same assumptions given in ([12], Theorem 2.5); the proof of linear convergence of the relaxed-inertial-forward-reflected-barckward algorithm ([12], Theorem 4.3), and the proof of linear convergence of the three operator splitting algorithm ([12], theorem 5.2) whenever the operator A is strongly monotone.

2.1 Forward Reflected Backward (FRB)

Theorem 2.1. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $C : \mathcal{H} \to \mathcal{H}$ be a monotone and β-Lipschitzian operator with $\beta \in \mathbb{R}_{++}$. Suppose that* $\operatorname{zer}(A+C) \neq \varnothing$ *and*

$$\{\lambda_n\}_{n\in\mathbb{N}} \subseteq \left[\varepsilon, \frac{1-2\varepsilon}{2\beta}\right], \tag{2.1}$$

for some

$$\varepsilon \in \left(0, \frac{1}{2(1+\beta)}\right]. \tag{2.2}$$

Given $x_0, x_1 \in \mathcal{H}$ consider the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by:

$$x_{n+1} = \mathrm{J}_{\lambda_n A}\big(x_n - [\lambda_n + \lambda_{n-1}]Cx_n + \lambda_{n-1}Cx_{n-1}\big). \tag{2.3}$$

Then the following hold:

(i) $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(A+C)$.

(ii) Suppose that A or C is strongly monotone. Then $\{x_n\}_{n\in\mathbb{N}}$ converges R-linearly to the unique point in $\operatorname{zer}(A+C)$.

Proof. (i) Let $\bar{x} \in \operatorname{zer}(A+C)$, then $\bar{x} = \mathrm{J}_{\lambda_n A}(\bar{x} - \lambda_n C\bar{x})$ and by Proposition 1.14 we have that $\mathrm{J}_{\lambda_n A}$ is a single-valued, firmly nonexpansive operator with full domain. Therefore the following holds:

$$\|x_{n+1}-\bar{x}\|^2 = \| \mathrm{J}_{\lambda_n A}(x_n - \lambda_n C x_n - \lambda_{n-1}(Cx_n - Cx_{n-1})) - \mathrm{J}_{\lambda_n A}(\bar{x} - \lambda_n C\bar{x})\|^2 \tag{2.4}$$

$$\underset{(1.9)}{\leqslant} \langle\, x_{n+1}-\bar{x}, x_n - \lambda_n C x_n - \lambda_{n-1}(Cx_n - Cx_{n-1}) - \bar{x} + \lambda_n C\bar{x} \,\rangle \tag{2.5}$$

$$= \langle\, x_{n+1}-\bar{x}, x_n - \bar{x} \,\rangle \tag{2.6}$$

$$+ \lambda_n \langle\, x_{n+1}-\bar{x}, Cx_{n+1} - Cx_n \,\rangle - \lambda_n \underbrace{\langle\, x_{n+1}-\bar{x}, Cx_{n+1} - C\bar{x} \,\rangle}_{\geqslant 0 \text{ by } (1.4)}$$

$$- \lambda_{n-1} \langle\, x_{n+1}-x_n, Cx_n - Cx_{n-1} \,\rangle \tag{2.7}$$

$$- \lambda_{n-1} \langle\, x_n-\bar{x}, Cx_n - Cx_{n-1} \,\rangle .$$

By the identity (1.1), we can write (2.6) as

$$\langle\, x_{n+1}-\bar{x}, x_n-\bar{x} \,\rangle = \frac{1}{2}\left(\|x_{n+1}-\bar{x}\|^2 + \|x_n-\bar{x}\|^2 - \|x_{n+1}-x_n\|^2\right). \tag{2.8}$$

and by the Lipschitz continuity of C, we can estimate (2.7) as follows

$$\begin{aligned} -\lambda_{n-1} \langle\, x_{n+1}-x_n, Cx_n - Cx_{n-1} \,\rangle &\leqslant \lambda_{n-1}\beta\|x_{n+1}-x_n\|\|x_n-x_{n-1}\| \\ &\leqslant \frac{\lambda_{n-1}\beta}{2}\left(\|x_{n+1}-x_n\|^2 + \|x_n-x_{n-1}\|^2\right) \\ &\leqslant \frac{1-2\varepsilon}{4}\left(\|x_{n+1}-x_n\|^2 + \|x_n-x_{n-1}\|^2\right). \end{aligned} \tag{2.9}$$

In light of (2.8) and (2.9), the inequality in (2.5) implies

$$\begin{aligned} &\|x_{n+1}-\bar{x}\|^2 - 2\lambda_n \langle\, x_{n+1}-\bar{x}, Cx_{n+1}-Cx_n \,\rangle + \left(\frac{1}{2}+\varepsilon\right)\|x_{n+1}-x_n\|^2 \\ &\leqslant \|x_n-\bar{x}\|^2 - 2\lambda_{n-1} \langle\, x_n-\bar{x}, Cx_n - Cx_{n-1} \,\rangle + \frac{1}{2}\|x_n-x_{n-1}\|^2. \end{aligned} \tag{2.10}$$

We will use this inequality many times, therefore it is worth capturing its information in a function which allows us to write the formulas in a short manner:

Proposition 2.2. *In the setting of Theorem 2.1, let* $\{x_n\}_{n\in\mathbb{N}}$ *be the sequence generated in* (2.3) *and let* $\Psi : \mathcal{H}\times\mathbb{N}_1 \to \mathbb{R}$ *be defined as:*

$$\Psi(s,n) = \frac{1}{2}\|x_n - s\|^2 - 2\lambda_{n-1}\langle\, x_n - s, Cx_n - Cx_{n-1} \,\rangle + \frac{1}{2}\|x_n - x_{n-1}\|^2.$$

Then the following hold:

(i) $\Psi(s,n) \geqslant 0$.

(ii) $\Psi(s,n) \leqslant (1-\varepsilon)\left(\|x_n - s\|^2 + \|x_n - x_{n-1}\|^2\right).$ (2.11)

Proof. In both cases we use that C is β-Lipschitzian:

$$(i)\ \Psi(s,n) \geqslant \frac{1}{2}\|x_n - s\|^2 - 2\lambda_{n-1}\beta\|x_n - s\|\|x_n - x_{n-1}\| + \frac{1}{2}\|x_n - x_{n-1}\|^2$$

$$\geqslant \underbrace{\frac{1 - 2\lambda_{n-1}\beta}{2}}_{\geqslant\varepsilon}(\|x_n - s\|^2 + \|x_n - x_{n-1}\|^2) \geqslant 0.$$

$$(ii)\ \Psi(s,n) \leqslant \frac{1}{2}\|x_n - s\|^2 + 2\lambda_{n-1}\beta\|x_n - s\|\|x_n - x_{n-1}\| + \frac{1}{2}\|x_n - x_{n-1}\|^2$$

$$\leqslant \underbrace{\frac{1 + 2\lambda_{n-1}\beta}{2}}_{\leqslant 1-\varepsilon}\left(\|x_n - s\|^2 + \|x_n - x_{n-1}\|^2\right).$$

□

For the rest of the proof, we set $\Psi_n := \Psi(\bar{x}, n)$. Combining Proposition 2.2(i) with (2.10) we obtain:

$$\begin{aligned}
\frac{1}{2}\|x_{n+1} - \bar{x}\|^2 &\leqslant \frac{1}{2}\|x_{n+1} - \bar{x}\|^2 + \Psi_{n+1} + \varepsilon\|x_{n+1} - x_n\|^2 \\
&\leqslant \frac{1}{2}\|x_n - \bar{x}\|^2 + \Psi_n \\
&= \frac{1}{2}\|x_n - \bar{x}\|^2 + \Psi_n + \varepsilon\|x_n - x_{n-1}\|^2 - \varepsilon\|x_n - x_{n-1}\|^2 \\
&\vdots \\
&\leqslant \frac{1}{2}\|x_1 - \bar{x}\|^2 + \Psi_1 - \varepsilon\sum_{j=0}^{n-1}\|x_{j+1} - x_j\|^2.
\end{aligned}$$

Rearranging the terms, we obtain:

$$\frac{1}{2}\|x_{n+1} - \bar{x}\|^2 + \varepsilon\sum_{j=0}^{n-1}\|x_{j+1} - x_j\|^2 \leqslant \frac{1}{2}\|x_1 - \bar{x}\|^2 + \Psi_1 < \infty \quad \forall n \in \mathbb{N}_1.$$

This implies that the sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded and that

$$\|x_{n+1} - x_n\| \to 0 \quad \text{as } n \to +\infty. \tag{2.12}$$

By the definition of the resolvent, (2.3) implies:

$$Cx_{n+1} - Cx_n - \frac{1}{\lambda_n}(x_{n+1} - x_n + \lambda_{n-1}(Cx_n - Cx_{n-1})) \in (A + C)x_{n+1}. \tag{2.13}$$

By Lemma 1.3 there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ of the bounded sequence $\{x_n\}_{n\in\mathbb{N}}$ which converges weakly to some point $\tilde{x} \in \mathcal{H}$; we now show that $\tilde{x} \in \operatorname{zer}(A+C)$.
Since λ_n is strictly positive and bounded $\forall n \in \mathbb{N}$, the limit of the LHS in (2.13) along the subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ converges strongly to 0:

$$\begin{aligned} &\lim_{k\to\infty} \left\| Cx_{n_{k+1}} - Cx_{n_k} - \frac{1}{\lambda_{n_k}}(x_{n_{k+1}} - x_{n_k} + \lambda_{n_{k-1}}(Cx_{n_k} - Cx_{n_{k-1}})) \right\| \\ &\leqslant \lim_{k\to\infty} \left(\beta + \frac{1}{\lambda_{n_k}}\right) \underbrace{\|x_{n_{k+1}} - x_{n_k}\|}_{(2.12)\to 0} + \frac{\beta\lambda_{n_k-1}}{\lambda_{n_k}} \underbrace{\|x_{n_k} - x_{n_{k-1}}\|}_{(2.12)\to 0} = 0. \end{aligned} \tag{2.14}$$

By Proposition 1.9 follows that $A + C$ is maximally monotone. Then, from (2.13), (2.14), and Proposition 1.11 follows that every weak cluster point of $\{x_n\}_{n\in\mathbb{N}}$ belongs to $\operatorname{zer}(A+C)$.
Moreover, $\left(\frac{1}{2}\|x_n - \bar{x}\|^2 + \Psi_n\right)_{n\in\mathbb{N}_1}$ is nonnegative and nonincreasing, therefore its limit exists for each $\bar{x} \in \operatorname{zer}(A+C)$ and by the sandwich rule it is equal to $\lim_{n\to\infty}\|x_n - \bar{x}\|$:

$$\begin{aligned} \lim_{n\to\infty}\|x_n - \tilde{x}\|^2 &= \lim_{n\to\infty}\|x_n - \tilde{x}\|^2 - 2\lambda_{n-1}\beta\underbrace{\|x_n - \tilde{x}\|}_{<\infty}\underbrace{\|x_n - x_{n-1}\|}_{\to 0} + \frac{1}{2}\underbrace{\|x_n - x_{n-1}\|^2}_{\to 0} \\ &\leqslant \lim_{n\to\infty}\|x_n - \tilde{x}\|^2 - 2\lambda_{n-1}\langle x_n - \tilde{x}, Cx_n - Cx_{n-1}\rangle + \frac{1}{2}\|x_n - x_{n-1}\|^2 \\ &\leqslant \lim_{n\to\infty}\|x_n - \tilde{x}\|^2 + 2\lambda_{n-1}\beta\underbrace{\|x_n - \tilde{x}\|}_{<\infty}\underbrace{\|x_n - x_{n-1}\|}_{\to 0} + \frac{1}{2}\underbrace{\|x_n - x_{n-1}\|^2}_{\to 0} \\ &= \lim_{n\to\infty}\|x_n - \tilde{x}\|^2. \end{aligned}$$

Since the limit of $\{\|x_n - \tilde{x}\|\}_{n\in\mathbb{N}}$ exists for each weak cluster point $\tilde{x}$ of the sequence $\{x_n\}_{n\in\mathbb{N}}$, and every such weak cluster point belongs to $\operatorname{zer}(A+C)$, it follows from Lemma 1.4 that the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated in (2.3) converges weakly to a point in $\operatorname{zer}(A+B)$ and the proof is complete.

(ii) In order to show linear convergence, we assume that A is ρ-strongly monotone (there is no loss of generallity, see remark 2.3). Using that $J_{\lambda_n A}$ is $(1+\lambda_n\rho)$-cocoercive guaranteed by Proposition 1.18, in place of firmly nonexpansivity in (2.4) we obtain:

$$\begin{aligned} (1+2\rho\lambda_n)\|x_{n+1} - \bar{x}\|^2 &- 2\lambda_n\langle x_{n+1} - \bar{x}, Cx_{n+1} - Cx_n\rangle + \left(\frac{1}{2} + \varepsilon\right)\|x_{n+1} - x_n\|^2 \\ &\leqslant \|x_n - \bar{x}\|^2 - 2\lambda_{n-1}\langle x_n - \bar{x}, Cx_n - Cx_{n-1}\rangle + \frac{1}{2}\|x_n - x_{n-1}\|^2. \end{aligned} \tag{2.15}$$

Next, set:

$$a_n := \frac{1}{2}\|x_n - \bar{x}\|^2,$$
$$\mu := \min\left\{\frac{\rho\varepsilon}{1-\varepsilon}, \frac{\varepsilon}{1-\varepsilon}\right\}.$$

Using the auxiliary function $\Psi_n = \Psi(\bar{x}, n)$ from Propositon 2.2, and replacing λ_n by ε in the first term of the LHS in (2.15), (recall that $\lambda_n \geqslant \varepsilon$ for all $n \in \mathbb{N}$ by assumption), the inequality (2.15) implies that:

$$\begin{aligned}
a_n + \Psi_n &\geqslant (1+4\varepsilon\rho)a_{n+1} + \Psi_{n+1} + \varepsilon\|x_{n+1} - x_n\|^2 \\
&= (1+4\varepsilon\rho)a_{n+1} + (1+\mu)\Psi_{n+1} - \mu\Psi_{n+1} + \varepsilon\|x_{n+1} - x_n\|^2 \\
&\underset{(2.11)}{\geqslant} (1+4\varepsilon\rho)a_{n+1} + (1+\mu)\Psi_{n+1} \\
&\quad - \mu(1-\varepsilon)\left(2a_{n+1} + \|x_{n+1} - x_n\|^2\right) + \varepsilon\|x_{n+1} - x_n\|^2 \\
&= (1+2\varepsilon\rho)a_{n+1} + (1+\mu)\Psi_{n+1} + 2\underbrace{(\rho\varepsilon - \mu[1-\varepsilon])}_{\geqslant 0}a_{n+1} \\
&\quad + \underbrace{(\varepsilon - \mu[1-\varepsilon])}_{\geqslant 0}\|x_{n+1} - x_n\|^2 \\
&\geqslant (1 + \min\{2\varepsilon\rho, \mu\})(a_{n+1} + \Psi_{n+1}) \\
&\underset{\varepsilon<\frac{1}{2}}{=} (1+\mu)(a_{n+1} + \Psi_{n+1}).
\end{aligned}$$

From here, it follows that:

$$a_{n+1} \leqslant a_{n+1} + \Psi_{n+1} \leqslant \frac{1}{1+\mu}(a_n + \Psi_n) \leqslant \cdots \leqslant \frac{1}{(1+\mu)^n}(a_1 + \Psi_1),$$

and according to Definition 1.2, the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges R-linearly to $\bar{x}$. Since A is strongly monotone, the set $\operatorname{zer}(A+C)$ has exactly one element. □

Remark 2.3. In the case where the operator C is ρ-strongly monotone, we can reformulate the equations using the operators $\tilde{A}$ and $\tilde{C}$ defined as follows:

$$\tilde{A} := A + \rho\mathrm{Id},$$
$$\tilde{C} := C - \rho\mathrm{Id}.$$

Then

$$\tilde{A} + \tilde{C} = A + C,$$

and by Proposition 1.16

$$\mathrm{J}_{\lambda\tilde{A}} \underset{(1.11)}{=} J_{\frac{\lambda}{1+\lambda\rho}A} \circ (1+\lambda\rho)^{-1}\mathrm{Id}\,.$$

where $\tilde{A}$ is ρ-strongly monotone and $\tilde{C}$ is $\tilde{\beta}$-Lipschitz continuous, with $\tilde{\beta} = (\beta+\rho)$. In this case the recursion (2.3) takes the form

$$\begin{aligned}
x_{n+1} &= \mathrm{J}_{\lambda_n\tilde{A}}\left(x_n - \lambda_n\tilde{C}x_n - \lambda_{n-1}(\tilde{C}x_n - \tilde{C}x_{n-1})\right)\\
&= J_{\frac{\lambda_n}{1+\lambda_n\rho}A}\left(\frac{x_n - \lambda_n\tilde{C}x_n - \lambda_{n-1}(\tilde{C}x_n - \tilde{C}x_{n-1})}{(1+\lambda_n\rho)}\right)\\
&= J_{\frac{\lambda_n}{1+\lambda_n\rho}A}\left(x_n - \frac{(\lambda_n+\lambda_{n-1})Cx_n - \lambda_{n-1}Cx_{n-1} - \lambda_{n-1}\rho(x_n - x_{n-1})}{1+\lambda_n\rho}\right)
\end{aligned}$$

and linear convergence is guaranteed by Theorem 2.1 as long as

$$\{\lambda_n\}_{n\in\mathbb{N}} \subset \left[\varepsilon, \frac{1-2\varepsilon}{2(\beta+\rho)}\right].$$

2.2 Relaxed Inertial Forward Reflected Backward (RIFRB)

Theorem 2.4. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $C : \mathcal{H} \to \mathcal{H}$ be monotone and β-Lipschitzian with $\beta \in \mathbb{R}_{++}$ such that* $\operatorname{zer}(A+C) \neq \varnothing$. *Let $\xi \in [0,1), \eta \in (0,1]$ and $\lambda \in \mathbb{R}_{++}$ such that*

$$\xi < \frac{2-\eta}{2+\eta}, \tag{2.16}$$

and

$$\lambda < \min\left\{\frac{2(1-\xi)-\eta(1+\xi)}{2\beta}, \frac{1-\xi(1+\eta)}{\eta\beta}\right\}. \tag{2.17}$$

Given $x_0, x_1 \in \mathcal{H}$, consider the sequence $\{x_n\}_{n\in\mathbb{N}}$ generated by:

$$\begin{cases} z_{n+1} & := \mathrm{J}_{\lambda A}\left(x_n - \lambda Cx_n - \frac{\lambda}{\eta}(Cx_n - Cx_{n-1}) + \frac{\xi}{\eta}(x_n - x_{n-1})\right), \\ x_{n+1} & := (1-\eta)x_n + \eta z_{n+1}. \end{cases} \tag{2.18}$$

Then the following hold:

(i) $\{x_n\}_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(A+C)$.

(ii) Suppose that A is strongly monotone, then $\{x_n\}_{n\in\mathbb{N}}$ converges R-linearly to the unique point in $\operatorname{zer}(A+C)$.

Proof. (i) Define the operator

$$C' := C - \frac{\xi}{\lambda}\mathrm{Id}\,, \tag{2.19}$$

which is β'-Lipschitz continuous, with

$$\beta' = \beta + \frac{\xi}{\lambda}. \tag{2.20}$$

Then the recursion scheme (2.18) is equivalent to

$$\begin{cases} z_{n+1} & := \mathrm{J}_{\lambda A}\left(x_n - \lambda C x_n - \frac{\lambda}{\eta}(C'x_n - C'x_{n-1})\right), \\ x_{n+1} & := (1-\eta)x_n + \eta z_{n+1}. \end{cases} \tag{2.21}$$

Let $\bar{x} \in \operatorname{zer}(A+C)$, then $\bar{x} = \mathrm{J}_{\lambda A}(\bar{x} - \lambda C\bar{x})$ and by Proposition 1.14 $\mathrm{J}_{\lambda A}$ is a single-valued, firmly nonexpansive operator, therefore the following holds

$$\|z_{n+1} - \bar{x}\|^2 = \|J_{\lambda A}[x_n - \lambda C x_n - \frac{\lambda}{\eta}(C'x_n - C'x_{n-1})] - \mathrm{J}_{\lambda A}(\bar{x} - \lambda C\bar{x})\|^2 \tag{2.22}$$

$$\underset{(1.9)}{\leqslant} \left\langle z_{n+1} - \bar{x}, x_n - \lambda C x_n - \frac{\lambda}{\eta}(C'x_n - C'x_{n-1}) - \bar{x} + \lambda C\bar{x} \right\rangle \tag{2.23}$$

$$= \langle z_{n+1} - \bar{x}, x_n - \bar{x} \rangle \tag{2.24}$$

$$- \lambda \langle z_{n+1} - \bar{x}, C x_n - C\bar{x} \rangle \tag{2.25}$$

$$- \frac{\lambda}{\eta} \langle z_{n+1} - \bar{x}, C'x_n - C'x_{n-1} \rangle. \tag{2.26}$$

Using the identity $z_{n+1} = \frac{1}{\eta}x_{n+1} + \frac{\eta-1}{\eta}x_n$ from (2.21) and the substitution

$$c_n := 2\lambda \langle x_n - \bar{x}, C x_n - C\bar{x} \rangle \underset{(1.4)}{\geqslant} 0,$$

we can compute and estimate the terms in the last equation as follows:

$$\begin{aligned} (2.22) &= \left\| \frac{x_{n+1} - \bar{x}}{\eta} + \frac{\eta - 1}{\eta}(x_n - \bar{x}) \right\|^2 \\ &= \frac{1}{\eta^2}(\|x_{n+1} - \bar{x}\|^2 + (\eta-1)^2\|x_n - \bar{x}\|^2 \\ &\quad + 2(\eta-1)\langle x_{n+1} - \bar{x}, x_n - \bar{x} \rangle) \\ &= \frac{1}{\eta^2}(\|x_{n+1} - \bar{x}\|^2 + (\eta-1)^2\|x_n - \bar{x}\|^2 \\ &\quad + (\eta-1)(\|x_{n+1} - \bar{x}\|^2 + \|x_n - \bar{x}\|^2 - \|x_{n+1} - x_n\|^2) \\ &= \frac{1}{\eta}\|x_{n+1} - \bar{x}\|^2 + \frac{\eta - 1}{\eta}\|x_n - \bar{x}\|^2 + \frac{1-\eta}{\eta^2}\|x_{n+1} - x_n\|^2. \end{aligned} \tag{2.27}$$

$$
\begin{aligned}
(2.24) &= \frac{1}{\eta}\left\langle x_{n+1}-\bar{x}, x_n-\bar{x}\right\rangle + \frac{\eta-1}{\eta}\left\langle x_n-\bar{x}, x_n-\bar{x}\right\rangle \\
&= \frac{1}{2\eta}\left(\|x_{n+1}-\bar{x}\|^2 + \|x_n-\bar{x}\|^2 - \|x_{n+1}-x_n\|^2\right) + \frac{\eta-1}{\eta}\|x_n-\bar{x}\|^2 \\
&= \frac{1}{2\eta}\|x_{n+1}-\bar{x}\|^2 + \frac{2\eta-1}{2\eta}\|x_n-\bar{x}\|^2 - \frac{1}{2\eta}\|x_{n+1}-x_n\|^2. \qquad (2.28)
\end{aligned}
$$

$$
\begin{aligned}
(2.25) &= -\frac{\lambda}{\eta}\left\langle x_{n+1}-\bar{x}, Cx_n - C\bar{x}\right\rangle + \frac{(1-\eta)}{\eta}\lambda\left\langle x_n-\bar{x}, Cx_n-C\bar{x}\right\rangle \\
&= -\frac{c_{n+1}}{2\eta} + \frac{\lambda}{\eta}\left\langle x_{n+1}-\bar{x}, Cx_{n+1}-Cx_n\right\rangle + \frac{1-\eta}{2\eta}c_n \\
&= -\frac{c_{n+1}}{2\eta} + \frac{1-\eta}{2\eta}c_n + \frac{\lambda}{\eta}\left\langle x_{n+1}-\bar{x}, C'x_{n+1}-C'x_n\right\rangle + \frac{\xi}{\eta}\left\langle x_{n+1}-\bar{x}, x_{n+1}-x_n\right\rangle \\
&= -\frac{c_{n+1}}{2\eta} + \frac{1-\eta}{2\eta}c_n + \frac{\lambda}{\eta}\left\langle x_{n+1}-\bar{x}, C'x_{n+1}-C'x_n\right\rangle \\
&\quad + \frac{\xi}{2\eta}\left(\|x_{n+1}-\bar{x}\|^2 + \|x_{n+1}-x_n\|^2 - \|x_n-\bar{x}\|^2\right). \qquad (2.29)
\end{aligned}
$$

$$
\begin{aligned}
(2.26) &= -\frac{\lambda}{\eta}\left\langle x_n-\bar{x}, C'x_n-C'x_{n-1}\right\rangle - \frac{\lambda}{\eta^2}\left\langle x_{n+1}-x_n, C'x_n-C'x_{n-1}\right\rangle \\
&\leqslant -\frac{\lambda}{\eta}\left\langle x_n-\bar{x}, C'x_n-C'x_{n-1}\right\rangle + \frac{\lambda\beta'}{2\eta^2}(\|x_{n+1}-x_n\|^2 + \|x_n-x_{n-1}\|^2). \qquad (2.30)
\end{aligned}
$$

Substituting (2.22), (2.24), (2.25), (2.26) by (2.27), (2.28), (2.29), (2.30) respectively, and multiplying every term by 2η; the inequality (2.23) implies:

$$
\begin{aligned}
&(1-\xi)\|x_{n+1}-\bar{x}\|^2 - 2\lambda\left\langle x_{n+1}-\bar{x}, C'x_{n+1}-C'x_n\right\rangle \\
&\qquad\qquad + c_{n+1} + \left(\frac{2-\lambda\beta'-\eta(1+\xi)}{\eta}\right)\|x_{n+1}-x_n\|^2 \\
\leqslant\ &(1-\xi)\|x_n-\bar{x}\|^2 - 2\lambda\left\langle x_n-\bar{x}, C'x_n-C'x_{n-1}\right\rangle \\
&\qquad\qquad + (1-\eta)c_n + \frac{\lambda\beta'}{\eta}\|x_n-x_{n-1}\|^2. \qquad (2.31)
\end{aligned}
$$

Observe that this inequality is very similar to (2.10). The idea is to show that this recursion generates a nonnegative and nonincreasing sequence. To this end set

$$
\varepsilon := \min\{2-2\lambda\beta'-\eta(1+\xi), 1-\xi-\lambda\beta'\eta\} \qquad (2.32)
$$

which is strictly positive by (2.17) and (2.20). Next, as we did in the last theorem, we define an auxilliary function Φ which captures the information in (2.31):

Proposition 2.5. *In the setting of Theorem 2.4, let* $\{x_n\}_{n\in\mathbb{N}}$ *be the sequence generated in* (2.21)*, let* $\varepsilon > 0$ *defined in* (2.32)*, and let* $\Psi : \mathcal{H} \times \mathbb{N}_1 \to \mathbb{R}$ *be defined as:*

$$\Phi : \mathcal{H} \times \mathbb{N}_1 \to \mathbb{R}_+$$

$$\Phi(s,n) = (1-\xi)\|x_n - s\|^2 - 2\lambda \left\langle x_n - s, C'x_n - C'x_{n-1} \right\rangle + \frac{\lambda\beta'}{\eta}\|x_n - x_{n-1}\|^2. \tag{2.33}$$

Then the following hold:

(i) $\Phi(s,n) \geqslant \varepsilon\|x_n - s\|^2,$ (2.34)

(ii) $\Phi(s,n) \leqslant (1-\xi+\lambda\beta'\eta)\|x_n - s\|^2 + \dfrac{2\lambda\beta'}{\eta}\|x_n - x_{n-1}\|^2.$ (2.35)

Proof. In both cases we use that C' is β'-Lipschitz continuos:

$$\begin{aligned}
(i)\ \Phi(s,n) \geqslant\ & (1-\xi)\|x_n - s\|^2 - 2\lambda\beta'\|x_n - s\|\|x_n - x_{n-1}\| + \frac{\lambda\beta'}{\eta}\|x_n - x_{n-1}\|^2 \\
&\geqslant (1-\xi)\|x_n - s\|^2 - \lambda\beta'\left(\eta\|x_n - s\|^2 + \frac{1}{\eta}\|x_n - x_{n-1}\|^2\right) + \frac{\lambda\beta'}{\eta}\|x_n - x_{n-1}\|^2 \\
&= \underbrace{(1-\xi-\lambda\beta'\eta)}_{\geqslant\varepsilon}\|x_n - s\|^2.
\end{aligned}$$

$$\begin{aligned}
(ii)\ \Phi(s,n) &\leqslant (1-\xi)\|x_n - s\|^2 + 2\lambda\beta'\|x_n - s\|\|x_n - x_{n-1}\| + \frac{\lambda\beta'}{\eta}\|x_n - x_{n-1}\|^2 \\
&\underset{\substack{(1.2)\\ \delta=\eta}}{\leqslant} (1-\xi+\lambda\beta'\eta)\|x_n - s\|^2 + \frac{2\lambda\beta'}{\eta}\|x_n - x_{n-1}\|^2.
\end{aligned}$$

□

For the rest of the proof, we set $\Phi_n := \Phi(\bar{x}, n)$ in (2.33). Combining Proposition 2.5(i) and (2.31) we obtain:

$$\begin{aligned}
\varepsilon\|x_{n+1} - \bar{x}\|^2 &\leqslant \Phi_{n+1} + c_{n+1} + \frac{\varepsilon}{\eta}\|x_{n+1} - x_n\|^2 \\
&\leqslant \Phi_n + c_n \\
&= \Phi_n + c_n + \frac{\varepsilon}{\eta}\|x_n - x_{n-1}\|^2 - \frac{\varepsilon}{\eta}\|x_n - x_{n-1}\|^2 \\
&\ \vdots \\
&\leqslant \Phi_1 + c_1 - \frac{\varepsilon}{\eta}\sum_{j=0}^{n-1}\|x_{j+1} - x_j\|^2.
\end{aligned}$$

or equivalently

$$\varepsilon\left(\|x_{n+1}-\bar{x}\|^2+\frac{1}{\eta}\sum_{j=0}^{n-1}\|x_{j+1}-x_j\|^2\right)\leqslant \Phi_1+c_1<\infty, \tag{2.36}$$

which implies that the sequence $\{x_n\}_{n\in\mathbb{N}}$ is bounded and

$$\|x_{n+1}-x_n\|\to 0, \qquad \text{as } n\to+\infty. \tag{2.37}$$

The reminder of the proof follows a similiar argument to (2.13).

(ii) Suppose that A is ρ-strongly monotone and using that $\mathrm{J}_{\lambda A}$ is $(1+\lambda\rho)$-cocoercive in place of firmly nonexpansivity in (2.22) we obtain:

$$\begin{aligned}(1+\lambda\rho)\|z_{n+1}-\bar{x}\|^2 &\leqslant \langle\, z_{n+1}-\bar{x}, x_n-\bar{x}\,\rangle-\lambda\,\langle\, z_{n+1}-\bar{x}, Cx_n-C\bar{x}\,\rangle\\ &\quad -\frac{\lambda}{\eta}\,\langle\, z_{n+1}-\bar{x}, C'x_n-C'x_{n-1}\,\rangle.\end{aligned}\tag{2.38}$$

Next, we replace the computations (2.27) - (2.30) in (2.38) and we get:

$$\begin{aligned}&(1-\xi+2\lambda\rho)\|x_{n+1}-\bar{x}\|^2-2\lambda\,\langle\, x_{n+1}-\bar{x}, C'x_{n+1}-C'x_n\,\rangle\\ &\qquad + c_{n+1}+\frac{2+2\lambda\rho(1-\eta)-\eta(1+\xi)-\lambda\beta'}{\eta}\|x_{n+1}-x_n\|^2\\ \leqslant\ &(1-\xi+2\lambda\rho(1-\eta))\|x_n-\bar{x}\|^2-2\lambda\,\langle\, x_n-\bar{x}, C'x_n-C'x_{n-1}\,\rangle\\ &\qquad +(1-\eta)c_n+\frac{\lambda\beta'}{\eta}\|x_n-x_{n-1}\|^2.\end{aligned}\tag{2.39}$$

Set

$$\mu:=\min\left\{\frac{\lambda\rho\eta}{1-\xi+\lambda\eta\beta'}, \frac{\varepsilon+2\rho\lambda(1-\eta)}{2\lambda\beta'}\right\}, \tag{2.40}$$

which is strictly positive by (2.17), (2.20) and (2.32).

If $\eta=1$ then the inequality (2.39) reads (recall $\Phi_n=\Phi(\bar{x},n)$ in (2.33))

$$\begin{aligned}\Phi_n &\geqslant 2\lambda\rho\|x_{n+1}-\bar{x}\|^2+\Phi_{n+1}+c_{n+1}+\varepsilon\|x_{n+1}-x_n\|^2\\ &= 2\lambda\rho\|x_{n+1}-\bar{x}\|^2+(1+\mu)\Phi_{n+1}-\mu\Phi_{n+1}+\varepsilon\|x_{n+1}-x_n\|^2\\ &\underset{(2.35)}{\geqslant} \lambda\rho\|x_{n+1}-\bar{x}\|^2+(1+\mu)\Phi_{n+1}\\ &\quad+\underbrace{(\lambda\rho-\mu[(1-\xi+\lambda\beta')])}_{\geqslant 0 \text{ by } (2.40)}\left(\|x_{n+1}-\bar{x}\|^2\right)+\underbrace{(\varepsilon-2\lambda\mu\beta')}_{\geqslant 0\text{ by }(2.40)}\|x_{n+1}-x_n\|^2.\end{aligned}$$

From here it follows that

$$\|x_{n+1}-\bar{x}\|^2 \leqslant \frac{1}{\lambda\rho}\Phi_n \leqslant \frac{1}{\lambda\rho}\frac{1}{(1+\mu)}\Phi_{n-1} \leqslant \cdots \leqslant \frac{1}{\lambda\rho}\frac{1}{(1+\mu)^{n-1}}\Phi_1, \tag{2.41}$$

and according to the Definition 1.2, the sequence $(x_n)_{n\in\mathbb{N}}$ converges R-linearly to $\bar{x}$.

If $\eta < 1$ set:

$$a_n := 2\lambda\rho(1-\eta)\|x_n-\bar{x}\|^2, \tag{2.42}$$

$$\tilde{c}_n := (1-\eta)c_n, \tag{2.43}$$

$$\eta^* := \frac{\eta}{2(1-\eta)}. \tag{2.44}$$

Substituting (2.42), (2.43), and (2.44) in (2.39) we get:

$$\begin{aligned}
a_n+\Phi_n+\tilde{c}_n \underset{(2.39)}{\geqslant} & (1+2\eta^*)(a_{n+1}+\tilde{c}_{n+1})+\Phi_{n+1}+\frac{\varepsilon+2\lambda\rho(1-\eta)}{\eta}\|x_{n+1}-x_n\|^2 \\
&= (1+2\eta^*)(a_{n+1}+\tilde{c}_{n+1})+(1+\mu)\Phi_{n+1}-\mu\Phi_{n+1} \\
&\quad + \frac{\varepsilon+2\lambda\rho(1-\eta)}{\eta}\|x_{n+1}-x_n\|^2 \qquad (2.45)\\
&= (1+\eta^*)(a_{n+1}+\tilde{c}_{n+1})+(1+\mu)\Phi_{n+1} \\
&\quad + \lambda\rho\eta\|x_{n+1}-\bar{x}\|^2-\mu\Phi_{n+1}+\frac{\varepsilon+2\lambda\rho(1-\eta)}{\eta}\|x_{n+1}-x_n\|^2+\eta^*\tilde{c}_{n+1} \\
&\underset{(2.35)}{\geqslant} (1+\underbrace{\min\{\eta^*,\mu\}}_{:=\omega})(a_{n+1}+\Phi_{n+1}+\tilde{c}_{n+1}) \qquad (2.46)\\
&\quad + \underbrace{(\lambda\rho\eta-\mu[1-\xi+\lambda\beta'\eta])}_{\geqslant 0 \text{ by } (2.40)}\|x_{n+1}-x_n\|^2 \\
&\quad + \underbrace{\frac{\varepsilon+2\lambda\rho(1-\eta)-2\mu\lambda\beta'}{\eta}}_{\geqslant 0 \text{ by } (2.40)}\|x_{n+1}-x_n\|^2+\frac{1}{2}\underbrace{\eta c_{n+1}}_{\geqslant 0 \text{ by } (1.4)}.
\end{aligned}$$

By (2.46), since $\omega = \min\{\eta^*,\mu\} > 0$, it follows that

$$\begin{aligned}
a_{n+1} &\leqslant a_{n+1}+\tilde{c}_{n+1}+\Phi_{n+1} \\
&\leqslant \frac{1}{1+\omega}(a_n+\Phi_n+\tilde{c}_n) \leqslant \cdots \leqslant \frac{1}{(1+\omega)^n}(a_1+\Phi_1+\tilde{c}_1).
\end{aligned} \tag{2.47}$$

and according to the Definition 1.2, the sequence x_n converges R-linearly to $\bar{x}$.

$\square$

Remark 2.6. Since β^{-1}-cocoercivity implies β-Lipschitz continuity for $\beta > 0$. This algorithm can be applied when C is β^{-1}-cocoercive with the following modifications:

$$\lambda < \min\left\{\frac{(2-\eta)(1+\xi)}{2\beta}, \frac{1-\xi(1-\eta)}{\eta\beta}\right\}, \tag{2.48}$$

in place of (2.17), and the proof is carried out with

$$\beta' = \begin{cases} \beta - \frac{\xi}{\lambda} & \text{if } 2\xi \leqslant \lambda\beta \\ \frac{\xi}{\lambda} & \text{if } 2\xi > \lambda\beta, \end{cases} \tag{2.49}$$

in place of (2.20). Observe that λ can be computed from the definition of ε in (2.32). The computations in order to obtain β' in (2.49) can be found in ([12], Lemma 4.1).

2.3 Three Operator splitting

We close this section showing the linear convergence of the algorithm proposed in ([12], section 5). Let us recall the theorem:

Theorem 2.7. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $B : \mathcal{H} \to \mathcal{H}$ be monotone and L_1-Lipschitz, and let $C : \mathcal{H} \to \mathcal{H}$ be $\frac{1}{L_2}$-cocoercive.*
Suppose that

$$\operatorname{zer}(A+B+C) \neq \varnothing,$$

and

$$\lambda \in \left(0, \frac{2}{4L_1 + L_2}\right). \tag{2.50}$$

Given $x_0, x_1 \in \mathcal{H}$ define the sequence $\{x_n\}_{n\in\mathbb{N}}$ according to

$$x_{n+1} = \mathrm{J}_{\lambda A}(x_n - 2\lambda Bx_n + \lambda Bx_{n-1} - \lambda Cx_n). \tag{2.51}$$

Then the following hold:

(i) $x_n \rightharpoonup \bar{x}$, for some $\bar{x} \in \operatorname{zer}(A+B+C)$.

(ii) If A is ρ-strongly monotone, then x_n converges $R-$linearly to the unique point $\bar{x} \in \operatorname{zer}(A+B+C)$.

Proof. (i) See ([12],Theorem 5.2).
(ii) From (2.50), choose $\varepsilon > 0$ such that

$$\varepsilon < 1 - 2\lambda L_1 - \lambda\frac{L_2}{2}, \tag{2.52}$$

and let

$$\bar{x} \in \operatorname{zer}(A+B+C).$$

By Proposition 1.18 the resolvent $J_{\lambda A}$ is $(1+\lambda\rho)$-cocoercive and

$$\bar{x} = J_{\lambda A}(\bar{x} - \lambda(B\bar{x} + C\bar{x})). \tag{2.53}$$

Combining cocoercivity, (2.51) and (2.53) we obtain

$$
\begin{aligned}
&(1+\lambda\rho)\|x_{n+1}-\bar{x}\|^2 && (2.54)\\
&= (1+\lambda\rho)\|\,\mathrm{J}_{\lambda A}(x_n - 2\lambda Bx_n + \lambda Bx_{n-1} - \lambda Cx_n) - J_{\lambda A}(\bar{x} - \lambda(B\bar{x}+C\bar{x}))\|^2 && (2.55)\\
&\underset{(1.10)}{\leqslant} \langle\, x_{n+1}-\bar{x}, x_n - \bar{x}\,\rangle && (2.56)\\
&\quad + \lambda\,\langle\, x_{n+1}-\bar{x}, B\bar{x} - Bx_n\,\rangle && (2.57)\\
&\quad - \lambda\,\langle\, x_{n+1}-\bar{x}, Bx_n - Bx_{n-1}\,\rangle && (2.58)\\
&\quad + \lambda\,\langle\, x_{n+1}-\bar{x}, C\bar{x} - Cx_n\,\rangle && (2.59)
\end{aligned}
$$

Doing some computations, we have that

$$(2.56) \underset{(1.1)}{=} \frac{1}{2}\left(\|x_{n+1}-\bar{x}\|^2 + \|x_n-\bar{x}\|^2 - \|x_{n+1}-x_n\|^2\right). \tag{2.60}$$

$$(2.57) = \lambda\underbrace{\langle\, x_{n+1}-\bar{x}, B\bar{x} - Bx_{n+1}\,\rangle}_{\leqslant 0} + \lambda\,\langle\, x_{n+1}-\bar{x}, Bx_{n+1} - Bx_n\,\rangle. \tag{2.61}$$

$$
\begin{aligned}
(2.58) &= -\lambda\,\langle\, x_n-\bar{x}, Bx_n - Bx_{n-1}\,\rangle - \lambda\,\langle\, x_{n+1}-x_n, Bx_n - Bx_{n-1}\,\rangle\\
&\leqslant -\lambda\,\langle\, x_n-\bar{x}, Bx_n - Bx_{n-1}\,\rangle + \lambda L_1\|x_{n+1}-x_n\|\|x_n-x_{n-1}\|\\
&\leqslant -\lambda\,\langle\, x_n-\bar{x}, Bx_n - Bx_{n-1}\,\rangle + \frac{\lambda L_1}{2}(\|x_{n+1}-x_n\|^2 + \|x_n-x_{n-1}\|^2). && (2.62)
\end{aligned}
$$

$$
\begin{aligned}
(2.59) &= -\lambda\,\langle\, x_n-\bar{x}, Cx_n - C\bar{x}\,\rangle + \lambda\,\langle\, x_{n+1}-x_n, C\bar{x} - Cx_n\,\rangle\\
&\underset{C\text{ cocoercive}}{\leqslant} -\frac{\lambda}{L_2}\|C\bar{x}-Cx_n\|^2 + \lambda\|x_{n+1}-x_n\|\|C\bar{x}-Cx_n\|\\
&\underset{\substack{(1.2)\\ \delta=\frac{2}{L_2}}}{\leqslant} -\frac{\lambda}{L_2}\|C\bar{x}-Cx_n\|^2 + \frac{\lambda}{L_2}\|C\bar{x}-Cx_n\|^2 + \frac{\lambda L_2}{4}\|x_{n+1}-x_n\|^2\\
&= \frac{\lambda L_2}{4}\|x_{n+1}-x_n\|^2. && (2.63)
\end{aligned}
$$

Multiplying each equation (2.56)-(2.59) by 2 and substituting them by (2.60)-(2.63) respectively, we obtain

$$
\begin{aligned}
(1+2\lambda\rho)\|x_{n+1}-\bar{x}\|^2 - 2\lambda\langle x_{n+1}-\bar{x}, Bx_{n+1}-Bx_n\rangle + \lambda L_1\|x_{n+1}-x_n\|^2 \\
+ \underbrace{\left(1-2\lambda\left(L_1+\frac{L_2}{4}\right)\right)}_{\geqslant \varepsilon \text{ by } (2.52)}\|x_{n+1}-x_n\|^2 \\
\leqslant \|x_n-\bar{x}\|^2 - 2\lambda\langle x_n-\bar{x}, Bx_n-Bx_{n-1}\rangle + \lambda L_1\|x_n-x_{n-1}\|^2
\end{aligned}
\tag{2.64}
$$

Next, let us capture the main information of this inequality in a function Φ:

$$
\Phi_n := \frac{1}{2}\|x_n-\bar{x}\|^2 - 2\lambda\langle x_n-\bar{x}, Bx_n-Bx_{n-1}\rangle + \lambda L_1\|x_n-x_{n-1}\|^2 \tag{2.65}
$$

then

$$
\begin{aligned}
\Phi_n &\geqslant \frac{1}{2}\|x_n-\bar{x}\|^2 - 2\lambda L_1\|x_n-\bar{x}\|\|x_n-x_{n-1}\| + \lambda L_1\|x_n-x_{n-1}\|^2 \\
&\geqslant \frac{1-2\lambda L_1}{2}\|x_n-\bar{x}\|^2 \\
&\geqslant \frac{\varepsilon}{2}\|x_n-\bar{x}\|^2.
\end{aligned}
\tag{2.66}
$$

$$
\begin{aligned}
\Phi_n &\leqslant \frac{1}{2}\|x_n-\bar{x}\|^2 + 2\lambda L_1\|x_n-\bar{x}\|\|x_n-x_{n-1}\| + \lambda L_1\|x_n-x_{n-1}\|^2 \\
&\leqslant \frac{1+2\lambda L_1}{2}\|x_n-\bar{x}\|^2 + 2\lambda L_1\|x_n-x_{n-1}\|^2 \\
&< \left(1-\frac{\varepsilon}{2}\right)\|x_n-\bar{x}\|^2 + (1-\varepsilon)\|x_n-x_{n-1}\|^2
\end{aligned}
\tag{2.67}
$$

Set

$$
\mu := \min\left\{\frac{2\lambda\rho}{2-\varepsilon}, \frac{\varepsilon}{1-\varepsilon}\right\} \tag{2.68}
$$

$$
a_n := \frac{1}{2}\|x_n-\bar{x}\|^2
$$

Substituting Φ_n and a_n in (2.64) we obtain:

$$
\begin{aligned}
a_n + \Phi_n \underset{(2.64)}{\geqslant} & \ (1+4\lambda\rho)a_{n+1} + \Phi_{n+1} + \varepsilon\|x_{n+1} - x_n\|^2 \\
& = (1+4\lambda\rho)a_{n+1} + (1+\mu)\Phi_{n+1} + \varepsilon\|x_{n+1} - x_n\|^2 - \mu\Phi_{n+1} \\
& \underset{(2.67)}{\geqslant} (1+2\lambda\rho)a_{n+1} + (1+\mu)\Phi_{n+1} \\
& + \left(\lambda\rho - \mu\left(1 - \frac{\varepsilon}{2}\right)\right)\|x_{n+1} - \bar{x}\|^2 + (\varepsilon - \mu(1-\varepsilon))\|x_{n+1} - x_n\|^2 \\
& \underset{(2.68)}{\geqslant} (1 + \min\{2\lambda\rho, \mu\})(a_{n+1} + \Phi_{n+1}) \\
& \underset{\varepsilon<1}{=} (1+\mu)(a_{n+1} + \Phi_{n+1})
\end{aligned}
$$

From here, it follows that:

$$
a_{n+1} \leqslant a_{n+1} + \Phi_{n+1} \leqslant \frac{1}{1+\mu}(a_n + \Phi_n) \leqslant \cdots \leqslant \frac{1}{(1+\mu)^n}(a_1 + \Phi_1), \tag{2.69}
$$

and according to the Definition 1.2, the sequence x_n converges R-linearly to $\bar{x}$

□

2.4 Special Cases of FRB

As mentioned in ([12], Remark 2.1), there are important cases where the recursion

$$
x_{n+1} = \mathrm{J}_{\lambda_n A}\big(x_n - [\lambda_n + \lambda_{n-1}]Cx_n + \lambda_{n-1}Cx_{n-1}\big). \tag{2.70}
$$

in Theorem 2.1, (FRB) reduces or is equivalent to known algorithms:

1. If $C = 0$, then (2.70) becomes

$$
x_{n+1} = \mathrm{J}_{\lambda_n A}\, x_n
$$

   which is the proximal point algorithm.

2. If $A = N_K$ is the normal cone to a nonempty closed convex set K, C is an affine operator and $\lambda_n = \lambda$ for all n, then we obtain the projected reflected gradient method (see Ref 26 in [12]):

$$
x_{n+1} = P_K\left(x_n - \lambda B(2x_n - x_{n-1})\right), \tag{2.71}
$$

3. If $A = 0$, $\lambda_n = \lambda$ for all n, and

$$
C = \begin{pmatrix} \nabla_x \Phi(x,y) \\ -\nabla_y \Phi(x,y) \end{pmatrix} \tag{2.72}
$$

for a smooth convex-concave function $\Phi : \mathcal{H} \times \mathcal{K} \to \mathbb{R}$, then (2.70) becomes the optimistic gradient descent ascent method (OGDA):

$$\begin{aligned} x_{n+1} &= x_n - 2\lambda \nabla_x \Phi(x_n, y_n) + \lambda \nabla_x \Phi(x_{n-1}, y_{n-1}) \\ y_{n+1} &= y_n + 2\lambda \nabla_y \Phi(x_n, y_n) - \lambda \nabla_y \Phi(x_{n-1}, y_{n-1}) \end{aligned} \tag{2.73}$$

which is widely-used in saddle point problems and machine learning.

3 Composite Inclusion Problems

The aim of this section is to solve monotone inclusion problems involving a mixture of sums, linear compositions and parallel sums of operators. The strategy is to reduce this new problem to the sum of two operators satisfying the assumptions of the algorithms from section 2.
It is important to remark that the resulting algorithms (for details we refer to [9] and [10]) act on each operator separately, offering a full splitting method which solves both, the primal and dual problem simultaneaously.

The problem under investigation is the following:

Problem 3.1. Let $\mathcal{H}$ be a real Hilbert space, let $A : \mathcal{H} \to 2^{\mathcal{H}}$ a maximally monotone operator and let $C : \mathcal{H} \to \mathcal{H}$ a monotone and μ-Lipschitzian operator with $\mu \in \mathbb{R}_{++}$. Let m be a strictly positive integer and for each $i = 1, \ldots, m$, let $\mathcal{G}_i$ be a real Hilbert, $B_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ a maximally monotone operator, $D_i : \mathcal{G}_i \to 2^{\mathcal{G}_i}$ a monotone operator such that D_i^{-1} is ν_i-Lipschitzian, for $\nu_i \in \mathbb{R}_{++}$. Suppose that $L_i : \mathcal{H} \to \mathcal{G}_i$ is a nonzero linear continuous operator and let $z \in \mathcal{H}, r_i \in \mathcal{G}_i$. The problem is to solve the primal inclusion

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } z \in A\bar{x} + \sum_{i=1}^{m} L_i^*((B_i \Box D_i)(L_i\bar{x} - r_i)) + C\bar{x}, \tag{3.1}$$

together with the dual inclusion

$$\text{find } \bar{v}_i \in \mathcal{G}_i, \text{ for } i = 1, \ldots, m, \text{ such that } \exists x \in \mathcal{H} : \begin{cases} z - \sum_{i=1}^{m} L_i^* \bar{v}_i \in Ax + Cx \\ \bar{v}_i \in (B_i \Box D_i)(L_i x - r_i). \end{cases} \tag{3.2}$$

where $A \Box B$ denotes the parallel sum of two operators defined as

$$A \Box B := (A^{-1} + B^{-1})^{-1}. \tag{3.3}$$

We say that $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m) \in \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$ is a primal-dual solution to Problem 3.1 if

$$z - \sum_{i=1}^{m} L_i^* \bar{v}_i \in A\bar{x} + C\bar{x} \quad \text{ and } \quad \bar{v}_i \in (B_i \Box D_i)(L_i\bar{x} - r_i), \quad i = 1, \ldots, m. \tag{3.4}$$

The solutions to (3.1) and (3.2) are denoted by $\mathscr{P}$ and $\mathscr{D}$ respectively. Note that if $\bar{x} \in \mathscr{P}$ then there exists $\bar{v}_i \in \mathcal{G}_i$, for each $i \in \{1, \ldots, m\}$, such that $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m)$ is a primal-dual solution. On the other hand, if exists $(\bar{v}_1, \ldots, \bar{v}_m) \in \mathscr{D}$ then there exists $\bar{x} \in \mathscr{P}$ such that $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m)$ is a primal-dual solution. Moreover, if $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m)$ is a primal-dual solution, then $\bar{x} \in \mathscr{P}$ and $(\bar{v}_1, \ldots, \bar{v}_m) \in \mathscr{D}$.

3.1 FRB adapted to composite inclusion problems

Theorem 3.2. *(FRB) In the setting of Problem 3.1. Let*

$$\beta = \max\{\mu, \nu_1, \ldots, \nu_m\} + \sqrt{\sum_{i=1}^{m} \|L_i\|^2}. \tag{3.5}$$

Suppose that

$$\{\lambda_n\}_{n\in\mathbb{N}} \subseteq \left[\varepsilon, \frac{1-2\varepsilon}{2\beta}\right], \tag{3.6}$$

for some

$$\varepsilon \in \left(0, \frac{1}{2(1+\beta)}\right], \tag{3.7}$$

and that

$$z \in \operatorname{ran}\left(A + \sum_{i=1}^{m} L_i^*(B_i \square D_i)(L_i \,.\, - r_i) + C\right). \tag{3.8}$$

Given $x_0, x_1 \in \mathcal{H}$ and $v_{i_0}, v_{i_1} \in \mathcal{G}_i$, for $i = 1, \ldots, m$, consider the sequence

$$\{(x_n, v_{1_n}, \ldots, v_{m_n})\}_{n\in\mathbb{N}_1}$$

generated by:

$$\left\lfloor \begin{array}{l} \alpha_n = \lambda_n + \lambda_{n-1} \\ x_{n+1} = \mathrm{J}_{\lambda_n A}\left(x_n + \lambda_n z - \alpha_n\left(Cx_n + \sum_{i=1}^{m} L_i^* v_{i_n}\right) + \lambda_{n-1}\left(Cx_{n-1} + \sum_{i=1}^{m} L_i^* v_{i_{n-1}}\right)\right) \\ \text{for } \ i = 1, \ldots, m. \\ \left\lfloor v_{i_{n+1}} = \mathrm{J}_{\lambda_n B_i^{-1}}\left(v_{i_n} - \lambda_n r_i - \alpha_n\left(D_i^{-1} v_{i_n} - L_i x_n\right) + \lambda_{n-1}\left(D_i^{-1} v_{i_{n-1}} - L_i x_{n-1}\right)\right) \right. \end{array} \right. \tag{3.9}$$

Then the following hold:

(i) $\{(x_n, v_{1_n}, \ldots, v_{m_n})\}_{n\in\mathbb{N}}$ *converges weakly to a primal-dual solution to Problem 3.1.*

(ii) Suppose that A and B_i^{-1} are strongly monotone for each $i \in \{1, \ldots, m\}$. Then $\{(x_n, v_{1_n}, \ldots, v_{m_n})\}_{n\in\mathbb{N}}$ *converges R-linearly to the unique primal-dual solution to Problem 3.1.*

Proof. (i) Define the Hilbert Space $\mathcal{K}$ as:

$$\mathcal{K} := \mathcal{H} \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_m,$$

and the operators

$$M : \mathcal{K} \to 2^{\mathcal{K}} : (x, v_1, \ldots, v_m) \mapsto (-z + Ax) \times (r_1 + B_1^{-1} v_1) \times \cdots \times (r_m + B_m^{-1} v_m) \tag{3.10}$$

$$Q : \mathcal{K} \to \mathcal{K} : (x, v_1, \ldots, v_m) \mapsto \left(Cx + \sum_{i=1}^{m} L_i^* v_i, -L_1 x + D_1^{-1} v_1, \ldots, -L_m x + D_m^{-1} v_m \right) \tag{3.11}$$

From Proposition 1.10 and Proposition 1.17 follows that M is maximally monotone. The operator Q is monotone and β-Lipschitzian (see proof in [10], Theorem 3.1). Moreover,

$$(3.8) \Leftrightarrow \exists x \in \mathcal{H} \text{ s.t. } z \in \operatorname{ran} \left(Ax + \sum_{i=1}^{m} L_i^* (B_i \square D_i)(L_i x - r_i) + Cx \right) \tag{3.12}$$

$$\Leftrightarrow \exists (x, v_1, \ldots, v_m) \in \mathcal{K} \text{ s.t. } \begin{cases} z \in Ax + \sum_{i=1}^{m} L_i^* v_i + Cx \\ v_i \in (B_i \square D_i)(L_i x - r_i), \quad i = 1, \ldots, m. \end{cases} \tag{3.13}$$

$$\Leftrightarrow \exists (x, v_1, \ldots, v_m) \in \mathcal{K} \text{ s.t. } \begin{cases} 0 \in -z + Ax + \sum_{i=1}^{m} L_i^* v_i + Cx \\ 0 \in r_i + B_i^{-1} v_i + D_i^{-1} v_i - L_i x, \quad i = 1, \ldots, m. \end{cases}$$

$$\begin{aligned} \Leftrightarrow \exists (x, v_1, \ldots, v_m) \in \mathcal{K} \text{ s.t. } 0 \in \quad & (-z + Ax) \times (r_1 + B_1^{-1} v_1) \times, \ldots, \times (r_m + B_m^{-1} v_m) \\ + \quad & \left(Cx + \textstyle\sum_{i=1}^{m} L_i^* v_i, D_1^{-1} v_1 - L_1 x, \ldots, D_m^{-1} - L_m x \right) \end{aligned}$$

$$\Leftrightarrow \exists (x, v_1, \ldots, v_m) \in \mathcal{K} \text{ s.t. } 0 \in (M + Q)(x, v_1, \ldots, v_m). \tag{3.14}$$

(3.14) implies that $\operatorname{zer}(M + Q) \neq \varnothing$ and we can apply Theorem 2.1 with M and Q. To this end, let

$$w_j = (x_j, v_{1_j}, \ldots, v_{m_j}) \in \mathcal{K}, \quad \text{for } j = 0, 1,$$

be our starting points. Applying the recursion (2.3) with M, Q and w_j, $j = 0, 1$ we obtain:

$$\forall n \in \mathbb{N}_1 \left\lfloor \begin{aligned} q_n &= w_n - (\lambda_n + \lambda_{n-1}) Q w_n + \lambda_{n-1} Q w_{n-1} \\ w_{n+1} &= \mathrm{J}_{\lambda_n M}(q_n) \end{aligned} \right. \tag{3.15}$$

where

$$q_n = (p_n, q_{1_n}, \dots, q_{m_n}).$$

From Proposition 1.17 and Proposition 1.16 follows that

$$\begin{aligned}
\mathrm{J}_{\lambda_n M}(q_n) &\underset{(1.17)}{=} \mathrm{J}_{\lambda_n(-z+A)}(p_n) \times \mathrm{J}_{\lambda_n(r_1+B_1^{-1})}(q_{1_n}) \times \cdots \times \mathrm{J}_{\lambda_n(r_m+B_m^{-1})}(q_{m_n}) \\
&\underset{(1.12)}{=} \mathrm{J}_{\lambda_n A}(p_n + \lambda_n z) \times \mathrm{J}_{\lambda_n B_1^{-1}}(q_{1_n} - \lambda_n r_1) \times \cdots \times \mathrm{J}_{\lambda_n B_m^{-1}}(q_{m_n} - \lambda_n r_m).
\end{aligned} \tag{3.16}$$

Substituting (3.16) in (3.15) we have that $\forall n \in \mathbb{N}_1$

$$\left\lfloor \begin{array}{l}
p_n = x_n - (\lambda_n + \lambda_{n-1}) \left(Cx_n + \sum_{i=1}^m L_i^* v_{i_n} \right) + \lambda_{n-1} \left(Cx_{n-1} + \sum_{i=1}^m L_i^* v_{i_{n-1}} \right) \\
x_{n+1} = \mathrm{J}_{\lambda_n A}(p_n + \lambda_n z) \\
\text{for } \ i = 1, \dots, m. \\
\left\lfloor \begin{array}{l}
q_{i_n} = v_{i_n} - (\lambda_n + \lambda_{n-1}) \left(D_i^{-1} v_{i_n} - L_i x_n \right) + \lambda_{n-1} \left(D_i^{-1} v_{i_{n-1}} - L_i x_{n-1} \right) \\
v_{i_{n+1}} = \mathrm{J}_{\lambda_n B_i^{-1}}(q_{i_n} - \lambda_n r_i)
\end{array} \right.
\end{array} \right.$$

Replacing p_n, q_{i_n} and setting $\alpha_n = \lambda_n + \lambda_{n-1}$ we obtain exactly the iterative scheme (3.9).

Next, by Theorem 2.1(i) follows that:

$$w_n \rightharpoonup \bar{w} = (\bar{x}, \bar{v}_1, \dots, \bar{v}_m), \tag{3.17}$$

for some $\bar{w} \in \operatorname{zer}(M + Q)$. Therefore, since (3.14) $\Rightarrow$ (3.12) and (3.13)

$$x_n \rightharpoonup \bar{x} \in \mathscr{P}, \tag{3.18}$$

and

$$(v_{1_n}, \dots, v_{m_n}) \rightharpoonup (\bar{v}_1, \dots, \bar{v}_m) \in \mathscr{D}. \tag{3.19}$$

(ii) This is straightforward, we only need to show that the operator M is strongly monotone with constant ϕ for some $\phi > 0$.

Suppose that A is ρ-strongly monotone, and B_i^{-1} is τ_i-strongly monotone, with $\rho, \tau_i > 0$ for $i = 1, \dots, m$. Set

$$\phi = \min\{\rho, \tau_1, \dots, \tau_m\}. \tag{3.20}$$

Let $(x, v_1, \ldots, v_m), (y, w_1, \ldots, w_m) \in \mathcal{K}$. Then

$$\begin{aligned}
&\langle M(x, v_1, \ldots, v_m) - M(y, w_1, \ldots, w_m), (x, v_1, \ldots, v_m) - (y, w_1, \ldots, w_m) \rangle_{\mathcal{K}} \\
&= \langle Ax - Ay, x - y \rangle_{\mathcal{H}} + \sum_{i=1}^{m} \langle B_i^{-1}(v_i) - B_i^{-1}(w_i), v_i - w_i \rangle_{\mathcal{G}_i} \\
&\underset{(1.6)}{\geqslant} \rho \|x - y\|_{\mathcal{H}}^2 + \sum_{i=1}^{m} \tau_i \|v_i - w_i\|_{\mathcal{G}_i}^2 \\
&\underset{(3.20)}{\geqslant} \phi \left(\|x - y\|_{\mathcal{H}}^2 + \sum_{i=1}^{m} \|v_i - w_i\|_{\mathcal{G}_i}^2 \right) \\
&= \phi \|(x, v_1, \ldots, v_m) - (y, w_1, \ldots, w_m)\|_{\mathcal{K}}^2
\end{aligned}$$

Therefore M is strongly monotone and the R-linear convergence follows from theorem 2.1(ii).

□

3.2 RIFRB adapted to composite inclusion problems

Theorem 3.3. *(RIFRB) In the setting of Problem 3.1. Let*

$$\beta = \max\{\mu, \nu_1, \ldots, \nu_m\} + \sqrt{\sum_{i=1}^{m} \|L_i\|^2}. \tag{3.21}$$

.

Suppose that $\eta \in (0, 1]$, $\xi \in [0, 1)$ *and* $\lambda \in \mathbb{R}_{++}$ *such that*

$$\xi < \frac{2 - \eta}{2 + \eta}, \tag{3.22}$$

and

$$\lambda < \min \left\{ \frac{2(1 - \xi) - \eta(1 + \xi)}{2\beta}, \frac{1 - \xi(1 + \eta)}{\eta\beta} \right\}. \tag{3.23}$$

Moreover, assume that

$$z \in \operatorname{ran} \left(A + \sum_{i=1}^{m} L_i^*(B_i \square D_i)(L_i \cdot - r_i) + C \right). \tag{3.24}$$

Given $x_0, x_1 \in \mathcal{H}$ *and* $v_{i_0}, v_{i_1} \in \mathcal{G}_i$, *for* $i = 1, \ldots, m$, *consider the sequence*

$$\{(x_n, v_{1_n}, \ldots, v_{m_n})\}_{n \in \mathbb{N}_1}$$

generated by:

$$\left|\begin{array}{l} p_n = x_n - \lambda\dfrac{1+\eta}{\eta}\left(Cx_n + \displaystyle\sum_{i=1}^{m} L_i^* v_{i_n}\right) + \dfrac{\lambda}{\eta}\left(Cx_{n-1} + \displaystyle\sum_{i=1}^{m} L_i^* v_{i_{n-1}}\right) + \dfrac{\xi}{\eta}(x_n - x_{n-1}) \\ x_{n+1} = (1-\eta)x_n + \eta\,\mathrm{J}_{\lambda A}(p_n + \lambda z) \\ \text{for } \ i = 1,\ldots,m. \\ \left\lfloor\begin{array}{l} q_{i_n} = v_{i_n} - \lambda\dfrac{1+\eta}{\eta}\left(D_i^{-1} v_{i_n} - L_i x_n\right) + \dfrac{\lambda}{\eta}\left(D_i^{-1} v_{i_{n-1}} - L_i x_{n-1}\right) + \dfrac{\xi}{\eta}(v_{i_n} - v_{i_{n-1}}) \\ v_{i_{n+1}} = (1-\eta)v_{i_n} + \eta\,\mathrm{J}_{\lambda B_i^{-1}}(q_{i_n} - \lambda r_i) \end{array}\right. \end{array}\right. \tag{3.25}$$

Then the following hold:

(i) $\{(x_n, v_{1_n}, \ldots, v_{m_n})\}_{n\in\mathbb{N}}$ *converges weakly to a primal-dual solution to Problem 3.1.*

(ii) Suppose that A *and* B_i^{-1} *are strongly monotone for each* $i \in \{1,\ldots,m\}$*. Then* $\{(x_n, v_{1_n}, \ldots, v_{m_n})\}_{n\in\mathbb{N}}$ *converges R-linearly to the unique primal-dual solution to Problem 3.1.*

Proof. For $j = 0, 1$, set

$$w_j = (x_j, v_{1_j}, \ldots, v_{m_j}) \in \mathcal{K},$$

Applying the recursion (2.18) in Theorem 2.4 to M and Q with w_j, $j = 0, 1$ as starting points, we obtain:

$$\forall n \in \mathbb{N}_1 \left\lfloor\begin{array}{l} q_n = w_n - \lambda Q w_n - \dfrac{\lambda}{\eta}(Qw_n - Qw_{n-1}) + \dfrac{\xi}{\eta}(w_n - w_{n-1}) \\ w_{n+1} = (1-\eta)w_n + \eta\,\mathrm{J}_{\lambda M}(q_n) \end{array}\right. \tag{3.26}$$

where

$$q_n := (p_n, q_{1_n}, \ldots q_{m_n}). \tag{3.27}$$

Substituting the resolvent of M, (computed in (3.16)) in (3.26) we obtain $\forall n \in \mathbb{N}_1$:

$$
\left\lfloor
\begin{aligned}
& p_n = x_n - \lambda \left(Cx_n + \sum_{i=1}^{m} L_i^* v_{i_n} \right) \\
& \qquad - \frac{\lambda}{\eta} \left(Cx_n + \sum_{i=1}^{m} L_i^* v_{i_n} - Cx_{n-1} - \sum_{i=1}^{m} L_i^* v_{i_{n-1}} \right) + \frac{\xi}{\eta}(x_n - x_{n-1}) \\
& x_{n+1} = (1-\eta)x_n + \eta \, \mathrm{J}_{\lambda A}(p_n + \lambda z) \\
& \text{for } \ i = 1, \dots, m. \\
& \quad \left\lfloor
\begin{aligned}
& q_{i_n} = v_{i_n} - \lambda \left(D_i^{-1} v_{i_n} - L_i x_n \right) \\
& \qquad - \frac{\lambda}{\eta} \left(D_i^{-1} v_{i_n} - L_i x_n - D_i^{-1} v_{i_{n-1}} + L_i x_{n-1} \right) + \frac{\xi}{\eta}(v_{i_n} - v_{i_{n-1}}) \\
& v_{i_{n+1}} = (1-\eta) v_{i_n} + \eta \, \mathrm{J}_{\lambda B_i^{-1}}(q_{i_n} - \lambda r_i)
\end{aligned}
\right.
\end{aligned}
\right. \tag{3.28}
$$

Factoring (3.28) we obtain exactly the iterative scheme (3.25). Next, by Theorem 2.4(i) follows that

$$
w_n \rightharpoonup \bar{w} = (\bar{x}, \bar{v}_1, \dots, \bar{v}_m), \tag{3.29}
$$

for some $\bar{w} \in \operatorname{zer}(M+Q)$. Therefore, since (3.14) $\Rightarrow$ (3.12) and (3.13)

$$
x_n \rightharpoonup \bar{x} \in \mathscr{P}, \tag{3.30}
$$

and

$$
(v_{1_n}, \dots, v_{m_n}) \rightharpoonup (\bar{v}_1, \dots, \bar{v}_m) \in \mathscr{D}. \tag{3.31}
$$

(ii) We have already shown that M is strongly monotone when A and B_i^{-1} are strongly monotone in the last theorem. Therefore R-linear convergence follows from theorem 2.4(ii). □

3.3 Related Problems

The following remarks are intended to describe the form that **FRB** and **RIFRB** take in Problems, which are special cases of Problem 3.1. Since the dual problem is a simple substitution, we will only consider the formulation of the primal problem.

Remark 3.4. In Problem 3.1, let

$$
D_i(v) = \begin{cases} G_i, & \text{if } v = 0, \\ \varnothing, & \text{otherwise.} \end{cases} \tag{3.32}
$$

Then

$$
B_i \,\square\, D_i = B_i, \qquad \text{with } D_i^{-1}(G_i) = 0, \text{ beeing 0-Lipschitzian.} \tag{3.33}
$$

Thus, the primal problem (3.1) reduces to

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } z \in A\bar{x} + \sum_{i=1}^{m} L_i^*(B_i(L_i\bar{x} - r_i)) + C\bar{x}. \tag{3.34}$$

- The update rule (3.9) of **FRB** becomes:

$$\left\lfloor \begin{array}{l} \alpha_n = \lambda_n + \lambda_{n-1} \\ x_{n+1} = \mathrm{J}_{\lambda_n A}\left(x_n + \lambda_n z - \alpha_n\left(Cx_n + \sum_{i=1}^{m} L_i^* v_{i_n}\right) + \lambda_{n-1}\left(Cx_{n-1} + \sum_{i=1}^{m} L_i^* v_{i_{n-1}}\right)\right) \\ \text{for } \; i = 1, \dots, m. \\ \left\lfloor v_{i_{n+1}} = \mathrm{J}_{\lambda_n B_i^{-1}}\left(v_{i_n} - \lambda_n r_i + \alpha_n L_i x_n - \lambda_{n-1} L_i x_{n-1}\right) \right. \end{array} \right. \tag{3.35}$$

- The update rule (3.25) of **RIFRB** becomes:

$$\left\lfloor \begin{array}{l} p_n = x_n - \lambda\dfrac{1+\eta}{\eta}\left(Cx_n + \sum_{i=1}^{m} L_i^* v_{i_n}\right) + \dfrac{\lambda}{\eta}\left(Cx_{n-1} + \sum_{i=1}^{m} L_i^* v_{i_{n-1}}\right) + \dfrac{\xi}{\eta}(x_n - x_{n-1}) \\ x_{n+1} = (1-\eta)x_n + \eta\,\mathrm{J}_{\lambda A}(p_n + \lambda z) \\ \text{for } \; i = 1, \dots, m. \\ \left\lfloor \begin{array}{l} q_{i_n} = v_{i_n} + \lambda\dfrac{1+\eta}{\eta}L_i x_n - \dfrac{\lambda}{\eta}L_i x_{n-1} + \dfrac{\xi}{\eta}(v_{i_n} - v_{i_{n-1}}) \\ v_{i_{n+1}} = (1-\eta)v_{i_n} + \eta\,\mathrm{J}_{\lambda B_i^{-1}}(q_{i_n} - \lambda r_i) \end{array} \right. \end{array} \right. \tag{3.36}$$

Remark 3.5. Under the assumption given in the above remark, further consider $z = 0$, $r_i = 0$ for each $i \in \{1, \dots, m\}$, and

$$Cx = 0 \quad \forall x \in \mathcal{H}. \tag{3.37}$$

Then the primal problem (3.1) reduces to

$$\text{find } \bar{x} \in \mathcal{H} \text{ such that } 0 \in A\bar{x} + \sum_{i=1}^{m} L_i^*(B_i(L_i\bar{x})). \tag{3.38}$$

- The update rule (3.9) of **FRB** becomes:

$$\left\lfloor \begin{array}{l} \alpha_n = \lambda_n + \lambda_{n-1} \\ x_{n+1} = \mathrm{J}_{\lambda_n A}\left(x_n - \alpha_n \sum_{i=1}^{m} L_i^* v_{i_n} + \lambda_{n-1}\sum_{i=1}^{m} L_i^* v_{i_{n-1}}\right) \\ \text{for } \; i = 1, \dots, m. \\ \left\lfloor v_{i_{n+1}} = \mathrm{J}_{\lambda_n B_i^{-1}}\left(v_{i_n} + \alpha_n L_i x_n - \lambda_{n-1} L_i x_{n-1}\right) \right. \end{array} \right. \tag{3.39}$$

- The update rule (3.25) of **RIFRB** becomes:

$$
\left\lfloor
\begin{array}{l}
p_n = x_n - \lambda \dfrac{1+\eta}{\eta} \displaystyle\sum_{i=1}^{m} L_i^* v_{i_n} + \dfrac{\lambda}{\eta} \displaystyle\sum_{i=1}^{m} L_i^* v_{i_{n-1}} + \dfrac{\xi}{\eta}(x_n - x_{n-1}) \\
x_{n+1} = (1-\eta)x_n + \eta \, \mathrm{J}_{\lambda A}(p_n) \\
\text{for } \ i = 1, \dots, m. \\
\left\lfloor
\begin{array}{l}
q_{i_n} = v_{i_n} + \lambda \dfrac{1+\eta}{\eta} L_i x_n - \dfrac{\lambda}{\eta} L_i x_{n-1} + \dfrac{\xi}{\eta}(v_{i_n} - v_{i_{n-1}}) \\
v_{i_{n+1}} = (1-\eta) v_{i_n} + \eta \, \mathrm{J}_{\lambda B_i^{-1}}(q_{i_n})
\end{array}
\right.
\end{array}
\right. \tag{3.40}
$$

4 Convex minimization problems

In this section we provide an application of the algorithms **FRB**, and **RIFRB** to convex minimization problems by revisiting ([10], Problem 4.1). Before we start, I would like to point out, that there is no original idea in this section, except for the adaptation of the algorithms to the problem below, which is straightforward, as soon as the necessary substitutions are justified. For the sake of completeness and the readers convenience, I prefer to include relevant proofs and analysis which can be found in its original work [10].

Problem 4.1. Let $\mathcal{H}$ be a real Hilbert space, let $z \in \mathcal{H}$, let $f \in \Gamma(\mathcal{H})$, and let $h : \mathcal{H} \to \mathbb{R}$ be convex and differentiable with μ-Lipschitzian gradient for some $\mu \in \mathbb{R}_{++}$. For each $i = 1, \dots m$, let $\mathcal{G}_i$ be a real Hilbert space, let $r_i \in \mathcal{G}_i$, let $g_i, l_i \in \Gamma(\mathcal{G}_i)$, where l_i is $\frac{1}{\nu_i}$-strongly convex with $\nu_i \in \mathbb{R}_{++}$ and suppose that $L_i \in \mathscr{B}(\mathcal{H}, \mathcal{G}_i)$ is nonzero operator. Consider the problem:

$$\underset{x \in \mathcal{H}}{\text{minimize}}\, f(x) + \sum_{i=1}^{m} (g_i \,\square\, l_i)(L_i x - r_i) + h(x) - \langle\, x, z \,\rangle\,, \tag{4.1}$$

and its dual problem

$$\underset{v_1 \in \mathcal{G}_1, \dots, v_m \in \mathcal{G}_m}{\text{minimize}} (f^* \,\square\, h^*) \left(z - \sum_{i=1}^{m} L_i^* v_i \right) + \sum_{i=1}^{m} (g_i^*(v_i) + l_i^*(v_i) + \langle\, v_i, r_i \,\rangle). \tag{4.2}$$

We start reducing Problem 4.1 to the form given in Definition 1.46 by a special choice of Hilbert spaces.

Proposition 4.2. *The primal problem* (4.1) *and its dual problem* (4.2) *are equivalent to*

$$\underset{x \in \mathcal{H}}{\text{minimize}}\, F(x) + (G \circ L)(x) \tag{4.3}$$

$$\underset{v \in \mathcal{G}}{\text{minimize}}\, F^*(-L^* v) + G^*(v) \tag{4.4}$$

respectively, with $F \in \Gamma(\mathcal{H})$, $G \in \Gamma(\mathcal{G})$ *and* $L \in \mathscr{B}(\mathcal{H}, \mathcal{G})$ *where*

$$\mathcal{G} := \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_m,$$

$$L : \mathcal{H} \to \mathcal{G} \quad : \quad x \mapsto \bigtimes_{i=1}^{m} L_i x, \tag{4.5}$$

$$L^* : \mathcal{G} \to \mathcal{H} \quad : \quad v \mapsto \sum_{i=1}^{m} L_i^* v_i, \tag{4.6}$$

$$F : \mathcal{H} \to \mathbb{R} \quad : \quad x \mapsto f(x) + h(x) - \langle x, z \rangle, \tag{4.7}$$

$$G : \mathcal{G} \to \mathbb{R} \quad : \quad (w_1, \dots, w_m) \mapsto \sum_{i=1}^{m} (g_i \,\Box\, l_i)(w_i - r_i). \tag{4.8}$$

Moreover,

$$\operatorname{dom} F = \operatorname{dom} f, \tag{4.9}$$

$$\operatorname{dom} G = \bigtimes_{i=1}^{m} \operatorname{dom} g_i + \operatorname{dom} l_i + r_i. \tag{4.10}$$

Proof. A simple substitution using (4.5), (4.7) and (4.8) shows the equivalence between (4.1) and (4.3).

Since F is a finite addtion of functions in $\Gamma(\mathcal{H})$, it follows that $F \in \Gamma(\mathcal{H})$ as well. Moreover, since $\operatorname{dom} h = \mathcal{H}$, we have that

$$\operatorname{dom} F = \operatorname{dom}(f + h - \langle \cdot, z \rangle) = \operatorname{dom} f. \tag{4.11}$$

The infinal convolution of convex functions is always convex, but it is not neccessarily lower-semicontinuous (see [1], Example 12.13). In this case, $g_i \,\Box\, l_i$ is lower-semicontinuous since l_i is strongly convex, thus by (1.59) $\operatorname{dom} l_i^* = \mathcal{G}_i$ and it follows that

$$(g_i^* + l_i^*)^* \underset{(1.41)}{=} g_i^{**} \boxdot l_i^{**} \in \Gamma(\mathcal{G}_i) \underset{(1.27)}{=} g_i \boxdot l_i \tag{4.12}$$

Thus, $G \in \Gamma(\mathcal{G})$ as a finite addition of functions in $\Gamma(\mathcal{G}_i)$, with

$$\operatorname{dom} G = \bigtimes_{i=1}^{m} \operatorname{dom}(g_i \,\Box\, l_i)(\cdot - r_i) = \bigtimes_{i=1}^{m} \operatorname{dom} g_i + \operatorname{dom} l_i + r_i. \tag{4.13}$$

Next, we compute their Fenchel conjugates:

$$\begin{aligned} F^*(u) &= \sup_{x \in \mathcal{H}} \{ \langle x, u + z \rangle - f(x) - h(x) \} \\ &= (f + h)^*(u + z) \\ &\underset{\substack{(1.41) \\ \operatorname{dom} h = \mathcal{H}}}{=} (f^* \boxdot h^*)(u + z) \end{aligned} \tag{4.14}$$

$$
\begin{aligned}
G^*(v) &\underset{(1.31)}{=} \sum_{i=1}^{m} (g_i \,\square\, l_i)^*(v_i) + \langle\, v_i, r_i \,\rangle \\
&\underset{(1.40)}{=} \sum_{i=1}^{m} g_i^*(v_i) + l_i^*(v_i) + \langle\, v_i, r_i \,\rangle
\end{aligned} \tag{4.15}
$$

where $v := (v_1, \ldots, v_m) \in \mathcal{G}$.

Finally, substituting (4.6), (4.14) and (4.15) shows the equivalence between (4.2) and (4.4).

□

By the Fermat's rule we know that

$$
\bar{x} \in \operatorname{Argmin}(F + G \circ L) \Leftrightarrow 0 \in \partial(F + G \circ L)(\bar{x}). \tag{4.16}
$$

On the other hand

$$
\partial F(\bar{x}) + (L^* \circ (\partial G) \circ L)(\bar{x}) \underset{(1.23)}{\subseteq} \partial(F + G \circ L)(\bar{x}). \tag{4.17}
$$

Observe that if there exists a solution to problem (4.3) and we can guarantee equality in (4.17), then we can apply **FRB** and **RIFRB** to find a point $\bar{x} \in \mathcal{H}$ such that

$$
0 \in \partial F(\bar{x}) + (L^* \circ (\partial G) \circ L)(\bar{x}). \tag{4.18}
$$

To this end, observe that, since $\operatorname{dom} h = \mathcal{H}$, Proposition 1.35 and Proposition 1.37 imply

$$
\partial F = \partial(f + h - \langle\, \cdot, z \,\rangle) \underset{\substack{(1.42)\\(1.22)}}{=} \partial f + \nabla h - z. \tag{4.19}
$$

Applying Proposition 1.35 again, this time to $g_i^* + l_i^*$, (recall that $\operatorname{dom} l_i^* = \mathcal{G}_i$) and combining with Proposition 1.33, follows that for each $i = 1, \ldots, m$:

$$
\begin{aligned}
(\partial(g_i \,\square\, l_i))^{-1} \underset{(1.29)}{=} \partial(g_i \,\square\, l_i)^* \underset{(1.40)}{=} \partial(g_i^* + l_i^*) \underset{(1.42)}{=} \partial g_i^* + \partial l_i^* \underset{(1.29)}{=} (\partial g_i)^{-1} + (\partial l_i)^{-1} \\
\Rightarrow \partial(g_i \,\square\, l_i) = \partial g_i \,\square\, \partial l_i.
\end{aligned} \tag{4.20}
$$

Thus, by Proposition 1.29(ii)

$$
\partial G(\cdot) \underset{(1.32)}{=} \bigtimes_{i=1}^{m} (\partial g_i \,\square\, \partial l_i)(\cdot - r_i) \tag{4.21}
$$

Finally

$$
\partial F + L^* \circ (\partial G) \circ L = -z + \partial f + \sum_{i=1}^{m} L_i^* (\partial g_i \,\square\, \partial l_i)(L_i x - r_i) + \nabla h \tag{4.22}
$$

The connection between Problem 3.1 and Problem 4.1 is established in the proof of ([10], Theorem 4.2) setting:

$$A = \partial f, \quad C = \nabla h, \quad \text{and for each } i = 1, \dots, m, \quad B_i = \partial g_i, \quad D_i = \partial l_i. \tag{4.23}$$

From Proposition 1.37 (i) follows that the operators A and B_i, for each $i = 1, \dots, m$, are maximally monotone. The $\frac{1}{\nu_i}$ strongly convexity of l_i implies by (1.59) that l_i^* is Fréchet differentiable on $\mathcal{G}_i$ with $\frac{1}{\nu_i}$-cocoercive (resp. ν_i-Lipschitz by (1.60)) gradient and by Corollary 1.27 follows that

$$D_i^{-1} = (\partial l_i)^{-1} \underset{(1.29)}{=} \partial l_i^* \underset{(1.22)}{=} \nabla l_i^*, \tag{4.24}$$

and

$$B_i^{-1} = \partial g_i^*. \tag{4.25}$$

Alltogether, under the assumption of the existence of a primal solution, we can apply Theorem 3.2 and Theorem 3.3 to obtain the existence of a point $(\bar{x}, \bar{v}) \in \mathcal{H} \oplus \mathcal{G}$ such that

$$z \in \partial f(\bar{x}) + \sum_{i=1}^{m} L_i^* \left((\partial g_i \,\square\, \partial l_i)(L_i \bar{x} - r_i) \right) + \nabla h(\bar{x}) \tag{4.26}$$

and

$$\exists x \in \mathcal{H} : \begin{cases} z - \sum_{i=1}^{m} L_i^* \bar{v}_i \in \partial f(x) + \nabla h(x) \\ \bar{v}_i \in (\partial g_i \,\square\, \partial l_i)(L_i x - r_i), \quad \forall i = 1, \dots, m. \end{cases} \tag{4.27}$$

Observe that:

$$\partial F^* = \partial(f^* \boxdot h^*)(\cdot + z) = (\partial F)^{-1} = (\partial f + \nabla h - z)^{-1} \tag{4.28}$$

and

$$\partial G^*(\bar{v}) \underset{(1.32)}{=} \bigtimes_{i=1}^{m} \partial(g_i^* + l_i^* + r_i)(\bar{v}_i) \underset{(1.42)}{=} r + \bigtimes_{i=1}^{m} \partial(g_i^* + l_i^*)(\bar{v}_i) \tag{4.29}$$

where $r = (r_1, \dots, r_m)$.

Hence, combining (4.27), (4.28) and (4.29), we obtain

$$\exists x \in \mathcal{H} \ : \ \begin{cases} x \in \partial(f^* \boxdot h^*)(z - L^*(\bar{v})) \\ L_i x - r_i \in \partial(g_i^* + l_i^*)(\bar{v}_i), \quad \forall i = 1, \dots, m. \end{cases} \tag{4.30}$$

which is equivalent to

$$\exists x \in \mathcal{H} \ : \ \begin{cases} x \in \partial F^*(-L^*(\bar{v})) \\ Lx \in \partial G^*(\bar{v}). \end{cases} \tag{4.31}$$

Applying $-L$ to the first term in (4.31) and summing up, we obtain

$$0 \in -L(\partial F^*(-L^*(\bar{v}))) + \partial G^*(\bar{v}) \underset{(1.23)}{\subseteq} \partial(F^*(-L^*(\bar{v})) + G^*(\bar{v})) \tag{4.32}$$

and the Fermat's rule implies

$$\bar{v} \in \operatorname{Argmin} F^* \circ (-L^*) + G^* \tag{4.33}$$

Thus, $\bar{v}$ is a dual solution to Problem 4.1.

Remark 4.3 (Existence of primal solution). The existence of a primal solution, is guaranteed, for instance by Proposition 1.48, i.e., if $f + h - \langle \cdot, z \rangle$ is coercive and g_i is bounded from below for each $i = 1, \ldots m$. Moreover, if $f + h$ is strongly convex, then so ist $F + G \circ L$, and (4.1) has a unique solution. The coercivity of F can be checked applying the **Moreau-Rockafellar**, Theorem ([1], Theorem 14.17) which asserst that

$$f + h - \langle \cdot, z \rangle \text{ is coercive } \Leftrightarrow z \in \operatorname{int} \operatorname{dom} (f + h)^*, \tag{4.34}$$

or verifying if its lower level sets are bounded (see [1], Proposition 11.12).

Remark 4.4. Suppose that (4.1) has at least one solution. Applying Proposition 1.35 (i) we have equality in (4.17) if, for instance,

$$\begin{aligned} & 0 \in \operatorname{sri} (\operatorname{dom} G - L(\operatorname{dom} F)) \\ \underset{\substack{(4.9)\\(4.10)}}{\Leftrightarrow} \; & 0 \in \operatorname{sri} \left(\bigtimes_{i=1}^{m} \operatorname{dom} g_i + \operatorname{dom} l_i + r_i - L_i(\operatorname{dom} f) \right) \end{aligned}$$

Moreover, if $\mathcal{H}$ and $\mathcal{G}_i$, are finite dimensional and there exists $x \in \operatorname{ri} \operatorname{dom} f$ such that for each $i = 1, \ldots, m$

$$L_i x - r_i \in \operatorname{ri} \operatorname{dom} g_i + \operatorname{ri} \operatorname{dom} l_i$$

Then, by Proposition 1.6 follows that

$$0 \in \operatorname{sri} (\operatorname{dom} G - L(\operatorname{dom} f))$$

Finally, by Proposition 1.37 we have that

$$\begin{aligned} J_{\lambda_n A} &\underset{(1.45)}{=} \operatorname{Prox}_{\lambda_n f}, \\ J_{\lambda_n B_i^{-1}} &\underset{(1.45)}{=} \operatorname{Prox}_{\lambda_n g_i^*}. \end{aligned}$$

Now, that all substitutions are justified, we can continue with the application of the algorithms from the previous section:

4.1 FRB in convex minimization

Theorem 4.5 (Solving Problem 4.1 via FRB)**.** *In Problem 4.1, suppose that*

$$z \in \operatorname{ran}\left(\partial f + \sum_{i=1}^{m} L_i^*(\partial g_i \Box \partial l_i)(L_i\,. - r_i) + \nabla h\right). \tag{4.35}$$

Set

$$\beta = \max\{\mu, \nu_1, \dots, \nu_m\} + \sqrt{\sum_{i=1}^{m} \|L_i\|^2}. \tag{4.36}$$

and suppose that

$$\{\lambda_n\}_{n\in\mathbb{N}} \subseteq \left[\varepsilon, \frac{1-2\varepsilon}{2\beta}\right], \tag{4.37}$$

for some

$$\varepsilon \in \left(0, \frac{1}{2(1+\beta)}\right]. \tag{4.38}$$

Given $x_0, x_1 \in \mathcal{H}$, and $v_{i_0}, v_{i_1} \in \mathcal{G}_i$ for $i = 1, \dots, m$, consider the sequence

$$\{(x_n, v_{1_n}, \dots, v_{m_n})\}_{n\in\mathbb{N}_1}$$

generated by the following recursive scheme:

$$\left\lfloor\begin{array}{l} \alpha_n = \lambda_n + \lambda_{n-1} \\ p_n = \alpha_n\left(\nabla h(x_n) + \sum_{i=1}^{m} L_i^* v_{i_n}\right) - \lambda_{n-1}\left(\nabla h(x_{n-1}) + \sum_{i=1}^{m} L_i^* v_{i_{n-1}}\right) \\ x_{n+1} = \operatorname{Prox}_{\lambda_n f}(x_n + \lambda_n z - p_n) \\ \text{for } \ i = 1, \dots, m. \\ \left\lfloor\begin{array}{l} y_{i_n} = \alpha_n(\nabla l_i^* v_{i_n} - L_i x_n) - \lambda_{n-1}(\nabla l_i^* v_{i_{n-1}} - L_i x_{n-1}) \\ v_{i_{n+1}} = \operatorname{Prox}_{\lambda_n g_i^*}(v_{i_n} - \lambda_n r_i - y_{i_n}) \end{array}\right. \end{array}\right. \tag{4.39}$$

Then the following hold:

(i) $\{(x_n, v_{1_n}, \dots, v_{m_n})\}_{n\in\mathbb{N}}$ *converges weakly to a primal-dual solution to Problem 4.1.*

(ii) Suppose that f and g_i^* are strongly convex for each $i \in \{1, \dots, m\}$. Then $\{(x_n, v_{1_n}, \dots, v_{m_n})\}_{n\in\mathbb{N}}$ converges R-linearly to the unique primal-dual solution to Problem 4.1.

Proof. (i) Since the substitutions has been justified at the beginning of this section, the assertion follows from Theorem 3.2 (i)

(ii) The strongly convexity of f, and g_i implies by Proposition 1.44 (iii) that their subdifferentials are strongly monotone and the assertion follows from Theorem 3.2 (ii) $\square$

4.2 RIFRB in convex minimization

Theorem 4.6 (Solving Problem 4.1 via Relaxed Inertial FRB). *In Problem 4.1, suppose that*

$$z \in \operatorname{ran}\left(\partial f + \sum_{i=1}^{m} L_i^*(\partial g_i \square \partial l_i)(L_i \,.\, - r_i) + \nabla h\right) \tag{4.40}$$

Set

$$\beta = \max\{\mu, \nu_1, \ldots, \nu_m\} + \sqrt{\sum_{i=1}^{m} \|L_i\|^2} \tag{4.41}$$

Moreover, suppose that $\eta \in (0,1], \xi \in [0,1)$ and $\lambda \in \mathbb{R}_{++}$ such that

$$\xi < \frac{2-\eta}{2+\eta} \tag{4.42}$$

and

$$\lambda < \min\left\{\frac{2(1-\xi) - \eta(1+\xi)}{2\beta}, \frac{1-\xi(1+\eta)}{\eta\beta}\right\}. \tag{4.43}$$

Given $x_0, x_1 \in \mathcal{H}$, and $v_{i_0}, v_{i_1} \in \mathcal{G}_i$ for $i = 1, \ldots, m$, consider the sequence

$$\{(x_n, v_{1_n}, \ldots, v_{m_n})\}_{n\in\mathbb{N}}$$

generated by the following recursive scheme:

$$\left\lfloor \begin{array}{l} p_n = x_n - \lambda\dfrac{1+\eta}{\eta}\left(\nabla h(x_n) + \displaystyle\sum_{i=1}^{m} L_i^* v_{i_n}\right) + \dfrac{\lambda}{\eta}\left(\nabla h(x_{n-1}) + \displaystyle\sum_{i=1}^{m} L_i^* v_{i_{n-1}}\right) + \dfrac{\xi}{\eta}(x_n - x_{n-1}) \\ x_{n+1} = (1-\eta)x_n + \eta \operatorname{Prox}_{\lambda f}(p_n + \lambda z) \\ \text{for } \ i = 1, \ldots, m. \\ \quad \left\lfloor \begin{array}{l} q_{i_n} = v_{i_n} - \lambda\dfrac{1+\eta}{\eta}\left(\nabla l_i^*(v_{i_n}) - L_i x_n\right) + \dfrac{\lambda}{\eta}\left(\nabla l_i^*(v_{i_{n-1}}) - L_i x_{n-1}\right) + \dfrac{\xi}{\eta}(v_{i_n} - v_{i_{n-1}}) \\ v_{i_{n+1}} = (1-\eta)v_{i_n} + \eta \operatorname{Prox}_{\lambda g_i^*}(q_{i_n} - \lambda r_i) \end{array} \right. \end{array} \right. \tag{4.44}$$

Then the following hold:

(i) $\{(x_n, v_{1_n}, \ldots, v_{m_n})\}_{n\in\mathbb{N}}$ *converges weakly to a primal-dual solution to Problem 4.1.*

(ii) If f and g_i^* are strongly convex, for each $i \in \{1, \dots, m\}$, then $\{(x_n, v_{1_n}, \dots, v_{m_n})\}_{n\in\mathbb{N}}$ converges R-linearly to the unique primal-dual solution to Problem 4.1.

Proof. (i) Since the substitutions has been justified at the beginning of this section, the assertion follows from Theorem (3.2) (i)

(ii) The strongly convexity of f, and g_i implies by Proposition 1.44 (iii) that their subdifferentials are strongly monotone and the assertion follows from Theorem 3.2 (ii) □

4.3 Naive stepsize strategy for FRB (FRBD)

The idea is very simple, since the FRB algorithm converges for any starting points x_0, x_1, as long as λ_n satisfies (4.37), which can be written as

$$\lambda_{\min} \leqslant \lambda_n \leqslant \lambda_{\max} \quad \forall n \in \mathbb{N}, \tag{4.45}$$

we can try to find a good λ_n for the current iteration. For simplicity, denote the recursive scheme (4.39) by

$$x_{n+1} = \mathrm{FRB}(x_n, x_{n-1}, \lambda_n, \lambda_{n-1})$$

Now, we will assume that when the algorithm starts, it is important to use a big stepsize, no matter in which direction we are moving. That means that we will break the above restriction for a while, making use of a constant $M > 1$ and settting

$$\lambda_1 = M\lambda_{\max}.$$

We also want to be able to either reduce or enlarge the stepsize depending on the behaviour of the objective function at x_n. In order to be clear, suppose that we want to minimize the function F, and we observe that

$$F(x_{n+1}) < F(x_n) < F(x_{n-1}) < \cdots < F(x_{n-p}),$$

for some fixed positive integer $p < n$, then we will enlarge the stepsize in the hope of avoid some iterations. If $F(x_{n+1}) > F(x_n) > F(x_{n-1}) > \cdots > F(x_{n-p})$, then we will reduce the stepsize assuming that the current one is too large. For this enlarge/reduce procedure we will use a finite sequence $\{D_1, \dots, D_m\}$ in $(0, 1]^m$ and two positive constants, called *DecMax* and *IncMax*, whose goal is to suggest when a new stepsize could be more efficient.

Finally, we will always check that $\lambda_n > \lambda_{\min}$, since if λ_n becomes too small the resolvent is nothing but the identity and we will get stuck. After a fixed number of iterations we will also check that $\lambda_n < \lambda_{\max}$, in order to ensure convergence.

This idea is better described with pseudocode below and in our numerical experiments FRB with this stepsize strategy will be called FRBD.

Algorithm 1 Current Stepsize determination in FRBD

procedure FRBD(x_0, x_1, $\lambda_{\min}$, $\lambda_{\max}$, MAXITER)
  $\lambda_0 = M_0\lambda_{\max}$ with $M_0 > 1$
  $\lambda_1 = M_1\lambda_{\max}$ with $M_1 > 1$
  $\{D_1, \ldots, D_m\} \in \times_{i=1}^{m}(0, 1]$
  $n := 2$
  Dec := 0
  Inc := 0
  Count := 1
  $x_{n+1} = \text{FRB}(x_n\ x_{n-1}\ \lambda_n\ \lambda_{n-1})$
  if $F(x_{n+1}) \geqslant F(x_n)$ **then**
    Inc = Inc +1
    Dec = 0
    if Count $\leqslant m$ **then**
      $\lambda_n = D_{\text{Count}}\lambda_n$
      Count = Count + 1
      goto 9
    else
      Count = 1
    end if
  else
    Count = 1
    Inc = 0
    Dec = Dec +1
  end if
  $\lambda_{n+1} = \lambda_n$
  $n = n + 1$
  if Dec = $\text{Dec}_{\max}$ **then** ▷ number of consecutive decreasing function evaluations
    $\lambda_{n+1} = \lambda_{n+1}D_1^{-1}D_2^{-1}\ldots D_s^{-1}$ ▷ where $1 \leqslant s \leqslant m$
  end if
  if Inc = $\text{Inc}_{\max}$ **then** ▷ number of consecutive increasing function evaluations
    $\lambda_{n+1} = \lambda_{n+1}D_1D_2\ldots D_p$ ▷ where $1 \leqslant p \leqslant m$
  end if
  if $\lambda_{n+1} < \lambda_{\min}$ **then**
    $\lambda_{n+1} = \frac{\lambda_{\min}+\lambda_{\max}}{2}$
  end if
  if $n \geqslant \frac{\text{MAXITER}}{N_0}$ **then** ▷ where $1 < N_0 \leqslant \text{MAXITER}$
    $\lambda_{n+1} = \max\{\lambda_{\min}, \min\{\lambda_{n+1}, \lambda_{\max}\}\}$
  end if
  goto 9
end procedure

4.4 Algorithms for comparison

In this subsection we breafly describe the algorithms, which we use to compare performances.

4.4.1 Forward-Backward-Forward (FBF)

The error-free forward-backward-forward method proposed in ([10], Theorem 4.1) solves Problem 4.1 as follows:

Set

$$\beta = \max\{\mu, \nu_1, \ldots, \nu_m\} + \sqrt{\sum_{i=1}^{m} \|L_i\|^2}, \tag{4.46}$$

let $x_0 \in \mathcal{H}$, let $(v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_m$, let $\delta \in \left(0, \frac{1}{\beta+1}\right)$, let $\{\gamma_n\}_{n\in\mathbb{N}}$ be a sequence in $\left[\delta, \frac{(1-\delta)}{\beta}\right]$ and set

$$\left\lfloor \begin{array}{l}
y_{1,n} = x_n - \gamma_n \left(\nabla h(x_n) + \sum_{i=1}^{m} L_i^* v_{i,n} \right) \\
p_{1,n} = \operatorname{Prox}_{\gamma_n f} (y_{1,n} + \gamma_n z) \\
\text{for } \; i = 1, \ldots, m. \\
\quad \left\lfloor \begin{array}{l}
y_{2,i,n} = v_{i,n} + \gamma_n (L_i x_n - \nabla l_i^*(v_{i,n})) \\
p_{2,i,n} = \operatorname{Prox}_{\gamma_n g_i^*} (y_{2,i,n} - \gamma_n r_i) \\
q_{2,i,n} = p_{2,i,n} + \gamma_n (L_i p_{1,n} - \nabla l_i^*(p_{2,i,n})) \\
v_{i,n+1} = v_{i,n} - y_{2,i,n} + q_{2,i,n}.
\end{array} \right. \\
q_{1,n} = p_{1,n} - \gamma_n (\nabla h(p_{1,n}) + \sum_{i=1}^{m} L_i^* p_{2,i,n}) \\
x_{n+1} = x_n - y_{1,n} + q_{1,n}
\end{array} \right. \tag{4.47}$$

Then the following statements are true:

i) The sequence $\{(x_n, v_{1,n}, \ldots, v_{m,n})\}_{n\in\mathbb{N}}$ converges weakly to a primal-dual solution $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m)$ to Problem 4.1.

ii) Suppose that f or h is uniformly convex at $\bar{x}$. Then $x_n \to \bar{x}$.

iii) Suppose that g_i^* or l_i^* is uniformly convex at $\bar{v}_i$ for some $i \in \{1, \ldots, m\}$. Then $v_{i,n} \to \bar{v}_i$.

Remark 4.7. The results in ([10], Theorem 4.1) remain valid in the special case when $l_i = \iota_{\{0\}}$ and even when $h \equiv 0$. (see [10], Remark 4.4).

4.4.2 Forward-Backward (FB)

We will also use the forward-backward method proposed in ([4], Theorem 2), which is an error-free, weights-free with $\lambda_n = 1$ constant, adaptation of the algorithm proposed by Vũ (see [14], Theorem 3.1) and solves Problem 4.1 as follows:

Let τ and σ_i for $i = 1, \ldots, m$, be strictly positive numbers such that

$$2\min\{\tau^{-1}, \sigma_i^{-1}, \ldots, \sigma_m^{-1}\} \min\{\mu^{-1}, \nu_1^{-1}, \ldots, \nu_m^{-1}\} \left(1 - \sqrt{\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2}\right) > 1. \tag{4.48}$$

Let $(x_0, v_{1,0}, \ldots, v_{m,0}) \in \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$ and for any $n \geqslant 0$ set:

$$\left\lfloor \begin{array}{l} p_n = x_n - \tau \left(\sum_{i=1}^{m} L_i^* v_{i,n} + \nabla h(x_n) \right) \\ x_{n+1} = \operatorname{Prox}_{\tau f}(p_n - \tau z) \\ y_n = 2x_{n+1} - x_n \\ \text{for } \ i = 1, \ldots, m. \\ \left\lfloor \begin{array}{l} q_{i,n} = v_{i,n} + \sigma_i [L_i y_n - \nabla l_i^*(v_{i,n})] \\ v_{i,n+1} = \operatorname{Prox}_{\sigma_i g_i^*}(q_{i,n} - \sigma_i r_i) \end{array} \right. \end{array} \right. \tag{4.49}$$

Then the following statements are true:

i) the sequence $\{(x_n, v_{1,n}, \ldots, v_{m,n})\}_{n \in \mathbb{N}}$ converges weakly to a primal-dual solution $(\bar{x}, \bar{v}_1, \ldots, \bar{v}_m)$ to Problem 4.1

ii) if h is strongly convex then $x_n \to \bar{x}$.

iii) if l_i^* is strongly convex for some $i \in \{1, \ldots, m\}$, then $v_{i,n} \to \bar{v}_i$.

Remark 4.8. For the special case where $z = 0, r_i = 0, h \equiv 0$ and $l_i = \iota_{\{0\}}$ for each $i = 1, \ldots, m$, the condition (4.48) must be replaced by

$$\tau \sum_{i=1}^{m} \sigma_i \|L_i\|^2 < 1. \tag{4.50}$$

as mentioned in ([4], Remark 6).

5 Numerical experiments

In the the next two subsections, we investigate the numerical perfomance of the proposed algorithnms in the context of image processing. To this end, we formulate the "real problem" as a convex optimization, fitting the assumptions of Problem 4.1 or one of its derived problems (see Remark 3.4, Remark 3.5, Remark 4.7, Remark 4.8). We proceed as follows, each algorithm will be run twice. The first time for a number $N = 10000$ of iterations, where $X^* = X_N$ will be stored as the minimizer. The second time, they will run until a given stopping criteria is achieved.

The the so-called *root-mean-square* error (RMSE), will be our measure of performance, and it is defined as

$$\mathrm{RMSE}_n = \sqrt{\frac{\sum_{i,j}\left((X_n)_{i,j} - X^*_{i,j}\right)^2}{\sum_{i,j} 1}} := \frac{\|X_n - X^*\|_F}{\sqrt{d}} \tag{5.1}$$

where X_n is the current iterate, $\|\cdot\|_F$ is the Frobenius norm and $d = MN$, is the dimension of the image $X \in \mathbb{R}^{M\times N}$.

The quality of the restored images will be measured based on the *improvement in signal-to-noise ratio* (ISNR), defined as

$$\mathrm{ISNR}_n = 20\log_{10}\left(\frac{\|X - B\|_F}{\|X - X_n\|_F}\right) \tag{5.2}$$

Our stopping criteria will be $\mathrm{RMSE}_n \leqslant \varepsilon$, for some small $\varepsilon > 0$.

Moreover, since the performance highly depends on the choosen parameters, when possible, the parameters for FBF and FB will be taken from known experiments.

5.1 TV-based image deblurring

The first numerical experiment is the image deblurring problem with l_2 data fidelity. For a linear operator $A \in \mathscr{B}(\mathbb{R}^{m\times n}, \mathbb{R}^{m\times n})$ describing the blur operator, and a matrix $B \in \mathbb{R}^{m\times n}$ representing the blurred and noisy image. Our task is to estimate the unkonwn orignal image $\bar{X} \in \mathbb{R}^{m\times n}$ fulfilling

$$A\bar{X} = B$$

To this end, we solve the following regularized convex minimization problem

$$\inf_{X\in[0,1]^{m\times n}} \left\{\|AX-B\|_1 + \mu\left(\mathrm{TV}_{\mathrm{iso}}(X) + \|X\|_F^2\right)\right\} \tag{5.3}$$

where:

- $\|\cdot\|_1$ is the sum of the absolute values of X

$$\|X\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n} |X_{i,j}|,$$

- $\mu > 0$ is a regularization parameters,
- $TV_{\mathrm{iso}} : \mathbb{R}^{m\times n} \to \mathbb{R}$ is the isotropic total variation functional defined as

$$\begin{aligned} TV_{\mathrm{iso}}(x) = &\sum_{i=1}^{m-1}\sum_{j=1}^{n-1} \sqrt{(x_{i+1,j}-x_{i,j})^2 + (x_{i,j}-x_{i,j+1})^2} \\ &+ \sum_{i=1}^{m-1} |x_{i+1,n}-x_{i,n}| + \sum_{j=1}^{n-1} |x_{m,j+1}-x_{m,j}| \end{aligned}$$

In our numerical experiments we used a Gaussian blur operator with odd kernel size. Due to the symmetry of this filter we have $A^* = A$ with $\|A\| = 1$.

Problem (5.1) fits the model of Problem 4.1:

- $\mathcal{H} = \mathcal{G}_1 = \mathbb{R}^{m\times n}$, with $\langle X, Y \rangle_{\mathcal{H}} := \mathrm{tr}\,(X^t Y)$.
- $\mathcal{G}_2 = \mathcal{H} \oplus \mathcal{H}$, with $\langle (V_1,V_2),(W_1,W_2) \rangle_{\mathcal{G}_2} := \sum_{i=1}^{2} \langle V_i, W_i \rangle_{\mathcal{H}}$.

where $\mathrm{tr}\,(X)$ denotes the trace of the matrix X.

The functions and operators are taken as follows:

- $f(X) = \iota_{[0,1]^{m\times n}}(X) + \mu\|X\|_F^2 \in \Gamma(\mathcal{H})$, with $\mathrm{dom}\, f = [0,1]^{m\times n}$
  $\mathrm{Prox}_{\gamma f}(X) \underset{\substack{(1.11)\\(1.37)\\(1.42)}}{=} P_{[0,1]^{m\times n}}\left(\frac{1}{1+2\mu\gamma}X\right)$
- $h(X) = 0$.
- $g_1(X) = \|X-B\|_1 \in \Gamma(\mathcal{H})$, with $\mathrm{dom}\, g_1 = \mathcal{H}$
  $\mathrm{Prox}_{\gamma g_1^*}(X) \underset{(1.54)}{=} P_{[-1,1]^{m\times n}}(X - \gamma B)$.

- $L_1(X) = AX$,
  $L_1^* = L_1$,
  $\|L_1\|_{\mathcal{H}} = 1$.
- $g_2(V,W) = \mu\|(V,W)\|_\times := \mu \sum_{i=1}^m \sum_{j=1}^n \sqrt{V_{ij}^2 + W_{ij}^2}$, with $\operatorname{dom} g_2 = \mathcal{G}_2$.

  $\operatorname{Prox}_{\gamma g_2^*} \underset{(1.36)}{=} P_C, \quad$ where $C = \left[\, \overline{B}_{\mathbb{R}^2}(0;\mu)\right]^{m\times n}$.
- $L_2(X) = (p^X, q^X)$, where
$$p_{i,j}^X = \begin{cases} X_{i+1,j} - X_{i,j} & \text{if } i < m \\ 0 & \text{if } i = m \end{cases}, \quad \textit{and} \quad q^X = \begin{cases} X_{i,j+1} - X_{i,j} & \text{if } j < n \\ 0 & \text{if } j = n \end{cases}$$
  $(L_2^*(V,W))_{i,j} = V_{i-1,j} - V_{i,j} + W_{i,j-1} - W_{i,j}$, with the convention (5.4),
  $\|L_2\|_{\mathcal{G}_2} = \sqrt{8}$.

The operator L_2 represents a discretization of the gradient using reflexive (Neumann) boundary conditions and standard finite differences. The following computations are based on ([2], page 368):

$$\begin{aligned}\langle\, L_2(X), (V,W)\,\rangle_{\mathcal{G}_2} &= \sum_{i=1}^{m-1}\sum_{j=1}^{n}(X_{i+1,j} - X_{i,j})V_{i,j} + \sum_{i=1}^{m}\sum_{j=1}^{n-1}(X_{i,j+1} - X_{i,j})W_{i,j} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} X_{i,j}(V_{i-1,j} - V_{i,j} + W_{i,j-1} - W_{i,j}) \\ &= \langle\, X, L_2^*(V,W)\,\rangle_{\mathcal{H}}\end{aligned}$$

where we assumed that

$$V_{0,j} = V_{m,j} = W_{i,0} = W_{i,n} = 0 \quad \text{for each } i = 1,\dots,m, j = 1,\dots n. \tag{5.4}$$

Therefore, with the convention (5.4)

$$(L_2^*(V,W))_{i,j} = V_{i-1,j} - V_{i,j} + W_{i,j-1} - W_{i,j},$$

and an upper bound on $\|L_2\|$ can be easily computed:

$$\begin{aligned}\|L_2(X)\|^2 &= \sum_{i=1}^{m-1}\sum_{j=1}^{n}(X_{i+,j} - X_{i,j})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n-1}(X_{i,j+1} - X_{i,j})^2 \\ &\leqslant 2\sum_{i=1}^{m-1}\sum_{j=1}^{n}(X_{i+1,j}^2 + X_{i,j}^2) + 2\sum_{i=1}^{m}\sum_{j=1}^{n-1}(X_{i,j+1}^2 + X_{i,j}^2) \\ &\leqslant 8\sum_{i=1}^{m}\sum_{j=1}^{m} x_{i,j}^2.\end{aligned}$$

Remark 4.3 implies te existence of a unique primal solution, since f is 2μ-strongly convex and g_1, g_2 are non-negative. Moreover, according to Remark 4.4, we only need to guarantee the existence of an $x \in \operatorname{ri}(\operatorname{dom} f)$ such that

$$L_i x \in \operatorname{ri}(\operatorname{dom} g_i)$$

which is trivial, since $\operatorname{ri}(\operatorname{dom} f) = (0,1)^{m\times n}$, $\operatorname{dom} g_1 = \mathcal{H}$ and $\operatorname{dom} g_2 = \mathcal{H} \times \mathcal{H}$.

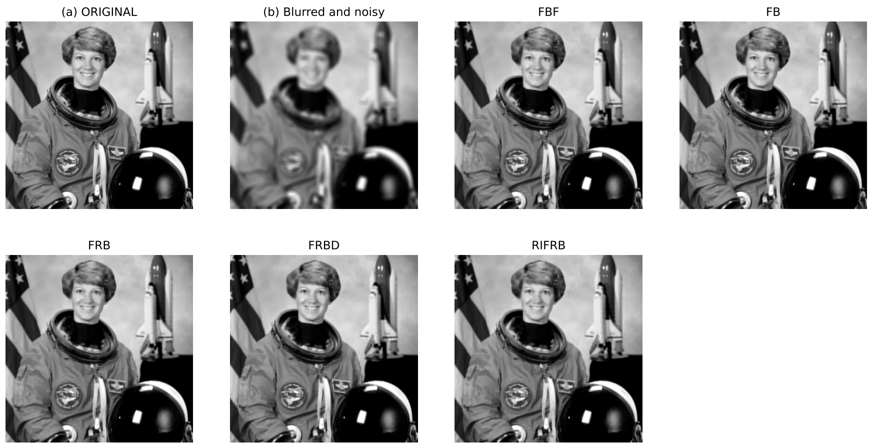

Figure 5.1: (a) Original 512×512 astronaut.png test image, Source scikit-image, (b) The obtained image after multiplying the original one with a blur operator and adding Gaussian noise with standard deviation 10^{-3}. The rest show the reconstructed image by each method after 300 iterations.

The next table shows the execution time of each method for a fixed number of iterations:

Number of Iterations	FBF	FB	FRB
100	2.415 (s)	1.271 (s)	1.325 (s)
200	4.843 (s)	2.557 (s)	2.686 (s)
300	7.373 (s)	3.881 (s)	3.994 (s)
400	9.735 (s)	5.326 (s)	5.308 (s)
500	12.276 (s)	6.574 (s)	6.614 (s)
600	14.64 (s)	7.84 (s)	8.0 (s)
700	17.143 (s)	9.161 (s)	9.559 (s)
800	19.515 (s)	10.655 (s)	10.785 (s)
900	22.046 (s)	11.958 (s)	12.273 (s)

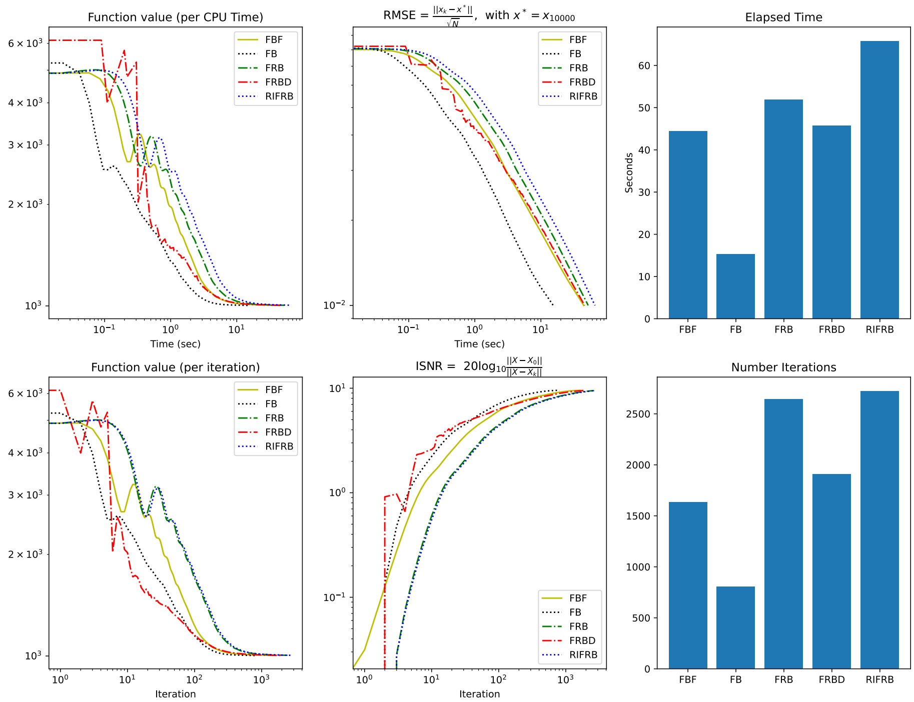

Figure 5.2: Progress of the different methods solving (5.3) for a gaussian blur operator of 13×13 kernel size, standard deviation 8 and $\mu = 0.01$. The parameters for the different methods are the following:
$\beta = 3$.
FBF: $\gamma_n = \frac{1-1e^{-12}}{\beta}$.
FB: $\tau = 0.49, \sigma_1 = 0.7$ and $\sigma_2 = 0.01$
FRB: $\lambda_n = \frac{1-1e^{-10}}{2\beta}$.
FRBD: $M = 8,\ D = \{0.97, 0.95, 0.93, 1\}$, and $\lambda_{\min} = \frac{1e-3}{\beta},\ \lambda_{\max} = \frac{1}{2\beta}$.
RIFRB: $\eta = 0.95$ and $\xi = 0.003$.
Stopping criteria: RMSE < 0.01.

5.2 TV-based image inpainting

We consider the following TV-regularized model

$$\inf_{\substack{X \in [0,1]^{m \times n} \\ M \odot X = B}} TV_{\text{iso}}(X) \tag{5.5}$$

where $M \in [0,1]^{m\times n}$, represents the missing pixels in the noisy image $B \in \mathbb{R}^{m\times n}$, i.e.

$$M_{i,j} = \begin{cases} 0 & \text{if the pixel in the } i^{th} \text{ row and the } j^{th} \text{ column of } B \text{ is missing,} \\ 1 & \text{ otherwise.} \end{cases}$$

$M \odot X$ is the Hadamard product of matrices (or pointwise product) definded for matrices $A, B \in \mathbb{R}^{m\times n}$ as follows:

$$\begin{aligned} &\odot : \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} \\ &(A \odot B)_{i,j} = A_{i,j} B_{i,j}. \end{aligned}$$

Problem (5.5) can be formulated as:

$$\inf_{X \in \mathcal{H}} \{ f(X) + g_1(L_1(X)) + g_2(L_2(X)) \} \tag{5.6}$$

where

- $\mathcal{H} = \mathcal{G}_1 = \mathcal{G}_2 = \mathbb{R}^{m\times n}$, with $\langle X, Y \rangle_{\mathcal{H}} := \operatorname{tr}(X^t Y)$.

The functions and operators are taken as follows:

- $f(X) = \iota_{[0,1]^{m\times n}}(X) \in \Gamma(\mathcal{H})$, with $\operatorname{dom} f = [0,1]^{m\times n}$,
  $\operatorname{Prox}_{\gamma f} \underset{(1.50)}{=} P_{[0,1]^{m\times n}}$.
- $h \equiv 0$.
- $g_1(X) = \iota_{\{B\}}(X) \in \Gamma(\mathcal{H})$, with $\operatorname{dom} g_1 = B$,
  $\operatorname{Prox}_{\gamma g_1}(X) \underset{(1.50)}{=} B$,
  $\operatorname{Prox}_{\gamma g_1^*}(X) \underset{(1.47)}{=} X - \gamma B$.
- $L_1(X) = M \odot X$,
  $L_1^* = L_1$,

$\|L_1\| = 1.$

$$
\begin{aligned}
\langle L_1(X), Y \rangle &= \operatorname{tr}\left((M \odot X)^t Y\right) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} M_{i,j} X_{i,j} Y_{i,j} \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} X_{i,j} M_{i,j} Y_{i,j} \\
&= \langle X, L_1(Y) \rangle
\end{aligned}
$$

- g_2 same as in deblurring problem.
- L_2 same as in deblurring problem.

The existence of a primal solution is guaranteed by Remark 4.3. In order to apply Remark 4.4 we need to find $X \in (0,1)^{m\times n}$ such that

$$M \odot X = B$$

and this only true if $M_{i,j} = 0$ whenever the value of the original image in the coefficients (i,j) is either 0 or 1. Anyway, any number closed enough to 0 or 1 will be good enough for our eyes. It means that we can assume that the original picture X_o lives in $(0,1)^{m\times n}$.

Figure 5.3: (a) Original 512×512 cameraman.png test image, Source scikit-image, (b) The obtained image after a 70% uniformly distributed missing pixel. The rest show the reconstructed image by each method after 300 iterations.

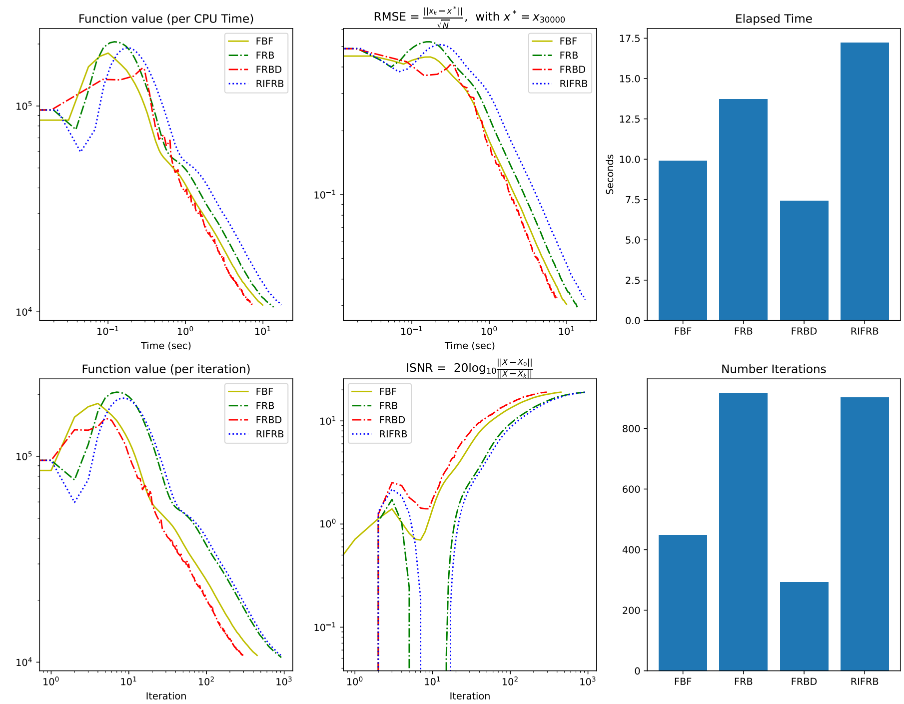

Figure 5.4: Progress of the different methods solving (5.6).
The parameters for the different methods are the following:
$\beta = 3$.
FBF: $\gamma = \frac{1-1e^{-12}}{1+\beta}$.
FRB: $\frac{1-1e^{-10}}{2\beta}$.
FRBD: $M = 4$, $D = \{0.97, 0.95, 0.93, 1\}$, $\lambda_{\min} = \frac{1e-2}{\beta}$, $\lambda_{\max} = \frac{1-1e^{-12}}{2\beta}$.
RIFRB: $\eta = 0.85$, $\xi = 0.0004$.
Stopping criteria: ISNR < 19.

The next table shows the execution time for a fixed number of iterations:

Number of Iterations	FBF	FB	FRB
200	3.086 (s)	1.999 (s)	1.981 (s)
300	4.584 (s)	2.91 (s)	2.925 (s)
400	6.128 (s)	3.893 (s)	3.891 (s)
500	7.666 (s)	4.757 (s)	4.84 (s)
600	9.188 (s)	5.664 (s)	5.766 (s)
700	10.697 (s)	6.593 (s)	6.764 (s)
800	11.981 (s)	7.417 (s)	7.562 (s)

Now, we solve the inpainting problem using the following model:

$$\inf_{X \in [0,1]^{m \times n}} \|M \odot X - B\|_1 + \mu TV_{\text{iso}}(X) \tag{5.7}$$

which can also be formulated in the form of (5.6). This time we take:

- $g_1(X) = \|X - B\|_1$,

  $\operatorname{Prox}_{\gamma g_1}(X) = P_{[-1,1]^{m \times n}}(X - \gamma B)$

and f, L_1, g_2, L_2 as in the previous example.

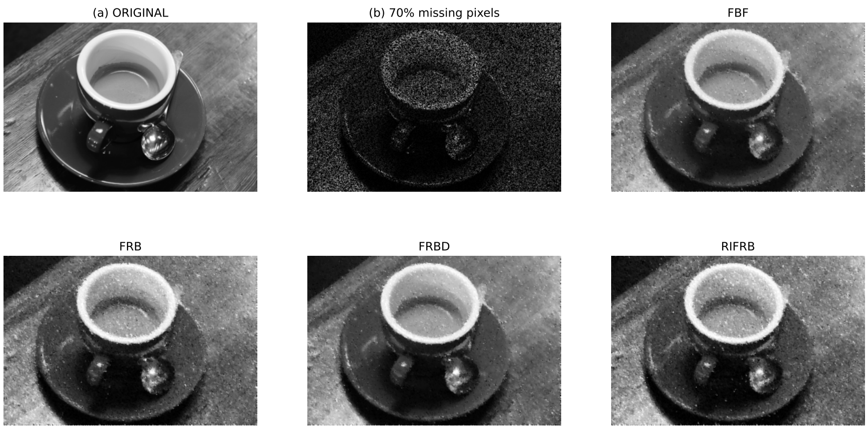

Figure 5.5: (a) Original 400×600 coffee.png test image, Source scikit-image, (b) The obtained image after a 70% uniformly distributed missing pixel. The rest show the reconstructed image by each method after 300 iterations.

The next table shows the execution time for a fixed number of iterations:

Number of Iterations	FBF	FB	FRB
200	3.406 (s)	2.073 (s)	2.055 (s)
400	6.794 (s)	4.038 (s)	4.057 (s)
600	10.069 (s)	5.992 (s)	6.024 (s)
800	13.285 (s)	7.916 (s)	8.022 (s)
1000	16.625 (s)	9.868 (s)	9.983 (s)
1200	20.104 (s)	11.768 (s)	11.961 (s)
1400	23.468 (s)	13.974 (s)	13.937 (s)
1600	26.782 (s)	15.558 (s)	15.88 (s)
1800	30.108 (s)	17.368 (s)	17.818 (s)

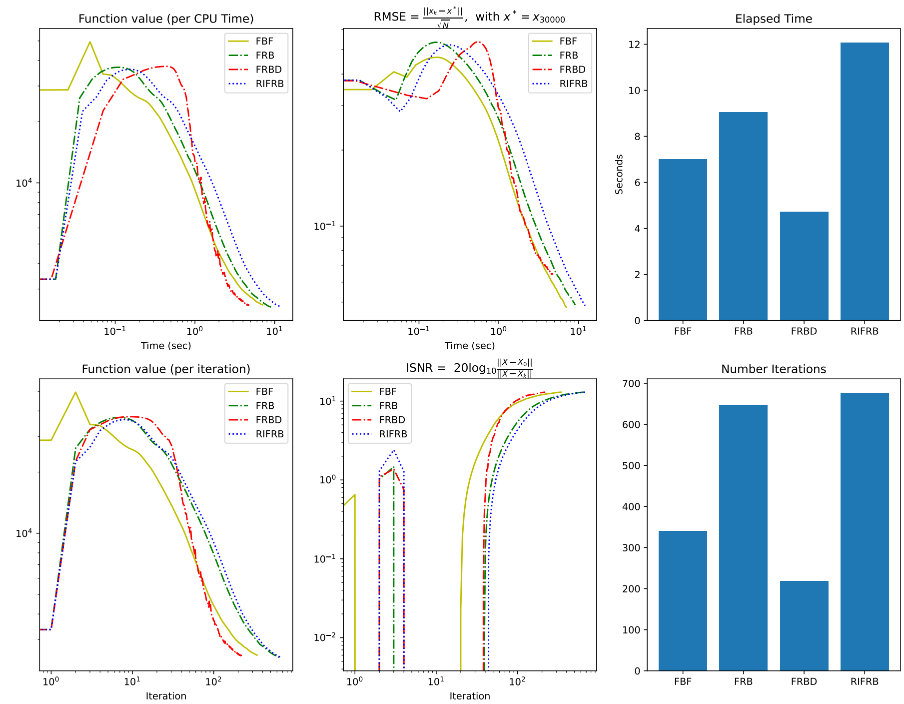

Figure 5.6: Progress of the different methods solving (5.7).
The parameters for the different methods are the following:
FBF,FB, FRB: as in the previous example.
FRBD: $M = 2$, $D = \{0.93, 0.89, 1\}$, $\lambda_{\min} = \frac{1e-2}{\beta}$, $\lambda_{\max} = \frac{1-1e^{-10}}{2\beta}$.
RIFRB: $\eta = 0.85$, $\xi = 0.0012$.
Stopping criteria: ISNR < 13.

5.3 Fermat-Weber problem

The last experiment is the the Fermat-Weber problem consider in [3], which can be expressed as the nondifferentiable convex minimization problem:

$$\inf_{x\in\mathbb{R}^m}\left\{\sum_{i=1}^{k}\lambda_i\|x-c_i\|\right\} \tag{5.8}$$

where $c_i \in \mathbb{R}^m$ are given points and $\lambda_i \in \mathbb{R}_{++}$, for $i = 1, \ldots, k$.

We will solve the following example, which was taken from ([3], Eq (41)):

$$\begin{aligned} &c_1 = (59, 0), \quad c_2 = (20, 0), \quad c_3 = (-20, 48), \quad c_4 = (-20, -48),\\ &\lambda_1 = \lambda_2 = 5, \quad \lambda_3 = \lambda_4 = 13. \end{aligned} \tag{5.9}$$

The optimal solution is $x^* = (0, 0)$ and as starting point we take $x^0 = (44, 0)$, which brings some algorithms into troubles (see [3]).

This particularly example, can obviously be represented as:

$$\lambda_j\|x-c_j\| + \sum_{\substack{i=1\\ i\neq j}}^{4}\lambda_i\|x-c_i\|, \qquad \text{with } j \in \{1,\ldots,4\}, \tag{5.10}$$

and we will solve this example setting:

$$\begin{aligned} f(x) &= \lambda_j\|x-c_j\|\\ g_i(x) &= \lambda_i\|x-c_i\| \end{aligned} \tag{5.11}$$

where

$$\operatorname{Prox}_{\gamma g_i^*}(x) = \frac{\lambda_i}{\max\{\lambda_i, \|x-\gamma c_i\|\}}(x-\gamma c_i) \tag{5.12}$$

with stopping criteria:

$$\|x_n - x^*\| < \varepsilon, \qquad \text{for } \varepsilon \in \mathbb{R}_{++}. \tag{5.13}$$

In the following figures, we show the different performances taking $j \in \{1, 2, 3, 4\}$ and the following parameters:

FRB : $\lambda = \dfrac{1}{2\beta}$.

FRBD : $M = 8,\ D = \{0.95, 0.93, 1\},\ \lambda_{\min} = \dfrac{10^{-2}}{\beta},\ \lambda_{\max} = \dfrac{1-10^{-12}}{2\beta}$.

RIFRB : $\eta = 0.85,\ \xi = 0.0121$.

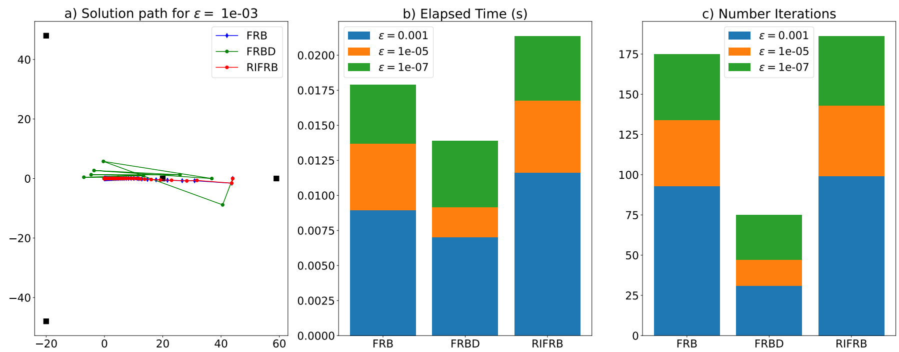

Figure 5.7: Performances solving (5.10) with $j = 1$ in (5.11).

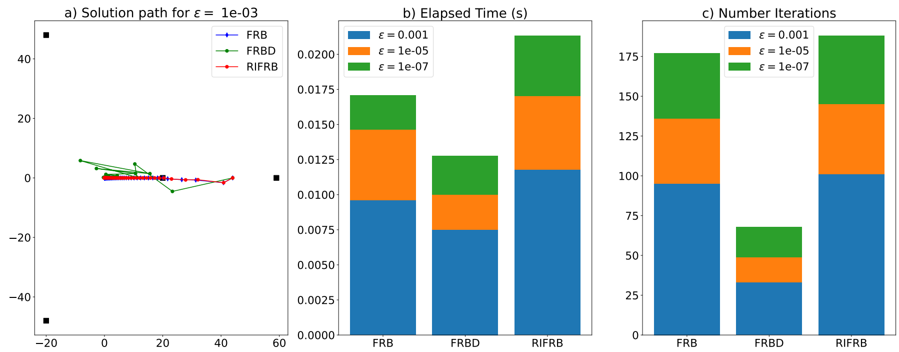

Figure 5.8: Performances solving (5.10) with $j = 2$ in (5.11).

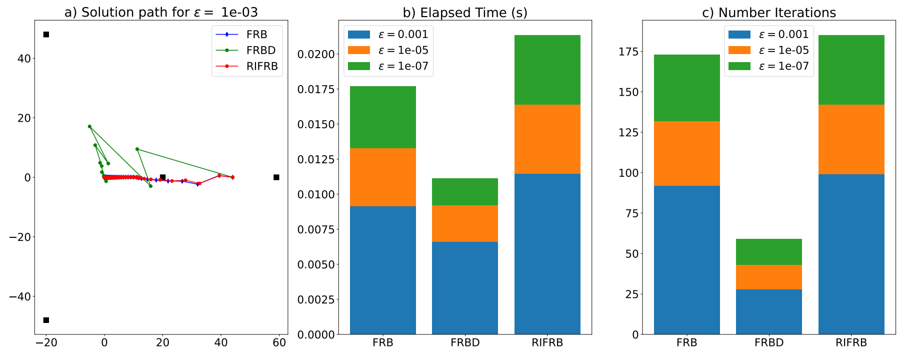

Figure 5.9: Performances solving (5.10) with $j = 3$ in (5.11).

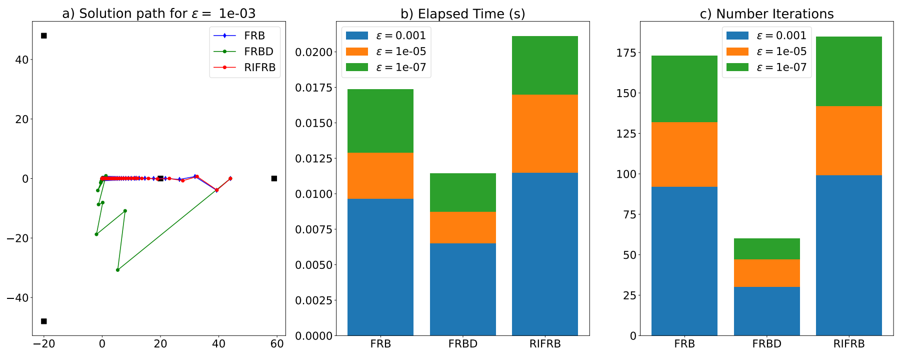

Figure 5.10: Performances solving (5.10) with $j = 4$ in (5.11).

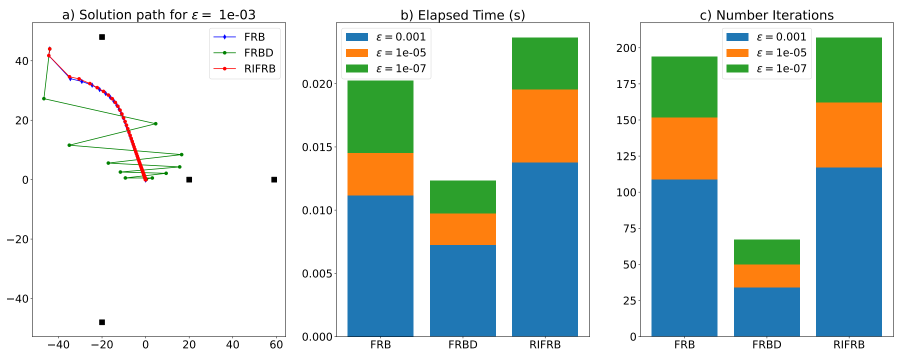

Figure 5.11: Performances solving (5.10) for $j = 3$ in (5.11) with $x_0 = (40, 40)$.

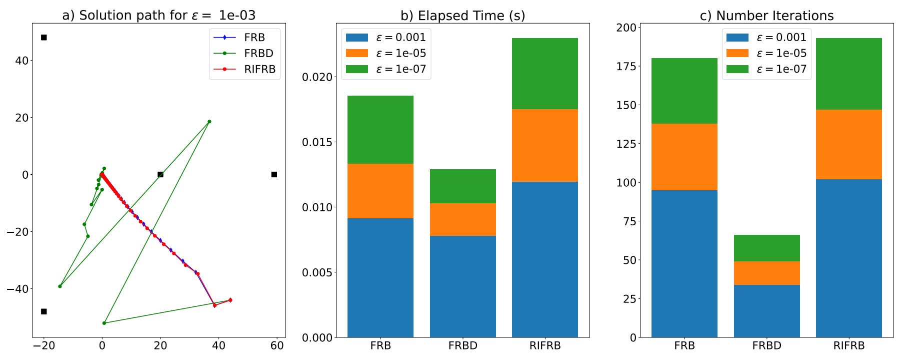

Figure 5.12: Performances solving (5.10) for $j = 3$ in (5.11) with $x_0 = (40, -40)$.

6 Conclusion and future work

Conclusion

1. In section 2 we showed the linear convergence of the FRB method with variable stepsize as well as the linear convergence of two of its variants (RIFRB and Three operator splitting) as long as one of the operators is strong monotone.

2. According to our numerical experiments in section 5, we conclude that the FRB algorithm is very competitive, when compared to FBF and FB. Its computational cost per iteration was similar to the FB's one and lower than the one of FBF.

3. In section 4 we described a known naive stepsize strategy to use with FRB.

4. We could not see any significant advantage applying the relaxed-intertial version of FRB.

5. Even if our stepsize strategy was very simple, the FRB method with variable stepsize performed better than its variant with fixed stepsize in all our experiments.

Future Work

1. It does not look difficult to show the linear convergence of the RIFRB method with variable stepsize.

2. It could be interesting to find out, if the strongly monotonicity of just one of the operators in Theorem 3.2 and Theorem 3.3, would imply linear convergence in its corresponding variable.

Bibliography

[1] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer International Publishing, Switzerland, second edition, 2017.

[2] Amir Beck. *First-Order Methods in Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.

[3] Radu Boţ, Ernö Csetnek, and Andre Heinrich. A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators. *SIAM Journal on Optimization*, 23:2011–2036, 01 2013.

[4] Radu Ioan Boţ and Ernö Robert Csetnek. On the convergence rate of a forward-backward type primal-dual splitting algorithm for convex optimization problems. *Optimization*, pages 5–23, 2015.

[5] Radu Ioan Boţ and Ernö Robert Csetnek. An inertial forward-barckward-forward primal-dual splitting algorithm for solving monotone inclusion problems. *Numerical Algorithms*, 71:519–540, 2016.

[6] Radu Ioan Boţ, Ernö Robert Csetnek, and Christopher Hendrich. Recent developments on primal–dual splitting methods with applications to convex minimization. *Mathematics Without Boundaries: Surveys in Interdisciplinary Research*, pages 57–99, 2014.

[7] Radu Ioan Boţ and Christopher Hendrich. Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization. *Journal of Mathematical Imaging and Vision*, pages 551–568, 2014.

[8] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer, New York, New York, 2011.

[9] Luis M. Briceño-Arias and Patrick L. Combettes. A monotone + skew splitting model for composite monotone inclusions in duality. *SIAM J. Optim*, 21(4):12301–1250, 2011.

[10] Patrick L. Combettes and Jean-Christophe Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal*, 20(2):307–330, 2012.

[11] Christopher Hendrich. *Proximal Splitting Methods in Nonsmooth convex Optimization*. dissertation, Technische Universität Chemnitz, 2014.

[12] Yura Malitsky and Matthew K. Tam. A forward-backward splitting method for monotone inclusions without cocoercivity. *SIAM J. Optim*, 30(2):1451–1472, 2020.

[13] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization.* Springer, New York, NY, USA, second edition, 2006.

[14] Băng Công Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38:667–681, 2011.

[15] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014.