DISSERTATION | DOCTORAL THESIS

Titel | Title

Extrapolation from the past in Tseng's algorithm for monotone inclusion problems and applications

verfasst von | submitted by

Buris Tongnoi

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

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 796 605 405
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Mathematik
Betreut von | Supervisor:	Dr. Ernö Robert Csetnek Privatdoz. MSc

Acknowledgements

First and foremost, I would like to thank my supervisor Ernö Robert Csetnek for their invaluable guidance, tolerance, encouragement, and insightful feedback. Your expertise and commitment to my research have been instrumental in shaping my work and fostering my academic growth. I would also like to thank Professor Radu Ioan Boț, who always supported me and provided many opportunities during my study period. I wish to express my gratitude to Professor Suthep Suantai for encouraging my studies abroad. I would like to thank Associate Professor Phan Tu Vuong for his guidance on the programming code. I am also grateful to Professor Francisco Javier Aragón Artacho and Professor Samir Adly for reviewing this PhD thesis. I dedicate this thesis to Professor Paul Tseng, who pioneered the central concept of this thesis.

This research was funded through the Development and Promotion of Science and Technology Talents Project (DPST), which receives financial support from the Thai government. I am grateful to the Office of the Civil Service Commission (OCSC) for their guidance and support during my studies. I am also deeply thankful to the Royal Thai Embassy in Vienna for assisting me, especially with various bureaucratic matters in Austria. Furthermore, I would like to extend my appreciation to the Vienna School of Mathematics (VSM) and the Doctoral Programme at the Vienna Graduate School on Computational Optimization (VGSCO) for their generous provision of resources and opportunities, including a wide range of educational courses and summer schools.

A special thanks goes to my colleagues and fellow students in the Faculty of Mathematics whom I was lucky to work with. It is my pleasure to thank Alex, Andres, Axel, Beryl, Bo, Chiara, David, Jikai, Khoa, Rossen, Michael, Mikhail, Minh, Morteza, Rossen, Yakun, and many others whom I cannot name individually. A special mention goes to Bettina for her invaluable assistance, particularly in handling general affairs. I would like to acknowledge the Thai Student&Researcher Club in Austria (THSRCA) and the Association of Thai Professionals in the European Region (ATPER), along with all the members of both groups. Their efforts in building vibrant Thai communities and creating opportunities to collaborate with exceptional individuals have made my learning journey even more rewarding.

My deepest appreciation goes to my family, who have been my constant source of encouragement and support. I especially want to thank my father for being a role model who always sacrificed everything for his family. Thank you to my mother for loving me unconditionally, no matter who I am. I am also grateful to my younger sister for always cheering me on and taking care of our parents while I was away studying. A special thanks to Yok, who has been one of my greatest sources of motivation, pushing me to reach this point. You helped me get through the early days of my studies, even during the challenging times of the pandemic. Most importantly, I want to thank myself for never giving up and pushing through to get here.

Finally, I would like to express my gratitude to anyone I may have inadvertently overlooked. Your contributions, whether large or small, have been integral to my achievements. Thank you all for your unwavering support and encouragement. This thesis would not have been possible without each and every one of you.

Abstrakt

Die vorliegende Dissertation widmet sich der Lösung von monotonen Inklusionsproblemen mit zusammengesetzter Struktur sowie besonderen Fällen, d. h. der Bestimmung der Nullstelle der Summe monotoner Operatoren, Minimax-Problemen, konvexen Optimierungsproblemen und Variationsungleichungen. Der Hauptfokus dieser Arbeit liegt auf der Untersuchung von Tsengs Forward-Backward-Forward-Algorithmus mit Extrapolation aus der Vergangenheit in verschiedenen Aspekten.

Ein Aspekt ist die Verallgemeinerung von Tsengs Forward-Backward-Forward-Algorithmus durch die Einführung variabler Metriken und Fehlertermen sowie dessen primal-duale Splitting-Algorithmen zur Lösung monotoner Inklusionsprobleme und konvexer Optimierungsprobleme. Darüber hinaus wird die Anwendung des vorgeschlagenen Algorithmus anhand der Bildentzerrung demonstriert. Diese Untersuchung ist in Kapitel 3 zu finden, und die detaillierten Ergebnisse wurden 2022 in Numerical Functional Analysis and Optimization veröffentlicht (siehe [89]).

In Kapitel 4 wird Tsengs Forward-Backward-Forward-Algorithmus mit Extrapolation aus der Vergangenheit im Rahmen eines Strafschemas zur Lösung monotoner Inklusionsprobleme betrachtet, bei denen die Summe dreier Operatoren eine Rolle spielt. Die schwache (ergodische) Konvergenz wird unter bestimmten Annahmen zur Fitzpatrick-Funktion nachgewiesen. Zudem enthält das Kapitel Anwendungen auf Minimax-Probleme, Probleme mit der Komposition linearer stetiger Operatoren sowie konvexe Minimierungsprobleme. Die numerischen Experimente umfassen die Anwendung des Algorithmus auf die TV-basierte Bildrekonstruktion. Diese Arbeit wurde 2024 in Numerical Algorithms veröffentlicht (siehe [88]).

Schließlich wird in Kapitel 5 Tsengs Forward-Backward-Forward-Algorithmus mit Extrapolation aus der Vergangenheit für pseudomonotone Variationsungleichungen untersucht. Das Kapitel enthält einen Konvergenznachweis sowie eine adaptive Schrittweitenstrategie für die vorgeschlagene Methode. Die Effektivität der Algorithmen wird anhand numerischer Experimente in endlich- und unendlichdimensionalen Räumen demonstriert. Diese Phase der Studie basiert auf [87] und wurde 2024 in der Taiwanese Journal of Mathematics veröffentlicht.

Diese wissenschaftliche Arbeit stellt flexible Werkzeuge auf der Grundlage von Tsengs Forward-Backward-Forward-Algorithmus mit Extrapolation aus der Vergangenheit zur Lösung monotoner Inklusionsprobleme in verschiedenen Kontexten und deren Anwendungen bereit. Darüber hinaus wird die Leistungsfähigkeit von Tsengs Forward-Backward-Forward-Algorithmus mit Extrapolation aus der Vergangenheit gegenüber der klassischen Methode von Tseng in einigen praxisnahen Anwendungsfällen begründet.

Schlüsselwörter: Konvergenzanalyse; Iterative Methoden; Monotone Inklusionsprobleme; Optimierungsprobleme; Minimax-Probleme; Tsengs Forward-Backward-Forward-Algorithmus; Variationsungleichungen; Pseudomonotonie.

Abstract

The thesis is devoted to the solving of monotone inclusion problems involving composite structure and particular cases, i.e., the zero of the sum of monotone operators, minimax problems, convex optimization problems and variational inequality problems. The main focus throughout this manuscript is to investigate Tseng's forward-backward-forward algorithm with extrapolation from the past in various aspects.

One is the generalization of Tseng's forward-backward-forward algorithm endowed with variable metrics and error terms, and its primal-dual splitting algorithms for solving monotone inclusion problems and convex optimization problems. In addition, the image deblurring shows the application of the proposed algorithm. This exploration can be found in Chapter 3 and the detail of the work is published in Numerical Functional Analysis and Optimization in 2022, see [89].

Next, in Chapter 4, Tseng's forward-backward-forward algorithm with extrapolation from the past is considered in the penalty scheme for solving the monotone inclusion problem in terms of the sum of three operators. The weak (ergodic) convergence is obtained under some hypothesis related to the Fitzpatrick function, and the chapter consists of applications in minimax problems, problems involving the composition of linear continuous operators and convex minimization problems. The numerical experiment is also demonstrated by applying the algorithm to TV-based image inpainting. The task is published in Numerical Algorithms in 2024, see [88].

Finally, in Chapter 5, Tseng's forward-backward-forward algorithm with extrapolation from the past is investigated for pseudo-monotone variational inequalities. The chapter contains the convergent result, and an adaptive stepsize strategy of the proposed method is also contributed. The effectiveness of the proposed algorithms is demonstrated through numerical experiments in finite and infinite dimensional spaces. This phase of the study is based on [87], which was published in Taiwanese Journal of Mathematics in 2024.

This academic work provides flexible tools based on Tseng's forward-backward-forward algorithm with extrapolation from the past for solving monotone inclusion problems in several contexts and their applications. Furthermore, it justifies the performance of Tseng's forward-backward-forward algorithm with extrapolation from the past over the classical Tseng method in some practical real-world problems.

Key words: Convergence analysis; Iterative methods; Monotone inclusion problems; Optimization problems; Minimax problems; Tseng's forward-backward-forward algorithm; Variational inequalities; Pseudo-monotonicity.

Contents

1 Introduction and Motivation

When we are talking about the generalized monotone inclusion problems involving set-valued operators in Hilbert space $\mathcal{H}$, it means the following problem:

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad z \in Fx, \tag{1.1}$$

where $F : \mathcal{H} \rightrightarrows \mathcal{H}$ is monotone and $z \in \mathcal{H}$; see, e.g. [23, 91, 92]. In terms of non-smooth optimization problems, this monotone inclusion problem links to various problems in real-world applications, such as signal and image processing [19], positron emission tomography [11] and machine learning [83]. In particular case, if $z = 0$, the problem (1.1) becomes a special case, namely

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Fx, \tag{1.2}$$

with a monotone operator F (see, e.g. [14, 24, 25, 32, 35, 62]).

In general, people are interested in finding an $x \in \mathcal{H}$ satisfying (1.2) for given maximal monotone mapping F on $\mathcal{H}$. A traditional method for doing this proposed by Martinet [63, 64] and generalized by Rockafellar [76, 79]. It is known as the proximal algorithm,

$$x^{k+1} = (\operatorname{Id} + \alpha_k F)^{-1}(x^k), \quad k = 0, 1, \ldots,$$

where $\alpha_k > 0$ and $\operatorname{Id} : \mathcal{H} \to \mathcal{H}$ is the identity operator. This approach, along with its dual variant in convex programming—the Hestenes and Powell method of multipliers—has been thoroughly explored (refer to [12, 45, 54, 58] and the associated references). It is recognized for producing several specialized decomposition techniques, including the method of partial inverses, the Douglas-Rachford splitting method, and the alternating direction method of multipliers [1, 38, 39, 40, 59, 81, 85].

In scenarios where $F = A + B$, with A and B being maximal monotone mappings on $\mathcal{H}$ and B being single-valued on $\operatorname{dom} B \supset \operatorname{dom} A$, we mention the classical forward-backward (FB) splitting method

$$x^{k+1} = (\operatorname{Id} + \alpha_k A)^{-1}(\operatorname{Id} - \alpha_k B)(x^k), \quad k = 0, 1, \ldots,$$

where $\alpha_k > 0$, for solving the problem

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Bx. \tag{1.3}$$

This approach is introduced by Lions and Mercier [59], further developed by Passty [69], and presented in a dual form for convex programming by Han and Lou [46]. When A is equal to normal cone N_C, with C representing a nonempty closed convex set in $\mathcal{H}$, this method simplifies to a projection technique suggested by Sibony [84] for monotone variational inequalities (VIs), i.e.,

$$\text{find} \quad x \in C \quad \text{such that} \quad \langle B(x), y - x\rangle \geq 0,\ \forall y \in C. \tag{1.4}$$

In general, the methods based on forward-backward splitting scheme require the cocoercivity of the single-valued operator B for the convergence results, but this property can be relaxed into the monotonicity and Lipschitz continuity as the extragradient method, Tseng's algorithm, and others like below.

When discussing saddle point and minimization problems, one of the simplest and most natural general method is the gradient method [3, 71]. While effective for minimization problems, the gradient method often fails to converge when applied to finding saddle points (or converges for saddle point problems only under extremely rigid assumptions). In 1976, Korpelevich [57] proposed an extragradient method. The idea is instead of directly updating the solution using the gradient at the current point, the extragradient method first takes a "trial" gradient step to an intermediate point. Then, it calculates the gradient at this extrapolated point and uses that gradient for the actual update. Let $C \subseteq \mathcal{H}$ be a nonempty closed convex set, P_C be the projection operator onto C (see Definition 2.4.1 in Chapter 2) and $B : \mathcal{H} \to \mathcal{H}$ a monotone and Lipschitz operator. The extragradient for solving (1.4) takes the form (see also [41])

$$\text{EG} \begin{cases} y_n = P_C(x_n - \gamma B(x_n)), \\ x_{n+1} = P_C(x_n - \gamma B(y_n)), \end{cases} \tag{1.5}$$

for $0 < \gamma < 1/L$ and $(x_n)_{n\in\mathbb{N}}$ converges weakly to some solution of problem (1.4). Notice that EG method has two projections.

In 2000, Tseng [90] proposed a simple modification to the FB method for solving the problem in (1.3). This approach draws inspiration from Korpelevich's extragradient method [57] for monotone variational inequalities, which enhances Sibony's projection method [84] by performing an additional forward step and only one projection step at each iteration. It can be expressed in the simple formula as shown below

$$\begin{aligned} y_n &= J_{\gamma A}(x_n - \gamma B x_n), \\ x_{n+1} &= y_n + \gamma(B x_n - B y_n), \end{aligned} \tag{1.6}$$

where γ is a constant parameters satisfying $0 < \gamma < 1/L$ and $J_{\gamma A} = (\mathrm{Id} + \gamma A)^{-1}$ is the resolvent operator.

From Tseng's algorithm [90], we can see that the algorithm must compute twice both $B(x_n)$ and $B(y_n)$, which wastes the algorithm process. To alter this issue, we follow Popov's idea [73], which provides a strategy in the extragradient method or EG (also known as Korpelevich [57]) that only requires a single gradient computation per update for the problem. We refer to this improved approach as extragradient with extrapolation from the past, namely EG-EP (the terminology "extrapolation from the past" is taken from [43]). Indeed, let $C \subseteq \mathcal{H}$ be a nonempty closed convex set, and $B : \mathcal{H} \to \mathcal{H}$ be a monotone Lipschitz operator. For some positive parameter γ we have the following scheme for solving (1.4)

$$\text{EG} \begin{cases} y_n = P_C(x_n - \gamma B(x_n)), \\ x_{n+1} = P_C(x_n - \gamma(B(y_n))), \end{cases} \quad \text{and} \quad \text{EG-EP} \begin{cases} y_n = P_C(x_n - \gamma B(y_{n-1})), \\ x_{n+1} = P_C(x_n - \gamma(B(y_n))). \end{cases} \tag{1.7}$$

Notably, we may use prior forward evaluations of B that have been previously computed from the earlier step in the EG-EP method. Then, we refer to this method as *extrapolation from the past*, which was proposed and detailed in [43] (see Section 3.4 of [43]).

We can combine this technique with Tseng's algorithm and call it *Tseng's algorithm with extrapolation from the past* (Tseng-EP) or *forward-backward-forward algorithm with extrapolation from the past* (FBF-EP). This approach is also suggested in [14]. One can consider a general scheme as

$$\text{Tseng-General} \begin{cases} y_n = J_{\gamma A}(x_n - \gamma B(z_n)), \\ x_{n+1} = y_n + \gamma(B(z_n) - B(y_n)). \end{cases} \tag{1.8}$$

1. For $z_n = x_n$, we obtain Tseng's algorithm, see [90] and Section 2.3 in [14].

   For $z_n = y_{n-1}$, we obtain Tseng's algorithm with extrapolation from the past (FBF-EP):

   $$\text{Tseng-EP or FBF-EP} \begin{cases} y_n = J_{\gamma A}(x_n - \gamma B(y_{n-1})), \\ x_{n+1} = y_n + \gamma(B(y_{n-1}) - B(y_n)). \end{cases} \tag{1.9}$$

   On the other hand, we obtain *forward-reflected-backward* (see [62]) if we insert x_{n+1} into the first step of Tseng-EP at y_{n+1}:

   $$\begin{aligned} & y_{n+1} = J_{\gamma A}(y_n + \gamma(B(y_{n-1}) - B(y_n)) - \gamma B(y_n)) \\ \text{or} \quad & y_{n+1} = J_{\gamma A}(y_n - 2\gamma B(y_n) + \gamma B(y_{n-1})). \end{aligned} \tag{1.10}$$

   In particular, both approaches are identical and become *optimistic mirror descent* (see [74, 75]) in the unconstrained case with $A = 0$, as

   $$y_{n+1} = y_n - 2\gamma B(y_n) + \gamma B(y_{n-1}), \tag{1.11}$$

   which, under the term *Optimistic Gradient Descent Ascent* (OGDA), has been suggested for the training of Generative Adversarial Network (GANs), see [36, 37] for more details.

The aim of this thesis is to inspect several algorithms based on Tseng's algorithm with extrapolation in the past (Tseng-EP or FBF-EP). The work is based on the publications [87, 88, 89]. The ordinary basic knowledge and useful preliminaries are introduced in Chapter 2.

In Chapter 3 (based on [89]), we engage to solve the monotone inclusion problem, **Problem 3.1**, involving composite structure. This kind of problem can be addressed with Tseng's algorithm endowed with error terms like in Briceño-Arias and Combettes [23]. The error terms make it more flexible in computation. These additional error terms fix situations when the resolvent operator cannot be computed efficiently. Furthermore, Vũ [92] introduced a variable metric extension of Tseng's algorithm endowed error terms, which enhances efficiency by using metrics. Due to our purpose of studying, we propose a variable metric extension of FBF-EP with error terms and prove its convergence in Section 3.2. This method can be represented as primal-dual solver for monotone inclusions as in Section 3.3. Moreover, a primal-dual splitting algorithm for convex optimization problems in the context of Tseng-EP (or FBF-EP) is proposed in Section 3.4. In order to confirm the effectiveness of our proposed algorithm in Chapter 3, numerical experiments in image deblurring are considered in Section 3.5.

In Chapter 4 (based on [88]), we study the general monotone inclusion problem in the term of the sum of three operators. In fact, it is the monotone inclusion problem rooting from the problem in (1.2) while in this occasion we consider $F = A + D + N_C$ where A is a maximally monotone operator, D is a monotone and Lipschitz continuous operator, and C is the set of zeros of another monotone and Lipschitz continuous operator B. As Tseng-EP (or FBF-EP) is our scope of the study direction, we provide Tseng's algorithm with extrapolation and penalization techniques and also prove the weak convergence of the sequence of weighted averages. The strong convergence can be assured by assuming the strong monotonicity of A. These aforesaid results belong in Section 4.2. Further, a penalty scheme version of Tseng-EP (or FBF-EP) for solving minimax problems is presented in Section 4.3. Apart from the problem of the sum of three operators, we also develop Tseng-EP (or FBF-EP) and penalty scheme for solving the problem involving composition of linear continuous operators in Section 4.4. In Section 4.5, the theoretical results were validated through a numerical experiment in TV-based image information by comparing the proposed method with the classical Tseng's algorithm.

In Chapter 5 (based on [87]), we turn our attention to the work in variational problem (VI) as (1.4) by applying Tseng-EP (or FBF-EP). In Section 5.2, we propose a modified version of Tseng's forward-backward-forward algorithm incorporating an extraplotion from the past procedure to solve pseudo-monotone variational inequalities in Hilbert space. Also, we provide a variational stepsize scheme for the algorithm and investigate an adaptive stepsize scenario improving the efficiency and applicability of the method, especially when the Lipschitz constant of the operator is unknown. In Section 5.3, we contribute numerical experiments to support the chapter's theoretical results, including examples in both finite-dimensional Euclidean space and the infinite-dimensional L_2 space. The numerical results suggest that Tseng-EP (or FBF-EP) can be more efficient than the classical Tseng (or FBF) algorithm in terms of CPU time, but its performance can be sensitive to computer errors or computational power.

2 Preliminaries

In this chapter, we present essential definitions, propositions, and lemmas, many of which can be found in [8].

Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces with *inner product* $\langle \cdot, \cdot \rangle$ and the asscociated norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. The notation $\mathbb{N}$ is the set of natural numbers, $\mathbb{R}$ represents the set of real numbers, and $\mathbb{R}_{++}$ serves as the set of all positive real numbers. We denote weak and strong convergence as symbols $\rightharpoonup$ and $\to$, respectively. Finally, ℓ_+ denotes the set of all sequences in $[0, +\infty)$, and ℓ^1 (resp. ℓ^2) denotes the space of all absolutely (resp. square) summable sequences in $\mathbb{R}$. Therefore ℓ^1_+ means the space of all absolutely summable sequences in $[0, \infty)$.

The *limit inferior* of a sequence (x_n) is defined by

$$\liminf_{n \to +\infty} x_n := \sup_{n \geq 0} \inf_{m \geq n} x_m = \sup\{\inf\{x_m \mid m \geq n\} \mid n \geq 0\},$$

and the *limit superior* of a sequence (x_n) is defined by

$$\limsup_{n \to +\infty} x_n := \inf_{n \geq 0} \sup_{m \geq n} x_m = \inf\{\sup\{x_m \mid m \geq n\} \mid n \geq 0\}.$$

Now, we will provide notations regarding sets.

Definition 2.0.1 (Affine Set and Convex Set)**.** A set $C \subseteq \mathcal{H}$ is said to be

(i) *affine* if $\lambda x + (1 - \lambda) y \in C$ for all $x, y \in C$ and all $\lambda \in \mathbb{R}$;

(ii) *convex* if $\lambda x + (1 - \lambda) y \in C$ and all $\lambda \in [0, 1]$;

(iii) a *cone* if $\lambda x \in$ for all $x \in C$ and for all $\lambda \geq 0$.

The intersection of an arbitrary family of convex sets (cones, affine sets, linear subspaces) is a convex set (cone, affine set, linear sunspace, respectively). Furthermore, we provide the following definition for the intersection of the set containing $C \subseteq \mathcal{H}$.

Definition 2.0.2 (Linear hull, Affine hull, Convex hull, Conical hull)**.**

(i) $\operatorname{lin} C := \cap\{D \subset \mathcal{H} \mid C \subseteq D, D \text{ is a linear subspace}\}$ is called the *linear hull* of C;

(ii) $\operatorname{aff} C := \cap\{D \subset \mathcal{H} \mid C \subseteq D, D \text{ is an affine set}\}$ is called the *affine hull* of C;

(iii) $\operatorname{co} C := \cap\{D \subset \mathcal{H} \mid C \subseteq D, D \text{ is a convex set}\}$ is called the *convex hull* of C;

(iv) $\operatorname{cone} C := \cap\{D \subset \mathcal{H} \mid C \subseteq D, D \text{ is a cone}\}$ is called the *conical hull* of C.

The interior of a subset C of $\mathcal{H}$ is the largest open set that is contained in C defined as

$$\operatorname{int} C = \{x \in C \mid \exists \rho > 0 \text{ such that } B(0,\rho) \subset C - x\},$$

where for all $\rho > 0$, $B(0,\rho) := \{y \in \mathcal{H} \mid \|y\| < \rho\}$ is the open ball with the center at 0 and radius ρ.

Definition 2.0.3. Let C be a convex subset of $\mathcal{H}$. The *core* of C is

$$\operatorname{core} C = \{x \in C \mid \operatorname{cone}(C - x) = \mathcal{H}\};$$

the *strong relative interior* of the set C is defined as

$$\operatorname{sri} C := \{x \in C \mid \operatorname{cone}(C - x) \text{ is a closed linear subspace}\};$$

moreover, it holds $\operatorname{int} C \subseteq \operatorname{sri} C$, where the inclusion is strict in general.

Definition 2.0.4 (Relative Interior)**.** Let $C \subseteq \mathbb{R}^n$ be a nonempty convex set. The *relative interior* of the set C is defined as

$$\operatorname{ri} C := \{x \in C \mid \exists \varepsilon > 0 \text{ such that } B(x,\varepsilon) \cap \operatorname{aff} C \subseteq C\}$$

and it represents the interior of the set C relative to its affine hull. If $\operatorname{int} C \neq \emptyset$, then $\operatorname{aff} C = \mathbb{R}^n$ and $\operatorname{ri} C = \operatorname{int} C$.

Theorem 2.0.5 ([15, 44])**.** *Let $C \subseteq \mathbb{R}^n$ be a convex set. It holds $\operatorname{sri} C = \operatorname{ri} C$.*

2.1 Useful Facts

Fact 2.1.1 (Cauchy-Schwarz Inequality, [8, Fact 2.11])**.** *Let x and y be in $\mathcal{H}$. Then*

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

Moreover, $\langle x, y\rangle = \|x\|\|y\|$ if and only if there is $\alpha \in \mathbb{R}$ such that $y = \alpha x$.

Proposition 2.1.2. *Let $x, y \in \mathcal{H}$ and $\alpha \in \mathbb{R}$. Then*

$$\|\alpha x + (1-\alpha)y\|^2 + \alpha(1-\alpha)\|x - y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2. \tag{2.1}$$

Next, we present the fertile lemma confirming the sequence's convergence and summability associated with a specific form of inequality. This lemma is essential and applied multiple times in this thesis, and it has been sourced from [5].

Lemma 2.1.3. *Let $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ be real sequences. Assume that $(a_n)_{n\in\mathbb{N}}$ is bounded from below, $(b_n)_{n\in\mathbb{N}}$ is a nonnegative sequence, $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1$ such that*

$$a_{n+1} \leq a_n - b_n + \varepsilon_n \quad \forall n \in \mathbb{N}$$

Then $(a_n)_{n\in\mathbb{N}}$ is convergent and $(b_n)_{n\in\mathbb{N}} \in \ell^1$.

Two more extended general versions of the Lemma above will be used further in Chapter 3, as shown below.

Lemma 2.1.4 ([33], [72, Lemma 2 in Section 2.2.1]). *Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $[0,+\infty)$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ and let $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ be such that $\alpha_{n+1} \leq (1+\eta_n)\alpha_n + \varepsilon_n, \forall n \in \mathbb{N}$. Then $(\alpha_n)_{n\in\mathbb{N}}$ converges.*

Lemma 2.1.5 ([30, Lemma 3.1]). *Let $\chi \in (0,1], (\alpha_n)_{n\geq 0} \in \ell_+, (\beta_n)_{n\geq 0} \in \ell_+$ and $(\varepsilon_n)_{n\geq 0} \in \ell^1_+$ be such that*

$$\alpha_{n+1} \leq \chi\alpha_n - \beta_n + \varepsilon_n, \quad \forall n \in \mathbb{N}. \tag{2.2}$$

Then

(i) $(\alpha_n)_{n\geq 0}$ *is bounded;*

(ii) $(\alpha_n)_{n\geq 0}$ *converges;*

(iii) $(\beta_n)_{n\geq 0} \in \ell^1$*;*

(iv) If $\chi \neq 1, (\alpha_n)_{n\geq 0} \in \ell^1$.

To demonstrate the weak convergence of our proposed algorithms that will appear in this thesis, we present a valid theorem known as Opial's Lemma.

Lemma 2.1.6 ([67, Opial's Lemma]). *Let S be a nonempty subset of $\mathcal{H}$, and $(x_n)_{n\geq 0}$ a sequence of elements of $\mathcal{H}$. Assume that*

(i) for every $z \in S$, $\lim\limits_{n\to+\infty} \|x_n - z\|$ exists;

(ii) every weak sequential limit point of $(x_n)_{n\in\mathbb{N}}$ belongs to S as $n \to +\infty$.

Then x_n converges weakly as $n \to +\infty$ to a point in S.

2.2 Functions and Their Properties

Now we consider definitions and properties of functions with values in the extended-real line $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$.

Definition 2.2.1. Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be a function. Then

(i) The *effective domain* of f is defined as $\operatorname{dom} f := \{x \in \mathcal{H} \mid f(x) < +\infty\}$.

(ii) The function f is said to be *proper* if $\operatorname{dom} f \neq \emptyset$ and $f(x) > -\infty$ for all $x \in \mathcal{H}$.

(iii) The function f is said to be *lower semicontinuous (l.s.c) at $x \in \mathcal{H}$* if

$$\liminf_{y\to x} f(y) = \sup_{\delta>0} \inf_{y\in B(x,\delta)} f(y) \geq f(x)$$

The function f is called *lower semicontinuous* if it is lower semicontinuous at every $x \in \mathcal{H}$.

Definition 2.2.2 (Convex Function)**.** A function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is called *convex* if for all $x, y \in \mathcal{H}$ and all $\lambda \in (0, 1)$ it holds

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

Theorem 2.2.3 ([8, Theorem 9.1 (or Theorem 10.23)])**.** *Let $f : \mathcal{H} \to (-\infty, +\infty]$ be convex. Then the following statements are equivalent:*

(i) f is weakly sequentially lower semicontinuous.

(ii) f is sequentially lower semicontinuous.

(iii) f is lower semicontinuous.

(iv) f is weakly lower semicontinuous.

Definition 2.2.4. (Separable Sum of Functions, [8, Example 2.1]) Let I be a nonempty index set and $\mathcal{H}_{i \in I}$ be real Hilbert spaces. Assume that for any $i \in I, f_i : \mathcal{H}_i \to (-\infty, \infty]$ and that, if I is infinite, $\inf_{i \in I} f_i \geq 0$. Then we define

$$\bigoplus_{i \in I} f_i : \bigoplus_{i \in I} \mathcal{H}_i \to (-\infty, \infty] : (x_i)_{i \in I} \mapsto \sum_{i \in I} f_i(x_i).$$

2.3 Conjugate Functions, Subdifferentiability and Other Useful Functions

Definition 2.3.1 (Subdifferential)**.** Let $f : \mathcal{H} \to (-\infty, \infty]$ be proper. The *subdifferential* of f is the set-valued operator defined as

$$\partial f : \mathcal{H} \rightrightarrows \mathcal{H}, x \mapsto \{u \in \mathcal{H} \mid \ \langle y - x, u \rangle + f(x) \leq f(y) \ \forall y \in \mathcal{H}\} \quad \text{for } f(x) \in \mathbb{R}, \tag{2.3}$$

and $\partial f(x) := \emptyset$ for $f(x) \notin \mathbb{R}$. The elements of $\partial f(x)$ are called *subgradients* of f at x.

Definition 2.3.2 (Conjugate Function and Infimal Convolution)**.** Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be a given function. The *conjugate function* of f is defined by

$$f^* : \mathcal{H} \to \overline{\mathbb{R}} : u \mapsto \sup_{x \in \mathcal{H}} \{\langle x, u \rangle - f(x)\}, \tag{2.4}$$

and the *infimal convolution* of f, $g : \mathcal{H} \to \overline{\mathbb{R}}$ is

$$f \Box g : \mathcal{H} \to \overline{\mathbb{R}} : x \mapsto \inf_{y \in \mathcal{H}} \{f(y) + g(x - y)\}. \tag{2.5}$$

Proposition 2.3.3. *(Characterization of Subdifferential, [8, Theorem 16.29]) Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be a given function, $x \in \mathcal{H}$ and $u \in \mathcal{H}$. Then*

$$u \in \partial f(x) \Leftrightarrow f^*(u) + f(x) = \langle u, x \rangle.$$

As the above characterization of the element in subgradient of f, we can see that the inverse of the subdifferential (in the sense of set-valued operators) can be described as [8, Corollary 16.30]

$$(\partial f)^{-1} = \partial f^*. \tag{2.6}$$

Definition 2.3.4. Let C be a subset of $\mathcal{H}$. Then

(i) The *indicator function* of C is defined by $\iota_C : \mathcal{H} \to \overline{\mathbb{R}}, x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C. \end{cases}$

(ii) The *distance function* of C is defined by $d_C = \iota_C \Box \| \cdot \| : \mathcal{H} \to \overline{\mathbb{R}}, x \mapsto \inf_{y \in C} \|x - y\|$.

(iii) The *support function* of C is $\sigma_C : \mathcal{H} \to \overline{\mathbb{R}} : u \mapsto \sup\langle C, u\rangle$. Moreover:

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

Definition 2.3.5 (Normal Cone)**.** Let C be a nonempty convex subset of $\mathcal{H}$ and let $x \in \mathcal{H}$. The *normal cone* to C at x is empty set, if $x \notin C$, and

$$N_C(x) = \left\{ u \in \mathcal{H} \;\middle|\; \sup_{y \in C}\langle y - x, u\rangle \leq 0 \right\}, \quad \text{if } x \in C.$$

It worth nothing that for $x \in C$, $u \in N_C(x)$ if and only if $\sigma_C(u) = \langle x, u\rangle$. Note that a *cone* is a set that is closed under positive scaling (see Definition 2.0.1 (iii)), and the normal cone is always a cone itself.

Example 2.3.6. *Let $C \subseteq \mathcal{H}$ be a given set. Then, the (convex) subdifferential of its indicator function can be written as*

$$\partial\iota_C(x) = \{u \in \mathcal{H} \mid \langle u, y - x\rangle \leq 0, \ \forall y \in C\} = N_C(x) \quad \textit{for } x \in C.$$

and $\partial\iota_C(x) = \emptyset$ for $x \notin C$.

We recall that a proper function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is said to be *Gâteaux differentiable* at $x \in \operatorname{dom} f$ if there exists $u \in \mathcal{H}$ such that

$$\langle u, a\rangle = \lim_{t \to 0} \frac{f(x + ta) - f(x)}{t}, \quad \forall a \in \mathcal{H}.$$

The element u is called *Gâteaux differential (derivative)* of f at x, and it is denoted by $\nabla f(x)$.

Remark 2.3.7.

(i) We recall that a proper function $f : \mathcal{H} \to \overline{\mathbb{R}}$ is said to be *Fréchet differentiable* at $x \in \operatorname{int}(\operatorname{dom} f)$ if there exists $u \in \mathcal{H}$ such that

$$\lim_{a \to 0} \frac{f(x+a) - f(x) - \langle u, a \rangle}{\|a\|} = 0.$$

In this case, f is continuous and Gâteaux differentiable at x. The element u is called *Fréchet differential (derivative)* of f at x, and is consequently denoted also by $\nabla f(x)$.

(ii) Let $f : \mathcal{H} \to \overline{\mathbb{R}}$ be proper, convex and lower semicontinuous. If f is continuous and Gâteaux differentiable on $B(x, \delta) \subset \operatorname{dom} f$, for $\delta > 0$, and the Gâteaux derivative $\nabla f : B(x, \delta) \to \mathcal{H}$ is continuous at x, then f is Fréchet differentiable at x (see also [95, Corollary 3.3.6], [8, Corollary 17.42]).

Proposition 2.3.8 ([8, Proposition 17.31]). *Let $f : \mathcal{H} \to (-\infty, \infty]$ be proper and convex and let $x \in \operatorname{dom} f$. Then the following hold:*

(i) Suppose that f is Gâteaux differentiable at x. Then $\partial f(x) = \{\nabla f(x)\}$;

(ii) Suppose that f is continuous at x, and $\partial f(x)$ consists of a single element u. Then f is Gâteaux differentiable at x and $u = \nabla f$.

2.4 Projection and Proximal Point Operator

Definition 2.4.1 (Best Approximation / Projection). Let C be a nonempty subset of $\mathcal{H}$, let $x \in \mathcal{H}$ and let $p \in C$. Then p is a *best approximation* to x from C (or a *projection* of x onto C) if $\|x - p\| = d_C(x)$. If every point in $\mathcal{H}$ has exactly one projection onto C, then C is called a *Chebyshev set*. In this case, the *projector* (or projection operator) onto C is the operator, denoted by P_C, that maps every point in $\mathcal{H}$ to its unique projection onto C.

Now, we let $\Gamma_0(\mathcal{H})$ be represented as the class of proper convex lower semicontinuous functions. We also know that $f^* \in \Gamma_0(\mathcal{H})$ if $f \in \Gamma_0(\mathcal{H})$.

Definition 2.4.2 (Proximal Point [8]). Let f be in $\Gamma_0(\mathcal{H})$. Then the *proximity operator (or proximal mapping)* is defined by

$$prox_f : \mathcal{H} \to \mathcal{H}, x \mapsto \arg\min_{y \in \mathcal{H}} (f(y) + \frac{1}{2}\|x - y\|^2),$$

$prox_f(x)$ is the unique minimizer of $f + \frac{1}{2}\| \cdot - x\|^2$.

Example 2.4.3 ([8, Example 12.25]). *Let C be a nonempty closed convex set of $\mathcal{H}$, and $f = \iota_c$. Then $prox_f = prox_{\iota_c} = P_C$.*

The characterization of the projection mapping is a useful tool that we will employ to present our main results, as demonstrated in the theorem below.

Theorem 2.4.4 ([8, Theorem 3.14]). *Let C be a nonempty closed convex subset of $\mathcal{H}$. Then for every x and p in $\mathcal{H}$,*

$$p = P_C x \quad \Leftrightarrow \quad [p \in C \text{ and } (\forall y \in C)\ \langle y - p, x - p\rangle \leq 0].$$

Theorem 2.4.5 ([8, Proposition 16.44]). *Let $f \in \Gamma_0(\mathcal{H})$, and let x and p be in $\mathcal{H}$. Then*

$$p = prox_f x \quad \text{if and only if} \quad x - p \in \partial f(p).$$

In other words, $prox_f = (\operatorname{Id} + \partial f)^{-1}$.

2.5 Operators and Their properties

Definition 2.5.1. An operator $T : \mathcal{H} \to \mathcal{H}$ is called

(i) *linear* if $T(\lambda x + \mu y) = \lambda T(x) + \mu T(y)$ for all $x, y \in \mathcal{H}$ and all $\lambda, \mu \in \mathbb{R}$.

(ii) *affine* if $T(\lambda x + (1-\lambda)y) = \lambda T(x) + (1-\lambda)T(y)$ for all $x, y \in \mathcal{H}$ and all $\lambda \in \mathbb{R}$.

Definition 2.5.2. Let D be a nonempty subset of $\mathcal{H}$ and let $A : D \to \mathcal{H}$. Then A is said to be

(i) *firmly nonexpansive* if

$$\|Ax - Ay\|^2 + \|(\operatorname{Id} - A)x - (\operatorname{Id} - A)y\|^2 \leq \|x - y\|^2 \quad \forall x, y \in D. \tag{2.7}$$

(ii) *nonexpansive* if

$$\|Ax - Ay\| \leq \|x - y\| \quad \forall x, y \in D,$$

which is nothing else than 1-Lipschitz continuous with constant 1. Also, we can notice that every firmly nonexpansive operator is always nonexpansive.

(iii) *Lipschitz continuous (or Lipschitzian)* if

$$\|Ax - Ay\| \leq \nu \|x - y\| \quad \forall x, y \in \mathcal{H}, \tag{2.8}$$

where $\nu > 0$ is called a Lipschitz constant.

(iv) *cocoercive (or inverse strongly monotone)* if it satisfies

$$\langle Ax - Ay, x - y\rangle \geq \beta \|Ax - Ay\|^2, \quad \forall x, y \in \mathcal{H} \text{ and } \beta > 0. \tag{2.9}$$

Example 2.5.3 (see [8, Corollary 18.17 (Baillon-Haddad)]). *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Fréchet differentiable. If ∇f is L-Lipschitz continuous (with $L > 0$), then ∇f is $(1/L)$-cocoercive.*

Definition 2.5.4 (Bounded Linear Operator)**.** Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces. The *space of bounded linear operator* from $\mathcal{H}$ to $\mathcal{G}$ is denoted by

$$\mathcal{B}(\mathcal{H},\mathcal{G}) = \{T : \mathcal{H} \to \mathcal{G} \mid T \text{ is linear and continuous}\}$$

and $\mathcal{B}(\mathcal{H}) = \mathcal{B}(\mathcal{H},\mathcal{H})$. We equipe $\mathcal{B}(\mathcal{H},\mathcal{G})$ with the norm

$$\|T\| = \sup \|T(\overline{B}(0,1))\| = \sup_{\substack{x\in\mathcal{H}\\ \|x\|\leq 1}} \|Tx\|, \quad \forall T \in \mathcal{B}(\mathcal{H},\mathcal{G}).$$

The *adjoint* of $T \in \mathcal{B}(\mathcal{H},\mathcal{G})$ is the unique operator $T^* \in \mathcal{B}(\mathcal{G},\mathcal{H})$ which satisfies

$$\langle Tx, y\rangle = \langle x, T^*y\rangle, \quad \forall x \in \mathcal{H}, \forall y \in \mathcal{G}.$$

2.6 Monotone and Maximally monotone operator

For any arbitrary set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$, we denote by

$$\begin{aligned}
\mathrm{Gra}(A) &:= \{(x,\xi) \in \mathcal{H}\times\mathcal{H} \mid \xi \in A(x)\};\\
\mathrm{Dom}(A) &:= \{x \in \mathcal{H} \mid Ax \neq \emptyset\};\\
\mathrm{Ran}(A) &:= \{u \in \mathcal{H} \mid \exists x \in \mathcal{H} \text{ such that } u \in Ax\};\\
\mathrm{zer}(A) &:= \{x \in \mathcal{H} \mid 0 \in Ax\};
\end{aligned}$$

its *graph*, *domain*, *range* and the set of *zeros* of A, respectively. The *inverse* of A, denoted by A^{-1}, is defined through its graph such that $\mathrm{Gra}(A^{-1}) = \{(u,x) \in \mathcal{H}\times\mathcal{H} \mid (x,u) \in \mathrm{Gra}(A)\}$. Notice that $\mathrm{zer}(A) = A^{-1}(0)$.

Definition 2.6.1. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-valued operator.

(i) The operator A is said to be *monotone* if

$$\langle x-y, u-v\rangle \geq 0, \quad \forall (x,u),(y,v) \in \mathrm{Gra}(A). \tag{2.10}$$

(ii) The monotone operator A is said to be *maximally monotone* (or *maximal monotone*) if there exits no monotone operator $B : \mathcal{H} \rightrightarrows \mathcal{H}$ such that $\mathrm{Gra}\, A \subset \mathrm{Gra}\, B$ and $A \neq B$. Furthermore, we have the following characterization: the monotone operator A is maximally monotone if and only if

$$(x,u) \in \mathcal{H}\times\mathcal{H} \text{ and } \langle v-u, y-x\rangle \geq 0,\ \forall (y,v) \in \mathrm{Gra}(A) \Rightarrow (x,u) \in \mathrm{Gra}(A). \tag{2.11}$$

Furthermore, a point $z \in \mathcal{H}$ belongs to $\mathrm{zer}(A)$ if and only if $\langle u-z, w\rangle \geq 0$ for all $(u,w) \in \mathrm{Gra}(A)$.

(iii) The operator A is said to be γ*-strongly monotone* with $\gamma > 0$ if $A-\gamma\,\mathrm{Id}$ is monotone, i.e.,

$$\langle x-y, u-v\rangle \geq \gamma\|x-y\|^2, \quad \forall (x,u),(y,v) \in \mathrm{Gra}(A). \tag{2.12}$$

We know that the subdifferential of f is maximally monotone if f is proper convex and lower semicontinuous function:

Theorem 2.6.2 (Moreau's theorem, [8, Theorem 20.25] or [80])**.** *Let $f \in \Gamma_0(\mathcal{H})$. Then ∂f is maximally monotone.*

Remark 2.6.3. A single-valued linear operator $S : \mathcal{H} \to \mathcal{H}$ is said to be *skew* if $\langle Sx, x\rangle = 0$ for all $x \in \mathcal{H}$ or, equivalently, if

$$\langle Sy, x\rangle = -\langle Sx, y\rangle \quad \forall x, y \in \mathcal{H}.$$

It is known that skew operators are maximally monotone, and any nonzero single-valued linear and skew operator is not a (convex) subdifferential. This example provides a maximally monotone operator, which is not a (convex) subdifferential.

If A is maximally monotone, then $\operatorname{zer}(A)$ is a closed and convex set (see [8, Proposition 23.39]), and conditions ensuring that $\operatorname{zer}(A)$ is nonempty can be found in Section 23.4 of [8]. Moreover, if A is maximally monotone and strongly monotone, then the set $\operatorname{zer}(A)$ is a singleton and thus nonempty (refer to [8, Corollary 23.37]).

Remark 2.6.4. Notice that every cocoercive operator is monotone and Lipschitz continuous but the converse is not the case. For example, when $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ for $\mathbf{x} = [x_1, x_2]^T, \mathbf{y} = [y_1, y_2]^T \in \mathbb{R}^2$ such that $\mathbf{x} \neq \mathbf{y}$, we can observe that if let $\mathbf{z} = \mathbf{x} - \mathbf{y}$, then $\langle A(\mathbf{x}-\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle = \mathbf{z}^T A\mathbf{z} = [z_1, z_2]\begin{bmatrix} z_2 \\ -z_1 \end{bmatrix} = z_1z_2 - z_2z_1 = 0 \ngeq \beta\|A(\mathbf{x}-\mathbf{y})\|^2, \forall \beta > 0$ (furthermore, we have $\|A(\mathbf{x}-\mathbf{y})\| = \|A\mathbf{z}\| = \left\| \begin{bmatrix} z_2 \\ -z_1 \end{bmatrix} \right\| = \sqrt{z_2^2 + (-z_1)^2} = \sqrt{z_1^2 + z_2^2} = \|\mathbf{z}\| = \|\mathbf{x}-\mathbf{y}\|$).

We now present a basic property of maximally monotone operators, which will be used in the proof of the main results of the thesis.

Proposition 2.6.5 ([8, Proposition 20.38])**.** *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone. Then the following statements hold:*

(i) $\operatorname{Gra}(A)$ *is sequentially closed in $\mathcal{H}^{strong} \times \mathcal{H}^{weak}$, i.e., for every sequence $(x_n, u_n)_{n\in\mathbb{N}}$ in* $\operatorname{Gra}(A)$ *and every $(x, u) \in \mathcal{H} \times \mathcal{H}$ if $x_n \to x$ and $u_n \rightharpoonup u$, then $(x, u) \in$* $\operatorname{Gra}(A)$*;*

(ii) $\operatorname{Gra}(A)$ *is sequentially closed in $\mathcal{H}^{weak} \times \mathcal{H}^{strong}$, i.e., for every sequence $(x_n, u_n)_{n\in\mathbb{N}}$ in* $\operatorname{Gra}(A)$ *and every $(x, u) \in \mathcal{H} \times \mathcal{H}$ if $x_n \rightharpoonup x$ and $u_n \to u$, then $(x, u) \in$* $\operatorname{Gra}(A)$*;*

(iii) $\operatorname{Gra}(A)$ *is closed in $\mathcal{H}^{strong} \times \mathcal{H}^{strong}$.*

Definition 2.6.6. (Resolvent Operator [8]) Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$. The *resolvent* of A is

$$J_A = (\operatorname{Id} + A)^{-1}. \tag{2.13}$$

In addition, for $f \in \Gamma_0(\mathcal{H})$ we have $J_{\partial f} = prox_f$, and for a given nonempty, closed, convex set C according to Definition 2.3.5, Example 2.4.3 and Proposition 2.4.5: we have $J_{\partial\iota_C} = (\operatorname{Id} + \partial\iota_C)^{-1} = (\operatorname{Id} + N_C)^{-1} = prox_{\iota_C} = P_C$.

Moreover, if A is maximally monotone, then the resolvent operator of A, $J_A : \mathcal{H} \to \mathcal{H}$, is single-valued and maximally monotone (see [8, Corrollary 23.10]).

The following proposition provides the closed relationship between firmly nonexpansive mappings and the monotone operators.

Proposition 2.6.7. *([8, Proposition 23.8]) Let D be a nonempty subset of the real Hilbert space $\mathcal{H}$, let $T : D \to \mathcal{H}$ and $A = T^{-1} - \mathrm{Id}$. Then the following hold:*

(i) $T = J_A$;

(ii) *T is firmly nonexpansive if and only if A is monotone;*

(iii) *T is firmly nonexpansive and $D = \mathcal{H}$ if and only if A is maximally monotone.*

Corollary 2.6.8. *([8, Corollary 23.9]) Let $T : \mathcal{H} \to \mathcal{H}$. Then T is firmly nonexpansive if and only if it is the resolvent of a maximally monotone operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$.*

3 Tseng's Algorithm with Extrapolation from the Past Endowed with Variable Metrics and Error Terms

Various problems in real-world applications, such as signal and image processing [19], positron emission tomography [11] and machine learning [83], can be expressed in term of non-smooth optimization problems. These problems can be modelled as inclusion problems involving monotone set-valued operators in Hilbert space $\mathcal{H}$ shown as **Problem 3.1** below. In many situations, the operator F in (1.1) can be represented as the finite sum of monotone operators, where compositions with linear transformations are allowed. In these cases, it is typically preferable to also solve the associated dual inclusion, see [2, 19, 23, 31, 91]. More precisely, the monotone inclusion problem has the following form:

Problem 3.1: Let $\mathcal{H}$ be a real Hilbert space, let m be a strictly positive integer, let $z \in \mathcal{H}$, let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, let $C : \mathcal{H} \to \mathcal{H}$ be monotone and v_0-Lipschitzian for some $v_0 \in (0, +\infty)$. For every $i \in \{1, \dots, m\}$, let $\mathcal{G}_i$ be a real Hilbert space, let $r_i \in \mathcal{G}_i$, let $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}$ be a maximally monotone operator, let $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero bounded linear operator. Suppose that

$$z \in \operatorname{Ran}\left(A + \sum_{i=1}^{m} L_i^* \left(B_i(L_i \cdot - r_i)\right) + C\right). \tag{3.1}$$

The problem is to solve the primal inclusion

$$\text{find} \quad \bar{x} \in \mathcal{H} \text{ such that } z \in A\bar{x} + \sum_{i=1}^{m} L_i^* \left(B_i(L_i\bar{x} - r_i)\right) + C\bar{x}, \tag{3.2}$$

and the dual inclusion

$$\text{find} \quad \bar{v}_1 \in \mathcal{G}_1, \dots, \bar{v}_m \in \mathcal{G}_m \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} z - \displaystyle\sum_{i=1}^{m} L_i^* \bar{v}_i \in Ax + Cx, \\ \bar{v}_i \in B_i(L_i x - r_i), \ \forall i \in \{1, \dots, m\}. \end{cases} \tag{3.3}$$

The particular case of convex optimization problems can be seen in Section 3.4.

In this chapter, we are interested in developing Tseng's algorithm with extrapolation from the past endowed with variable metrics and error terms. Moreover, when the resolvent operator cannot be computed efficiently, it is allowed to have errors. For example, such modification in the classical Tseng's algorithm has been proposed in [23]: the algorithm is more flexible if we concede it has error terms. Additionally, some works proposed the use of variable metrics to get more efficient proximal algorithms (see [24, 25, 33, 68]), which can be applied to the forward-backward splitting algorithm in [32] and Tseng's algorithm in [92]. Therefore, we round up the modification algorithm's benefits and put them into our scheme, as shown in this chapter's details.

In the layout of this chapter, we propose the variable metric Tseng's algorithm with extrapolation from the past and error terms shown in Section 3.2. The basic knowledge and useful tools are presented in Section 3.1, along with various notations. After that, in Sections 3.3 and 3.4, we employ our main result to propose a variable metric primal-dual algorithm for the various kinds of composite inclusions for **Problem 3.1** (and **Problem 3.2**, respectively). Moreover, we illustrate the application of our algorithm to image deblurring in Section 3.5. All aspects of this work were published in Numerical Functional Analysis and Optimization in 2022, see [89].

3.1 Preliminaries and Useful Tools

In order to effectively navigate the results and proofs presented in the next sections, it is essential to establish a solid foundation of key concepts, useful lemmas and propositions. This section outlines the preliminary materials that underpin our analysis. Together with the main preliminaries section (see Chapter 2), articulating these essential components will ensure a comprehensive understanding of the context and framework necessary for the subsequent discussions.

Let $\mathcal{S}(\mathcal{H}) = \{L \in \mathcal{B}(\mathcal{H}) \mid L = L^*\}$ be the linear subspace of $\mathcal{B}(\mathcal{H})$ of self-adjoint operators. The Loewner partial ordering on $\mathcal{S}(\mathcal{H})$ is denoted by

$$U \succcurlyeq V \quad \forall\, U, V \in \mathcal{S}(\mathcal{H}) \;\Leftrightarrow\; \langle Ux, x\rangle \geq \langle Vx, x\rangle \quad \forall x \in \mathcal{H}. \tag{3.4}$$

Now let $\alpha \in [0, +\infty)$. We set

$$\mathcal{P}_\alpha(\mathcal{H}) = \{U \in \mathcal{S}(\mathcal{H}) \mid U \succcurlyeq \alpha\,\mathrm{Id}\}, \tag{3.5}$$

and we denote by $\sqrt{U}$ the square root of $U \in \mathcal{P}_\alpha(\mathcal{H})$. Moreover, for every $U \in \mathcal{P}_\alpha(\mathcal{H})$, we define a semi-scalar product and a semi-norm (a scalar product and a norm if $\alpha > 0$) by

$$\langle x, y\rangle_U = \langle Ux, y\rangle \text{ and } \|x\|_U = \sqrt{\langle Ux, x\rangle} \quad \forall x \in \mathcal{H}\ \forall y \in \mathcal{H}. \tag{3.6}$$

The *proximity (proximal) operator* of $f \in \Gamma_0(\mathcal{H})$ relative to the metric induced by $U \in \mathcal{P}_\alpha(\mathcal{H})$ is [48, Section XV.4]

$$prox_f^U : \mathcal{H} \to \mathcal{H} : x \mapsto \underset{y \in \mathcal{H}}{\arg\min}(f(y) + \frac{1}{2}\|x - y\|_U^2), \tag{3.7}$$

and the projector onto a nonempty closed convex subset C of $\mathcal{H}$ relative to the norm $\|\cdot\|_U$ is denoted by P_C^U. We have

$$prox_f^U = J_{U^{-1}\partial f} \text{ and } P_C^U = prox_{\iota_C}^U, \tag{3.8}$$

and we write $prox_f^{\mathrm{Id}} = prox_f$. Note that if $U = \mathrm{Id}$, the projector operator onto C becomes the standard projector operator: P_C (see also Definition 2.4.1).

Next, we introduce the useful lemmas giving properties of operators in $\mathcal{S}(\mathcal{H})$, properties of the operator UA and J_{UA}, and the behavior of $(U_n)_{n\in\mathbb{N}}$ in $\mathcal{P}_\alpha(\mathcal{H})$.

Lemma 3.1.1 ([52, Section VI.2.6], [32], [33]). *Let $\alpha, \mu \in (0,+\infty)$ and let A and B be operators in $\mathcal{S}(\mathcal{H})$ such that $\mu\,\mathrm{Id} \succcurlyeq A \succcurlyeq B \succcurlyeq \alpha\,\mathrm{Id}$. Then the following hold:*

(i) $\alpha^{-1}\,\mathrm{Id} \succcurlyeq B^{-1} \succcurlyeq A^{-1} \succcurlyeq \mu^{-1}\,\mathrm{Id}$;

(ii) for all $x \in \mathcal{H}$, $\langle A^{-1}x, x\rangle \geq \|A\|^{-1}\|x\|^2$;

(iii) $\|A^{-1}\| \leq \alpha^{-1}$.

Lemma 3.1.2 ([32]). *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $\alpha \in (0,+\infty)$, let $U \in \mathcal{P}_\alpha(\mathcal{H})$, and let $\mathcal{G}$ be the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x,y) \mapsto \langle x,y\rangle_{U^{-1}} = \langle x, U^{-1}y\rangle$. Then the following statements hold:*

(i) $UA : \mathcal{G} \rightrightarrows \mathcal{G}$ is maximally monotone;

(ii) $J_{UA} : \mathcal{G} \to \mathcal{G}$ is 1-cocoercive, i.e., firmly nonexpansive, hence nonexpansive;

(iii) $J_{UA} = (U^{-1} + A)^{-1} \circ U^{-1}$.

Lemma 3.1.3 ([33, Lemma 2.2]). *Let $\alpha \in (0,+\infty)$ let $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, and let $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$ such that $\mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty$. Suppose that one of the following statements holds:*

(i) $(1+\eta_n)U_n \succcurlyeq U_{n+1} \quad \forall n \in \mathbb{N}$;

(ii) $(1+\eta_n)U_{n+1} \succcurlyeq U_n \quad \forall n \in \mathbb{N}$.

Then there exists $U \in \mathcal{P}_\alpha(\mathcal{H})$ such that $U_n \to U$ pointwise.

The definition of quasi-Fejér sequence with respect to some target set relative to a sequence $(U_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}_\alpha(\mathcal{H})$ is provided as follows.

Definition 3.1.4 ([33] Definition 3.1). Let $\alpha \in (0,+\infty)$, let $\phi : [0,+\infty) \to [0,+\infty)$, let $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$, let C be a nonempty subset of $\mathcal{H}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Then $(x_n)_{n\in\mathbb{N}}$ is ϕ-quasi-Féjer monotone with respect to the target set C relative to $(U_n)_{n\in\mathbb{N}}$ if $(\exists(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))$ $(\forall z \in C)$ $(\exists(\varepsilon_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N}))$ $(\forall n \in \mathbb{N})$,

$$\phi(\|x_{n+1} - z\|_{U_{n+1}}) \leq (1+\eta_n)\phi(\|x_n - z\|_{U_n}) + \varepsilon_n. \tag{3.9}$$

The following proposition is a effective tool to guarantee the boundedness and convergence of the sequence.

Proposition 3.1.5 ([33, Proposition 3.2]). *Let $\alpha \in (0, +\infty)$, let $\phi : [0, +\infty) \to [0, +\infty)$ be strictly increasing and such that $\lim_{t\to+\infty} \phi(t) = +\infty$, let $(U_n)_{n\in\mathbb{N}}$ be in $\mathcal{P}_\alpha(\mathcal{H})$, let C be a nonempty subset of $\mathcal{H}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that (3.9) is satisfied. Then the following statements hold:*

(i) Let $z \in C$. Then $(\|x_n - z\|_{U_n})_{n\in\mathbb{N}}$ converges;

(ii) $(x_n)_{n\in\mathbb{N}}$ is bounded.

We require some of the previous lemmas and propositions to provide the necessary context for the following theorem, which confirms the weak convergence of our method in Section 3.2. The theorem states as follows.

Theorem 3.1.6 ([33, Theorem 3.3]). *Let $\alpha \in (0, +\infty)$, let $\phi : [0, +\infty) \to [0, +\infty)$ be strictly increasing and such that $\lim_{t\to+\infty} \phi(t) = +\infty$, let $(U_n)_{n\in\mathbb{N}}$ and U be operators on $\mathcal{P}_\alpha(\mathcal{H})$ such that $U_n \to U$ pointwise, let C be nonempty subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be sequence in $\mathcal{H}$ such that (3.9) is satisfied. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C if and only if every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in C.*

3.2 A Variable Metric Tseng's Algorithm with Extrapolation from the Past and Error Terms

Using the product space approach, the primal-dual inclusion problems (3.2) and (3.3) can be written in an appropriate product space as finding $\bar{x} \in \boldsymbol{\mathcal{H}}$ with $0 \in \mathbf{A}(\bar{x}) + \mathbf{B}(\bar{x})$, where $\mathbf{A}$ is maximally monotone and $\mathbf{B}$ is monotone and Lipschitz continuous, see Section 3.3. This is the reason why we start by proposing a method for the problem $0 \in A(x) + B(x)$ in the following theorem.

Theorem 3.2.1. *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone, let $\alpha, \beta \in (0, +\infty)$, $B : \mathcal{H} \to \mathcal{H}$ be a monotone and β-Lipschitz operator on $\mathcal{H}$ such that $zer(A + B) \neq \emptyset$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$ and $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$ such that*

$$\mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty \quad \text{and} \qquad (1 + \eta_n)U_{n+1} \succcurlyeq U_n. \tag{3.10}$$

Let $(\gamma_n)_{n\in\mathbb{N}} \leq \lambda$ with $\lambda < \frac{1}{\sqrt{10}\mu\beta}$ and $\liminf\limits_{n\to+\infty} \gamma_n > 0$. Let $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{H}$. Let $x_0, p_{-1} \in \mathcal{H}$ and set

$$\begin{cases} y_n = x_n - \gamma_n U_n(B(p_{n-1}) + a_n), \\ p_n = J_{\gamma_n U_n A}(y_n) + b_n, \\ q_n = p_n - \gamma_n U_n(B(p_n) + c_n), \\ x_{n+1} = x_n - y_n + q_n. \end{cases} \tag{3.11}$$

Then the following statements hold for some $\bar{x} \in zer(A + B)$:

(i) $\sum_{n\in\mathbb{N}} \|x_n - p_n\|^2 < +\infty$ *and* $\sum_{n\in\mathbb{N}} \|y_n - q_n\|^2 < +\infty$*;*

(ii) $x_n \rightharpoonup \bar{x}$ *and* $p_n \rightharpoonup \bar{x}$*.*

Proof. The proof is divided into the steps below because it contains many details.

Step 1: We introduce a new set of sequences in an error-free case, and give their properties relating to semi-scalar product and semi-norm.

Step 2: We prove a suitable inequality (shown as in (3.42)) and derive $\sum_{n\in\mathbb{N}} \|p_{n-1}-\tilde{p}_n\|^2 < +\infty$ by using Lemma 2.1.4 and Lemma 2.1.5, respectively.

Step 3: We prove the inequality (shown as in (3.47)) to assure that the sequence $(x_n)_{n\in\mathbb{N}}$ is $|\cdot|^2$-quasi-Fejér monotone (see Definition 3.1.4, when $\phi = |\cdot|^2$) with respect to the target set $zer(A+B)$ relative to $(U_n^{-1})_{n\in\mathbb{N}}$, and later we obtain that $\sum_{n\in\mathbb{N}} \|x_n - \tilde{p}_n\|^2 < +\infty$.

Step 4: Finally, (i) can be shown with above bounded summable results (in the previous steps of the proof). Consequently, we will apply the $|\cdot|^2$-quasi-Fejér monotone setting together with Theorem 3.1.6 to prove (ii) as required.

Now let us present the proof.

Step 1: We introduce new (sequence) parameters in terms of the error-free case of the algorithm and show the summable properties of involved sequences needed for the rest of the proof.

It follows from Lemma 3.1.2 that the sequences $(x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}, (p_n)_{n\in\mathbb{N}}$ and $(q_n)_{n\in\mathbb{N}}$ are well defined. From (3.10), we obtain that for each $x \in \mathcal{H}$

$$\langle U_n x, x\rangle \leq \|U_n x\|\|x\| \leq \|U_n\|\|x\|^2 \leq \mu\|x\|^2 = \langle \mu x, x\rangle \text{ implies that } U_n \preccurlyeq \mu\,\mathrm{Id},$$

and since $U_n \in \mathcal{P}_\alpha(\mathcal{H})$, then $U_n \succcurlyeq \alpha\,\mathrm{Id}$. Hence we have that

$$\begin{cases} \mu\,\mathrm{Id} \succcurlyeq U_n \succcurlyeq \alpha\,\mathrm{Id}, \\ \alpha^{-1}\,\mathrm{Id} \succcurlyeq U_n^{-1} \succcurlyeq \mu^{-1}\,\mathrm{Id}, \quad \text{by Lemma 3.1.1.} \end{cases} \tag{3.12}$$

For all $g_n \in \mathcal{H}$, $n \in \mathbb{N}$,

$$\begin{aligned} \|g_n\|\sqrt{\frac{1}{\alpha}} &= \sqrt{\langle g_n, \alpha^{-1}\,\mathrm{Id}\, g_n\rangle} \geq \sqrt{\langle g_n, U_n^{-1} g_n\rangle} \\ &= \|g_n\|_{U_n^{-1}} = \sqrt{\langle g_n, U_n^{-1} g_n\rangle} \geq \sqrt{\langle g_n, \mu^{-1}\,\mathrm{Id}\, g_n\rangle} = \|g_n\|\sqrt{\frac{1}{\mu}}. \end{aligned}$$

Thus we have that $\sqrt{\frac{1}{\mu}}\|g_n\| \leq \|g_n\|_{U_n^{-1}} \leq \|g_n\|\sqrt{\frac{1}{\alpha}}$, $\forall (g_n)_{n\in\mathbb{N}} \subset \mathcal{H}$. Similarly, we also have that $\sqrt{\alpha}\|g_n\| = \sqrt{\langle \alpha g_n, g_n\rangle} \leq \sqrt{\langle U_n g_n, g_n\rangle} = \|g_n\|_{U_n} = \sqrt{\langle U_n g_n, g_n\rangle} \leq \sqrt{\langle \mu g_n, g_n\rangle} = \|g_n\|\sqrt{\mu}$, $\forall (g_n)_{n\in\mathbb{N}} \subset \mathcal{H}$ and then

$$\sum_{n\in\mathbb{N}} \|g_n\| < +\infty \;\Leftrightarrow\; \sum_{n\in\mathbb{N}} \|g_n\|_{U_n^{-1}} < +\infty \;\Leftrightarrow\; \sum_{n\in\mathbb{N}} \|g_n\|_{U_n} < +\infty,\ \forall (g_n)_{n\in\mathbb{N}} \subset \mathcal{H}. \tag{3.13}$$

Let us set for all $n \in \mathbb{N}$,

$$\begin{cases} \tilde{y}_n = x_n - \gamma_n U_n B(p_{n-1}), \\ \tilde{p}_n = J_{\gamma_n U_n A}(\tilde{y}_n), \\ \tilde{q}_n = \tilde{p}_n - \gamma_n U_n B(\tilde{p}_n), \\ \tilde{x}_{n+1} = x_n - \tilde{y}_n + \tilde{q}_n, \end{cases} \quad \text{and} \quad \begin{cases} u_n = \gamma_n^{-1} U_n^{-1}(x_n - \tilde{p}_n) + B(\tilde{p}_n) - B(p_{n-1}), \\ e_n = \tilde{x}_{n+1} - x_{n+1} = y_n - q_n - \tilde{y}_n + \tilde{q}_n. \end{cases} \tag{3.14}$$

Since $\tilde{p}_n = J_{\gamma_n U_n A}(\tilde{y}_n)$, then $\tilde{y}_n \in \tilde{p}_n + \gamma_n U_n A(\tilde{p}_n)$ and therefore

$$\gamma_n^{-1} U_n^{-1}(\tilde{y}_n - \tilde{p}_n) \in A(\tilde{p}_n). \tag{3.15}$$

From (3.14) and (3.15), we have that $\forall n \in \mathbb{N}$,

$$\begin{aligned} u_n &= \gamma_n^{-1} U_n^{-1}\left(x_n - \gamma_n U_n B(p_{n-1}) - \tilde{p}_n\right) + B(p_{n-1}) + B(\tilde{p}_n) - B(p_{n-1}) \\ &\overset{(3.14)}{=} \gamma_n^{-1} U_n^{-1}(\tilde{y}_n - \tilde{p}_n) + B(\tilde{p}_n) \overset{(3.15)}{\in} A(\tilde{p}_n) + B(\tilde{p}_n) = (A+B)(\tilde{p}_n). \end{aligned} \tag{3.16}$$

We observe that for all $x \in \mathcal{H}$:

$$\|U_n x\|_{U_n^{-1}} = \sqrt{\langle U_n^{-1} U_n x, U_n x\rangle} = \sqrt{\langle x, U_n x\rangle} = \|x\|_{U_n}, \quad \forall n \in \mathbb{N}. \tag{3.17}$$

Applying (3.11), (3.14), (3.6), (3.17) and Lemma 3.1.2 yield $\forall n \in \mathbb{N}$,

$$\begin{aligned} \|y_n - \tilde{y}_n\|_{U_n^{-1}} &\overset{(3.11),(3.14)}{=} \|x_n - \gamma_n U_n\left(B(p_{n-1}) + a_n\right) - \left(x_n - \gamma_n U_n B(p_{n-1})\right)\|_{U_n^{-1}} \\ &= \gamma_n \|U_n a_n\|_{U_n^{-1}} \overset{(3.17)}{=} \gamma_n \|a_n\|_{U_n} \leq \lambda \|a_n\|_{U_n}, \end{aligned}$$

and

$$\begin{aligned} \|p_n - \tilde{p}_n\|_{U_n^{-1}} &\overset{(3.11),(3.14)}{=} \|J_{\gamma_n U_n A}(y_n) + b_n - J_{\gamma_n U_n A}(\tilde{y}_n)\|_{U_n^{-1}} \\ &\leq \|J_{\gamma_n U_n A}(y_n) - J_{\gamma_n U_n A}(\tilde{y}_n)\|_{U_n^{-1}} + \|b_n\|_{U_n^{-1}} \\ &\overset{Lem.3.1.2}{\leq} \|y_n - \tilde{y}_n\|_{U_n^{-1}} + \|b_n\|_{U_n^{-1}} \\ &\leq \lambda \|a_n\|_{U_n} + \|b_n\|_{U_n^{-1}}, \end{aligned} \tag{3.18}$$

and

$$\begin{aligned} \|q_n - \tilde{q}_n\|_{U_n^{-1}} &\overset{(3.11),(3.14)}{=} \|p_n - \gamma_n U_n\left(B(p_n) + c_n\right) - \tilde{p}_n + \gamma_n U_n B(\tilde{p}_n)\|_{U_n^{-1}} \\ &\leq \|p_n - \tilde{p}_n\|_{U_n^{-1}} + \|\gamma_n U_n B(\tilde{p}_n) - \gamma_n U_n B(p_n)\|_{U_n^{-1}} + \gamma_n \|U_n c_n\|_{U_n^{-1}}. \end{aligned} \tag{3.19}$$

From the β-Lipschitz continuity of B, we obtain $\forall n \in \mathbb{N}$,

$$
\begin{aligned}
\|\gamma_n U_n \left(B(\tilde{p}_n) - B(p_n)\right)\|^2_{U_n^{-1}} &\overset{(3.17)}{=} \|\gamma_n \left(B(\tilde{p}_n) - B(p_n)\right)\|^2_{U_n} \\
&\overset{(3.6)}{=} \gamma_n^2 \langle B(\tilde{p}_n) - B(p_n), U_n B(\tilde{p}_n) - U_n B(p_n)\rangle \\
&\leq \gamma_n^2 \|U_n\| \|B(\tilde{p}_n) - B(p_n)\|^2 \\
&\leq \gamma_n^2 \mu \beta^2 \|\tilde{p}_n - p_n\|^2 \\
&\quad \left(\text{since } \mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty, \beta\text{-Lipschitz continuity of } B\right) \\
&\leq \gamma_n^2 \mu^2 \beta^2 \|\tilde{p}_n - p_n\|^2_{U_n^{-1}} \ (\text{since } \mu^{-1}\|g_n\|^2 \leq \|g_n\|^2_{U_n^{-1}}, \forall g_n \in \mathcal{H}) \\
&\leq \|\tilde{p}_n - p_n\|^2_{U_n^{-1}} \ \left(\text{since } \gamma_n \leq \lambda < \frac{1}{\sqrt{10}\mu\beta} \leq \frac{1}{\beta\mu}\right).
\end{aligned} \tag{3.20}
$$

Then, it follows from (3.19), (3.20), (3.17) and (3.18) that $\forall n \in \mathbb{N}$,

$$
\begin{aligned}
\|q_n - \tilde{q}_n\|_{U_n^{-1}} &\overset{(3.19),(3.20)}{\leq} 2\|p_n - \tilde{p}_n\|_{U_n^{-1}} + \gamma_n \|U_n c_n\|_{U_n^{-1}} \\
&\overset{(3.17),(3.18)}{\leq} 2\left[\|b_n\|_{U_n^{-1}} + \lambda\|a_n\|_{U_n}\right] + \gamma_n \|c_n\|_{U_n} \\
&\leq 2\left[\|b_n\|_{U_n^{-1}} + \lambda\|a_n\|_{U_n}\right] + \lambda\|c_n\|_{U_n} \ (\text{since } \gamma_n \leq \lambda, \forall n \in \mathbb{N}).
\end{aligned}
$$

Hence, we have that $\forall n \in \mathbb{N}$,

$$
\begin{cases}
\|y_n - \tilde{y}_n\|_{U_n^{-1}} \leq \lambda \|a_n\|_{U_n}, \\
\|p_n - \tilde{p}_n\|_{U_n^{-1}} \leq \|b_n\|_{U_n^{-1}} + \lambda\|a_n\|_{U_n}, \\
\|q_n - \tilde{q}_n\|_{U_n^{-1}} \leq 2\left[\|b_n\|_{U_n^{-1}} + \lambda\|a_n\|_{U_n}\right] + \lambda\|c_n\|_{U_n}.
\end{cases} \tag{3.21}
$$

Since $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are absolutely summable sequences in $\mathcal{H}$, we derive from (3.13), (3.14) and (3.21) that

$$
\begin{cases}
\sum_{n\in\mathbb{N}} \|y_n - \tilde{y}_n\| < +\infty \quad \text{and} \quad \sum_{n\in\mathbb{N}} \|y_n - \tilde{y}_n\|_{U_n^{-1}} < +\infty, \\
\sum_{n\in\mathbb{N}} \|p_n - \tilde{p}_n\| < +\infty \quad \text{and} \quad \sum_{n\in\mathbb{N}} \|p_n - \tilde{p}_n\|_{U_n^{-1}} < +\infty, \\
\sum_{n\in\mathbb{N}} \|q_n - \tilde{q}_n\| < +\infty \quad \text{and} \quad \sum_{n\in\mathbb{N}} \|q_n - \tilde{q}_n\|_{U_n^{-1}} < +\infty,
\end{cases} \tag{3.22}
$$

According to (3.11), (3.14) and (3.21), we can derive that $\forall n \in \mathbb{N}$,

$$
\begin{aligned}
\|e_n\| &\overset{(3.14)}{=} \|y_n - q_n - \tilde{y}_n + \tilde{q}_n\| \\
&\leq \|y_n - \tilde{y}_n\| + \|q_n - \tilde{q}_n\| \\
&\overset{(3.21)}{\leq} (\lambda\|a_n\|_{U_n}) + 2\left(\|b_n\|_{U_n^{-1}} + \lambda\|a_n\|_{U_n}\right) + \lambda\|c_n\|_{U_n}.
\end{aligned} \tag{3.23}
$$

Since $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are absolutely summable sequences in $\mathcal{H}$, we derive from (3.13), (3.22) and (3.23) that $\sum_{n\in\mathbb{N}} \|e_n\| < +\infty$ and

$$\sum_{n\in\mathbb{N}} \|e_n\| < +\infty \Leftrightarrow \sum_{n\in\mathbb{N}} \|e_n\|_{U_n^{-1}} < +\infty \Leftrightarrow \sum_{n\in\mathbb{N}} \|e_n\|_{U_n} < +\infty. \tag{3.24}$$

Step 2: In this part of the proof, we prove some essential inequalities to show the inequality (3.42). Afterwards, we can show that $\sum\limits_{n\in\mathbb{N}} \|p_{n-1}-\tilde{p}_n\|^2 < +\infty$ by using Lemma 2.1.5. This summable sequence will be used in the next steps of the proof.

Let $x \in zer(A+B)$. Then, for every $n \in \mathbb{N}$, $(x, -\gamma_n U_n B(x)) \in Gra(\gamma_n U_n A)$, and (3.15) yields $(\tilde{p}_n, \tilde{y}_n - \tilde{p}_n) \in Gra(\gamma_n U_n A)$. Hence by monotonicity of $U_n A$ with respect to the scalar product $\langle\cdot,\cdot\rangle_{U_n^{-1}}$ in Lemma 3.1.2 (i), we obtain that

$$\langle \tilde{p}_n - x, \tilde{p}_n - \tilde{y}_n - \gamma_n U_n B(x)\rangle_{U_n^{-1}} \leq 0;$$

moreover, by monotonicity of $\gamma_n U_n B$ with respect to the scalar product $\langle\cdot,\cdot\rangle_{U_n^{-1}}$, we also have

$$\langle \tilde{p}_n - x, \gamma_n U_n B(x) - \gamma_n U_n B(\tilde{p}_n)\rangle_{U_n^{-1}} \leq 0.$$

By the last two inequalities, we obtain that $\forall n \in \mathbb{N}$,

$$\langle \tilde{p}_n - x, \tilde{p}_n - \tilde{y}_n - \gamma_n U_n B(\tilde{p}_n)\rangle_{U_n^{-1}} \leq 0. \tag{3.25}$$

Thus, we derive from (3.14) and (3.25) that $\forall n \in \mathbb{N}$,

$$\begin{aligned}
2\gamma_n\langle \tilde{p}_n - x, U_n B(p_{n-1}) - U_n B(\tilde{p}_n)\rangle_{U_n^{-1}} &= 2\overbrace{\langle \tilde{p}_n - x, \tilde{p}_n - \tilde{y}_n - \gamma_n U_n B(\tilde{p}_n)\rangle_{U_n^{-1}}}^{\leq 0} \\
&\quad + 2\langle \tilde{p}_n - x, \gamma_n U_n B(p_{n-1}) + \tilde{y}_n - \tilde{p}_n\rangle_{U_n^{-1}} \\
&\overset{(3.25)}{\leq} 2\langle \tilde{p}_n - x, \gamma_n U_n B(p_{n-1}) + \tilde{y}_n - \tilde{p}_n\rangle_{U_n^{-1}} \\
&\overset{(3.14)}{=} 2\langle \tilde{p}_n - x, x_n - \tilde{p}_n\rangle_{U_n^{-1}} \\
&= \|x_n - x\|^2_{U_n^{-1}} - \|\tilde{p}_n - x\|^2_{U_n^{-1}} - \|x_n - \tilde{p}_n\|^2_{U_n^{-1}}.
\end{aligned} \tag{3.26}$$

Next, using (3.12), (3.14), (3.26), (3.17), the β-Lipschitz continuity of B and Lemma 3.1.1, for every $n \in \mathbb{N}$, we obtain that

$$\begin{aligned}
\|\tilde{x}_{n+1} - x\|^2_{U_n^{-1}} &\overset{(3.14)}{=} \|\tilde{q}_n + x_n - \tilde{y}_n - x\|^2_{U_n^{-1}} \\
&\overset{(3.14)}{=} \|(\tilde{p}_n - x) + \gamma_n U_n(B(p_{n-1}) - B(\tilde{p}_n))\|^2_{U_n^{-1}} \\
&= \|\tilde{p}_n - x\|^2_{U_n^{-1}} + 2\gamma_n\langle \tilde{p}_n - x, U_n(B(p_{n-1}) - B(\tilde{p}_n))\rangle_{U_n^{-1}} + \gamma_n^2\|U_n(B(p_{n-1}) - B(\tilde{p}_n))\|^2_{U_n^{-1}} \\
&\overset{(3.26),(3.17)}{\leq} \|\tilde{p}_n - x\|^2_{U_n^{-1}} + \Big[\|x_n - x\|^2_{U_n^{-1}} - \|\tilde{p}_n - x\|^2_{U_n^{-1}} - \|x_n - \tilde{p}_n\|^2_{U_n^{-1}}\Big] + \gamma_n^2\|B(p_{n-1}) - B(\tilde{p}_n)\|^2_{U_n}
\end{aligned}$$

$$
\begin{aligned}
&\overset{(3.12)}{\leq} \|x_n - x\|^2_{U_n^{-1}} - \|x_n - \tilde{p}_n\|^2_{U_n^{-1}} + \gamma_n^2\beta^2\mu\|p_{n-1} - \tilde{p}_n\|^2 \text{ (since } \alpha\,\mathrm{Id} \preccurlyeq U_n \preccurlyeq \mu\,\mathrm{Id})\\
&\overset{(3.12)}{\leq} \|x_n - x\|^2_{U_n^{-1}} - \mu^{-1}\|x_n - \tilde{p}_n\|^2 + \gamma_n^2\beta^2\mu\|p_{n-1} - \tilde{p}_n\|^2 \text{ (since } \alpha^{-1}\mathrm{Id} \succcurlyeq U_n^{-1} \succcurlyeq \mu^{-1}\mathrm{Id}).
\end{aligned} \tag{3.27}
$$

By Parallelogram law, we have that $2\|x_n - \tilde{p}_n\|^2 + 2\|x_n - p_{n-1}\|^2 = \|p_{n-1} - \tilde{p}_n\|^2 + \|(x_n - \tilde{p}_n) + (x_n - p_{n-1})\|^2$, $\forall n \in \mathbb{N}$, then $\|p_{n-1} - \tilde{p}_n\|^2 \leq 2\|x_n - \tilde{p}_n\|^2 + 2\|x_n - p_{n-1}\|^2$ and so $\|x_n - \tilde{p}_n\|^2 \geq -\|x_n - p_{n-1}\|^2 + \frac{1}{2}\|p_{n-1} - \tilde{p}_n\|^2$. This means that $\forall n \in \mathbb{N}$,

$$
-\|x_n - \tilde{p}_n\|^2 \leq \|x_n - p_{n-1}\|^2 - \frac{1}{2}\|p_{n-1} - \tilde{p}_n\|^2. \tag{3.28}
$$

Now, it follows from (3.27) and (3.28) that $\forall n \in \mathbb{N}$,

$$
\begin{aligned}
\|\tilde{x}_{n+1} - x\|^2_{U_n^{-1}} &\overset{(3.27)}{\leq} \|x_n - x\|^2_{U_n^{-1}} - \mu^{-1}\|x_n - \tilde{p}_n\|^2 + \gamma_n^2\beta^2\mu\|p_{n-1} - \tilde{p}_n\|^2\\
&\overset{(3.28)}{\leq} \|x_n - x\|^2_{U_n^{-1}} + \mu^{-1}\left[\|x_n - p_{n-1}\|^2 - \frac{1}{2}\|p_{n-1} - \tilde{p}_n\|^2\right] + \gamma_n^2\beta^2\mu\|p_{n-1} - \tilde{p}_n\|^2\\
&= \|x_n - x\|^2_{U_n^{-1}} + \mu^{-1}\|x_n - p_{n-1}\|^2 + \left(\gamma_n^2\beta^2\mu - \frac{1}{2\mu}\right)\|p_{n-1} - \tilde{p}_n\|^2.
\end{aligned} \tag{3.29}
$$

Then we find that $\forall n \in \mathbb{N}$,

$$
\|\tilde{x}_{n+1} - x\|^2_{U_n^{-1}} + \left(\frac{1}{2\mu} - \gamma_n^2\beta^2\mu\right)\|p_{n-1} - \tilde{p}_n\|^2 \leq \|x_n - x\|^2_{U_n^{-1}} + \mu^{-1}\|x_n - p_{n-1}\|^2. \tag{3.30}
$$

Since (3.11) gives us that for all $n \in \mathbb{N}$, $x_{n+1} = x_n - y_n + q_n = \gamma_n U_n\left(B(p_{n-1}) + a_n\right) + p_n - \gamma_n U_n\left(B(p_n) + c_n\right)$, then we have $x_n = \gamma_{n-1}U_{n-1}(B(p_{n-2}) + a_{n-1}) + p_{n-1} - \gamma_{n-1}U_{n-1}(B(p_{n-1}) + c_{n-1})$. Therefore $\forall n \in \mathbb{N}$,

$$
\begin{aligned}
\|x_n - p_{n-1}\| &= \|\gamma_{n-1}U_{n-1}\left(B(p_{n-2}) + a_{n-1}\right) + p_{n-1} - \gamma_{n-1}U_{n-1}\left(B(p_{n-1}) + c_{n-1}\right) - p_{n-1}\|\\
&\leq \gamma_{n-1}\|U_{n-1}B(p_{n-2}) - U_{n-1}B(p_{n-1})\| + \gamma_{n-1}\|U_{n-1}\left(a_{n-1} - c_{n-1}\right)\|\\
&\leq \gamma_{n-1}\mu\beta\|p_{n-2} - p_{n-1}\| + \gamma_{n-1}\mu\left(\|a_{n-1}\| + \|c_{n-1}\|\right).
\end{aligned} \tag{3.31}
$$

It follows from (3.30), (3.31) and Cauchy–Schwarz inequality that $\forall n \in \mathbb{N}$,

$$\begin{aligned}
&\|\tilde{x}_{n+1} - x\|^2_{U_n^{-1}} + \left(\frac{1}{2\mu} - \gamma_n^2\beta^2\mu\right)\|p_{n-1} - \tilde{p}_n\|^2 \\
&\overset{(3.30)}{\le} \|x_n - x\|^2_{U_n^{-1}} + \mu^{-1}\|x_n - p_{n-1}\|^2 \\
&\overset{(3.31)}{\le} \|x_n - x\|^2_{U_n^{-1}} + \mu^{-1}\left[\gamma_{n-1}\mu\beta\|p_{n-2} - p_{n-1}\| + \gamma_{n-1}\mu\left(\|a_{n-1}\| + \|c_{n-1}\|\right)\right]^2 \\
&\le \|x_n - x\|^2_{U_n^{-1}} + \mu^{-1}\left[2(\gamma_{n-1}\mu\beta)^2\|p_{n-2} - p_{n-1}\|^2 + 2(\gamma_{n-1}\mu)^2(\|a_{n-1}\| + \|c_{n-1}\|)^2\right] \\
&\le \|x_n - x\|^2_{U_n^{-1}} + 2\gamma_{n-1}^2\beta^2\mu\|p_{n-2} - p_{n-1}\|^2 + 2\gamma_{n-1}^2\mu(\|a_{n-1}\| + \|c_{n-1}\|)^2 \\
&\le \|x_n - x\|^2_{U_n^{-1}} + 2\gamma_{n-1}^2\beta^2\mu\left[2(\|p_{n-2} - \tilde{p}_{n-1}\|^2 + \|\tilde{p}_{n-1} - p_{n-1}\|^2)\right] \\
&\quad + 2\gamma_{n-1}^2\mu(\|a_{n-1}\| + \|c_{n-1}\|)^2 \\
&= \|x_n - x\|^2_{U_n^{-1}} + 4\gamma_{n-1}^2\mu\beta^2\|p_{n-2} - \tilde{p}_{n-1}\|^2 + 4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1} - p_{n-1}\|^2 \\
&\quad + 2\gamma_{n-1}^2\mu(\|a_{n-1}\| + \|c_{n-1}\|)^2.
\end{aligned} \tag{3.32}$$

Let $z_n := \tilde{x}_{n+1} - x = x_n - \tilde{y}_n + \tilde{q}_n - x$ and $M_n := \frac{1}{2\mu} - \gamma_n^2\beta^2\mu > 0$ (since $\gamma_n < \frac{1}{\sqrt{2}\beta\mu}$, $\forall n \in \mathbb{N}$), then we derive from (3.32) that $\forall n \in \mathbb{N}$,

$$\begin{aligned}
\|z_n\|^2_{U_n^{-1}} + M_n\|p_{n-1} - \tilde{p}_n\|^2 &= \|\tilde{x}_{n+1} - x\|^2_{U_n^{-1}} + \left(\frac{1}{2\mu} - \gamma_n^2\beta^2\mu\right)\|p_{n-1} - \tilde{p}_n\|^2 \\
&\le \|x_n - x\|^2_{U_n^{-1}} + 4\gamma_{n-1}^2\mu\beta^2\|p_{n-2} - \tilde{p}_{n-1}\|^2 \\
&\quad + 4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1} - p_{n-1}\|^2 + 2\gamma_{n-1}^2\mu(\|a_{n-1}\| + \|c_{n-1}\|)^2.
\end{aligned} \tag{3.33}$$

Applying (3.11) and (3.14), we obtain that $x_{n+1} - x = \tilde{x}_{n+1} - x - \tilde{x}_{n+1} + x_{n+1} = z_n - e_n$. Then we get that

$$\|x_{n+1} - x\|^2_{U_n^{-1}} = \|z_n\|^2_{U_n^{-1}} - 2\langle z_n, e_n\rangle_{U_n^{-1}} + \|e_n\|^2_{U_n^{-1}}. \tag{3.34}$$

From (3.10), we know that $(1+\eta_n)U_{n+1} \succcurlyeq U_n$, $\forall n \in \mathbb{N}$. It follows from (3.34) that

$$\begin{aligned}
\|x_{n+1} - x\|^2_{U_{n+1}^{-1}} &\le (1+\eta_n)\|x_{n+1} - x\|^2_{U_n^{-1}} \\
&= (1+\eta_n)\left(\|z_n\|^2_{U_n^{-1}} - 2\langle z_n, e_n\rangle_{U_n^{-1}} + \|e_n\|^2_{U_n^{-1}}\right).
\end{aligned} \tag{3.35}$$

By using (3.33) and (3.35) yield $\forall n \in \mathbb{N}$,

$$\begin{aligned}
\|x_{n+1} - x\|^2_{U_{n+1}^{-1}} + M_n\|p_{n-1} - \tilde{p}_n\|^2 &\overset{(3.35)}{\le} \left(\|z_n\|^2_{U_n^{-1}} + M_n\|p_{n-1} - \tilde{p}_n\|^2\right) + \eta_n\|z_n\|^2_{U_n^{-1}} \\
&\quad + (1+\eta_n)(-2\langle z_n, e_n\rangle_{U_n^{-1}}) + (1+\eta_n)\|e_n\|^2_{U_n^{-1}} \\
&\overset{(3.33)}{\le} \big[\|x_n - x\|^2_{U_n^{-1}} + 4\gamma_{n-1}^2\mu\beta^2\|p_{n-2} - \tilde{p}_{n-1}\|^2
\end{aligned}$$

$$
\begin{aligned}
&+4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1}-p_{n-1}\|^2+2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2\big]\\
&+\eta_n\big[\|x_n-x\|_{U_n^{-1}}^2+4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2\\
&+4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1}-p_{n-1}\|^2+2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2\big]\\
&+(1+\eta_n)(-2\langle z_n,e_n\rangle_{U_n^{-1}})+(1+\eta_n)\|e_n\|_{U_n^{-1}}^2\\
=&\,(1+\eta_n)\|x_n-x\|_{U_n^{-1}}^2+(1+\eta_n)4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2\\
&+(1+\eta_n)4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1}-p_{n-1}\|^2\\
&+(1+\eta_n)2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2\\
&+(1+\eta_n)(-2\langle z_n,e_n\rangle_{U_n^{-1}})+(1+\eta_n)\|e_n\|_{U_n^{-1}}^2. \qquad (3.36)
\end{aligned}
$$

Now, from (3.33), we obtain that $\forall n \in \mathbb{N}$,

$$
\begin{aligned}
-2\langle z_n,e_n\rangle_{U_n^{-1}} &\leq 2\|z_n\|_{U_n^{-1}}\|e_n\|_{U_n^{-1}}\\
&\leq (\|z_n\|_{U_n^{-1}}^2+2\|z_n\|_{U_n^{-1}}+1)\|e_n\|_{U^{-1}}\\
&=(\|z_n\|_{U_n^{-1}}^2+1)\|e_n\|_{U_n^{-1}}\\
&\overset{(3.33)}{\leq} \big[\|x_n-x\|_{U_n^{-1}}^2+4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2+4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1}-p_{n-1}\|^2\\
&\quad+2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2+1\big]\|e_n\|_{U_n^{-1}}\\
&=\|e_n\|_{U_n^{-1}}\|x_n-x\|_{U_n^{-1}}^2+\|e_n\|_{U_n^{-1}}4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2\\
&\quad+\|e_n\|_{U_n^{-1}}4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1}-p_{n-1}\|^2\\
&\quad+\|e_n\|_{U_n^{-1}}2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2+\|e_n\|_{U_n^{-1}}. \qquad (3.37)
\end{aligned}
$$

Consider (3.36) together with (3.37), then $\forall n \in \mathbb{N}$,

$$
\begin{aligned}
\|x_{n+1}-x\|_{U_{n+1}^{-1}}^2+M_n\|p_{n-1}-\tilde{p}_n\|^2 \overset{(3.36),(3.37)}{\leq}&\,(1+\eta_n)\|x_n-x\|_{U_n^{-1}}^2+(1+\eta_n)4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2\\
&+(1+\eta_n)4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1}-p_{n-1}\|^2\\
&+(1+\eta_n)2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2\\
&+(1+\eta_n)\Big[\|e_n\|_{U_n^{-1}}\|x_n-x\|_{U_n^{-1}}^2\\
&+\|e_n\|_{U_n^{-1}}4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2\\
&+\|e_n\|_{U_n^{-1}}4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1}-p_{n-1}\|^2\\
&+\|e_n\|_{U_n^{-1}}2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2+\|e_n\|_{U_n^{-1}}\Big]\\
&+(1+\eta_n)\|e_n\|_{U_n^{-1}}^2\\
=&\,(1+\eta_n)\Big[\|x_n-x\|_{U_n^{-1}}^2+4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2\\
&+4\gamma_{n-1}^2\mu\beta^2\|\tilde{p}_{n-1}-p_{n-1}\|^2+2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2\Big]
\end{aligned}
$$

$$
\begin{aligned}
&+(1+\eta_n)\|e_n\|_{U_n^{-1}}\Big[\|x_n-x\|^2_{U_n^{-1}}+4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde p_{n-1}\|^2\\
&+4\gamma_{n-1}^2\mu\beta^2\|\tilde p_{n-1}-p_{n-1}\|^2+2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2+1\Big]\\
&+(1+\eta_n)\|e_n\|^2_{U_n^{-1}}\\
\le&\Big[(1+\eta_n)\Big(1+\|e_n\|_{U_n^{-1}}\Big)\Big]\Big[\|x_n-x\|^2_{U_n^{-1}}+4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde p_{n-1}\|^2\Big]\\
&+\Big[(1+\eta_n)\Big(1+\|e_n\|_{U_n^{-1}}\Big)\Big]\Big[4\gamma_{n-1}^2\mu\beta^2\|\tilde p_{n-1}-p_{n-1}\|^2\Big]\\
&+\Big[(1+\eta_n)\Big(1+\|e_n\|_{U_n^{-1}}\Big)\Big]\Big[2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2\Big]\\
&+2(1+\eta_n)\|e_n\|^2_{U_n^{-1}}\\
=&\Big[1+\Big(\eta_n+\|e_n\|_{U_n^{-1}}+\eta_n\|e_n\|_{U_n^{-1}}\Big)\Big]\Big[\|x_n-x\|^2_{U_n^{-1}}+4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde p_{n-1}\|^2\Big]\\
&+\Big[1+\Big(\eta_n+\|e_n\|_{U_n^{-1}}+\eta_n\|e_n\|_{U_n^{-1}}\Big)\Big]\Big[4\gamma_{n-1}^2\mu\beta^2\|\tilde p_{n-1}-p_{n-1}\|^2\Big]\\
&+\Big[1+\Big(\eta_n+\|e_n\|_{U_n^{-1}}+\eta_n\|e_n\|_{U_n^{-1}}\Big)\Big]\Big[2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2\Big]\\
&+2(1+\eta_n)\|e_n\|^2_{U_n^{-1}}. \qquad (3.38)
\end{aligned}
$$

Since $\gamma_n\le\lambda<\frac{1}{\sqrt{10}\mu\beta}\left(<\frac{1}{\sqrt{2}\beta\mu}\right)$, $\forall n\in\mathbb{N}$, we have $\lambda^2<\frac{1}{10\mu^2\beta^2}$ and so $5\lambda^2\mu\beta^2<\frac{1}{2\mu}$. Thus, $4\gamma_{n-1}^2\mu\beta^2+\gamma_{n-1}^2\mu\beta^2<4\lambda^2\mu\beta^2+\lambda^2\mu\beta^2<\frac{1}{2\mu}$ $\left(\text{or } 4\gamma_{n-1}^2\mu\beta^2+\gamma_n^2\mu\beta^2<4\lambda^2\mu\beta^2+\lambda^2\mu\beta^2<\frac{1}{2\mu}\right)$. Therefore $4\gamma_{n-1}^2\mu\beta^2<\frac{1}{2\mu}-\gamma_{n-1}^2\mu\beta^2=M_{n-1}$.
From (3.38), we let $D_n:=\|x_{n+1}-x\|^2_{U_{n+1}^{-1}}+M_n\|p_{n-1}-\tilde p_n\|^2$ and $D_{n-1}:=\|x_n-x\|^2_{U_n^{-1}}+M_{n-1}\|p_{n-2}-\tilde p_{n-1}\|^2$, then we have $\forall n\in\mathbb{N}$,

$$
\begin{aligned}
D_n&=\|x_{n+1}-x\|^2_{U_{n+1}^{-1}}+M_n\|p_{n-1}-\tilde p_n\|^2\\
&\overset{(3.38)}{\le}\Big[1+\Big(\eta_n+\|e_n\|_{U_n^{-1}}+\eta_n\|e_n\|_{U_n^{-1}}\Big)\Big]\left[\|x_n-x\|^2_{U_n^{-1}}+\overbrace{4\gamma_{n-1}^2\mu\beta^2}^{<M_{n-1}}\|p_{n-2}-\tilde p_{n-1}\|^2\right]\\
&\quad+\Big[1+\Big(\eta_n+\|e_n\|_{U_n^{-1}}+\eta_n\|e_n\|_{U_n^{-1}}\Big)\Big]\Big[4\gamma_{n-1}^2\mu\beta^2\|\tilde p_{n-1}-p_{n-1}\|^2\Big]\\
&\quad+\Big[1+\Big(\eta_n+\|e_n\|_{U_n^{-1}}+\eta_n\|e_n\|_{U_n^{-1}}\Big)\Big]\Big[2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2\Big]\\
&\quad+2(1+\eta_n)\|e_n\|^2_{U_n^{-1}}\\
&\le(1+\tilde\eta_n)\,D_{n-1}+E_n, \qquad (3.39)
\end{aligned}
$$

where $\tilde\eta_n:=\eta_n+\|e_n\|_{U_n^{-1}}+\eta_n\|e_n\|_{U_n^{-1}}$, and

$$
\begin{aligned}
E_n=&[1+\tilde\eta_n]\Big[4\gamma_{n-1}^2\mu\beta^2\|\tilde p_{n-1}-p_{n-1}\|^2\Big]\\
&+[1+\tilde\eta_n]\Big[2\gamma_{n-1}^2\mu(\|a_{n-1}\|+\|c_{n-1}\|)^2\Big]\\
&+2(1+\eta_n)\|e_n\|^2_{U_n^{-1}}. \qquad (3.40)
\end{aligned}
$$

Since $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are absolutely summable sequences in $\mathcal{H}$, (3.22), (3.23), (3.24), $(\gamma_n)_{n\in\mathbb{N}}$ is bounded and $\eta_n \in \ell^1_+(\mathbb{N})$, then we can conclude that

$$\tilde{\eta}_n \in \ell^1_+(\mathbb{N}) \quad \text{and} \quad \sum_{n\in\mathbb{N}} E_n < +\infty. \tag{3.41}$$

From (3.39) and (3.41), we know that

$$D_n \leq (1+\tilde{\eta}_n)D_{n-1} + E_n \quad \text{with} \quad \tilde{\eta}_n \in \ell^1_+(\mathbb{N}), \sum_{n\in\mathbb{N}} E_n < +\infty. \tag{3.42}$$

Applying Lemma 2.1.4, we have that $(D_n)_{n\in\mathbb{N}}$ converges. This means that $(D_n)_{n\in\mathbb{N}}$ is bounded and therefore $(\|x_{n+1}-x\|^2_{U^{-1}_{n+1}})_{n\in\mathbb{N}}$ and $(M_n\|p_{n-1}-\tilde{p}_n\|^2)_{n\in\mathbb{N}}$ are bounded. Since $M_n = \frac{1}{2\mu} - \gamma_n^2\beta^2\mu \geq \frac{1}{2\mu} - \lambda^2\beta^2\mu$ and $\lambda < \frac{1}{\sqrt{10}\mu\beta} < \frac{1}{\sqrt{2}\beta\mu}$, then $\liminf_{n\to+\infty} M_n > 0$. Therefore we also have that $(\|p_{n-1}-\tilde{p}_n\|^2)_{n\in\mathbb{N}}$ is bounded. Consequently, there are θ and ζ in $\mathbb{R}$ such that $\theta = \sup_{n\in\mathbb{N}} \|x_n - x\|^2_{U_n^{-1}}$ and $\zeta = \sup_{n\in\mathbb{N}} \|p_{n-1}-\tilde{p}_n\|^2$, respectively. Now, we consider again (3.39) and obtain that $\forall n \in \mathbb{N}$,

$$\begin{aligned}
\|x_{n+1}-x\|^2_{U^{-1}_{n+1}} + M_n\|p_{n-1}-\tilde{p}_n\|^2 &\overset{(3.39)}{\leq} (1+\tilde{\eta}_n)\left[\|x_n - x\|^2_{U_n^{-1}} + 4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2\right] + E_n \\
&= \|x_n - x\|^2_{U_n^{-1}} + 4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2 \\
&\quad + \tilde{\eta}_n\|x_n - x\|^2_{U_n^{-1}} + \tilde{\eta}_n 4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2 + E_n \\
&\leq \|x_n - x\|^2_{U_n^{-1}} + 4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2 \\
&\quad + \tilde{\eta}_n\theta + \tilde{\eta}_n 4\gamma_{n-1}^2\mu\beta^2\zeta + E_n.
\end{aligned} \tag{3.43}$$

For convenience, we let $\tilde{M}_{n-1} := 4\gamma_{n-1}^2\mu\beta^2$ and we can show that $\liminf_{n\to+\infty}(M_n - \tilde{M}_n) > 0$. Since $\gamma_n \leq \lambda < \frac{1}{\sqrt{10}\mu\beta}$, $\forall n \in \mathbb{N}$, we have that $4\gamma_n^2\mu\beta^2 + \gamma_n^2\mu\beta^2 \leq 4\lambda^2\mu\beta^2 + \lambda^2\mu\beta^2 < \frac{1}{2\mu}$. Hence $\limsup_{n\to+\infty}\left(4\gamma_n^2\mu\beta^2 + \gamma_n^2\mu\beta^2\right) < \frac{1}{2\mu}$ and so $\liminf_{n\to+\infty}\left[\frac{1}{2\mu} - \left(4\gamma_n^2\mu\beta^2 + \gamma_n^2\mu\beta^2\right)\right] > 0$, this means that $\frac{1}{2\mu} - \left(4\gamma_n^2\mu\beta^2 + \gamma_n^2\mu\beta^2\right) > \varepsilon > 0$ for some $0 < \varepsilon \in \mathbb{R}$, or equivalently, $M_n - \tilde{M}_n = \left[\frac{1}{2\mu} - \gamma_n^2\mu\beta^2\right] - \left[4\gamma_n^2\mu\beta^2\right] = \frac{1}{2\mu} - \left(4\gamma_n^2\mu\beta^2 + \gamma_n^2\mu\beta^2\right) > \varepsilon > 0$ for some $0 < \varepsilon \in \mathbb{R}$. Then it follows from (3.43) that $\forall n \in \mathbb{N}$,

$$\begin{aligned}
&\|x_{n+1}-x\|^2_{U^{-1}_{n+1}} + \left(M_n - \tilde{M}_n\right)\|p_{n-1}-\tilde{p}_n\|^2 + \tilde{M}_n\|p_{n-1}-\tilde{p}_n\|^2 \\
&\overset{(3.43)}{\leq} \|x_n - x\|^2_{U_n^{-1}} + 4\gamma_{n-1}^2\mu\beta^2\|p_{n-2}-\tilde{p}_{n-1}\|^2 + \tilde{\eta}_n\theta + \tilde{\eta}_n 4\gamma_{n-1}^2\mu\beta^2\zeta + E_n \\
&= \|x_n - x\|^2_{U_n^{-1}} + \tilde{M}_{n-1}\|p_{n-2}-\tilde{p}_{n-1}\|^2 + \tilde{\eta}_n\theta + \tilde{\eta}_n\tilde{M}_{n-1}\zeta + E_n.
\end{aligned} \tag{3.44}$$

We apply Lemma 2.1.5 *(iii)* with the setting of $\chi = 1$, $\alpha_n = \|x_n - x\|^2_{U_n^{-1}} + \tilde{M}_{n-1}\|p_{n-2}-\tilde{p}_{n-1}\|^2$, $\beta_n = \left(M_n - \tilde{M}_n\right)\|p_{n-1}-\tilde{p}_n\|^2$, $\varepsilon_n = \tilde{\eta}_n\theta + \tilde{\eta}_n\tilde{M}_{n-1}\zeta +$

E_n. It follows from (3.41), the fact that for all $n \in \mathbb{N}$, $M_n, \tilde{M}_n$ are bounded $\Big($since $M_n = \frac{1}{2\mu} - \gamma_n^2\beta^2\mu, \tilde{M}_n = 4\gamma_n^2\mu(1+\beta^2)\Big)$, and $M_n - \tilde{M}_n > \varepsilon > 0$ for some $0 < \varepsilon \in \mathbb{R}$. We can conclude that $\sum_{n\in\mathbb{N}} \Big(M_n - \tilde{M}_n\Big) \|p_{n-1} - \tilde{p}_n\|^2 < +\infty$ and so

$$\sum_{n\in\mathbb{N}} \|p_{n-1} - \tilde{p}_n\|^2 < +\infty. \tag{3.45}$$

Step 3: This part gives us the $|\cdot|^2$-quasi-Fejér monotone property of the sequence generated by our method and provides the summable sequence $\|x_n - \tilde{p}_n\|^2$, which plays an essential role in the next step of the proof.

It follows from (3.10), (3.27), Lemma 3.1.1, (3.45), $(\gamma_n)_{n\in\mathbb{N}} \le \lambda$, and (η_n), $(x_n)_{n\in\mathbb{N}}$ are bounded that $\forall n \in \mathbb{N}$,

$$\begin{aligned}
\|z_n\|^2_{U^{-1}_{n+1}} = \|\tilde{x}_{n+1} - x\|^2_{U^{-1}_{n+1}} &\overset{(3.10),\text{Lem } 3.1.1}{\le} (1+\eta_n)\|\tilde{x}_{n+1} - x\|^2_{U_n^{-1}} \\
&\overset{(3.27)}{\le} (1+\eta_n)\Big[\|x_n - x\|^2_{U_n^{-1}} - \mu^{-1}\|x_n - \tilde{p}_n\|^2 + \gamma_n^2\beta^2\mu\|p_{n-1} - \tilde{p}_n\|^2\Big] \\
&\le (1+\eta_n)\|x_n - x\|^2_{U_n^{-1}} + (1+\eta_n)\gamma_n^2\beta^2\mu\|p_{n-1} - \tilde{p}_n\|^2 \\
&= \|x_n - x\|^2_{U_n^{-1}} + \eta_n\|x_n - x\|^2_{U_n^{-1}} + \gamma_n^2\beta\mu\|p_{n-1} - \tilde{p}_n\|^2 \\
&\quad + \eta_n\gamma_n^2\beta\mu\|p_{n-1} - \tilde{p}_n\|^2,
\end{aligned} \tag{3.46}$$

hence $\sup_{n\in\mathbb{N}} \|z_n\|^2_{U^{-1}_{n+1}} < +\infty$. It follows from $x_{n+1} - x = z_n - e_n$, (3.46) and Cauchy-Schwarz inequality that $\forall n \in \mathbb{N}$,

$$\begin{aligned}
\|x_{n+1} - x\|^2_{U^{-1}_{n+1}} &= \|z_n\|^2_{U^{-1}_{n+1}} - 2\langle z_n, e_n\rangle_{U^{-1}_{n+1}} + \|e_n\|^2_{U^{-1}_{n+1}} \\
&\overset{(3.46)}{\le} \Big[(1+\eta_n)\Big(\|x_n - x\|^2_{U_n^{-1}} - \mu^{-1}\|x_n - \tilde{p}_n\|^2 + \gamma_n^2\beta^2\mu\|p_{n-1} - \tilde{p}_n\|^2\Big)\Big] \\
&\quad + 2\|z_n\|_{U^{-1}_{n+1}}\|e_n\|_{U^{-1}_{n+1}} + \|e_n\|^2_{U^{-1}_{n+1}} \\
&\le (1+\eta_n)\|x_n - x\|^2_{U_n^{-1}} - (1+\eta_n)\mu^{-1}\|x_n - \tilde{p}_n\|^2 + \tilde{E}_n \\
&\le (1+\eta_n)\|x_n - x\|^2_{U_n^{-1}} + \tilde{E}_n,
\end{aligned} \tag{3.47}$$

where $\tilde{E}_n = \gamma_n^2\beta^2\mu\|p_{n-1} - \tilde{p}_n\|^2 + \eta_n\gamma_n^2\beta^2\mu\|p_{n-1} - \tilde{p}_n\|^2 + 2\|z_n\|_{U^{-1}_{n+1}}\|e_n\|_{U^{-1}_{n+1}} + \|e_n\|^2_{U^{-1}_{n+1}}$, in which $\sum_{n\in\mathbb{N}} \tilde{E}_n < +\infty$, because (3.24), (3.45), (3.46), $\beta \in (0, +\infty)$, $(\gamma_n)_{n\in\mathbb{N}} < \lambda$ and $\eta_n \in \ell^1_+(\mathbb{N})$.

From Definition 3.1.4, the inequality (3.47) shows that $(x_n)_{n\in\mathbb{N}}$ is $|\cdot|^2$- quasi-Fejér monotone with respect to the target set $zer(A+B)$ relative to $(U_n^{-1})_{n\in\mathbb{N}}$. Moreover, by Proposition 3.1.5, we have $(\|x_n - x\|_{U_n^{-1}})_{n\in\mathbb{N}}$ is bounded.

Now, it follows from (3.31) that

$$\|x_n - \tilde{p}_n\|^2 \le 2\|x_n - p_{n-1}\|^2 + 2\|p_{n-1} - \tilde{p}_n\|^2$$

$$
\begin{aligned}
&\overset{(3.31)}{\leq} 2\left[\gamma_{n-1}\mu\beta\|p_{n-2}-p_{n-1}\| + \gamma_{n-1}\mu\left(\|a_{n-1}\| + \|c_{n-1}\|\right)\right]^2 + 2\|p_{n-1}-\tilde{p}_n\|^2 \\
&\leq 2\Big(2\Big(\gamma_{n-1}^2\mu^2\beta^2(\|p_{n-2}-\tilde{p}_{n-1}\|+\|\tilde{p}_{n-1}-p_{n-1}\|)^2+\gamma_{n-1}^2\mu^2\left(\|a_{n-1}\|+\|c_{n-1}\|\right)^2\Big)\Big) \\
&\quad + 2\|p_{n-1}-\tilde{p}_n\|^2 \\
&\leq 4\big(2\gamma_{n-1}^2\mu^2\beta^2\left(\|p_{n-2}-\tilde{p}_{n-1}\|^2+\|\tilde{p}_{n-1}-p_{n-1}\|^2\right)\big)+2\gamma_{n-1}^2\mu^2 2\big(\|a_{n-1}\|^2+\|c_{n-1}\|^2\big) \\
&\quad + 2\|p_{n-1}-\tilde{p}_n\|^2 \\
&\leq 8\gamma_{n-1}^2\mu^2\beta^2\left(\|p_{n-2}-\tilde{p}_{n-1}\|^2+\|\tilde{p}_{n-1}-p_{n-1}\|^2\right)+4\gamma_{n-1}^2\mu^2\left(\|a_{n-1}\|^2+\|c_{n-1}\|^2\right) \\
&\quad + 2\|p_{n-1}-\tilde{p}_n\|^2,
\end{aligned}
$$

and therefore we use (3.21) and (3.45) to conclude that

$$\sum_{n\in\mathbb{N}} \|x_n - \tilde{p}_n\|^2 < +\infty. \tag{3.48}$$

Step 4: In this step, we will prove our desired results.
(i) It follows from (3.48) and (3.22) that

$$\sum_{n\in\mathbb{N}} \|x_n - p_n\|^2 \leq 2\sum_{n\in\mathbb{N}} \|x_n - \tilde{p}_n\|^2 + 2\sum_{n\in\mathbb{N}} \|p_n - \tilde{p}_n\|^2 < +\infty. \tag{3.49}$$

Futhermore, we can derive form (3.14), (3.22), (3.24), (3.45) and (3.48) that

$$
\begin{aligned}
\sum_{n\in\mathbb{N}} \|y_n - q_n\|^2 &= \sum_{n\in\mathbb{N}} \|\tilde{q}_n - \tilde{y}_n + q_n - \tilde{q}_n + \tilde{y}_n - y_n\|^2 \\
&\overset{(3.14)}{=} \sum_{n\in\mathbb{N}} \|\tilde{q}_n - \tilde{y}_n - e_n\|^2 \\
&\overset{(3.14)}{=} \sum_{n\in\mathbb{N}} \|\tilde{p}_n - \gamma_n U_n B(\tilde{p}_n) - (x_n - \gamma_n U_n B(p_{n-1})) - e_n\|^2 \\
&= \sum_{n\in\mathbb{N}} \|\tilde{p}_n - x_n + \gamma_n U_n\left(B(p_{n-1}) - B(\tilde{p}_n)\right) - e_n\|^2 \\
&\leq 3\sum_{n\in\mathbb{N}} \left(\|\tilde{p}_n - x_n\|^2 + \gamma_n^2\|U_n\|^2\|B(p_{n-1}) - B(\tilde{p}_n)\|^2 + \|e_n\|^2\right) \\
&\leq 3\sum_{n\in\mathbb{N}} \left(\|\tilde{p}_n - x_n\|^2 + \gamma_n^2\mu^2\beta^2\|p_{n-1} - \tilde{p}_n\|^2 + \|e_n\|^2\right) \\
&< +\infty \quad \text{(by using(3.24), (3.45) and (3.48))}.
\end{aligned} \tag{3.50}
$$

(ii) We want to show that $x_n \rightharpoonup \bar{x}$ and $p_n \rightharpoonup \bar{x}$ for some $\bar{x} \in zer(A+B)$. Let x be a weak cluster point of $(x_n)_{n\in\mathbb{N}}$. Then there exists a subsequence $(x_{k_n})_{n\in\mathbb{N}}$ that converges weakly to x. By (3.48), we know that $\sum\limits_{n\in\mathbb{N}} \|x_n - \tilde{p}_n\|^2 < +\infty$, then $\tilde{p}_{k_n} \rightharpoonup x$. Furthermore, it follows from (3.14), (3.16), (3.45), (3.48) and $\liminf\limits_{n\to\infty} \gamma_n > 0$ that $u_{k_n} = \gamma_{k_n}^{-1}U_{k_n}^{-1}(x_{k_n} - \tilde{p}_{k_n}) + B(\tilde{p}_{k_n}) - B(p_{k_{n-1}}) \to 0$. By (3.16) we also know that

$(\tilde{p}_{k_n}, u_{k_n}) \in Gra(A+B)$. By Proposition 2.6.5, we obtain that $(x, 0) \in Gra(A+B)$ and then $x \in zer(A+B)$. Altogether, it follows from (3.47), Lemma 3.1.3 and Theorem 3.1.6 that $x_n \rightharpoonup \bar{x}$ and hence $p_n \rightharpoonup \bar{x}$, by using (i) $\left(\text{namely, } \sum_{n\in\mathbb{N}} \|x_n - p_n\|^2 < +\infty\right)$. □

Remark 3.2.2. We give some remarks below.

(i) From the proposed algorithm, if we put $a_n = c_n = 0$ but b_n still remains, the algorithm turns into $\forall n \in \mathbb{N}$,

$$\begin{cases} p_n = J_{\gamma_n U_n A}(x_n - \gamma_n U_n B(p_{n-1})) + b_n, \\ x_{n+1} = p_n + \gamma_n U_n(B(p_{n-1}) - B(p_n)). \end{cases}$$

Then, the method is a variation of forward-reflected-backward (1.10), which has variable metrics and error terms. In fact, this algorithm is nothing else than OGDA (1.11) when setting $A = 0$, $a_n = b_n = c_n = 0$ and $U_n = Id$. Although the shadow Douglus-Rachford (SDR) method in [35] can be reduced to OGDA (in case $A = 0$), our algorithm is different from SDR.

(ii) Because the error terms and variable metrics appearing in this algorithm, they make our method more flexible to handle. Indeed, they can generate a more alternative variable metric algorithm with error by using a different error model and involved iteration-dependent variable metrics.

(iii) In the error-free case ($a_n = b_n = c_n = 0$), we can observe that the results hold when the stepsize fulfills $0 < \gamma_n < \frac{1}{2\mu\beta}$ and $\liminf_{n\to+\infty} \gamma_n > 0$. In fact, we have that $y_n = \tilde{y}_n$, $p_n = \tilde{p}_n$, $q_n = \tilde{q}_n$, and $e_n = 0$, then the inequality (3.38) can be reduced to $\forall n \in \mathbb{N}$,

$$\|x_{n+1} - x\|^2_{U^{-1}_{n+1}} + M_n\|p_{n-1} - \tilde{p}_n\|^2 \le (1+\eta_n)\|x_n - x\|^2_{U_n^{-1}}.$$

Thus, we can prove the inequality: $\forall n \in \mathbb{N}$,

$$\begin{aligned} \|x_{n+1} - x\|^2_{U^{-1}_{n+1}} + \overbrace{\left(\frac{1}{2\mu} - \gamma_n^2\beta^2\mu - \gamma_n^2\beta^2\mu\right)}^{\hat{M}_n} \|p_{n-1} - \tilde{p}_n\|^2 + \left(\gamma_n^2\beta^2\mu\right)\|p_{n-1} - \tilde{p}_n\|^2 \\ \le (1+\eta_n)\|x_n - x\|^2_{U_n^{-1}} + \left(\gamma_{n-1}^2\beta^2\mu\right)\|p_{n-2} - \tilde{p}_{n-1}\|^2. \end{aligned}$$

To apply directly Lemma 2.1.5, we need to assure that $\liminf_{n\to+\infty} \hat{M}_n > 0$. This means that it is sufficient to assume that $0 < \gamma_n < \frac{1}{2\mu\beta}$ and $\liminf_{n\to+\infty} \gamma_n > 0$ for this case.

(iv) Although our algorithm becomes forward-reflected-backward (1.10) in the error-free scenario ($a_n = b_n = c_n = 0$) and $U_n = Id$, the analysis is not the same. Indeed, we do not have the inner product terms in the energy as in [62]. Moreover, we observe that the inequality (3.42) becomes $D_n \le D_{n-1}$ in this situation. Therefore, Lemma 2.1.4 is no longer required. Our energy is based on the Tseng's method adapted to the scenario with extrapolation from the part.

3.3 A Primal-Dual Solver for Monotone Inclusion Problems

Many non-smooth optimization problems can be written as monotone inclusion primal-dual problems. By using the product space approach (see the technique used by Briceño-Arias and Combettes [23]), we are able now to solve **Problem 3.1**. Therefore we derive a result which follows from Theorem 3.2.1 as shown below.

Theorem 3.3.1. *Let α be in $(0,+\infty)$, let $(\eta_{0,n})_{n\in\mathbb{N}}$ be a sequence in $\ell^1_+(\mathbb{N})$, let $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$, and for every $i\in\{1,\dots,m\}$, let $(\eta_{i,n})_{n\in\mathbb{N}}$ be a sequence in $\ell^1_+(\mathbb{N})$, let $(U_{i,n})_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{G}_i)$ such that $\mu = sup_{n\in\mathbb{N}}\{\|U_n\|,\|U_{1,n}\|,\dots,\|U_{m,n}\|\} < +\infty$ and*

$$
\begin{aligned}
&(1+\eta_{0,n})U_{n+1} \succcurlyeq U_n,\ \forall n\in\mathbb{N},\\
\text{and}\quad &(1+\eta_{i,n})U_{i,n+1} \succcurlyeq U_{i,n}\ \forall i\in\{1,\dots,m\}.
\end{aligned}
\tag{3.51}
$$

Let $(a_{1,n})_{n\in\mathbb{N}}, (b_{1,n})_{n\in\mathbb{N}}$ and $(c_{1,n})_{n\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{H}$, and for every $i\in\{1,\dots,m\}$, let $(a_{2,i,n})_{n\in\mathbb{N}}, (b_{2,i,n})_{n\in\mathbb{N}}$ and $(c_{2,i,n})_{n\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{G}_i$. Furthermore, set

$$
\beta = v_0 + \sqrt{\sum_{i=1}^m \|L_i\|^2},
\tag{3.52}
$$

let $x_0\in\mathcal{H}$, let $(v_{1,0},\dots,v_{m,0})\in\mathcal{G}_1\bigoplus\cdots\bigoplus\mathcal{G}_m$, let $(\gamma_n)_{n\in\mathbb{N}}\leq\lambda$ with $\lambda<\frac{1}{\sqrt{10}\mu\beta}$ and $\liminf\limits_{n\to+\infty}\gamma_n>0$. Set $\forall n\in\mathbb{N}$,

$$
\left\lfloor
\begin{array}{l}
y_{1,n} = x_n - \gamma_n U_n\left(C(p_{1,n-1}) + \sum\limits_{i=1}^m L_i^*(p_{2,i,n-1}) + a_{1,n}\right),\\
\text{for } i=1,\dots,m\\
\quad\left\lfloor
\begin{array}{l}
y_{2,i,n} = v_{i,n} + \gamma_n U_{i,n}\left(L_i(p_{1,n-1}) + a_{2,i,n}\right),\\
p_{2,i,n} = J_{\gamma_n U_{i,n} B_i^{-1}}(y_{2,i,n} - \gamma_n U_{i,n} r_i) + b_{2,i,n},
\end{array}
\right.\\
p_{1,n} = J_{\gamma_n U_n A}(y_{1,n} + \gamma_n U_n z) + b_{1,n},\\
\text{for } i=1,\dots,m\\
\quad\left\lfloor
\begin{array}{l}
q_{2,i,n} = p_{2,i,n} + \gamma_n U_{i,n}\left(L_i(p_{1,n}) + c_{2,i,n}\right),\\
v_{i,n+1} = v_{i,n} - y_{2,i,n} + q_{2,i,n},
\end{array}
\right.\\
q_{1,n} = p_{1,n} - \gamma_n U_n\left(C(p_{1,n}) + \sum\limits_{i=1}^m L_i^*(p_{2,i,n}) + c_{1,n}\right),\\
x_{n+1} = x_n - y_{1,n} + q_{1,n}.
\end{array}
\right.
\tag{3.53}
$$

Then the following statements hold:

(i) $\sum\limits_{n\in\mathbb{N}}\|x_n - p_{1,n}\|^2 < +\infty$ *and* $\sum\limits_{n\in\mathbb{N}}\|v_{i,n} - p_{2,i,n}\|^2 < +\infty$, $\forall i\in\{1,\dots,m\}$;

(ii) There exists a solution $\bar{x}$ to (3.2) and a solution $(\bar{v}_1,\dots,\bar{v}_m)$ to (3.3) such that the following statements hold:

(1) $x_n \rightharpoonup \bar{x}$ *and* $p_{1,n} \rightharpoonup \bar{x}$*;*

(2) $v_{i,n} \rightharpoonup \bar{v}_i$ *and* $p_{2,i,n} \rightharpoonup \bar{v}_i$*,* $\forall i \in \{1, \dots, m\}$*.*

Proof. All sequences generated by algorithm (3.53) are well defined by Lemma 3.1.2. We define $\boldsymbol{\mathcal{H}} = \mathcal{H} \bigoplus \mathcal{G}_1 \bigoplus \cdots \bigoplus \mathcal{G}_m$, the Hilbert direct sum of the Hilbert space $\mathcal{H}$ and $(\mathcal{G}_i)_{1 \le i \le m}$, the scalar product and the associated norm of $\boldsymbol{\mathcal{H}}$ respectively defined by

$$\langle\langle\langle \cdot, \cdot \rangle\rangle\rangle : ((x, \mathbf{v}), (y, \mathbf{w})) \mapsto \langle x, y \rangle + \sum_{i=1}^{m} \langle v_i, w_i \rangle, \textit{and}$$

$$||| \cdot ||| : (x, \mathbf{v}) \mapsto \sqrt{\|x\|^2 + \sum_{i=1}^{m} \|v_i\|^2}, \tag{3.54}$$

where $\mathbf{v} = (v_1, \dots, v_m)$ and $\mathbf{w} = (w_1, \dots, w_m)$ are generic elements in $\mathcal{G}_1 \bigoplus \cdots \bigoplus \mathcal{G}_m$. Set

$$\begin{cases} \mathbf{A} : \boldsymbol{\mathcal{H}} \rightrightarrows \boldsymbol{\mathcal{H}} : (x, v_1, \dots, v_m) \mapsto (-z + Ax) \times (r_1 + B_1^{-1} v_1) \times \cdots \times (r_m + B_m^{-1} v_m), \\ \mathbf{B} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} : (x, v_1, \dots, v_m) \mapsto \left(Cx + \sum\limits_{i=1}^{m} L_i^* v_i, -L_1 x, \dots, -L_m x \right), \\ \mathbf{U}_n : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} : (x, v_1, \dots, v_m) \mapsto (U_n x, U_{1,n} v_1, \dots, U_{m,n} v_m), \ \forall n \in \mathbb{N}. \end{cases} \tag{3.55}$$

Since $\mathbf{A}$ is maximally monotone (see Proposition 20.22 and 20.23 in [8]), $\mathbf{B}$ is monotone β-Lipschitzian (see Equation (3.10) in [31]) with $\operatorname{dom} \mathbf{B} = \boldsymbol{\mathcal{H}}$, $\mathbf{A} + \mathbf{B}$ is maximally monotone (see Corolarry 24.24(i) in [31]). Now we set $\eta_n = \max\{\eta_{0,n}, \eta_{1,n}, \dots, \eta_{m,n}\}$, $\forall n \in \mathbb{N}$, then $(\eta_n)_{n \in \mathbb{N}} \in \ell_+^1(\mathbb{N})$. Moreover, we derive from our assumptions on the sequences $(U_n)_{n \in \mathbb{N}}$ and $(U_{1,n})_{n \in \mathbb{N}}, \dots, (U_{m,n})_{n \in \mathbb{N}}$ that

$$\mu = \sup_{n \in \mathbb{N}} \|\mathbf{U}_n\| < +\infty, \quad \text{and} \quad (1 + \eta_n) \mathbf{U}_{n+1} \succcurlyeq \mathbf{U}_n \in \mathcal{P}_\alpha(\boldsymbol{\mathcal{H}}), \ \forall n \in \mathbb{N}. \tag{3.56}$$

In addition, Proposition 23.17(ii) and 23.18 in [8] yield $\forall \gamma \in (0, +\infty), \forall n \in \mathbb{N}, \forall (x, v_1, \dots, v_m) \in \boldsymbol{\mathcal{H}}$,

$$J_{\gamma \mathbf{U}_n \mathbf{A}}(x, v_1, \dots, v_m) = \left(J_{\gamma U_n A}(x + \gamma U_n z), (J_{\gamma U_{i,n} B_i^{-1}}(v_i - \gamma U_{i,n} r_i))_{1 \le i \le m} \right). \tag{3.57}$$

It is shown in [31] (see Equation (3.12) and Equation (3.13) of [31]) that under the condition (3.1), $zer(\mathbf{A} + \mathbf{B}) \neq \emptyset$. Moreover, Equation (3.21) and Equation (3.22) in [31] yield

$$(\bar{x}, \bar{v}_1, \dots, \bar{v}_m) \in zer(\mathbf{A} + \mathbf{B}) \Rightarrow \bar{x} \text{ solves (3.2) and } (\bar{v}_1, \dots, \bar{v}_m) \text{ solves (3.3)}. \tag{3.58}$$

Let us next set $\forall n \in \mathbb{N}$,

$$\begin{cases} \mathbf{x}_n = (x_n, v_{1,n}, \dots, v_{m,n}), \\ \mathbf{y}_n = (y_{1,n}, y_{2,1,n}, \dots, y_{2,m,n}), \\ \mathbf{p}_n = (p_{1,n}, p_{2,1,n}, \dots, p_{2,m,n}), \\ \mathbf{q}_n = (q_{1,n}, q_{2,1,n}, \dots, q_{2,m,n}), \end{cases} \text{and} \begin{cases} \mathbf{a}_n = (a_{1,n}, a_{2,1,n}, \dots, a_{2,m,n}), \\ \mathbf{b}_n = (b_{1,n}, b_{2,1,n}, \dots, b_{2,m,n}), \\ \mathbf{c}_n = (c_{1,n}, c_{2,1,n}, \dots, c_{2,m,n}). \end{cases} \tag{3.59}$$

Then, our assumptions imply that

$$\sum_{n\in\mathbb{N}} |||\mathbf{a}_n||| < \infty, \sum_{n\in\mathbb{N}} |||\mathbf{b}_n||| < \infty, \text{and} \sum_{n\in\mathbb{N}} |||\mathbf{c}_n||| < \infty. \tag{3.60}$$

Furthermore, it follows from the definition of $\mathbf{B}$, (3.57), and (3.59) that (3.53) can be written in $\boldsymbol{\mathcal{H}}$ as

$$\left\lfloor \begin{array}{l} \mathbf{y_n} = \mathbf{x}_n - \gamma_n \mathbf{U}_n(\mathbf{B}(\mathbf{p}_{n-1}) + \mathbf{a}_n), \\ \mathbf{p}_n = J_{\gamma_n \mathbf{U}_n \mathbf{A}} \mathbf{y}_n + \mathbf{b}_n, \\ \mathbf{q}_n = \mathbf{p}_n - \gamma_n \mathbf{U}_n(\mathbf{B}(\mathbf{p}_n) + \mathbf{c}_n), \\ \mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{y}_n + \mathbf{q}_n, \end{array} \right. \tag{3.61}$$

which is (3.11). Moreover, every specific conditions in Theorem 3.2.1 are satisfied.

(i) By Theorem 3.2.1(i), $\sum_{n\in\mathbb{N}} |||\mathbf{x}_n - \mathbf{p}_n|||^2 < +\infty$ with the norm in (3.54). This implies that $\sum_{n\in\mathbb{N}} \|x_n - p_{1,n}\|^2 < +\infty$ and $\sum_{n\in\mathbb{N}} \|v_{i,n} - p_{2,i,n}\|^2 < +\infty$, $\forall i \in \{1, \ldots, m\}$;

(ii) There exists a solution $\bar{x}$ to (3.2) and a solution $(\bar{v}_1, \ldots, \bar{v}_m)$ to (3.3) such that the following statements hold:

(a) $x_n \rightharpoonup \bar{x}$ and $p_{1,n} \rightharpoonup \bar{x}$;

(b) $v_{i,n} \rightharpoonup \bar{v}_i$ and $p_{2,i,n} \rightharpoonup \bar{v}_i$, $\forall i \in \{1, \ldots, m\}$.

□

3.4 A Primal-Dual Splitting Algorithm for Convex Optimization Problems

Let us consider the case of convex optimization problems with structure where the optimality conditions lead to **Problem 3.1**. More precisely, we consider:

Problem 3.2: Let $\mathcal{H}$ be a real Hilbert space, let $z \in \mathcal{H}$, let m be a strictly positive integer, let $f \in \Gamma_0(\mathcal{H})$, and let $h : \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a v_0-Lipschitzian gradient for some $v_0 \in (0, +\infty)$. For every $i \in \{1, \ldots, m\}$, let $\mathcal{G}_i$ be a real Hilbert space, let $r_i \in \mathcal{G}_i$, and let $g_i \in \Gamma_0(\mathcal{G}_i)$. Suppose that $L_i : \mathcal{H} \to \mathcal{G}_i$ is a nonzero bounded linear operator. Consider the problem

$$\underset{x\in\mathcal{H}}{minimize}\ f(x) + \sum_{i=1}^{m} g_i(L_i x - r_i) + h(x) - \langle x, z \rangle, \tag{3.62}$$

and the Fenchel-Rockafellar dual problem [78]:

$$\underset{v_i \in \mathcal{G}_i (\forall i=1,\ldots,m)}{minimize} \ (f^* \Box h^*) \left(z - \sum_{i=1}^m L_i^* v_i \right) + \sum_{i=1}^m \left(g_i^*(v_i) + \langle v_i, r_i \rangle \right). \tag{3.63}$$

In this section, we further introduce the primal-dual splitting algorithm that can decouple the involved functions to solve **Problem 3.2**, as shown in the algorithm's structure below.

Theorem 3.4.1. *In **Problem 3.2**, we suppose that*

$$z \in \operatorname{Ran} \left(\partial f + \sum_{i=1}^m L_i^* (\partial g_i)(L_i \cdot - r_i) + \nabla h \right). \tag{3.64}$$

Let α be in $(0, +\infty)$, let $(\eta_{0,n})_{n\in\mathbb{N}}$ be a sequence in $\ell_+^1(\mathbb{N})$, let $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$, and for every $i \in \{1, \ldots, m\}$, let $(\eta_{i,n})_{n\in\mathbb{N}}$ be a sequence in $\ell_+^1(\mathbb{N})$, let $(U_{i,n})_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{G}_i)$ such that $\mu = sup_{n\in\mathbb{N}}\{\|U_n\|, \|U_{1,n}\|, \ldots, \|U_{m,n}\|\} < +\infty$ and

$$\begin{aligned} &(1 + \eta_{0,n})U_{n+1} \succcurlyeq U_n, \ \forall n \in \mathbb{N}, \\ \text{and } &(1 + \eta_{i,n})U_{i,n+1} \succcurlyeq U_{i,n}, \ \forall i \in \{1, \ldots, m\}. \end{aligned} \tag{3.65}$$

Let $(a_{1,n})_{n\in\mathbb{N}}, (b_{1,n})_{n\in\mathbb{N}}$ and $(c_{1,n})_{n\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{H}$, and for every $i \in \{1, \ldots, m\}$, let $(a_{1,i,n})_{n\in\mathbb{N}}, (b_{1,i,n})_{n\in\mathbb{N}}$ and $(c_{1,i,n})_{n\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{G}_i$. Furthermore, set

$$\beta = v_0 + \sqrt{\sum_{i=1}^m \|L_i\|^2}, \tag{3.66}$$

let $x_0 \in \mathcal{H}$, let $(v_{1,0}, \ldots, v_{m,0}) \in \mathcal{G}_1 \bigoplus \cdots \bigoplus \mathcal{G}_m$, let $(\gamma_n)_{n\in\mathbb{N}} \leq \lambda$ with $\lambda < \frac{1}{\sqrt{10}\mu\beta}$ and $\liminf_{n\to+\infty} \gamma_n > 0$. Set $\forall n \in \mathbb{N}$,

$$\left\lfloor \begin{array}{l} y_{1,n} = x_n - \gamma_n U_n \left(\nabla h(p_{1,n-1}) + \sum_{i=1}^m L_i^*(p_{2,i,n-1}) + a_{1,n} \right), \\ for\, i = 1, \ldots, m \\ \left\lfloor \begin{array}{l} y_{2,i,n} = v_{i,n} + \gamma_n U_{i,n} \left(L_i(p_{1,n-1}) + a_{2,i,n} \right), \\ p_{2,i,n} = prox_{\gamma_n g_i^*}^{U_{i,n}^{-1}} (y_{2,i,n} - \gamma_n U_{i,n} r_i) + b_{2,i,n}, \end{array} \right. \\ p_{1,n} = prox_{\gamma_n f}^{U_n^{-1}} (y_{1,n} + \gamma_n U_n z) + b_{1,n}, \\ for\, i = 1, \ldots, m \\ \left\lfloor \begin{array}{l} q_{2,i,n} = p_{2,i,n} + \gamma_n U_{i,n} \left(L_i(p_{1,n}) + c_{2,i,n} \right), \\ v_{i,n+1} = v_{i,n} - y_{2,i,n} + q_{2,i,n}, \end{array} \right. \\ q_{1,n} = p_{1,n} - \gamma_n U_n \left(\nabla h(p_{1,n}) + \sum_{i=1}^m L_i^*(p_{2,i,n}) + c_{1,n} \right), \\ x_{n+1} = x_n - y_{1,n} + q_{1,n}. \end{array} \right. \tag{3.67}$$

Then, the following statements hold:

(i) $\sum_{n\in\mathbb{N}} \|x_n - p_{1,n}\|^2 < +\infty$ *and* $(\forall i \in \{1,\dots,m\}) \sum_{n\in\mathbb{N}} \|v_{i,n} - p_{2,i,n}\|^2 < +\infty$*;*

(ii) *There exists a solution* $\bar{x}$ *to (3.62) and a solution* $(\bar{v}_1,\dots,\bar{v}_m)$ *to (3.63) such that the following statements hold:*

(a) $z - \sum_{j=1}^{m} L_j^*\bar{v}_j \in \partial f(\bar{x}) + \nabla h(\bar{x})$ *and* $(\forall i \in \{1,\dots,m\})\ L_i\bar{x} - r_i \in \partial g_i^*(\bar{v}_i)$*;*

(b) $x_n \rightharpoonup \bar{x}$ *and* $p_{1,n} \rightharpoonup \bar{x}$*;*

(c) $(\forall i \in \{1,\dots,m\})\ v_{i,n} \rightharpoonup \bar{v}_i$ *and* $p_{2,i,n} \rightharpoonup \bar{v}_i$*.*

Proof. Let us define

$$A = \partial f, \quad C = \nabla h \quad \text{and} \quad (\forall i = \{1,\dots,m\})\ B_i = \partial g_i. \tag{3.68}$$

It is clear that (3.64) yields (3.1). Using (2.6), (3.8) and (3.67) yield (3.53). Moreover, it follows from Theorem 2.6.2 that the operators A and $(B_i)_{1\le i\le m}$ are maximally monotone, and Proposition 17.7 in [8] that C is monotone which is a Lipschitzian operator by the hypothesis of **Problem 3.2**. Altogether, we can apply Theorem 3.3.1 to obtain the existence of a point $\bar{x} \in \mathcal{H}$ such that

$$z \in \partial f(\bar{x}) + \sum_{i=1}^{m} L_i^* \left(\partial g_i(L_i\bar{x} - r_i)\right) + \nabla h(\bar{x}), \tag{3.69}$$

and of an m-tuple $(\bar{v}_1,\dots,\bar{v}_m) \in \mathcal{G}_1 \bigoplus \cdots \bigoplus \mathcal{G}_m$ such that for some $x \in \mathcal{H}$,

$$\begin{cases} z - \sum_{j=1}^{m} L_j^*\bar{v}_j \in \partial f(x) + \nabla h(x) \\ (\forall i \in \{1,\dots,m\})\ \bar{v}_i \in (\partial g_i)(L_i x - r_i), \end{cases} \tag{3.70}$$

with the property that $\sum_{n\in\mathbb{N}} \|x_n - p_{1,n}\|^2 < +\infty$, $(\forall i \in \{1,\dots,m\}) \sum_{n\in\mathbb{N}} \|v_{i,n} - p_{2,i,n}\|^2 < +\infty$, $x_n \rightharpoonup \bar{x}$, $p_{1,n} \rightharpoonup \bar{x}$, $(\forall i \in \{1,\dots,m\})$ $v_{i,n} \rightharpoonup \bar{v}_i$ and $p_{2,i,n} \rightharpoonup \bar{v}_i$, are satisfied (see Theorem 3.3.1). Now we can follow the proof in [31] with our setting above and some tools in [8] to obtain that $\bar{x}$ solves (3.62) and $(\bar{v}_1,\dots,\bar{v}_m)$ solves (3.63).

□

Remark 3.4.2. In order to assure (3.64), we need some similar conditions as given in [31, Proposition 4.3]: suppose that (3.62) has at least one solution and set

$$\mathbb{S} = \{(L_i x - y_i)_{1\le i\le m} \mid x \in \operatorname{dom} f \text{ and } (\forall i \in \{1,\dots,m\}) y_i \in \operatorname{dom} g_i\}. \tag{3.71}$$

Then, the equation (3.64) is satisfied if one of the following statements holds:

(i) $(r_1,\dots,r_m) \in \operatorname{sri}\ \mathbb{S}$;

(ii) For every $i \in \{1,\dots,m\}$, g_i is real-valued;

(iii) $\mathcal{H}$ and $(\mathcal{G}_i)_{1\le i\le m}$ are finite-dimentional, and there exists $x \in \operatorname{ri}\ \operatorname{dom} f$ such that

$$L_i x - r_i \in \operatorname{ri}\ \operatorname{dom}\ g_i,\ \forall i \in \{1,\dots,m\}. \tag{3.72}$$

3.5 Numerical Experiment in Imaging

For this section, we intend to illustrate the numerical experiments in image deblurring which is correlated with our proposed primal-dual problem. Throughout this part, we implemented the numerical codes in MATLAB and performed all computations on a Window desktop with an Intel(R) Core(TM) i5-8250U processor at 1.6 GHz up to 1.8 GHz and 8.00 GB of RAM. Accordingly, the theoretical result obtained in the previous section indicates that the algorithm is applicable. It should be noted that we use the grayscale image which have been normalized in order to make their pixels range in the closed interval from 0 to 1 for the experiments.

Given matrix $A \in \mathbb{R}^{n\times n}$ describing a blur operator and vector $b \in \mathbb{R}^n$ representing the blurred and noisy image, the task is to approximate the unknown original image $\bar{x} \in \mathbb{R}^n$ fulfilling

$$A\bar{x} = b.$$

Then, we solve the following regularized convex minimization problem as shown here,

$$\inf_{x\in[0,1]^n} \left\{\|Ax-b\|_1 + \lambda(TV_{iso}(x) + \|x\|^2)\right\}, \tag{3.73}$$

where $\lambda > 0$ is a regularization parameter and $TV_{iso} : \mathbb{R}^n \to \mathbb{R}$ is the discrete isotropic total variation functional. In this context, $x \in \mathbb{R}^n$ represents the vectorized image $X \in \mathbb{R}^{M\times N}$, where $n = M \cdot N$, $x_{i,j}$ denotes the normalized value of the pixel located in the ith row, and the jth column, when $i = 1,\dots,M$, and $j = 1,\dots,N$. The *isotropic total variation* $TV_{iso} : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$TV_{iso}(x) = \sum_{i=1}^{M-1}\sum_{i=1}^{N-1}\sqrt{(x_{i+1,j}-x_{i,j})^2+(x_{i,j+1}-x_{i,j})^2} + \sum_{i=1}^{M-1}|x_{i+1,N}-x_{i,N}| + \sum_{j=1}^{N-1}|x_{M,j+1}-x_{M,j}|.$$

The optimization problem (3.73) can be written in the framework of Problem (3.62). We denote $\mathcal{Y} = \mathbb{R}^n \times \mathbb{R}^n$ and define the linear operator $\tilde{L} : \mathbb{R}^n \to \mathcal{Y}$, $x_{i,j} \mapsto (\tilde{L}_1 x_{i,j}, \tilde{L}_2 x_{i,j})$, where

$$\tilde{L}_1 x_{i,j} = \begin{cases} x_{i+1,j} - x_{i,j}, & \text{if } i < M \\ 0, & \text{if } i = M \end{cases} \textit{ and } \tilde{L}_2 x_{i,j} = \begin{cases} x_{i,j+1} - x_{i,j}, & \text{if } j < N \\ 0, & \text{if } j = N \end{cases}.$$

The operator $\tilde{L}$ represents a discretization of the gradient using reflexive (Neumann) boundary conditions and standard finite differences, and fulfills $\|\tilde{L}\|^2 \leq 8$. For its adjoint operator $\tilde{L}^* : \mathcal{Y} \to \mathbb{R}^n$, we refer to [27].

For $(y,z),(p,q) \in \mathcal{Y}$, we introduce the inner product following this equation,

$$\langle (y,z),(p,q)\rangle = \sum_{i=1}^{M}\sum_{j=1}^{N} y_{i,j}p_{i,j} + z_{i,j}q_{i,j},$$

and define $\|(y,z)\|_\times = \sum_{i=1}^M \sum_{j=1}^N \sqrt{y_{i,j}^2 + z_{i,j}^2}$. Once we can check that $\|\cdot\|_\times$ is a norm on $\mathcal{Y}$ and that for every $x \in \mathbb{R}^n$, it holds $TV_{iso}(x) = \|\tilde{L}x\|_\times$. The conjugate function $(\|\cdot\|_\times)^* : \mathcal{Y} \to \bar{\mathbb{R}}$ of $\|\cdot\|_\times$ is for every $(p,q) \in \mathcal{Y}$ given by

$$(\|\cdot\|)^*(p,q) = \begin{cases} 0, & \text{if } \|(p,q)\|_{\times *} \leq 1; \\ +\infty, & \text{otherwise;} \end{cases}$$

where

$$\|(p,q)\|_{\times *} = \sup_{\|(y,z)\|_\times \leq 1} \langle (p,q),(y,z) \rangle = \max_{\substack{1\leq i\leq M \\ 1\leq j\leq N}} \sqrt{p_{i,j}^2 + q_{i,j}^2}.$$

Therefore, the optimization problem (3.73) can be written in the form of

$$\inf_{x\in\mathcal{H}} \{f(x) + g_1(Ax) + g_2(\tilde{L}x) + h(x)\}, \tag{3.74}$$

where $f : \mathbb{R}^n \to \bar{\mathbb{R}}, f(x) = \iota_{[0,1]^n}(x), g_1(y) = \|y-b\|_1, g_2 : \mathcal{Y} \to \mathbb{R}, g_2(y,z) = \lambda\|(y,z)\|_\times$ and $h : \mathbb{R}^n \to \mathbb{R}, h(x) = \lambda\|x\|^2$ (notice that terms r_i and z are taken to be the zero vectors for $i = 1,2$). For every $p \in \mathbb{R}^n$, it holds $g_1^*(p) = \iota_{[-1,1]^n}(p) + p^T b$, while for every $(p,q) \in \mathcal{Y}$, we have $g_2^*(p,q) = \iota_S(p,q)$, with $S = \{(p,q) \in \mathcal{Y} : \|(p,q)\|_{\times *} \leq \lambda\}$. Moreover, h is differentiable function with $\nu_0 := 2\lambda$-Lipschitz continuous gradient. To solve the problem (3.74), we require the following formulae including

$$prox_{\gamma f}(x) = \arg\min_{y\in\mathbb{R}^n}\{\gamma f(y) + \frac{1}{2}\|y-x\|^2\} = \arg\min_{y\in[0,1]^n}\{\frac{1}{2}\|y-x\|^2\} = P_{[0,1]^n}(x) \forall x \in \mathbb{R}^n,$$

$$\begin{aligned} prox_{\gamma g_1^*}(p) &= \arg\min_{y\in\mathbb{R}^n}\{\gamma g_1^*(y) + \frac{1}{2}\|y-x\|^2\} = \arg\min_{y\in\mathbb{R}^n}\{\gamma(\iota_{[-1,1]^n}(y) + y^T b) + \frac{1}{2}\|y-x\|^2\} \\ &= \arg\min_{y\in[-1,1]^n}\{\gamma(y^T b) + \frac{1}{2}\|y-x\|^2\} = P_{[-1,1]^n}(p - \gamma b) \ \forall p \in \mathbb{R}^n, \end{aligned}$$

$$prox_{\gamma g_2^*}(p,q) = P_S(p,q) \ \forall (p,q) \in \mathcal{Y},$$

where $\gamma > 0$ and the projection operator $P_S : \mathcal{Y} \to S$ is defined as (see [15])

$$(p_{i,j}, q_{i,j}) \mapsto \lambda \frac{(p_{i,j}, q_{i,j})}{\max\{\lambda, \sqrt{p_{i,j}^2 + q_{i,j}^2}\}}, 1 \leq i \leq M, 1 \leq j \leq N.$$

According to the definition of the proximity operator of f relative to the variable matrices (3.7) and Lemma 3.1.2 (*iii*), for $\gamma_n > 0$, we obtain that (see also (3.8))

$$prox_{\gamma_n f}^{U_n^{-1}}(x) = J_{(U_n^{-1})^{-1}\partial\gamma_n f}(x) = J_{U_n\partial\gamma_n f}(x) = (U_n^{-1} + \partial\gamma_n f)^{-1} \circ U_n^{-1},$$

and similarly for $i = 1, 2$;

$$prox_{\gamma_n g_i^*}^{U_n^{-1}}(x) = J_{(U_n^{-1})^{-1}\partial\gamma_n g_i^*}(x) = J_{U_n\partial\gamma_n g_i^*}(x) = (U_n^{-1} + \partial\gamma_n g_i^*)^{-1} \circ U_n^{-1}.$$

In Theorem 3.4.1, we choose $(\tau_n)_{n\in\mathbb{N}}$ and $(\sigma_i)_{1\leq i\leq m}$ in $(0, +\infty)$ such that $U_n = \tau_n Id$ and $(\forall i \in \{1, 2\})$ $U_{i,n} = \sigma_{i,n} Id$. Then, (3.67) reduce to the fixed metric methods (see also [91]). Then the proximity operators turn into as follows

$$\begin{aligned}
prox_{\gamma_n f}^{(\tau_n \operatorname{Id})^{-1}}(x) &= \big[((\tau_n \operatorname{Id})^{-1} + \partial\gamma_n f)^{-1} \circ (\tau_n \operatorname{Id})^{-1}\big](x) \\
&= \big[(\frac{1}{\tau_n} \operatorname{Id} + \partial\gamma_n f)^{-1} \circ (\frac{1}{\tau_n} \operatorname{Id})\big](x) \\
&= \big[(\frac{1}{\tau_n})^{-1} J_{\tau_n\partial\gamma_n f} \circ (\frac{1}{\tau_n} \operatorname{Id})\big](x) \\
&= \tau_n prox_{\tau_n\gamma_n f}\big(\frac{1}{\tau_n} x\big);\ (\forall x \in \mathbb{R}^n).
\end{aligned}$$

Similarly, we obtain that

$$\begin{aligned}
prox_{\gamma_n g_1^*}^{(\sigma_{1,n} \operatorname{Id})^{-1}}(p) &= \big[((\sigma_{1,n} \operatorname{Id})^{-1} + \partial\gamma_n g_1^*)^{-1} \circ (\sigma_{1,n} \operatorname{Id})^{-1}\big](x) \\
&= \big[(\frac{1}{\sigma_{1,n}} \operatorname{Id} + \partial\gamma_n g_1^*)^{-1} \circ (\frac{1}{\sigma_{1,n}} \operatorname{Id})\big](x) \\
&= \big[(\frac{1}{\sigma_{1,n}})^{-1} J_{\sigma_{1,n}\partial\gamma_n g_1^*} \circ (\frac{1}{\sigma_{1,n}} \operatorname{Id})\big](x) \\
&= \sigma_{1,n} prox_{\sigma_{1,n}\gamma_n g_1^*}\big(\frac{1}{\sigma_{1,n}} x\big);\ (\forall p \in \mathbb{R}^n),
\end{aligned}$$

and

$$\begin{aligned}
prox_{\gamma_n g_2^*}^{(\sigma_{2,n} \operatorname{Id})^{-1}}(p, q) &= \sigma_{2,n} prox_{\sigma_{2,n}\gamma_n g_2^*}\big(\frac{1}{\sigma_{2,n}}(p, q)\big) \\
&= \sigma_{2,n}\lambda \frac{(\frac{p_{\bar{i},\bar{j}}}{\sigma_{2,n}}, \frac{q_{\bar{i},\bar{j}}}{\sigma_{2,n}})}{\max\big\{\lambda, \sqrt{(\frac{p_{\bar{i},\bar{j}}}{\sigma_{2,n}})^2 + (\frac{q_{\bar{i},\bar{j}}}{\sigma_{2,n}})^2}\big\}} \\
&= \frac{(p_{\bar{i},\bar{j}}, q_{\bar{i},\bar{j}})}{\max\big\{1, \frac{1}{\lambda}\sqrt{(\frac{p_{\bar{i},\bar{j}}}{\sigma_{2,n}})^2 + (\frac{q_{\bar{i},\bar{j}}}{\sigma_{2,n}})^2}\big\}} 1 \leq \bar{i} \leq M, 1 \leq \bar{j} \leq N.
\end{aligned}$$

To measure the quality of the restored image, we use the tool known as *signal-to-noise ratio* (ISNR), which is given by (see [28])

$$\mathrm{ISNR}_n = 10 \log_{10}\left(\frac{\|x - b\|^2}{\|x - x_n\|^2}\right),$$

where x, b, and x_n are vectors referred to the original picture, the observed noisy image, and the reconstructed image at iteration $n \in \mathbb{N}$, respectively.

For the experiment, we consider the cameraman image with 256×256 pixels and construct the blurred image using a Gaussian blur operator of size 9×9 and a standard deviation 4. In order to obtain the blurred and noisy image, we add a zero-mean white Gaussian noise with a standard deviation of 10^{-3}. Figure 3.1 shows the original cameraman image and the blurred and noisy one. It also shows the image reconstructed by the algorithm after 1000 iterations, when the regularization parameter $\lambda = 0.003$, all error terms are zero, the variable metrics $U_n = \tau_n Id$, $U_{i,n} = \sigma_{i,n} Id$ and by choosing as parameters $\tau_n = 1$, $\sigma_{1,n} = 0.1$, $\sigma_{2,n} = 1$, a starting point $p_{1,-1} = p_{2,2,-1} = \bar{1} \times 0.4660$,

$$p_{2,1,-1} = (\bar{1}, \bar{1}) \times 0.4660 \text{ where } \bar{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{256 \times 256}, \ v_0 = 2\lambda,\ v_{1,0} = v_{2,0} = \bar{0}$$

(where $\bar{0}$ is a 256×256 zero matrix), and $\gamma = \frac{1}{\sqrt{10}\mu(\beta+1)}$ where $\mu = 1$, $\beta = 2\lambda + \sqrt{9}$ for $i \in \{1, 2\}$.

(a) Original image (b) blurred and noisy image (c) Reconstructed image

Figure 3.1: The figure depicts the original image (a), which turns out to be a blurred and noisy image (b) by a Gaussian blur operator. Furthermore, it shows the result after 1000 iterations of our algorithm for deblurring as the reconstructed image (c).

In the error-free case, the variable matrix is replaced by the identity matrix, and we consider the cameraman image with the same method of blurring with stopping criteria that is less than 10^{-2}. For $n \geq 0$, $\|x_n - x_{n+1}\|$, $|\mathrm{fval}_{x_n} - \mathrm{fval}_{x^{**}_{10000}}|$ and $\|x_n - x^{**}_{10000}\|$ are the examined criteria, where fval_{x_n} is the objective value at the point x_n, and x^{**}_{10000} is the solution point of the Tseng-EP algorithm after 10000 iterations. Table 3.1 shows the performance between the classical Tseng's algorithm and the Tseng's algorithm with extrapolation from the past (Tseng-EP) when taking as regularization parameter $\lambda = 0.003$, a starting point $p_{1,-1} = p_{2,2,-1} = \bar{1} \times 0.4660$, $p_{2,1,-1} = (\bar{1}, \bar{1}) \times 0.4660$, $v_0 = 2\lambda$, $v_{1,0} = v_{2,0} = \bar{0}$ and $\gamma_n = 1/(2\beta + 0.1)$ where $\beta = 2\lambda + \sqrt{9}$. We have noticed that our proposed Tseng-EP algorithm requires the CPU-Time less than the classical Tseng's algorithm.

To generalize the algorithm, we firstly choose $U_n = \tau_n Id$ and $(\forall i \in \{1, \ldots, m\})$ $U_{i,n} = \sigma_{i,n} Id$ when $\tau_n = 1$ and $\sigma_{i,n}$ is different values with the same setting of regularization parameter and initial points includes $p_{1,-1}$, $p_{2,2,-1}$, $p_{2,1,-1}$, v_0, $v_{1,0}$, and $v_{2,0}$. Inputting

Criteria	Algorithm	No.Iteration	ISNR	CPU-Time*
$\|x_n - x_{n+1}\| < 10^{-2}$	Tseng	728	7.822241	6.55718
	Tseng-EP	728	7.822187	4.53896
$\|\text{fval}_{x_n} - \text{fval}_{x^{**}_{10000}}\| < 10^{-2}$	Tseng	4802	7.423491	41.7736
	Tseng-EP	4802	7.423493	28.72458
$\|x_n - x^{**}_{10000}\| < 10^{-2}$	Tseng	4729	7.424201	44.25134
	Tseng-EP	4800	7.424102	29.23266

Table 3.1: The results of experiment for three different stopping criteria which are less than 10^{-2}. By using $\gamma_n = 1/(2\beta + 0.1)$ where $\beta = 2\lambda + \sqrt{9}$ and $\lambda = 0.003$ in error free case of the algorithm (the variable metrics $U_n = Id$ and $U_{i,n} = Id$).

$\gamma_n = 1/(2\beta + 0.1)$ where $\beta = 2\lambda + \sqrt{9}$, we obtain the results as shown in Table 3.2 ($\forall n \geq 0$ and $i \in \{1, 2\}$). The proposed algorithm provides the error free case when $|\text{fval}_{x_n} - \text{fval}_{x^{**}_{10000}}| < 10^{-2}$ is a stopping criteria. Moreover, we notice that the choice of $\sigma_{i,n}$ for $i \in \{1, 2\}$ can be any constant which is very close to 1. However, if we replace them by 1.004 and 0.996, then we have seen that the algorithm does not converge easily even after more than 8000 iterations. Furthermore, Table 3.2 gives us the idea to change $\sigma_{i,n}$ to the convergent sequences that converges to 1 instead of the constant values. We consider such sequences as follows: $\frac{1}{k}$, $\frac{1}{k^2}$, $\frac{1}{k^5}$, $\frac{1}{k^k}$ and $\frac{k}{k+1}$; subsequently, we demonstrate some results in Table 3.3.

Criteria	No.Iteration	τ_n	$\sigma_{i,n}$	$\text{fval}_{x_{iteration}}$	ISNR	CPU-Time(s)
$\|\text{fval}_{x_n} - \text{fval}_{x^{**}_{10000}}\| < 10^{-2}$	7232	1	1.003	97.7279	6.958979	46.0178
	5589	1	1.002	97.727654	7.213882	35.0158
	5000	1	1.001	97.727931	7.364356	31.1526
	4802	1	1	97.727766	7.423493	27.87
	4992	1	0.9999	97.72766	7.367581	34.8569
	5562	1	0.998	97.72766	7.222046	45.7339
	7283	1	0.997	97.727572	6.956446	58.8615

Table 3.2: The result of experiment when $\tau_n = 1$ and $\sigma_{i,n}$ ($i \in \{1, 2\}$) are some constants with stopping criteria $|\text{fval}_{x_n} - \text{fval}_{x^{**}_{10000}}| < 10^{-2}$ of the algorithm using $\gamma_n = 1/(2\beta + 0.1)$ where $\beta = 2\lambda + \sqrt{9}$, and $\lambda = 0.003$.

Since $\sigma_{i,n}$ ($\forall n \geq 0, i \in \{1, 2\}$) can be independent of choice, then we start the experiment with $\tau_n = 1$, and $\sigma_{1,n}$, $\sigma_{2,n}$ are equal to 1, $\frac{k}{k+1}$, $(1 + \frac{1}{k})^k$, 1-($\frac{1}{k}$), 1-($\frac{1}{k^2}$), 1-($\frac{1}{k^5}$). The experiment results are shown as in Table 3.4. Even though some results give us slightly better ISNR, they consume the CPU-Time more than the case of $\sigma_{i,n} = 1$, $\forall i \in \{1, 2\}$.

In the error case, we solve this problem by using the same setting of initial points, regularization parameter λ, $U_n = \tau_n Id$ and ($\forall n \geq 0, i \in \{1, 2\}$) $U_{i,n} = \sigma_{i,n} Id$. The error terms are absolutely summable sequences which include $a_{1,n}$, $b_{1,n}$, $c_{1,n}$, $a_{2,i,n}$, $b_{2,i,n}$, $c_{2,i,n}$. Then, we need to select γ_n which satisfies condition in Theorem 3.4.1 $\big((\gamma_n)_{n\in\mathbb{N}} \leq \lambda$ with $\lambda < \frac{1}{\sqrt{10\mu\beta}}$ and $\liminf_{n\to+\infty} \gamma_n > 0\big)$. Thus, we choose $\gamma_n = \frac{1}{\sqrt{10\mu(\beta+1)}}$ where $\beta = 2\lambda + \sqrt{9}$.

Iteration	τ_n	$\sigma_{i,n}$	$\text{fval}_{x_{iteration}}$	ISNR	CPU-Time(s)
4802	1	1	97.727766	7.423493	27.87
4804	1	1-($\frac{1}{k^2}$)	97.727863	7.423559	31.015
4803	1	1-($\frac{1}{k^5}$)	97.727786	7.423521	30.6754
4802	$\frac{k}{k+1}$	1	97.727619	7.423555	31.1954
4803	$\frac{k}{k+1}$	1-($\frac{1}{k^5}$)	97.72767	7.423575	29.1343
4802	1-($\frac{1}{k^2}$)	1	97.727403	7.42298	30.0258
4802	1-($\frac{1}{k^5}$)	1	97.727357	7.422935	30.1102
4804	1-($\frac{1}{k^5}$)	1-($\frac{1}{k^2}$)	97.727494	7.42299	30.055
4802	1-($\frac{1}{k}$)	1	97.727629	7.423016	30.9404
4803	1-($\frac{1}{k}$)	1-($\frac{1}{k^5}$)	97.727648	7.422963	29.8768
4803	1-($\frac{1}{k}$)	1-($\frac{1}{k^k}$)	97.72797	7.423079	30.7497

Table 3.3: The results of experiment when τ_n, $\sigma_{i,n}$ (for $i \in \{1,2\}$) are shuffled between a constant 1 and the sequences which converges to 1 with stopping criteria $|\text{fval}_{x_n} - \text{fval}_{x^{**}_{10000}}| < 10^{-2}$ of the algorithm using $\gamma_n = 1/(2\beta + 0.1)$ where $\beta = 2\lambda + \sqrt{9}$ and $\lambda = 0.003$ for all $n \geq 0$.

Iteration	τ_n	$\sigma_{1,n}$	$\sigma_{2,n}$	$\text{fval}_{x_{iteration}}$	ISNR	CPU-Time(s)
5431	1	1	$\frac{k}{k+1}$	97.727452	7.256485	33.3473
4804	1	1	1-($\frac{1}{k^2}$)	97.727866	7.423558	29.1967
5432	1	1	1-($\frac{1}{k}$)	97.727455	7.256488	33.4189
4803	1	1	1-($\frac{1}{k^5}$)	97.727787	7.42352	28.7536
4802	1	$\frac{k}{k+1}$	1	97.72749	7.423239	30.2019
4802	1	1-($\frac{1}{k^2}$)	1	97.727763	7.423495	29.4369
4802	1	1-($\frac{1}{k}$)	1	97.727488	7.42324	28.8945
4802	1	1-($\frac{1}{k^5}$)	1	97.727765	7.423495	29.1847
4802	1	$\frac{k}{k+1}$	1	97.72749	7.423239	30.5509

Table 3.4: the result of experiment when we fixed $\tau_n = 1$ and shuffled $\sigma_{1,n}$ and $\sigma_{2,n}$ between 1, $\frac{k}{k+1}$, $1-(\frac{1}{k})$, $1-(\frac{1}{k^2})$, $1-(\frac{1}{k^5})$ and $(1+\frac{1}{k})^k$ with stopping criteria $|\text{fval}_{x_n} - \text{fval}_{x^{**}_{10000}}| < 10^{-2}$ and $\gamma_n = 1/(2\beta + 0.1)$ where $\beta = 2\lambda + \sqrt{9}$, and $\lambda = 0.003$ for all $n \geq 0$, $i \in \{1,2\}$.

Table 3.5 shows the result when all of error terms equal to the following sequences: $1/k^2$, $1/k^5$, $1/k^k$, $(1/2)^k$ by fixing $\tau_n = \sigma_{i,n} = 1$ for all $n \geq 0$, $i \in \{1,2\}$. We observe that their performances are not much significantly different and they also use double time of the error-free case.

If we fix the iteration number of 5000, $\tau_n = 1$, $\sigma_{i,n} = 1$, then we differ the error terms as $1/k^2$, $1/k^5$, $1/k^k$, and $(1/2)^k$. We can see in in Table 3.6 that the modification of error in our experiment hardly affect on the result. However, when we consider the ISNR values, they deliver more than 8 with the highest being 8.344218. In addition, the function values are slightly higher than previous results for $\gamma_n = 1/(\sqrt{10}\mu(\beta + 1)$.

From the last trial, we plot the graph for 10000 iterations using fixed values as follows: $\tau_n = 1$, $\sigma_{i,n} = 1$ ($i \in \{1,2\}$), the error term to be $1/k^2$, $\gamma_n = 1/(\sqrt{10}\mu(\beta + 1)$ where

Iteration	τ_n	$\sigma_{i,n}$	Error	$\text{fval}_{x_{iteration}}$	ISNR	CPU-Time(s)
9976	1	1	$1/k^2$	97.727773	7.415526	65.3116
9972	1	1	$1/k^5$	97.727751	7.415609	64.946
9973	1	1	$1/k^k$	97.727652	7.415548	65.146
9971	1	1	$(1/2)^k$	97.727843	7.415333	65.8735

Table 3.5: the result of experiment when we fixed $\tau_n = 1$, $\sigma_{i,n} = 1$ (for $n \geq 0$, $i \in \{1, 2\}$) and various errors with stopping criteria $|\text{fval}_{x_n} - \text{fval}_{x^{**}_{10000}}| < 10^{-2}$ and $\gamma_n = 1/(\sqrt{10}\mu(\beta + 1)$ where $\beta = 2\lambda + \sqrt{9}$, and $\lambda = 0.003$.

Iteration	τ_n	$\sigma_{i,n}$	Error	$\text{fval}_{x_{iteration}}$	ISNR	CPU-Time(s)
5000	1	1	$\frac{1}{k^2}$	99.465867	8.344218	32.5193
5000	1	1	$\frac{1}{k^5}$	99.459621	8.342828	32.4141
5000	1	1	$\frac{1}{k^k}$	99.459901	8.34304	32.5323
5000	1	1	$(\frac{1}{2})^k$	99.460987	8.342093	32.5181

Table 3.6: the result of experiment after 5,000 iterations by fixing $\tau_n = 1$, $\sigma_{i,n} = 1$ ($i \in \{1, 2\}$) and varying errors as $1/k^2$, $1/k^5$, $1/k^k$, $(1/2)^k$ of the algorithm using $\gamma_n = 1/(\sqrt{10}\mu(\beta + 1)$ where $\beta = 2\lambda + \sqrt{9}$ and $\lambda = 0.003$ for all $n \geq 0$.

$\beta = 2\lambda + \sqrt{9}$ and $\lambda = 0.003$ for all $n \geq 0$ shown in Figure 3.2. We detect the peak point by using *findpeaks* in MATLAB to find the local maximum point. Lastly, we can find that the maximum point is presented at 3736 iterations with the ISNR = 8.467. However, we cannot confirm that this is the highest value of ISNR because the highest ISNR values probably change depending on the control parameters such as error terms, τ_n , $\sigma_{i,n}$ ($\forall n \geq 0, i \in \{1, 2\}$), and the stepsize γ_n.

Remark 3.5.1. In all our examples, we investigate only the cases $U_n = \text{Id}$ or $U_n = \tau_n \text{Id}$ and $U_{i,n} = \sigma_{i,n} \text{Id}$ for $n \geq 0$, $i \in \{1, 2\}$. Typically, the non-trivial case of U_n might be examined. The non-trivial choices for the operators U_n can be analyzed within the framework of optimization methods that combine proximal methods with quasi-Newton techniques. This is an interesting research approach, and because both the theory and the investigations are involved (as shown in [10, 93]), This can be regarded as a future research topic.

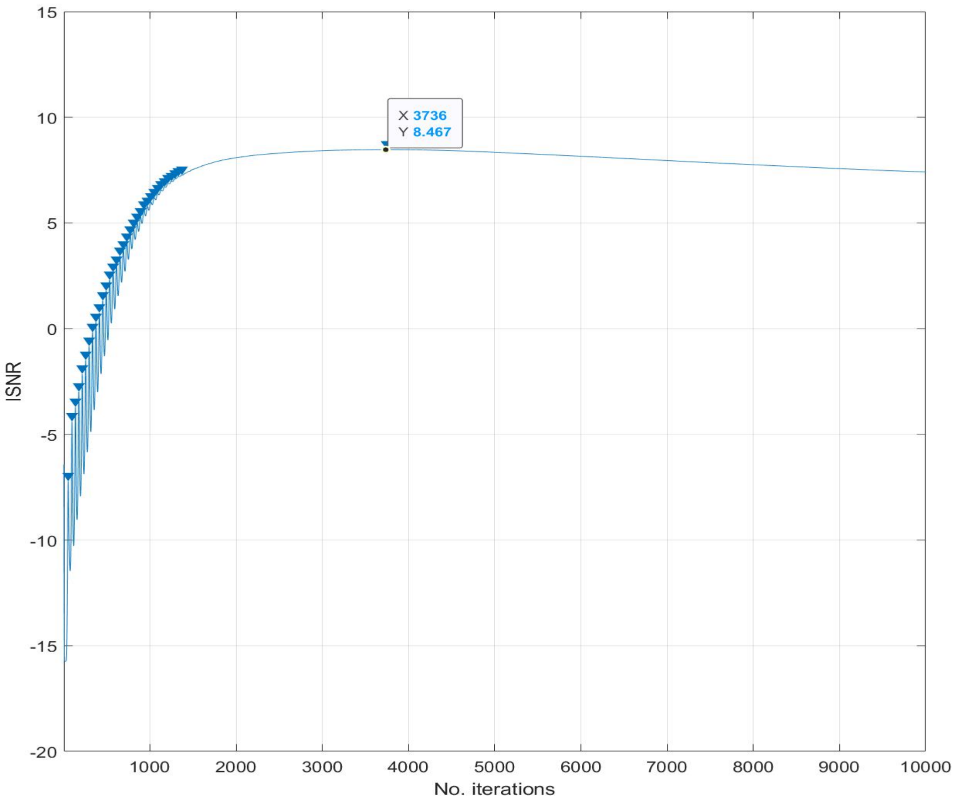


Figure 3.2: The graph illustrates the ISNR value after 10000 iterations when we fixed $\tau_n = 1$, $\sigma_{i,n} = 1$, and error terms is $1/k^2$. By setting $\gamma_n = 1/(\sqrt{10}\mu(\beta+1)$ where $\beta = 2\lambda + \sqrt{9}$ and $\lambda = 0.003$ for all $n \geq 0$, $i \in \{1, 2\}$.

4 The Forward-Backward-Forward Algorithm with Extrapolation from the Past and Penalty Scheme for Solving Monotone Inclusion Problems and Applications

In the last decade, penalty schemes have become a popular approach for studying constrained or hierarchical optimization problems in Hilbert spaces, particularly for solving complex problems (see [5, 6, 7, 17, 18, 20, 66]). In this chapter, we intend to study the general monotone inclusion problem in the term of the sum of three operators as follows:

$$0 \in A(x) + D(x) + N_C(x), \tag{4.1}$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator, $D : \mathcal{H} \to \mathcal{H}$ a (single-valued) monotone and η^{-1}-Lipschitz continuous operator with $\eta > 0$, and $N_C : \mathcal{H} \rightrightarrows \mathcal{H}$ is the normal cone of the closed convex set $C \subseteq \mathcal{H}$ which is the nonempty set of zeros of another operator $B : \mathcal{H} \to \mathcal{H}$, which is a (single-valued) μ^{-1}-Lipschitz continuous operator with $\mu > 0$. If the involved operators are subdifferentials, the situation can be framed as a specialized convex bilevel optimization problem. In this context, the feasible set of the upper-level problem corresponds to the set of solutions for the lower-level problem, and both sets are characterized by convexity (see (4.3) and (4.4) below).

If $D = 0$, the problem (4.1) is the generalized version of the monotone inclusion problem

$$0 \in A(x) + N_C(x), \tag{4.2}$$

where $C = \arg\min \Psi \neq \emptyset$, and $\Psi : \mathcal{H} \to \mathbb{R}$ is a convex differentiable function with a Lipschitz gradient satisfying $\min \Psi = 0$, which introduces a penalization function with respect to the constraint $x \in C$. Implicit and explicit discretized methods to solve the problem in (4.2) were proposed in Attouch et al.'s work [5,6]. Several subsequent penalty type numerical algorithms in the literature are inspired by the continuous nonautonomous differential inclusion investigated in Attouch and Czarnecki's paper [4].

In the case where A and D are the convex subdifferentials of the proper, convex, and lower semicontinuous functions $\Phi : \mathcal{H} \to \bar{\mathbb{R}}$ and $\phi : \mathcal{H} \to \bar{\mathbb{R}}$, respectively, then any solution of (4.1) solves the convex minimization problem

$$\min_{x \in \mathcal{H}} \{\Phi(x) + \phi(x) : x \in \arg\min \Psi\}, \tag{4.3}$$

and a solution of (4.2) solves the problem

$$\min_{x\in\mathcal{H}}\{\Phi(x) : x \in \arg\min \Psi\}, \tag{4.4}$$

respectively. Consequently, a solution to these convex minimizations can be approximated by using the same iterative scheme as in [5, 6, 18]. Furthermore, there are other methods for solving such optimization problems which are studied by several authors in the literature (see, for instance, [7, 20, 66, 70]).

The iterative algorithms for solving both monotone inclusion problems (4.1) and (4.2) in the penalty scheme have been proposed by several researchers (see [5, 6, 7, 17, 18]). They could provide the weak ergodic convergence of the iterates for their proposed algorithms. In particular, they must suppose the strong monotonicity of A to prove the strong convergence. For the (ergodic) convergent results, one has to assume the following hypothesis:

- For solving the problem (4.2) in case $C = \arg\min \Psi \neq \emptyset$ with the algorithm proposed by Attouch et al. [5, 6], the Fenchel conjugate function of Ψ (namely, Ψ^*) needs to fulfill some key hypotheses in this context, as shown below:

$$(H)\begin{cases} (i) \quad A + N_C \text{ is maximally monotone and } zer(A + N_C) \neq \emptyset; \\ (ii) \ \text{ For every } p \in \operatorname{Ran}(N_C), \sum_{n\in\mathbb{N}} \lambda_n\beta_n \left[\Psi^*\left(\frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] < +\infty; \\ (iii)\ (\lambda_n)_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1, \end{cases}$$

  where $(\lambda_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ are sequences of positive real numbers. Note that the hypothesis (ii) of (H) is satisfied when $\sum_{n\in\mathbb{N}} \frac{\lambda_n}{\beta_n} < +\infty$ and Ψ is bounded below by a multiple of the square of distances to C, as described in [6]. One such case is when $C = zer(L)$, where $L : \mathcal{H} \to \mathcal{H}$ is a linear continuous operator with closed range and $\Psi : \mathcal{H} \to \mathbb{R}$ is defined as $\Psi(x) = \frac{1}{2}\|Lx\|^2$, see further in [5, Section 4.1].

- For solving the problem (4.1) with a forward-backward type or Tseng's type (or forward-backward-forward) algorithm proposed by Boț et al. in [7, 17, 18], the required hypotheses involve the Fitzpatrick function associated to the maximally monotone operator B and reads as:

$$(H_{fitz})\begin{cases} (i) \quad A + N_C \text{ is maximally monotone and } zer(A + D + N_C) \neq \emptyset; \\ (ii) \ \text{ For every } p\in\operatorname{Ran}(N_C), \displaystyle\sum_{n\in\mathbb{N}} \lambda_n\beta_n \left[\sup_{u\in C} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] < +\infty; \\ (iii)\ (\lambda_n)_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1. \end{cases}$$

  where $(\lambda_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ are sequences of positive real numbers. Note that when $B = \partial\Psi$ and $\Psi(x) = 0$ for all $x \in C$, then by (4.5) one can see that condition (ii) of (H) implies the second condition of (H_{fitz}), see [17, 18].

For solving the monotone inclusion problem (4.1), if we use the approach involved with the forward-backward algorithm, then some of the operators in the iterates are cocoercive. For instance, (see [18])

$$\text{Choose } x_1 \in \mathcal{H}. \text{ For } n \in \mathbb{N}, \ \text{ set } x_{n+1} = J_{\lambda_n A}(x_n - \lambda_n D(x_n) - \lambda_n\beta_n B(x_n)),$$

with the sequences of positive real number, $(\lambda_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$, and the operators D and B are cocoercive. We know from the definition of the cocoercivity (see Section 2.5) that every cocoercive operator is a monotone and Lipschitz operator, but the converse argument is not the case, see Remark 2.6.4. One of the research directions is to develop an algorithm for inclusions involving monotone and Lipschitz operators, which has known as Tseng's type algorithm (or forward-backward-forward algorithm), see further [90]. In the penalty scheme, Boț and Csetnek [18] relaxed the cocoercivity of B and D to monotonicity and Lipschitzian and the convergence properties of Boț and Csetnek's algorithm are studied under (H_{fitz}) and the condition $\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}\right)<1$.

Motivated by the extrapolation technique used for Tseng's algorithm (see introduction Chapter 1), we investigate Tseng's forward-backward-forward algorithm with extrapolation from the past involving a penalty scheme under certain hypotheses for solving the generalized inclusion problem in terms of the sum of three operators (4.1). We propose the algorithm and prove its weak ergodic convergence. Moreover, strong convergence is also demonstrated when the operator A is a strongly monotone operator in Section 4.2. Furthermore, we can extend our results in Section 4.2 to solve minimax problems as elaborated in Section 4.3. In Section 4.4, our proposed algorithm can be extended to solving more intricate problems involving finite sums of monotone operators and composition with linear operators, by using the product space approach, and the convergence results for our modified iterative method are also provided.

To broaden the applicability of our scheme, we also provide the iterative scheme and its convergence result for the convex minimization problem expressed in Section 4.5. Finally, we demonstrate a numerical experiment in TV-based image inpainting to ensure theoretical convergence results in Section 4.6. This chapter is based on [88], published in Numerical Algorithms.

4.1 Preliminaries

The following notation was introduced by Fitzpatrick in [42] and has proved to be a useful tool for investigating the maximality of monotone operators using convex analysis techniques (see [8, 9]). For a monotone operator $A:\mathcal{H}\rightrightarrows\mathcal{H}$, we define its associated *Fitzpatrick function* $\varphi_A:\mathcal{H}\times\mathcal{H}\to\overline{\mathbb{R}}$ as follows:

$$\varphi_A(x,u)=\sup_{(y,v)\in GrA}\{\langle x,v\rangle+\langle y,u\rangle-\langle y,v\rangle\}.$$

This function is convex and lower semicontinuous, and it will be an important tool for investigating convergence in this scheme.

If A is a maximally monotone operator, then its Fitzpatrick function is proper and satisfies $\varphi_A(x,u)\geq\langle x,u\rangle$ for all $(x,u)\in\mathcal{H}\times\mathcal{H}$, with equality if and only if $(x,u)\in GrA$. Furthermore, for $f\in\Gamma_0(\mathcal{H})$ we have the inequality (see [9])

$$\varphi_{\partial f}(x,u)\leq f(x)+f^*(u)\quad\forall(x,u)\in\mathcal{H}\times\mathcal{H},\tag{4.5}$$

where f^* is the convex conjugate of f.

Before we enter on a detailed analysis of the convergence result, we introduce a definition of a particular sequence and an Opial's lemma for the weighted averages. Let $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ be a sequence in $\mathcal{H}$ and $(\lambda_k)_{k\in\mathbb{N}\cup\{0\}}$ be a sequence of positive numbers such that $\sum_{k\in\mathbb{N}\cup\{0\}}\lambda_k = +\infty$. Let $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ be the sequence of weighted averages defined as shown in [6, 17, 18]:

$$z_n = \frac{1}{\tau_n}\sum_{k=0}^{n}\lambda_k x_k, \quad \text{where } \tau_n = \sum_{k=0}^{n}\lambda_k \ \forall n \in \mathbb{N}. \tag{4.6}$$

Lemma 4.1.1 ([67, Opial's Lemma]). *Let S be a nonempty subset of $\mathcal{H}$, and $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ a sequence of elements of $\mathcal{H}$. Assume that*

(i) for every $z \in S$, $\lim_{n\to+\infty}\|x_n - z\|$ exists;

(ii) every weak sequential limit point of $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ belongs to S as $n \to +\infty$.

Then $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ converges weakly as $n \to +\infty$ to a point in S.

4.2 Forward-backward-forward with extrapolation and penalty schemes

In this section, we will begin by presenting a problem introduced in [17, 18], which can be formulated as follows:

Problem 4.1: Let $\mathcal{H}$ be a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $D : \mathcal{H} \to \mathcal{H}$ be a monotone and η^{-1}-Lipschitz continuous operator with $\eta > 0$, $B : \mathcal{H} \to \mathcal{H}$ be a monotone and μ^{-1}-Lipschitz continuous operator with $\mu > 0$, and suppose that $C = zer(B) \neq \emptyset$. The monotone inclusion problem that we want to solve is

$$0 \in A(x) + D(x) + N_C(x). \tag{4.7}$$

Some methods to solve **Problem 4.1** have been already studied in [17, 18] by using Tseng's algorithm and penalty scheme. The iterative scheme presented below for solving **Problem 4.1** draws inspiration from Tseng's forward-backward-forward algorithm with extrapolation from the past as explained in the introduction (see [14, 62, 89]).

Algorithm 4.1:

$$\begin{aligned}
\text{Initialization}: &\ \text{Choose } y_{-1}, x_0 \in \mathcal{H} \text{ with } y_{-1} = x_0.\\
\text{For } n \in \mathbb{N}\cup\{0\} \text{ set}: &\ y_n = J_{\lambda_n A}\left[x_n - \lambda_n D(y_{n-1}) - \lambda_n\beta_n B(y_{n-1})\right];\\
&\ x_{n+1} = \lambda_n\beta_n\left[(B(y_{n-1}) - B(y_n))\right] + \lambda_n\left[D(y_{n-1}) - D(y_n)\right] + y_n,
\end{aligned} \tag{4.8}$$

where $(\lambda_n)_{n\in\mathbb{N}\cup\{0\}}$ and $(\beta_n)_{n\in\mathbb{N}\cup\{0\}}$ are sequences of positive real numbers.

Of course, in order to demonstrate the (ergodic) convergence results we need to assume the hypotheses (H_{fitz}) for every $n \in \mathbb{N} \cup \{0\}$.

Remark 4.2.1.

(i) When the operator B satisfies $B(x) = 0$ for all $x \in \mathcal{H}$ (which implies $N_C(x) = 0$ for every $x \in \mathcal{H}$), the algorithm outlined in **Algorithm 4.1** reduces to the error-free scenario with the identity variable matrices of the forward-backward-forward with extrapolation method, as introduced in Chapter 3.

(ii) Actually, we can choose y_{-1} as any starting point in $\mathcal{H}$. This starting point affects the algorithm's performance, but we can not specify which vector is the best choice for the problem. This freedom of choice has the pros and cons to consider. As in Section 4.6, we will compare our proposed method with the classical Tseng's algorithm in the penalty scheme, which requires only one starting point x_0. To be fair, the starting point y_{-1} is selected the same as x_0.

Prior to establishing the weak ergodic convergence result, we shall prove the following lemma, which serves as a valuable tool for our main outcome.

Lemma 4.2.2. *Let $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ and $(y_n)_{n\in\mathbb{N}\cup\{0,-1\}}$ be the sequences generated by **Algorithm 4.1** and let $(u, w) \in \operatorname{Gra}(A + D + N_C)$ be such that $w = v + p + D(u)$, where $v \in A(u)$ and $p \in N_C(u)$. Then the following inequality holds for $n \geq 1$:*

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \left(\frac{1}{2} - \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2\right)\|y_{n-1} - y_n\|^2$$
$$\leq \left(\frac{\lambda_{n-1}\beta_{n-1}}{\mu} + \frac{\lambda_{n-1}}{\eta}\right)^2 \|y_{n-2} - y_{n-1}\|^2 + 2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] + 2\lambda_n\langle u - y_n, w\rangle. \tag{4.9}$$

Proof. From **Algorithm 4.1** and the definition of the resolvent, we can derive that we have $\frac{x_n - y_n}{\lambda_n} - \beta_n B(y_{n-1}) - D(y_{n-1}) \in A(y_n)$ for ever $n \in \mathbb{N}$. Since $v \in A(u)$, then we can guarantee by using the monotonicity of A that

$$\begin{aligned} &\langle y_n - u, \frac{x_n - y_n}{\lambda_n} - \beta_n B(y_{n-1}) - D(y_{n-1}) - v\rangle \geq 0 \\ &\Leftrightarrow \langle y_n - u, x_n - y_n - \lambda_n\beta_n B(y_{n-1}) - \lambda_n D(y_{n-1}) - \lambda_n v\rangle \geq 0. \end{aligned} \tag{4.10}$$

By using the definition of x_{n+1}, we obtain that $\forall n \in \mathbb{N}$,

$$\begin{aligned} &\langle u - y_n, x_n - y_n\rangle \\ &\overset{(4.10)}{\leq} \langle u - y_n, \lambda_n\beta_n B(y_{n-1}) + \lambda_n D(y_{n-1}) + \lambda_n v\rangle \\ &= \langle u - y_n, [\lambda_n\beta_n(B(y_{n-1}) - B(y_n)) + \lambda_n(D(y_{n-1}) - D(y_n)) + y_n] \\ &\quad + \lambda_n\beta_n B(y_n) + \lambda_n D(y_n) + \lambda_n v - y_n\rangle \\ &= \langle u - y_n, x_{n+1} - y_n + \lambda_n\beta_n B(y_n) + \lambda_n D(y_n) + \lambda_n v\rangle \\ &= \langle u - y_n, x_{n+1} - y_n\rangle + \langle u - y_n, \lambda_n\beta_n B(y_n)\rangle + \langle u - y_n, \lambda_n D(y_n)\rangle + \langle u - y_n, \lambda_n v\rangle. \end{aligned} \tag{4.11}$$

For any $a_1, a_2, a_3 \in \mathcal{H}$, we know that $\frac{1}{2}\|a_1 - a_2\|^2 - \frac{1}{2}\|a_3 - a_1\|^2 + \frac{1}{2}\|a_3 - a_2\|^2 = \langle a_1 - a_2, a_3 - a_2\rangle$. Then, it follows from (4.11) that

$$\begin{aligned}\frac{1}{2}\|u - y_n\|^2 - \frac{1}{2}\|x_n - u\|^2 + \frac{1}{2}\|x_n - y_n\|^2 \overset{(4.11)}{\le} & \frac{1}{2}\|u - y_n\|^2 - \frac{1}{2}\|x_{n+1} - u\|^2 + \frac{1}{2}\|x_{n+1} - y_n\|^2 \\ & + \langle u - y_n, \lambda_n\beta_n B(y_n)\rangle + \langle u - y_n, \lambda_n D(y_n)\rangle \\ & + \langle u - y_n, \lambda_n v\rangle.\end{aligned}$$

Because $v = w - p - D(u)$ and the fact that D is monotone, we have that for every $n \in \mathbb{N}$

$$\begin{aligned}\|x_{n+1} - u\|^2 - \|x_n - u\|^2 \le\ & \|x_{n+1} - y_n\|^2 - \|x_n - y_n\|^2 + 2\lambda_n\beta_n\langle u - y_n, B(y_n)\rangle + 2\lambda_n\langle u - y_n, D(y_n)\rangle \\ & + 2\lambda_n\langle u - y_n, w - p - D(u)\rangle \\ =\ & \|x_{n+1} - y_n\|^2 - \|x_n - y_n\|^2 + 2\lambda_n\beta_n\langle u - y_n, B(y_n)\rangle \\ & + 2\lambda_n\beta_n\langle u - y_n, \frac{-p}{\beta_n}\rangle + 2\lambda_n\langle u - y_n, D(y_n) - D(u)\rangle + 2\lambda_n\langle u - y_n, w\rangle \\ =\ & \|x_{n+1} - y_n\|^2 - \|x_n - y_n\|^2 \\ & + 2\lambda_n\beta_n\Big(\langle u, B(y_n)\rangle + \langle y_n, \frac{p}{\beta_n}\rangle - \langle y_n, B(y_n)\rangle - \langle u, \frac{p}{\beta_n}\rangle\Big) \\ & + 2\lambda_n \underbrace{\langle u - y_n, D(y_n) - D(u)\rangle}_{\le 0} + 2\lambda_n\langle u - y_n, w\rangle \\ \le\ & \|x_{n+1} - y_n\|^2 - \|x_n - y_n\|^2 \\ & + 2\lambda_n\beta_n\left[\sup_{u\in C}\left\{\sup_{(s,B(s))\in \mathrm{Gr}(B)}\{\langle u, B(s)\rangle + \langle s, \frac{p}{\beta_n}\rangle - \langle s, B(s)\rangle\}\right\} - \sigma_C\left(\frac{p}{\beta_n}\right)\right] \\ & + 2\lambda_n\langle u - y_n, w\rangle \\ =\ & \|x_{n+1} - y_n\|^2 - \|x_n - y_n\|^2 + 2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] \\ & + 2\lambda_n\langle u - y_n, w\rangle.\end{aligned} \tag{4.12}$$

By the Lipschitz continuity of B and D, it yields

$$\begin{aligned}\|x_{n+1} - y_n\| &= \|\lambda_n\beta_n(B(y_{n-1}) - B(y_n)) + \lambda_n(D(y_{n-1}) - D(y_n))\| \\ &\le \lambda_n\beta_n\|B(y_{n-1}) - B(y_n)\| + \lambda_n\|D(y_{n-1}) - D(y_n)\| \\ &= \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)\|y_{n-1} - y_n\|.\end{aligned} \tag{4.13}$$

It follows from (4.12) and above inequality that

$$\begin{aligned}\|x_{n+1} - u\|^2 - \|x_n - u\|^2 \le & \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2\|y_{n-1} - y_n\|^2 - \|x_n - y_n\|^2 \\ & + 2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] + 2\lambda_n\langle u - y_n, w\rangle.\end{aligned} \tag{4.14}$$

By Parallelogram law, we know that $2\|x_n - y_n\|^2 + 2\|x_n - y_{n-1}\|^2 = \|y_{n-1} - y_n\|^2 + \|(x_n - y_n) + (x_n - y_{n-1})\|^2$. Then we have

$$-\|x_n - y_n\|^2 \leq \|x_n - y_{n-1}\|^2 - \frac{1}{2}\|y_{n-1} - y_n\|^2.$$

So, we can derive from (4.14) that

$$\begin{aligned}\|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq& \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2 \|y_{n-1} - y_n\|^2 + \left[\|x_n - y_{n-1}\|^2 - \frac{1}{2}\|y_{n-1} - y_n\|^2\right] \\ &+ 2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] + 2\lambda_n\langle u - y_n, w\rangle.\end{aligned}$$

It follows from (4.13) that $\|x_n - y_{n-1}\| \leq \left(\frac{\lambda_{n-1}\beta_{n-1}}{\mu} + \frac{\lambda_{n-1}}{\eta}\right)\|y_{n-2} - y_{n-1}\|$. Thus,

$$\begin{aligned}\|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq& \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2 \|y_{n-1} - y_n\|^2 \\ &+ \left[\left(\frac{\lambda_{n-1}\beta_{n-1}}{\mu} + \frac{\lambda_{n-1}}{\eta}\right)^2 \|y_{n-2} - y_{n-1}\|^2 - \frac{1}{2}\|y_{n-1} - y_n\|^2\right] \\ &+ 2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] + 2\lambda_n\langle u - y_n, w\rangle.\end{aligned}$$

Therefore

$$\begin{aligned}&\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \left(\frac{1}{2} - \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2\right)\|y_{n-1} - y_n\|^2 \\ &\leq \left(\frac{\lambda_{n-1}\beta_{n-1}}{\mu} + \frac{\lambda_{n-1}}{\eta}\right)^2 \|y_{n-2} - y_{n-1}\|^2 + 2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] + 2\lambda_n\langle u - y_n, w\rangle.\end{aligned}$$

□

Subsequently, we state the convergence of **Algorithm 4.1** below.

Theorem 4.2.3. *Let $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ and $(y_n)_{n\in\mathbb{N}\cup\{0,-1\}}$ be the sequences generated by **Algorithm 4.1** and $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ be the sequence defined in (4.6). If (H_{fitz}) is fulfilled and $\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right) < \frac{1}{2}$, then $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ converges weakly to an element in $zer(A + D + N_C)$ as $n \to +\infty$.*

Proof. The proof of the theorem is divided into three parts as the following statements:

(a) $\sum_{n\in\mathbb{N}\cup\{0\}}\|y_n - y_{n-1}\|^2 < +\infty$ and the sequence $(\|x_n - u\|)_{n\in\mathbb{N}\cup\{0\}}$ is convergent for every $u \in zer(A + D + N_C)$;

(b) every weak cluster point of $(z'_n)_{n\in\mathbb{N}\cup\{0\}}$ lies in $zer(A + D + N_C)$, where

$$z'_n = \frac{1}{\tau_n}\sum_{k=0}^{n}\lambda_k y_k \quad \text{and} \quad \tau_n = \sum_{k=0}^{n}\lambda_k \ \forall n \in \mathbb{N};$$

(c) every weak cluster point of $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ lies in $zer(A+D+N_C)$.

Every step of proof can be shown as follows.
(a) Let $u \in zer(A+D+N_C)$. For $n \in \mathbb{N}$, we take $w=0$ in (4.9), so we have

$$\begin{aligned}&\|x_{n+1}-u\|^2 - \|x_n-u\|^2 + \left(\frac{1}{2} - M_n^2\right)\|y_{n-1}-y_n\|^2 \\ &\leq M_{n-1}^2\|y_{n-2}-y_{n-1}\|^2 + 2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right],\end{aligned} \tag{4.15}$$

where $M_n = \frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}$. Rearranging the inequality, we have that

$$\begin{aligned}&\|x_{n+1}-u\|^2 + M_n^2\|y_{n-1}-y_n\|^2 + \left(\frac{1}{2} - 2M_n^2\right)\|y_{n-1}-y_n\|^2 \\ &\leq \|x_n-u\|^2 + M_{n-1}^2\|y_{n-2}-y_{n-1}\|^2 + 2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right].\end{aligned} \tag{4.16}$$

By the hypothesis we have that $\liminf_{n\to+\infty}\left(\frac{1}{2} - 2M_n^2\right) > 0$, hence it follows from Lemma 2.1.3 that the sequence $E_n := (\|x_n-u\|^2 + M_{n-1}^2\|y_{n-2}-y_{n-1}\|^2)_{n\in\mathbb{N}}$ converges and $\sum_{n\in\mathbb{N}}\|y_{n-1}-y_n\|^2 < +\infty$. This means that $(\|y_{n-1}-y_n\|)_{n\in\mathbb{N}\cup\{0\}} \to 0$ and so $(\|x_n-u\|)_{n\in\mathbb{N}\cup\{0\}}$ is a convergent sequence.

(b) Let z be a weak cluster point of $(z'_n)_{n\in\mathbb{N}}$. Take $(u,w)\in \mathrm{Gra}(A+D+N_C)$ such that $w = v+p+D(u)$, where $v\in A(u)$ and $p\in N_C(u)$. Let $K\in\mathbb{N}$ with $\forall n_0\in\mathbb{N}$, $K\geq n_0+2$. Summing up for $n=n_0+1,\dots,K$ the inequalities in (4.9) (it is nothing else than (4.15) with the additional term $2\lambda_n\langle u-y_n, w\rangle$) with $\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\mu}\right) < \frac{1}{2}$, for $n\in\mathbb{N}$ and $M_n = \frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}$ we get that

$$\begin{aligned}\sum_{n=n_0+1}^{K}\|x_{n+1}-u\|^2 - \sum_{n=n_0+1}^{K}\|x_n-u\|^2 \leq &\sum_{n=n_0+1}^{K} M_{n-1}^2\|y_{n-2}-y_{n-1}\|^2 \\ &+ 2\sum_{n=n_0+1}^{K}\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] \\ &+ 2\left\langle \sum_{n=0}^{K}\lambda_n u - \sum_{n=0}^{K}\lambda_n y_n - \sum_{n=0}^{n_0}\lambda_n u + \sum_{n=0}^{n_0}\lambda_n y_n, w\right\rangle.\end{aligned}$$

This leads to the following inequality.

$$\begin{aligned}\|x_{K+1}-u\|^2 - \|x_{n_0+1}-u\|^2 \leq &\sum_{n=n_0+1}^{+\infty} M_{n-1}^2\|y_{n-2}-y_{n-1}\|^2 \\ &+ 2\sum_{n=n_0+1}^{+\infty}\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] \\ &+ 2\left\langle \sum_{n=0}^{K}\lambda_n u - \sum_{n=0}^{K}\lambda_n y_n - \sum_{n=0}^{n_0}\lambda_n u + \sum_{n=0}^{n_0}\lambda_n y_n, w\right\rangle\end{aligned}$$

$$= R + 2\left\langle \sum_{n=0}^{K} \lambda_n u - \sum_{n=0}^{K} \lambda_n y_n - \sum_{n=0}^{n_0} \lambda_n u + \sum_{n=0}^{n_0} \lambda_n y_n, w \right\rangle,$$

where $R = \sum_{n=n_0+1}^{+\infty} M_{n-1}^2 \|y_{n-2} - y_{n-1}\|^2 + 2 \sum_{n \geq n_0+1}^{+\infty} \lambda_n \beta_n \left[\sup_{u \in C} \varphi\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] \in \mathbb{R}$.

Discarding the nonnegative term $\|x_{K+1} - u\|^2$ and dividing by $2\tau_K = 2\sum_{n=1}^{K} \lambda_n$ we obtain

$$\begin{aligned}
-\frac{\|x_{n_0+1} - u\|^2}{2\tau_K} &\leq \frac{R + 2\langle -\sum_{n=1}^{n_0} \lambda_n u + \sum_{n=1}^{n_0} \lambda_n y_n, w\rangle}{2\tau_K} + \frac{2\langle \sum_{n=1}^{K} \lambda_n u - \sum_{n=1}^{K} \lambda_n y_n, w\rangle}{2\sum_{n=1}^{K} \lambda_n} \\
&= \frac{\tilde{R}}{2\tau_K} + \langle u - z'_K, w\rangle,
\end{aligned}$$

where $\tilde{R} := R + 2\langle -\sum_{n=1}^{n_0} \lambda_n u + \sum_{n=1}^{n_0} \lambda_n x, w\rangle \in \mathbb{R}$. By passing to the limit as $K \to +\infty$ and using that $\lim_{K\to+\infty} \tau_K = +\infty$, we get

$$\liminf_{K\to+\infty} \langle u - z'_K, w\rangle \geq 0.$$

Since z is a weak cluster point of $(z'_n)_{n\in\mathbb{N}\cup\{0\}}$, we obtain that $\langle u - z, w\rangle \geq 0$. Finally as this inequality holds for arbitrary $(u, w) \in \operatorname{Gra}(A + D + N_C)$, z lies in $zer(A + D + N_C)$ (since $A + D + N_C$ is maximally monotone, see also (2.11)).

(c) Let $n \in \mathbb{N} \cup \{0\}$. It is enough to prove that $\lim_{n\to+\infty} \|z_n - z'_n\| = 0$ and the statement of the theorem will be a consequence of Lemma 4.1.1. Taking $u \in zer(A + D + N_C)$ and $w = 0 = v + p + D(u)$ where $v \in A(u)$ and $p \in N_C(u)$, it follows from the proof of (a) and the proof of Lemma 4.2.2 (see (4.13)) that $\sum_{n\in\mathbb{N}\cup\{0\}} \|y_{n-1} - y_n\|^2 < +\infty$ and $\|x_{n+1} - y_n\| \leq \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right) \|y_{n-1} - y_n\|$, respectively. For any $n \in \mathbb{N} \cup \{0\}$ it holds

$$\begin{aligned}
\|z_n - z'_n\|^2 &= \frac{1}{\tau_n^2} \left\| \sum_{k=0}^{n} \lambda_k (x_k - y_k) \right\|^2 \\
&\leq \frac{1}{\tau_n^2} \left(\sum_{k=0}^{n} \lambda_k \|x_k - y_k\| \right)^2 \\
&\leq \frac{1}{\tau_n^2} \left(\sum_{k=0}^{n} \lambda_k^2 \right) \left(\sum_{k=0}^{n} \|x_k - y_k\|^2 \right) \\
&\leq \frac{1}{\tau_n^2} \left(\sum_{k=0}^{n} \lambda_k^2 \right) \left(\sum_{k=0}^{n} (2\|x_k - y_{k-1}\|^2 + 2\|y_{k-1} - y_k\|^2) \right) \\
&= \frac{1}{\tau_n^2} \left(\sum_{k=0}^{n} \lambda_k^2 \right) \left(2\sum_{k=0}^{n} \|x_k - y_{k-1}\|^2 + 2\sum_{k=0}^{n} \|y_{k-1} - y_k\|^2 \right) \\
&\leq \frac{1}{\tau_n^2} \left(\sum_{k=0}^{n} \lambda_k^2 \right) \left(2\sum_{k=0}^{n} \left(\frac{\lambda_{k-1}\beta_{k-1}}{\mu} + \frac{\lambda_{k-1}}{\eta} \right)^2 \|y_{k-2} - y_{k-1}\|^2 + 2\sum_{k=0}^{n} \|y_{k-1} - y_k\|^2 \right).
\end{aligned}$$

We know that $\limsup\limits_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}\right)<\frac{1}{2}$, $(\lambda_n)_{n\in\mathbb{N}\cup\{0\}}\in\ell^2\setminus\ell^1$, and $\sum_{n\in\mathbb{N}\cup\{0\}}\|y_n-y_{n-1}\|^2<+\infty$. Taking into consideration that $\tau_n=\sum_{k=0}^n\lambda_k\to+\infty$ $(n\to+\infty)$, we obtain that $\|z_n-z'_n\|\to 0$ $(n\to+\infty)$. □

As observed in the forward-backward-forward penalty scheme discussed in [17, 18], strong monotonicity of the operator A guarantees the strong convergence of the sequence $(x_n)_{n\in\mathbb{N}\cup\{0\}}$. However, it remains to be seen whether this result extends to the forward-backward-forward with extrapolation from the past and the penalty scheme under consideration. The following result provides the necessary clarification.

Theorem 4.2.4. *Let $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ and $(y_n)_{n\in\mathbb{N}\cup\{0,-1\}}$ be the sequences generated by* ***Algorithm 4.1****. If (H_{fitz}) is fulfilled, $\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}\right)<\frac{1}{2}$ and the operator A is γ-strongly monotone with $\gamma>0$, then $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ converges strongly to the unique element in $zer(A+D+N_C)$ as $n\to+\infty$.*

Proof. Let u be the unique element of $zer(A+D+N_C)$ and $w=0=v+p+D(u)$, where $v\in A(u)$ and $p\in N_C(u)$. We can follow the proof of Lemma 4.2.2 under A is γ-strongly monotone with $\gamma>0$. This means that we obtain $\frac{x_n-y_n}{\lambda_n}-D(y_{n-1})-\beta_nB(y_{n-1})\in A(y_n)$, and one simply shows that for each $n\in\mathbb{N}$

$$2\gamma\lambda_n\|y_n-u\|^2+\|x_{n+1}-u\|^2-\|x_n-u\|^2+\left(\frac{1}{2}-\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}\right)^2\right)\|y_{n-1}-y_n\|^2$$
$$\leq\left(\frac{\lambda_{n-1}\beta_{n-1}}{\mu}+\frac{\lambda_{n-1}}{\eta}\right)^2\|y_{n-2}-y_{n-1}\|^2+2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right)-\sigma_C\left(\frac{p}{\beta_n}\right)\right],$$

equivalently,

$$2\gamma\lambda_n\|y_n-u\|^2+\|x_{n+1}-u\|^2-\|x_n-u\|^2+\left(\frac{1}{2}-2\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}\right)^2\right)\|y_{n-1}-y_n\|^2$$
$$+\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}\right)^2\|y_{n-1}-y_n\|^2$$
$$\leq\left(\frac{\lambda_{n-1}\beta_{n-1}}{\mu}+\frac{\lambda_{n-1}}{\eta}\right)^2\|y_{n-2}-y_{n-1}\|^2+2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right)-\sigma_C\left(\frac{p}{\beta_n}\right)\right].$$

Let $\bar{E_n}=\left(\|x_{n+1}-u\|^2+\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}\right)^2\|y_{n-1}-y_n\|^2\right)_{n\in\mathbb{N}}$, $\quad\forall n\in\mathbb{N}$. Then:

$$2\gamma\lambda_n\|y_n-u\|^2+\bar{E}_n+\left(\frac{1}{2}-2\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}\right)^2\right)\|y_{n-1}-y_n\|^2$$
$$\leq\bar{E}_{n-1}+2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right)-\sigma_C\left(\frac{p}{\beta_n}\right)\right].$$

The hypothesis $\left(\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu}+\frac{\lambda_n}{\eta}\right)<\frac{1}{2}\right)$ implies the existence of $n_0\in\mathbb{N}$ such that for every $n\geq n_0$

$$2\gamma\lambda_n\|y_n-u\|^2+\bar{E}_n-\bar{E}_{n-1}\leq 2\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right)-\sigma_C\left(\frac{p}{\beta_n}\right)\right].$$

Then, the inequality can be used for finite summation as: for some $N\geq n_0$

$$2\gamma\sum_{n=n_0}^{N}\lambda_n\|y_n-u\|^2+\bar{E}_N\leq\bar{E}_{n_0-1}+2\sum_{n=n_0}^{N}\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right)-\sigma_C\left(\frac{p}{\beta_n}\right)\right].$$

Since $\bar{E}_N\geq 0$, then we can omit this term on the left-hand side of above inequality. Hence, we derive the following inequality:

$$2\gamma\sum_{n\geq n_0}\lambda_n\|y_n-u\|^2\leq\bar{E}_{n_0-1}+2\sum_{n\geq n_0}\lambda_n\beta_n\left[\sup_{u\in C}\varphi_B\left(u,\frac{p}{\beta_n}\right)-\sigma_C\left(\frac{p}{\beta_n}\right)\right]<+\infty. \tag{4.17}$$

It follows that

$$\sum_{n\in\mathbb{N}}\lambda_n\|y_n-u\|^2<+\infty. \tag{4.18}$$

From (4.13), we can derive

$$\begin{aligned}\|x_n-y_n\|^2&\leq 2\|x_n-y_{n-1}\|^2+2\|y_n-y_{n-1}\|^2\\&\overset{(4.13)}{\leq}2\left(\frac{\lambda_{n-1}\beta_{n-1}}{\mu}+\frac{\lambda_{n-1}}{\eta}\right)^2\|y_{n-2}-y_{n-1}\|^2+2\|y_n-y_{n-1}\|^2.\end{aligned}$$

Since we have already known from Theorem 4.2.3 (a) that $\sum_{n\in\mathbb{N}}\|y_n-y_{n-1}\|^2<+\infty$ and $\left(\frac{\lambda_{n-1}\beta_{n-1}}{\mu}+\frac{\lambda_{n-1}}{\eta}\right)^2$ is bounded from above by the hypothesis, then

$$\sum_{n\in\mathbb{N}}\|x_n-y_n\|^2\leq 2\sum_{n\in\mathbb{N}}\left(\frac{\lambda_{n-1}\beta_{n-1}}{\mu}+\frac{\lambda_{n-1}}{\eta}\right)^2\|y_{n-2}-y_{n-1}\|^2+2\sum_{n\in\mathbb{N}}\|y_n-y_{n-1}\|^2<+\infty. \tag{4.19}$$

It follows from (4.18), (4.19), and λ_n is bounded from above that

$$\sum_{n\in\mathbb{N}}\lambda_n\|x_n-u\|^2\leq 2\sum_{n\in\mathbb{N}}\lambda_n\|x_n-y_n\|^2+2\sum_{n\in\mathbb{N}}\lambda_n\|y_n-u\|^2<+\infty.$$

Because $\sum_{n\in\mathbb{N}}\lambda_n=+\infty$, and $\lim_{n\to+\infty}\|x_n-u\|$ exists (proof of Theorem 4.2.3 (a)), we can conclude that $\lim_{n\to+\infty}\|x_n-u\|=0$.

□

4.3 Minimax problems

The minimax problem is a classical problem where the goal is to find a saddle point for a function in two variables. This problem arises in many applications, including game theory [82], control theory [60], and robust optimization [50,61]. In recent years, there has been a growing interest in the study of minimax problems in modern machine learning, such as Generative Adversarial Networks(GANs) [14], data-driven optimization [94] and furthermore in [96].

In this section, we will consider a class of minimax problems with a convex-concave structure, which can be solved using Tseng's forward-backward-forward method with extrapolation from the past and penalty scheme. We will present the problem formulation, discuss the properties of the objective function, and introduce some numerical methods for solving the problem.

Let $f : \mathcal{H} \times \mathcal{G} \to \mathbb{R}$ be convex-concave and differentiable, where $f(\cdot, y)$ is convex for all $y \in \mathcal{G}$ and $f(x, \cdot)$ is concave for all $x \in \mathcal{H}$ ($\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces). The *Min-Max problem* (or minimax problem) is the problem in the form:

$$\min_{x \in \mathcal{H}} \max_{y \in \mathcal{G}} f(x, y). \tag{4.20}$$

A saddle point of (4.20) is a vector $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ fulfilling:

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq f(x, \bar{y}) \quad \forall x \in \mathcal{H},\ \forall y \in \mathcal{G}. \tag{4.21}$$

Let $F : \mathcal{H} \times \mathcal{G} \to \mathcal{H} \times \mathcal{G}$ defined by $F \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} \nabla_1 f(x, y) \\ -\nabla_2 f(x, y) \end{pmatrix}$, where $\nabla_1 f,\ \nabla_2 f$ are the gradients of f with respect to the first and second component, respectively. Then, writing the optimal conditions for the inequalities in (4.21), we get

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \nabla_1 f(\bar{x}, \bar{y}) \\ -\nabla_2 f(\bar{x}, \bar{y}) \end{pmatrix}, \quad \text{i.e.,} \quad (\bar{x}, \bar{y}) \in zer(F). \tag{4.22}$$

Moreover, F is monotone and also is Lipschitz continuous (if we impose Lipschitz continuous properties on $\nabla_1 f, \nabla_2 f$), see in [8, Proposition 17.7] and [77, Theorem 35.1].

In the constrained case, the problem in (4.20) can be written as

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y), \tag{4.23}$$

where $\mathcal{X} \subset \mathcal{H}, \mathcal{Y} \subset \mathcal{G}$ are nonempty closed convex sets and a saddle point $(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}$ satisfies:

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq f(x, \bar{y}) \quad \forall x \in \mathcal{X},\ \forall y \in \mathcal{Y}. \tag{4.24}$$

The characterization of saddle points of (4.23) follows from the considerations:

$$f(x, \bar{y}) \geq f(\bar{x}, \bar{y})\ \forall x \in \mathcal{X} \Leftrightarrow 0 \in \partial(f(\cdot, \bar{y}) + \delta_{\mathcal{X}})(\bar{x}) = \nabla_1 f(\bar{x}, \bar{y}) + N_{\mathcal{X}}(\bar{x}),$$

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \; \forall y \in \mathcal{Y} \Leftrightarrow 0 \in \partial(-f(\bar{x}, \cdot) + \delta_{\mathcal{Y}})(\bar{y}) = -\nabla_2 f(\bar{x}, \bar{y}) + N_{\mathcal{Y}}(\bar{y}),$$

equivalently,

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} N_{\mathcal{X}}(\bar{x}) \\ N_{\mathcal{Y}}(\bar{y}) \end{pmatrix} + \begin{pmatrix} \nabla_1 f(\bar{x}, \bar{y}) \\ -\nabla_2 f(\bar{x}, \bar{y}) \end{pmatrix} = \underbrace{N_{\mathcal{X} \times \mathcal{Y}}(\bar{x}, \bar{y})}_{\text{maximally monotone}} + \underbrace{F(\bar{x}, \bar{y})}_{\text{monotone (Lipschitzian)}} . \tag{4.25}$$

Further we consider more involved minimax problems with constrained sets described by linear equations:

$$\min_{\substack{x \in \mathcal{X} \\ K_1(x) = b}} \max_{\substack{y \in \mathcal{Y} \\ K_2(y) = b'}} f(x, y) \Leftrightarrow \min_{\{x \in \mathcal{X} \mid K_1(x) = b\}} \max_{\{y \in \mathcal{Y} \mid K_2(y) = b'\}} f(x, y), \tag{4.26}$$

where $\mathcal{Q}$, $\mathcal{Q}'$ are other real Hilbert spaces, $K_1 : \mathcal{H} \to \mathcal{Q}, K_2 : \mathcal{G} \to \mathcal{Q}'$ are linear continuous operators with closed ranges such that $b \in \mathcal{Q}, b' \in \mathcal{Q}'$ and

$$\{x \in \mathcal{X} \mid K_1(x) = b\} \neq \emptyset, \quad \{y \in \mathcal{Y} \mid K_2(y) = b'\} \neq \emptyset \tag{4.27}$$

are nonempty sets. This kind of minimax constrained problem was studied with plenty of applications, for instance in [94]. In this case, we have the following characterization of saddle points of (4.26)

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} N_{\mathcal{X} \cap \{x \in \mathcal{H} \mid K_1(x) = b\}}(\bar{x}) \\ N_{\mathcal{Y} \cap \{y \in \mathcal{G} \mid K_2(y) = b'\}}(\bar{y}) \end{pmatrix} + \underbrace{\begin{pmatrix} \nabla_1 f(\bar{x}, \bar{y}) \\ -\nabla_2 f(\bar{x}, \bar{y}) \end{pmatrix}}_{:= F(\bar{x}, \bar{y})} . \tag{4.28}$$

Note that $N_{C_1 \cap C_2} = \partial(\delta_{C_1} + \delta_{C_2}) = N_{C_1} + N_{C_2}$, if a constraint qualification is satisfied, for example $\emptyset \neq C_1 \cap \operatorname{int} C_2$ (see in [8, Corollary 16.38]). Then

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} N_{\mathcal{X}}(\bar{x}) \\ N_{\mathcal{Y}}(\bar{y}) \end{pmatrix}}_{\text{maximally monotone}} + \begin{pmatrix} N_{\{x \in \mathcal{H} \mid K_1(x) = b\}}(\bar{x}) \\ N_{\{y \in \mathcal{G} \mid K_2(y) = b'\}}(\bar{y}) \end{pmatrix} + \underbrace{\begin{pmatrix} \nabla_1 f(\bar{x}, \bar{y}) \\ -\nabla_2 f(\bar{x}, \bar{y}) \end{pmatrix}}_{\substack{:= F(\bar{x}, \bar{y}), \\ \text{monotone (Lipschitzian)}}} . \tag{4.29}$$

The solution of the linear equations

$$K_1(x) = b, \quad K_2(y) = b', \tag{4.30}$$

can be described by means of the functions $\psi_1 : \mathcal{H} \to \mathbb{R}, \psi_2 : \mathcal{G} \to \mathbb{R}$ defined by

$$\psi_1(x) = \frac{1}{2}\|K_1(x) - b\|^2 \quad \text{and} \quad \psi_2(y) = \frac{1}{2}\|K_2(y) - b'\|^2. \tag{4.31}$$

Namely, the solution of (4.30) can be obtained by solving the following optimization problem:

$$x \in \arg\min_{x\in\mathcal{H}} \frac{1}{2}\|K_1(x)-b\|^2 = \arg\min_{x\in\mathcal{H}} \psi_1(x) \Leftrightarrow \nabla\psi_1(x)=0,$$
$$y \in \arg\min_{y\in\mathcal{G}} \frac{1}{2}\|K_2(y)-b'\|^2 = \arg\min_{y\in\mathcal{G}} \psi_2(y) \Leftrightarrow \nabla\psi_2(y)=0.$$

For every $x\in\mathcal{H}$ and $y\in\mathcal{G}$, let

$$\bar{A}(x,y)=\begin{pmatrix} N_{\mathcal{X}}(x) \\ N_{\mathcal{Y}}(y)\end{pmatrix},\ \bar{B}(x,y)=\begin{pmatrix} \nabla\psi_1(x) \\ -\nabla\psi_2(y)\end{pmatrix},\ \bar{C}=zer\bar{B},\ \bar{D}(x,y)=\begin{pmatrix} \nabla_1 f(x,y) \\ -\nabla_2 f(x,y)\end{pmatrix} \tag{4.32}$$

Therefore, (4.29) can be written as

$$\begin{pmatrix}0\\0\end{pmatrix}\in\begin{pmatrix} N_{\mathcal{X}}(\bar{x}) \\ N_{\mathcal{Y}}(\bar{y})\end{pmatrix}+\begin{pmatrix} N_{zer(\nabla\psi_1)}(\bar{x}) \\ N_{zer(\nabla\psi_2)}(\bar{y})\end{pmatrix}+\begin{pmatrix} \nabla_1 f(\bar{x},\bar{y}) \\ -\nabla_2 f(\bar{x},\bar{y})\end{pmatrix}\Leftrightarrow\begin{pmatrix}0\\0\end{pmatrix}\in\bar{A}(\bar{x},\bar{y})+N_{\bar{C}}(\bar{x},\bar{y})+\bar{D}(\bar{x},\bar{y}). \tag{4.33}$$

Due to the above setting, the problem (4.29) turns into the monotone inclusion in (4.7). The iterative pattern in **Algorithm 4.1** can be transformed into for each $n\geq 0$ as

$$\begin{pmatrix} r_n \\ s_n\end{pmatrix}=J_{\lambda_n\begin{pmatrix} N_{\mathcal{X}} \\ N_{\mathcal{Y}}\end{pmatrix}}\left[\begin{pmatrix} x_n \\ y_n\end{pmatrix}-\lambda_n\begin{pmatrix} \nabla_1 f(r_{n-1},s_{n-1}) \\ -\nabla_2 f(r_{n-1},s_{n-1})\end{pmatrix}-\lambda_n\beta_n\begin{pmatrix} \nabla\psi_1(r_{n-1}) \\ \nabla\psi_2(s_{n-1})\end{pmatrix}\right];$$
$$\begin{pmatrix} x_{n+1} \\ y_{n+1}\end{pmatrix}=\lambda_n\beta_n\left[\begin{pmatrix} \nabla\psi_1(r_{n-1}) \\ \nabla\psi_2(s_{n-1})\end{pmatrix}-\begin{pmatrix} \nabla\psi_1(r_n) \\ \nabla\psi_2(s_n)\end{pmatrix}\right]+\lambda_n\left[\begin{pmatrix} \nabla_1 f(r_{n-1},s_{n-1}) \\ -\nabla_2 f(r_{n-1},s_{n-1})\end{pmatrix}-\begin{pmatrix} \nabla_1 f(r_n,s_n) \\ -\nabla_2 f(r_n,s_n)\end{pmatrix}\right]+\begin{pmatrix} r_n \\ s_n\end{pmatrix},$$

where $\nabla\psi_1(x)=K_1^*(K_1(x)-b)$ and $\nabla\psi_2(y)=K_2^*(K_2(y)-b')$.

Recall from Example 2.3.6, Example 2.4.3 and Definition 2.6.6 , we can notice that if G is a nonempty closed convex set subset of $\mathcal{H}$, then $J_{N_G}=(\mathrm{Id}+N_G)^{-1}=prox_{\iota_G}=P_G$. Therefore, we are able to construct a relevant algorithm for solving the monotone inclusion scheme in (4.33) (i.e., for solving problem (4.26)) demonstrated as:

Algorithm 4.2:

Initialization : Choose $r_{-1},x_0\in\mathcal{H}, s_{-1},y_0\in\mathcal{G}$ (with $r_{-1}=x_0, s_{-1}=y_0$).

For $n\in\mathbb{N}\cup\{0\}$ set :

$$\begin{aligned} r_n &= P_{\mathcal{X}}[x_n-\lambda_n\nabla_1 f(r_{n-1},s_{n-1})-\lambda_n\beta_n K_1^*(K_1(r_{n-1})-b)];\\ s_n &= P_{\mathcal{Y}}[y_n+\lambda_n\nabla_2 f(r_{n-1},s_{n-1})-\lambda_n\beta_n K_1^*(K_2(s_{n-1})-b')];\\ x_{n+1} &= \lambda_n\beta_n[K_1^*(K_1(r_{n-1})-b)-K_1^*(K_1(r_n)-b)]\\ &\quad+\lambda_n\nabla_1 f(r_{n-1},s_{n-1})-\lambda_n\nabla_1 f(r_n,s_n)+r_n;\\ y_{n+1} &= \lambda_n\beta_n[K_1^*(K_2(s_{n-1})-b')-K_1^*(K_2(s_n)-b')]\\ &\quad+\lambda_n\nabla_2 f(r_n,s_n)-\lambda_n\nabla_2 f(r_{n-1},s_{n-1})+s_n, \end{aligned} \tag{4.34}$$

where $(\lambda_n)_{n\in\mathbb{N}\cup\{0\}}$ and $(\beta_n)_{n\in\mathbb{N}\cup\{0\}}$ are positive real numbers.

Thus, we can use condition (ii) in (H) (in the introduction of this Chapter) to get the convergent results. We define $\Psi : \mathcal{H} \times \mathcal{G} \to \mathbb{R}$,

$$\Psi(x,y) := \psi_1(x) + \psi_2(y) = \frac{1}{2}\|K_1(x) - b\|^2 + \frac{1}{2}\|K_2(y) - b'\|^2. \tag{4.35}$$

Then, we have (see in [8, Proposition 13.30])

$$\Psi^*(x,y) = \psi_1^*(x) + \psi_2^*(y). \tag{4.36}$$

From the considerations above (see also (4.27)), we have:

$$\bar{C} = \left\{(x,y) \in \mathcal{H} \times \mathcal{G} : \begin{pmatrix} \nabla\psi_1(x) \\ -\nabla\psi_2(y) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right\} = \left\{(x,y)\mathcal{H} \times \mathcal{G} : \begin{matrix} K_1(x) = b \\ K_2(y) = b' \end{matrix}\right\}$$

and

$$\bar{C} = \arg\min(\psi_1) \times \arg\min(\psi_2) = \arg\min(\Psi).$$

Furthermore, we also have

$$\sigma_{\bar{C}} = \sigma_{\arg\min(\psi_1)\times\arg\min(\psi_2)} = \sigma_{\arg\min(\psi_1)} + \sigma_{\arg\min(\psi_2)}$$

and

$$N_{\bar{C}} = N_{\arg\min(\psi_1)} \times N_{\arg\min(\psi_2)} = N_{\arg\min(\Psi)}.$$

Thus, if we impose for $i = 1, 2$

$$\forall p_i \in \operatorname{ran} N_{\arg\min(\psi_i)}, \qquad \sum_{n\in\mathbb{N}\cup\{0\}} \lambda_n\beta_n \left[\psi_i^*\left(\frac{p_i}{\beta_n}\right) - \sigma_{\arg\min(\psi_i)}\left(\frac{p_i}{\beta_n}\right)\right] < +\infty, \tag{4.37}$$

then, it follows that

$$\forall (p_1,p_2) \in \operatorname{Ran}(N_{\bar{C}}), \qquad \sum_{n\in\mathbb{N}\cup\{0\}} \lambda_n\beta_n \left[\Psi^*\left(\frac{(p_1,p_2)}{\beta_n}\right) - \sigma_{\bar{C}}\left(\frac{(p_1,p_2)}{\beta_n}\right)\right] < +\infty,$$

is exactly the condition (ii) in (H) that we need for convergence.

In this context, we assume the following condition:

$$\text{(Condition A)} \qquad (\operatorname{int}\mathcal{X} \times \operatorname{int}\mathcal{Y}) \cap \left\{(x,y) \in \mathcal{H} \times \mathcal{G} : \begin{matrix} K_1(x) = b \\ K_2(y) = b' \end{matrix}\right\} \neq \emptyset.$$

which is a constraint qualification of our considered problem (see in [8, Corollary 16.48] and (4.28)-(4.29)).

Hence, we can propose the ergodic convergence results which follow from Theorem 4.2.3 as below.

Corollary 4.3.1. *In the framework of (4.26), let $\mathcal{X} \subset \mathcal{H}, \mathcal{Y} \subset \mathcal{G}$ be nonempty closed convex sets and the operators $\bar{A}, \bar{B}, \bar{C}$ and $\bar{D}$ be given as in (4.32) where $\bar{A}$ is maximally monotone, $\bar{D}$ is monotone and η^{-1}-Lipschitz continuous with $\eta > 0$, $\bar{B}$ is monotone and μ^{-1}-Lipschitz continuous operator with $\mu > 0$ (i.e., $\mu^{-1} := \bar{K}$, where $\bar{K} = \max\{\|K_1\|^2, \|K_2\|^2\}, K_1 \neq 0, K_2 \neq 0$), $\bar{C} := zer(\bar{B}) \neq \emptyset$, and (Condition A) is satisfied. Consider the sequences generated by **Algorithm 4.2**, and the sequence $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ defined as*

$$z_n = \frac{1}{\tau_n}\sum_{k=0}^{n} \lambda_k \begin{pmatrix} x_k \\ y_k \end{pmatrix}, \quad \textit{where } \tau_n = \sum_{k=0}^{n} \lambda_k \ \forall n \in \mathbb{N},$$

where $(\lambda_k)_{k\in\mathbb{N}\cup\{0\}}$ is a positive sequence such that $\sum_{k\in\mathbb{N}\cup\{0\}} \lambda_k = +\infty$. If we suppose the following hypotheses:

$$(H^{min-max}) \begin{cases} (i) & \bar{A} + N_{\bar{C}} \textit{ is maximally monotone and } (4.33) \textit{ has a solution;} \\ (ii) & \textit{For every } p_i \in \operatorname{Ran}(N_{\arg\min(\psi_i)}), \\ & \sum_{n\in\mathbb{N}\cup\{0\}} \lambda_n\beta_n \left[\psi_i^*(\frac{p_i}{\beta_n}) - \sigma_{\arg\min(\psi_i)}(\frac{p_i}{\beta_n})\right] < +\infty, \\ & i = 1, 2, \textit{ where } \psi_i \textit{ are defined in } (4.31); \\ (iii) & (\lambda_n)_{n\in\mathbb{N}\cup\{0\}} \in \ell^2 \setminus \ell^1, \end{cases}$$

and $\limsup_{n\to+\infty} \left(\frac{\lambda_n\beta_n}{\mu} + \frac{\lambda_n}{\eta}\right) < \frac{1}{2}$, *then $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ converges weakly to a saddle point of (4.26).*

Remark 4.3.2.

(i) Considering the condition (i) in $(H^{min-max})$, notice that $\bar{A} + N_{\bar{C}}$ is maximally monotone if a regularity condition is fulfilled, for example (Condition A) (see in [8, Corollary 24.4]). In addition, (4.33) has a solution if the (Condition A) holds and (4.26) has saddle points.

(ii) Notice that (ii) in $(H^{min-max})$ is fulfilled when $\sum_{n\in\mathbb{N}\cup\{0\}} \frac{\lambda_n}{\beta_n} < +\infty$ and $b = b' = 0$.

(iii) We can show the Lipschitz property of $\bar{B}$ as follows:

$$\begin{aligned} \|\bar{B}(x,y) - \bar{B}(u,v)\| &= \left\| \begin{pmatrix} \nabla\psi_1(x) \\ -\nabla\psi_2(y) \end{pmatrix} - \begin{pmatrix} \nabla\psi_1(u) \\ -\nabla\psi_2(v) \end{pmatrix} \right\| \\ &= \left\| \begin{pmatrix} \nabla\psi_1(x) - \nabla\psi_1(u) \\ \nabla\psi_2(v) - \nabla\psi_2(y) \end{pmatrix} \right\| \\ &= \left\| \begin{pmatrix} K_1^*(K_1x - b) - K_1^*(K_1u - b) \\ K_2^*(K_2v - b') - K_2^*(K_2y - b') \end{pmatrix} \right\| \\ &= \left\| \begin{pmatrix} K_1^*K_1\,(x-u) \\ -K_2^*K_2\,(y-v) \end{pmatrix} \right\| \\ &\leq \bar{K}\ \|(x,y) - (u,v)\|, \end{aligned}$$

where $\bar{K} = \max\{\|K_1\|^2, \|K_2\|^2\}$.

4.4 The Forward-Backward-Forward Algorithm with Extrapolation from the Past and Penalty Scheme for the Problem involving Composition of Linear Continuous Operators

In this section, we propose forward-backward-forward algorithm with extrapolation from the past and penalty scheme utilized to address the inclusion problem involving the finite sum of the compositions of monotone operators with linear continuous operators. To this end, we begin with the following problem:

Problem 4.3: Let $\mathcal{H}$ be a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $D : \mathcal{H} \to \mathcal{H}$ be a monotone and ν-Lipschitz continuous operator with $\nu > 0$. Let m be strictly positive integer and for any $i \in \{1, \ldots, m\}$, let $\mathcal{G}_i$ be a real Hilbert space, $B_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be a maximally monotone operator, and $L_i : \mathcal{H} \to \mathcal{G}_i$ be a nonzero linear continuous operator. Assume that $B : \mathcal{H} \to \mathcal{H}$ is a monotone and μ^{-1}-Lipschitz continuous operator with $\mu > 0$, and suppose $C = zer(B) \neq \emptyset$. The monotone inclusion problem to solve is

$$0 \in A(x) + \sum_{i=1}^{m} L_i^*(B_i(L_i(x))) + D(x) + N_C(x). \tag{4.38}$$

The monotone inclusion problem **Problem 4.3** can be reformulated into the same form as **Problem 4.1** by using the product space approach with a pertinent setting. To address this problem, we work with the product space $\mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$, where the inner product and associated norm are defined for every element $(x, v_1, \ldots, v_m)$ by $\langle (x, v_1, \ldots, v_m), (y, w_1, \ldots, w_m) \rangle = \langle x, y \rangle + \sum_{i=1}^{m} \langle v_i, w_i \rangle$ and $\|(x, v_1, \ldots, v_m)\| = \sqrt{\|x\|^2 + \sum_{i=1}^{m} \|v_i\|^2}$. We proceed to define the operators on the product space $\mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$ as follows: for every $(x, v_1, \ldots, v_m), (y, w_1, \ldots, w_m) \in \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$, $\tilde{A} : \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m \rightrightarrows \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$,

$$\tilde{A}(x, v_1, \ldots, v_m) = A(x) \times B_1^{-1}(v_1) \times \cdots \times B_m^{-1}(v_m);$$

$\tilde{D} : \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m \to \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$,

$$\tilde{D}(x, v_1, \ldots, v_m) = \left(\sum_{i=1}^{m} L_i^*(v_i) + D(x), -L_1(x), \ldots, -L_m(x) \right);$$

and $\tilde{B} : \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m \to \mathcal{H} \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m$,

$$\tilde{B}(x, v_1, \ldots, v_m) = (B(x), \underbrace{0, \ldots, 0}_{m-\text{elements}}).$$

Note that the maximal monotonicity of A and B_i, $i = 1, \ldots, m$, implies that the operator $\tilde{A}$ is also maximally monotone, as stated in [8, Proposition 20.22 and 20.23]. Moreover, according to in [31, Theorem 3.1], it is possible to demonstrate that the operator $\tilde{D}$ is monotone and β-Lipschitz continuous where $\beta = \nu + \sqrt{\sum_{i=1}^{m} \|L_i\|^2}$ (in this context, the monotonicity and Lipschitzian of D is a special case of [17, 31]). Moreover, we also have that $\tilde{B}$ is monotone and μ^{-1}-Lipschitz continuous with

$$\tilde{C} := zer(\tilde{B}) := zer(B) \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m = C \times \mathcal{G}_1 \times \cdots \times \mathcal{G}_m,$$

and

$$N_{\tilde{C}}(x, v_1, \ldots, v_m) = N_C(x) \times \underbrace{\{0\} \times \cdots \times \{0\}}_{m\text{-elements}}.$$

One can show (see [17]) that x is a solution to **Problem 4.3** if and only if there exists $v_1 \in \mathcal{G}_1, \ldots, v_m \in \mathcal{G}_m$ such that $(x, v_1, \ldots, v_m) \in zer(\tilde{A} + \tilde{D} + N_{\tilde{C}})$. On the other hand, when $(x, v_1, \ldots, v_m) \in zer(\tilde{A}+\tilde{D}+N_{\tilde{C}})$, then $x \in zer(A+\sum_{i=1}^m L_i^* B_i L_i + D + N_C)$. This means that determining the zeros of $\tilde{A} + \tilde{D} + N_{\tilde{C}}$ will automatically provide a solution to **Problem 4.3**.

By using the identity given in [8, Proposition 23.18], which provides that for any $(x, v_1, \ldots, v_m) \in \mathcal{H} \times \mathcal{G}_1, \times \cdots \times \mathcal{G}_m$ and $\lambda > 0$, $J_{\lambda\tilde{A}}(x, v_1, \ldots, v_m) = (J_{\lambda_n A}(x), J_{\lambda B_1^{-1}}(v_1), \ldots, J_{\lambda B_m^{-1}}(v_m))$, it can be observed that the iterations of **Algorithm 4.1** are given for any $n \in \mathbb{N} \cup \{0\}$ as follows:

$$\begin{aligned}
(y_n, q_{1,n}, \ldots, q_{m,n}) &= J_{\lambda_n \tilde{A}}[(x_n, v_{1,n}, \ldots, v_{m,n}) - \lambda_n \tilde{D}(y_{n-1}, q_{1,n-1}, \ldots, q_{m,n-1}) \\
&\quad - \lambda_n \beta_n \tilde{B}(y_{n-1}, q_{1,n-1}, \ldots, q_{m,n-1})], \\
(x_{n+1}, v_{1,n+1}, \ldots, v_{m,n+1}) &= \lambda_n \beta_n [\tilde{B}(y_{n-1}, q_{1,n-1}, \ldots, q_{m,n-1}) - \tilde{B}(y_n, q_{1,n}, \ldots, q_{m,n})] \\
&\quad + \lambda_n [\tilde{D}(y_{n-1}, q_{1,n-1}, \ldots, q_{m,n-1}) - \tilde{D}(y_n, q_{1,n}, \ldots, q_{m,n})] \\
&\quad + (y_n, q_{1,n}, \ldots, q_{m,n}).
\end{aligned}$$

This leads to the algorithm below:

Algorithm 4.3:

Initialization : Choose $(x_0, v_{1,0}, \ldots, v_{m,0}), (y_{-1}, q_{1,-1}, \ldots, q_{m,-1}) \in \mathcal{H} \times \mathcal{G}_1, \times \ldots \times \mathcal{G}_m$ with $y_{-1} = x_0, q_{i,-1} = v_{i,0}\ \forall i = 1, \ldots, m$.

For $n \in \mathbb{N} \cup \{0\}$,

$$\begin{aligned}
\text{set} :\ & y_n = J_{\lambda_n A}[x_n - \lambda_n (D(y_{n-1}) + \sum_{i=1}^m L^*(q_{i,n-1})) - \lambda_n \beta_n B(y_{n-1})]; \\
& q_{i,n} = J_{\lambda_n B_i^{-1}}[v_{i,n} + \lambda_n L_i(y_{n-1})],\ i = 1, \ldots, m,; \\
& x_{n+1} = \lambda_n \beta_n [B(y_{n-1}) - B(y_n)] + \lambda_n [D(y_{n-1}) - D(y_n)] \\
& \qquad + \lambda_n \sum_{i=1}^m L_i^*(q_{i,n-1} - q_{i,n}) + y_n; \\
& v_{i,n+1} = \lambda_n L_i(y_n - y_{n-1}) + q_{i,n},\ i = 1, \ldots, m,
\end{aligned} \tag{4.39}$$

where $(\lambda_n)_{n\in\mathbb{N}\cup\{0\}}$ and $(\beta_n)_{n\in\mathbb{N}\cup\{0\}}$ are positive real numbers.

To consider the convergence of this iterative scheme, we need the following additional hypotheses, similar to the hypotheses in [17, Section 2]:

$$(H^{sum}_{fitz})\begin{cases}(i) & A+N_C \text{ is maximally monotone and } zer(A+\sum_{i=1}^m L^*B_iL+D+N_C)\neq\emptyset;\\ (ii) & \text{For every } p\in\operatorname{ran}N_C, \sum_{n\in\mathbb{N}\cup\{0\}}\lambda_n\beta_n[\sup\limits_{u\in C}\varphi_B(u,\frac{p}{\beta_n})-\sigma_C(\frac{p}{\beta_n})]<+\infty;\\ (iii) & (\lambda_n)_{n\in\mathbb{N}\cup\{0\}}\in\ell^2\setminus\ell^1.\end{cases}$$

The Fitzpatrick function of $\tilde{B}$ (i.e., $\varphi_{\tilde{B}}$) and the support function of $\tilde{C}$ (i.e., $\sigma_{\tilde{C}}$) can be computed in the same way as demonstrated in [17, Section 2] for arbitrary elements $(x_1,v_1,\ldots,v_m),(x'_1,v'_1,\ldots,v'_m)\in\mathcal{H}\times\mathcal{G}_1,\times\cdots\times\mathcal{G}_m$. Moreover, the satisfaction of condition (i) in (H^{sum}_{fitz}) guarantees that $\tilde{A}+N_{\tilde{C}}$ is maximally monotone and $zer(\tilde{A}+\tilde{D}+N_{\tilde{C}})\neq\emptyset$. As a result, we can apply Theorem 4.2.3 and 4.2.4 to the problem of finding the zeros of $\tilde{A}+\tilde{D}+N_{\tilde{C}}$. Thus, we can establish the convergence results for the scheme of the sum of monotone operators and linear continuous operators, which are given below.

Theorem 4.4.1. *Let $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ and $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ be sequences generated by **Algorithm 4.3**. Assume (H^{sum}_{fitz}) is fulfilled and $\limsup_{n\to+\infty}(\frac{\lambda_n\beta_n}{\mu}+\lambda_n\beta)<\frac{1}{2}$, where*

$$\beta=\nu+\sqrt{\sum_{i=1}^m\|L_i\|^2}.$$

Then $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ converges weakly to an element in $zer(A+\sum_{i=1}^m L_i^*B_iL_i+D+N_C)$ as $n\to\infty$. If, additionally, A and $B_i^{-1}, i=1,\ldots,m$ are strongly monotone, then $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ converges strongly to the unique element in $zer(A+\sum_{i=1}^m L_i^*B_iL_i+D+N_C)$ as $n\to\infty$.

4.5 Convex Minimization Problems

Let us discuss the case of constrained optimization problem as follows:

Problem 4.4: Let $\mathcal{H}$ be a real Hilbert space, $f\in\Gamma_0(\mathcal{H})$, and $h:\mathcal{H}\to\mathbb{R}$ be a convex and differentiable function with ν-Lipschitz continuous gradient for $\nu>0$. Let m be a strictly positive integer, and for any $i=1,\ldots,m$, let $\mathcal{G}_i$ be a real Hilbert space, $g_i\in\Gamma_0(\mathcal{G}_i)$, and $L_i:\mathcal{H}\to\mathcal{G}_i$ be a nonzero linear continuous operator. Furthermore, let $\Psi:\mathcal{H}\to\mathbb{R}$ be convex and differentiable with a μ^{-1}-Lipschitz continuous gradient, fulfilling $\min\Psi=0$. The convex minimization problem under investigation is

$$\inf_{x\in\arg\min\Psi}\left\{f(x)+\sum_{i=1}^m g_i(L_i(x))+h(x)\right\}. \tag{4.40}$$

In order to solve this problem in our context, we can follow **Problem 4.3** in section 4.4. We know that any element belonging to $zer(\partial f+\sum_{i=1}^m L_i^*\partial g_iL_i+\nabla h+N_C)$ is an optimal solution for (4.40) if we substitute all of operators with

$$A = \partial f,\ B = \nabla\Psi,\ C = \arg\min \Psi = zer(B),\ D = \nabla h, \text{ and } B_i = \partial g_i,\ i = 1,\dots,m, \tag{4.41}$$

where B is a monotone and μ^{-1}-Lipschitz continuous operator (see in [8, Proposition 17.7 and Theorem 18.15]). However, the converse is true only if a suitable qualification condition is satisfied. Necessary conditions, namely that (4.40) leads to **Problem 4.3** with (4.38), have been previously given in [17] (for example, see in [31, Proposition 4.3 and Remark 4.4])

$$(0,\dots,0) \in \operatorname{sri}\left(\prod_{i=1}^{m} \operatorname{dom} g_i - \{(L_1(x),\dots,L_m(x)) \mid x \in \operatorname{dom}\ f \cap C\}\right). \tag{4.42}$$

Moreover, the condition in (4.42) is fulfilled if

- dom $g_i = \mathcal{G}_i, i = 1,\dots,m$ or
- $\mathcal{H}$ and $\mathcal{G}_i$ are finite dimensional and there exists $x \in \operatorname{ri}\operatorname{dom} f \cap \operatorname{ri} C$ such that $L_i(x) \in \operatorname{ri}\operatorname{dom}\ g_i, i = 1,\dots,m$ (see [31, Proposition 4.3]).

At this point, we are equipped to state the iterative method based on Tseng's forward-backward-forward algorithm with extrapolation from the past and penalty scheme and its convergent result.

Algorithm 4.4:

Initialization : Choose $(x_0, v_{1,0},\dots,v_{m,0}), (y_{-1}, q_{1,-1},\dots,q_{m,-1}) \in \mathcal{H}\times\mathcal{G}_1,\times\dots\times\mathcal{G}_m$ with $y_{-1} = x_0, q_{i,-1} = v_{i,0}\ \forall i = 1,\dots,m.$

For $n \in \mathbb{N}\cup\{0\}$,

$$\begin{aligned}
\text{set}:\ & y_n = prox_{\lambda_n f}[x_n - \lambda_n(\nabla h(y_{n-1}) + \sum_{i=1}^{m} L_i^*(q_{i,n-1})) - \lambda_n\beta_n\nabla\Psi(y_{n-1})];\\
& q_{i,n} = prox_{\lambda_n g_i^*}[v_{i,n} + \lambda_n(L_i(y_{n-1})],\ i = 1,\dots,m,\\
& x_{n+1} = \lambda_n\beta_n(\nabla\Psi(y_{n-1}) - \nabla\Psi(y_n)) + \lambda_n(\nabla h(y_{n-1}) - \nabla h(y_n))\\
& \qquad + \lambda_n\sum_{i=1}^{m} L_i^*(q_{i,n-1} - q_{i,n}) + y_n;\\
& v_{i,n+1} = \lambda_n L_i(y_n - y_{n-1}) + q_{i,n},\ i = 1,\dots,m,
\end{aligned} \tag{4.43}$$

where $(\lambda_n)_{n\in\mathbb{N}\cup\{0\}}$ and $(\beta_n)_{n\in\mathbb{N}\cup\{0\}}$ are positive real numbers.

To analyze the convergence, we need to assume the hypotheses similar to those in [17, Section 3], which are as follows:

$$(H^{opt}) \begin{cases} (i) & \partial f + N_C \text{ is maximally monotone and (4.40) has an optimal solution;}\\ (ii) & \text{For every } p \in \operatorname{Ran} N_C, \sum_{n\in\mathbb{N}\cup\{0\}} \lambda_n\beta_n[\Psi^*(\frac{p}{\beta_n}) - \sigma_C(\frac{p}{\beta_n})] < +\infty;\\ (iii) & (\lambda_n)_{n\in\mathbb{N}\cup\{0\}} \in \ell^2 \setminus \ell^1. \end{cases}$$

For the hypothesis (H^{opt}), some observations have already been presented in [17], which are as follows:

Remark 4.5.1.

(i) Due to [8, Corollary 24.4], $\partial f + N_C$ is maximally monotone, if $0 \in \operatorname{sri}(\operatorname{dom} f - C)$. This condition is fulfilled if the conditions in [8, Proposition 6.19 and Proposition 15.24] hold, for instance, $0 \in \operatorname{int}(\operatorname{dom} f - C)$ or f is continuous at a point in $\operatorname{dom} f \cap C$ or $\operatorname{int} C \cap \operatorname{dom} f \neq \emptyset$.

(ii) The condition (ii) of (H^{opt}) is similar to the condition (ii) of (H) which implies to the second contion of (H_{fitz}). Hence, we can apply our proposed convergent results in Section 4.2 for this context.

With the same effort as in Section 4.4, we may continue and provide the convergence results for **Algorithm 4.4** as shown below.

Theorem 4.5.2. *Let $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ be the sequences generated by **Algorithm 4.4** and $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ be the sequence defined in (4.6). If (H^{opt}) and (4.42) are fulfilled and $\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu} + \lambda_n\beta\right) < \frac{1}{2}$, where*

$$\beta = \nu + \sqrt{\sum_{i=1}^{m} \|L_i\|^2},$$

then $(z_n)_{n\in\mathbb{N}\cup\{0\}}$ converges weakly to an optimal solution to (4.40) as $n \to +\infty$. If, additionally, f and $g_i^*, i = 1, \ldots, m$ are strongly convex, then $(x_n)_{n\in\mathbb{N}\cup\{0\}}$ converges strongly to the unique optimal solution of the problem in (4.40) as $n \to +\infty$.

Remark 4.5.3.

1. According to in [8, Proposition 17.10 and Theorem 18.15], g^* is strongly convex whenever $g : \mathcal{H} \to \mathbb{R}$ is differentiable with Lipschitz continuous gradient.

2. If $\Psi(x) = 0$ for all $x \in \mathcal{H}$, then **Algorithm 4.4** reduces to the error-free version of the iterative scheme investigated in Chapter 3 where all of the variable matrices involved are identity matrices. This scheme is used to solve the convex minimization problem given by

$$\inf_{x\in\mathcal{H}} \left\{ f(x) + \sum_{i=1}^{m} g_i(L_i x) + h(x) \right\}.$$

4.6 A Numerical Experiment in TV-Based Image Inpainting

In this section, we demonstrate the application of **Algorithm 4.4** for solving an image inpainting problem, which involves recovering lost information in an image. We compare the classical Tseng's type (or the forward-backwards-forward (FBF) algorithm) penalty system, which Boț and Csetnek proposed in [17], with our modified Tseng's type penalty scheme. Because our algorithm can reduce the redundant computed terms in the typical Tseng's method (see introduction), then we expect to gain less time consumption in the next numerical experiment. The computations presented in this part were carried out with Python (version 3.7.9) on a Windows desktop computer powered by an Intel(R) Core(TM) i5-8250U processor that operated at speeds between 1.6 GHz and 1.8 GHz, and was equipped with 8.00 GB of RAM.

We represent images of size $M \times N$ as vectors $x \in \mathbb{R}^{\bar{n}}$, where $\bar{n} = M \cdot N$, and each pixel denoted by $x_{i,j}, 1 \leq i \leq M, 1 \leq j \leq N$, takes values in the closed interval from 0 (pure black) to 1 (pure white). Given an image $b \in \mathbb{R}^{\bar{n}}$ with missing pixels (which are set to black in this case), we define $P \in \mathbb{R}^{\bar{n}\times\bar{n}}$ as the diagonal matrix with $P_{i,i} = 0$ if the i-th pixel in the noisy image b is missing, and $P_{i,i} = 1$ otherwise, for $i = 1, \ldots, \bar{n}$ (noting that $\|P\| = 1$). The original image is reconstructed by solving the following TV-regularized model:

$$\inf\{TV_{iso}(x) \mid Px = b, x \in [0, 1]^{\bar{n}}\}. \tag{4.44}$$

The function $TV_{iso} : \mathbb{R}^{\bar{n}} \to \mathbb{R}$, which we use as our objective function, is defined as the isotropic total variation:

$$TV_{iso}(x)=\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\sqrt{(x_{i+1,j}-x_{i,j})^2+(x_{i,j+1}-x_{i,j})^2}+\sum_{i=1}^{M-1}|x_{i+1,N}-x_{i,N}|+\sum_{j=1}^{N-1}|x_{M,j+1}-x_{M,j}|.$$

It is possible to express the problem presented in (4.44) as an optimization problem of the form of **Problem 4.4** in (4.40). This can be achieved by introducing the set $\mathcal{Y} = \mathbb{R}^{\bar{n}} \times \mathbb{R}^{\bar{n}}$, and defining the linear operator $L : \mathbb{R}^{\bar{n}} \to \mathcal{Y}$ as $L(x_{i,j}) = (L^1(x_{i,j}), L^2(x_{i,j}))$,where

$$L^1(x_{i,j}) = \begin{cases} x_{i+1,j} - x_{i,j} & \text{if } i < M, \\ 0 & \text{if } i = M \end{cases} \quad \text{and} \quad L^2(x_{i,j}) = \begin{cases} x_{i,j+1} - x_{i,j} & \text{if } j < N, \\ 0 & \text{if } j = N. \end{cases}$$

The operator L represents a discretization of the gradient in the horizontal and vertical directions. It is worth noting that $\|L\|^2 < 8$, and its adjoint $L^* : \mathcal{Y} \to \mathbb{R}^{\bar{n}}$ is as easy to implement as L itself (see [27, 89]). Furthermore, the inner product on $\mathcal{Y}$, given by $\langle(y, z), (p, q)\rangle = \sum_{i=1}^{M}\sum_{j=1}^{N}(y_{i,j}p_{i,j} + z_{i,j}q_{i,j})$, induces a norm defined as $\|(y, z)\|_\times = \sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{y_{i,j}^2 + z_{i,j}^2}$ for $(y, z), (p, q) \in \mathcal{Y}$. It can be shown that $TV_{iso}(x) = \|L(x)\|_\times$ for every $x \in \mathbb{R}^{\bar{n}}$.

Moreover, by considering the function $\Psi : \mathbb{R}^{\bar{n}} \to \mathbb{R}, \Psi(x) = \frac{1}{2}\|Px - b\|^2$, problem in (4.44) can be reformulated as

$$\inf_{x\in\arg\min\Psi}\{f(x) + g_1(L(x))\}, \tag{4.45}$$

where $f : \mathbb{R}^{\bar{n}} \to \bar{\mathbb{R}}, f = \delta_{[0,1]^{\bar{n}}}$, $g_1 : \mathcal{Y} \to \mathbb{R}, g_1(y_1, y_2) = \|(y_1, y_2)\|_\times$. The problem given in (4.45) can be written as **Problem 4.4** (or (4.40)) when we set $m = 1, L_1 = L$, and $h = 0$. It is worth noting that $\nabla\Psi(x) = P(P(x)-b) = P(x-b)$ for all $x \in \mathbb{R}^{\bar{n}}$, which makes $\nabla\Psi$

Lipschitz continuous with a Lipschitz constant of $\mu = 1$. Therefore, the iterative scheme in **Algorithm 4.4** can be written as follows for every $n \geq 0$ in this particular case:

$$\begin{cases} y_n = prox_{\lambda_n f}[x_n - \lambda_n L^*(q_{1,n-1})) - \lambda_n \beta_n P(y_{n-1} - b)], \\ q_{1,n} = prox_{\lambda_n g_1^*}[v_{1,n} + \lambda_n L(y_{n-1})], \\ x_{n+1} = \lambda_n \beta_n P(y_{n-1} - y_n) + \lambda_n L^*(q_{1,n-1} - q_{1,n}) + y_n, \\ v_{1,n+1} = \lambda_n L(y_n - y_{n-1}) + q_{i,n}, \end{cases} \tag{4.46}$$

To implement this iterative method, we need the following formulas:

$$prox_{\gamma f}(x) = P_{[0,1]^n}(x) \quad \forall \gamma > 0 \quad \text{and} \quad \forall x \in \mathbb{R}^{\bar{n}},$$

and

$$prox_{\gamma g_1^*}(p,q) = P_S(p,q) \quad \forall \gamma > 0 \quad \text{and} \quad \forall (p,q) \in \mathcal{Y},$$

where $S = \left\{ (p,q) \in \mathcal{Y} \mid \max_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} \sqrt{p_{i,j}^2 + q_{i,j}^2} \leq 1 \right\}$. The projection operator $P_S : \mathcal{Y} \to S$ is defined via

$$(p_{i,j}, q_{i,j}) \mapsto \frac{(p_{i,j}, q_{i,j})}{\max\{1, \sqrt{p_{i,j}^2 + q_{i,j}^2}\}}, \ 1 \leq i \leq M, \ 1 \leq j \leq N.$$

The quality of the reconstructed images was compared using the Improvement in Signal-to-Noise Ratio (ISNR), defined as follows:

$$\text{ISNR}(n) = 10 log_{10} \left(\frac{\|x - b\|^2}{\|x - x_n\|^2} \right),$$

where x, b and x_n denote the original, the image with missing pixels and the recovered image at iteration n, respectively.

We evaluated the performance of the algorithm on a 256×256 pixel image of the Pisa Tower, using the parameters $\lambda_n = 0.9 \cdot n^{-0.75}$ (see [17]), $\lambda_n = 0.9 \cdot 2^{-1} \cdot n^{-0.75}$ and $\beta_n = n^{0.75}$ for all $n \in \mathbb{N}$. The original image, as well as an image with 80% randomly blacked-out pixels, were used in the test. Fig. 4.1 displays the original image, the noisy image, the non-averaged reconstructed image x_n, and the averaged reconstructed image z_n after 2000 iterations using both the Forward-Backward-Forward (FBF) in the penalty scheme with $\lambda_n = 0.9 \cdot n^{-0.75}$, FBF in penalty scheme with $\lambda_n = 0.9 \cdot 2^{-1} \cdot n^{-0.75}$ and Forward-Backward-Forward with Extrapolation from the Past(FBF-EP) in penalty scheme with $\lambda_n = 0.9 \cdot 2^{-1} \cdot n^{-0.75}$. Note that $\lambda_n = 0.9 \cdot n^{-0.75}$ can not be used for FBF-EP in penalty scheme because there is no theorem to support its convergence. In addition, since the parameters λ_n and β_n play a role in the methods as the stepsize of the algorithm, then we decided to choose these parameters, which have value close to

Table 4.1: The table presents the results of ISNR and CPU time (in seconds) for both averaged and non-averaged reconstructed images obtained through 2000 iterations of the FBF and FBF-EP methods.

	FBF $\lambda_n = 0.9 \cdot n^{-0.75}$		FBF $\lambda_n = 0.9 \cdot 2^{-1} \cdot n^{-0.75}$		FBF-EP $\lambda_n = 0.9 \cdot 2^{-1} \cdot n^{-0.75}$	
	average	non-average	average	non-average	average	non-average
ISNR	11.41243	10.92022	11.36461	8.78285	11.39179	8.77024
CPU-time (sec)	91.59540	88.66624	91.33789	88.41237	77.89488	74.59753

the upper bound of their conditions (namely, either $\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu} + \lambda_n\beta\right) < 1$ or $\limsup_{n\to+\infty}\left(\frac{\lambda_n\beta_n}{\mu} + \lambda_n\beta\right) < \frac{1}{2}$).

Fig. 4.2 and 4.3 depict the evolution of the ISNR values for both the averaged and non-averaged reconstructed images using FBF and FBF-EP algorithms (in both two values of λ_n in FBF, respectively). The theoretical outcomes concerning the sequences involved in Theorem 4.5.2 are illustrated in the figure, indicating that the averaged sequence exhibits better convergence properties than the non-averaged sequence. We notice that the behavior of ISNR values in both algorithms carries out similarly when we use the same value of λ_n. Table 4.1 reveals that the FBF-EP approach produced the averaged reconstructed image with an ISNR value of 11.39179, while it was 11.36461 for the averaged case of the FBF with $\lambda_n = 0.9 \cdot 2^{-1} \cdot n^{-0.75}$ and 11.41243 for the averaged case of the FBF with $\lambda_n = 0.9 \cdot n^{-0.75}$. However, the non-averaged reconstructed image using the FBF-EP method had the lowest ISNR value of 8.77024. Furthermore, the FBF-EP method was faster than the FBF method by around 13 seconds (compared with both two various of λ_n in FBF), indicating that it can offer time-saving benefits in image reconstruction.

Figure 4.1: TV image inpainting was performed on four images, including the original image, an image with 80% of its pixels missing, the non-averaged reconstructed image denoted as x_n, and the reconstructed image z_n obtained after 2000 iterations.

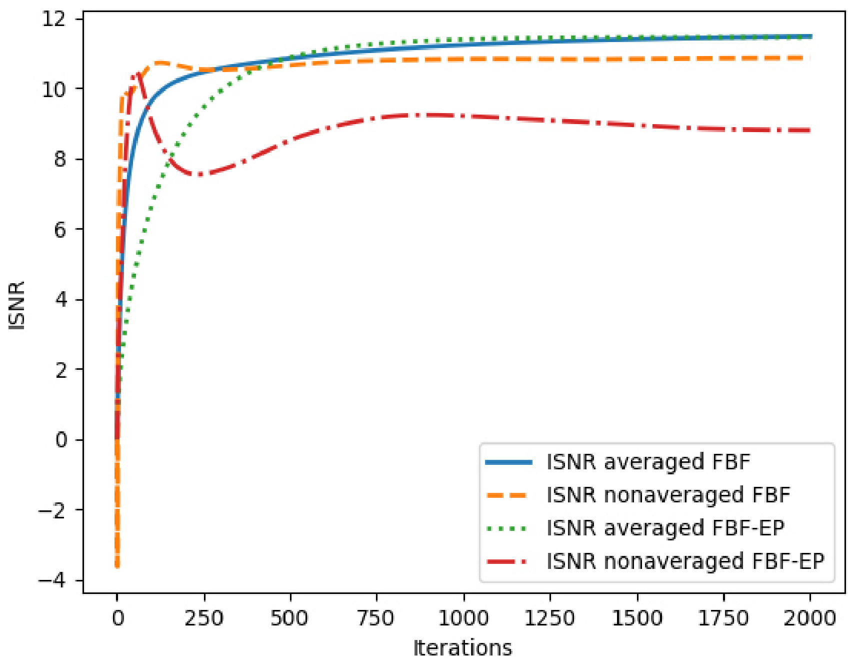


Figure 4.2: The figure illustrates the ISNR curves for both the averaged and non-averaged reconstructed images using the FBF method with $\lambda_n = 0.9 \cdot n^{-0.75}$ and FBF-EP method with $\lambda_n = 0.9 \cdot 2^{-1} \cdot n^{-0.75}$.

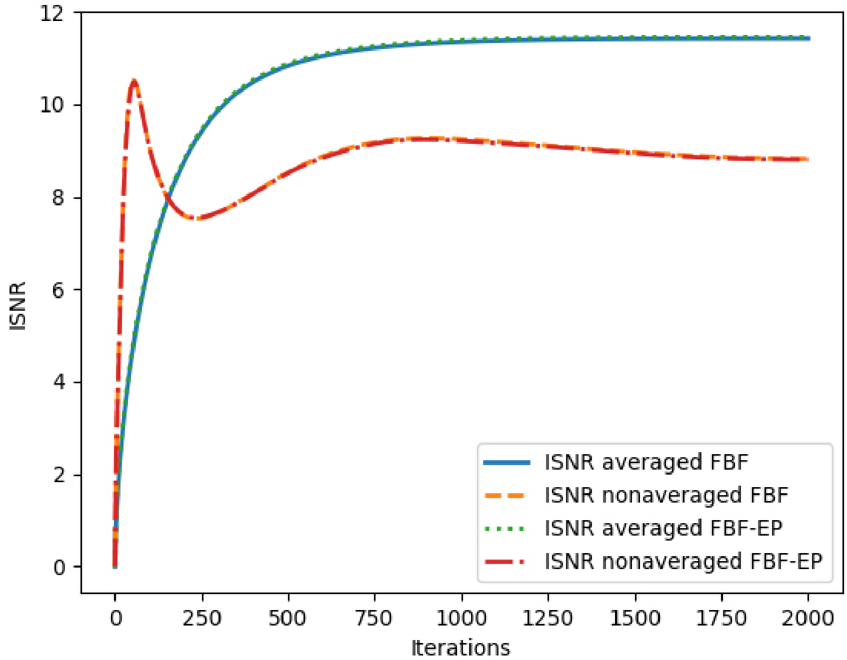


Figure 4.3: The figure illustrates the ISNR curves for both the averaged and non-averaged reconstructed images using the FBF and FBF-EP methods with the same value of $\lambda_n = 0.9 \cdot 2^{-1} \cdot n^{-0.75}$.

5 A Modified Tseng's Algorithm with Extrapolation from the Past for Pseudo-monotone Variational Inequalities

Variational inequalities (VIs) are beneficial mathematical models for solving various problems, like saddle problems, equilibrium problems, obstacle problems, and others; see [41, 53] and reference therein.

In this chapter, we are concerned with the *variational inequality (VI)* in the type of Hartman-Stampacchia (Stampacchia type) [47] :

$$\text{find } x^* \in C \quad \langle F(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C, \tag{5.1}$$

where C is a nonempty closed convex subset of Hilbert space $\mathcal{H}$ endowed with an inner product $\langle \cdot, \cdot \rangle$ and the corresponding norm $\| \cdot \|$ and F is monotone and L-Lipschitz continuous operator for $L \geq 0$. We denote the inequality of (5.1) by $VI(F, C)$ and assume that its solution set is represented by $\Omega \neq \emptyset$.

The projected-gradient algorithm is the simplest method for solving variational inequalities, and it is defined as follows: for a starting point $x_0 \in \mathcal{H}$,

$$x_{n+1} = P_C(x_n - \lambda F(x_n)), \quad \forall n \geq 0,$$

where P_C denotes the projection operator onto the closed convex set $C \subseteq \mathcal{H}$ and $\lambda > 0$. However, this method does not necessarily provide the convergence if F is only monotone (see [41, Example 12.1.3], for example). In order to converge, the method requires additional assumptions like F being cocoercive (or inverse strongly monotone, see also (2.9) or [8, 97]) or strongly (pseudo) monotone (see [41, 55]).

Recall the extragradient method proposed by Korpelevich in [57], which is used to solve variational inequalities governed by Lipschitz continuous and pseudo-monotone operators, reads as (1.5) (see introduction, Chapter 1) where $\lambda \in (0, 1/L)$. We can see that the algorithm needs to compute the projection onto C two times, which sometimes affects the method's efficiency if the projection is not easy to calculate. The extragradient method has gained overwhelming attention from several authors to modify this method in multiple ways; see, e.g., [26, 49, 56]. The work of Popov [73] gives us a sophisticated idea that we can reuse and store the computation of the projection term in each iterative scheme for extragradient shown as (1.7) where $\lambda \in (0, 1/3L)$. This modern concept called the *extrapolated technique or extrapolation from the past* is given in [43] and note that the latter approach is working within a smaller stepsize.

As an alternative extragradient method applied to solving the monotone inclusions, Tseng [90] proposed the forward-backward-forward (FBF) algorithm. Tseng's forward-backward-forward method has been studied in variational inequality in various aspects by many researchers, see [21,29]. Furthermore, the combination of Tseng's algorithm and the extrapolated technique is also suggested in [14,89] as known as Tseng's algorithm with extrapolation from the past (FBF-EP), see also (1.8). Both FBF and FBF-EP require only one projection per each iterative computation and provide the weak convergence of $(x_n)_{n\geq 0}$ to a solution of $VI(F,C)$ for monotone Lipschitz operator F, see [23,89,90]. Because of its simplicity and generality, this algorithm has attracted a lot of attention from many researchers, for example in [16,23,89]. The improvement of Tseng's method (FBF) goes further in the relaxation version for the pseudo-monotone and sequentially weak-to-weak continuous operator F based on FBF, proposed by Boț et al. in [21]. This work has shown that the convergence result holds when F is a pseudo-monotone (not necessary to be monotone), Lipschitz continuous and sequentially weak-to-weak-continuous operator. Sometimes, the Lipschitz constant L is unknown or is not simple to calculate. Therefore, Boț et al. [21] introduced an adaptive stepsize scenario for the their method.

The work of Boț et al. [21] motivates us to investigate the convergence of Tseng's algorithm with extrapolation from the past (FBF-EP) in case the operator F is pseudo-monotone, Lipschitz continuous and sequentially weak-to-weak-continuous. In addition, we propose an adaptive stepsize approach for FBF-EP, which does not depend on the knowledge of the Lipschitz constant. We also prove that the convergence statement holds in finite dimensional spaces under weaker assumptions on F. In the final section of this work, we additionally perform a numerical experiment on pseudo-monotone variational inequalities.

We would like to note that the full scope of this topic was published in Taiwanese Journal of Mathematics in 2024, as detailed in [87].

5.1 Preliminaries on Pseudo-Monotone Variational Inqualities

Before we present our main results for this chapter, let us give the relevant background knowledge.

We describe various properties of the operator F as follows:

Definition 5.1.1. Let C be a nonempty subset of the real Hilbert space $\mathcal{H}$. The mapping $F : \mathcal{H} \to \mathcal{H}$ is said to be *pseudo-monotone* on C if it holds that for every $x, y \in C$

$$\langle F(x), y - x\rangle \geq 0 \quad \Rightarrow \quad \langle F(y), y - x\rangle \geq 0.$$

Note that every monotone operator is pseudo-monotone but the pseudo-monotone is not necessarily monotone see for example [51, Definition 3.1.] the operator $F(x) = 1/(1 + x), \quad C = \{x \in \mathbb{R} \mid x \geq 0\}$, which is a pseudo-monotone operator on C, but not necessarily a monotone operator (we provide other examples in Section 5.3).

Recall that the operator $F : \mathcal{H} \to \mathcal{H}$ is called *Lipschitz continuous* with Lipschitz constant $L > 0$, if for every $x, y \in \mathcal{H}$ it holds that $\|F(x) - F(y)\| \leq L\|x - y\|$, and we say that the operator F is *sequentially weak-to-weak continuous*, if for every sequence $(x_n)_{n\geq 0}$ that converges weakly to x the sequence $(F(x_n))_{n\geq 0}$ also converges weakly to $F(x)$.

Now, we present a result for continuous pseudo-monotone operators on a nonempty, convex, and closed set. Additionally, this proposition establishes an identical relationship between the Stampacchia and Minty types of variational inequality (VI).

Proposition 5.1.2 ([34, Lemma 2.1]). *Let C be a nonempty, convex and closed subset of the real Hilbert space $\mathcal{H}$ and let $F : \mathcal{H} \to \mathcal{H}$ be an operator which is pseudo-monotone on C and continuous. Then for every $x \in C$ we have*

$$\langle F(x), y - x\rangle \geq 0, \quad \forall y \in C \quad \Leftrightarrow \quad \langle F(y), y - x\rangle \geq 0, \quad \forall y \in C.$$

Another version of variational inequality is called *Minty type*:

$$\text{Find } x^* \in C \quad \text{such that} \quad \langle F(x), x - x^*\rangle \geq 0, \quad \forall x \in C.$$

Notably, Proposition 5.1.2 implies that the solution sets of two variational inequalities in both Stampacchia type (5.1) and Minty type are the same when they are performed over a nonempty closed convex set and governed by pseudo-monotone and continuous operators.

5.2 Tseng's Algorithm with Extrapolation from the Past for Pseudo-Monotone Variational Inequalities

In this section, we introduce Tseng's algorithm with extrapolation from the past for pseudo-monotone variational inequalities, and we show a weak convergence result using Opial's lemma.

Theorem 5.2.1. *Let $\Omega \neq \emptyset$ be the solution set of the variational inequality problem of F on C, namely $VI(F, C)$, let $F : \mathcal{H} \to \mathcal{H}$ be pseudo-monotone on $\mathcal{H}$, Lipschitz continuous with constant L and sequentially weak-to-weak continuous. Let C be a nonempty, convex and closed subset of the real Hilbert space $\mathcal{H}$. Let $x_0, y_{-1} \in \mathcal{H}$. Assume that the sequence $(x_n)_{n\geq 0}$ is generated by the following algorithm*

$$\begin{aligned} y_n &= P_C(x_n - \lambda_n F(y_{n-1})), \\ x_{n+1} &= y_n + \lambda_n(F(y_{n-1}) - F(y_n)), \end{aligned} \tag{5.2}$$

with $0 < \liminf_{n\to\infty} \lambda_n \leq \limsup_{n\to\infty} \lambda_n < \frac{1}{2L}$. Then the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Ω.

Proof. The idea of the proof: we first show the inequality (5.12) by using $\Omega \neq \emptyset$, the pseudo-monotonicity of F, Theorem 2.4.4 (property of the projection operator) and the

Lipschitz continuity of F. Then we rearrange formula (5.12) and apply Lemma 2.1.3 to conclude that $\lim_{n\to+\infty} \|x_n - x^*\|^2$ exists which is the first condition of Opial's lemma (Lemma 2.1.6). For the rest of the proof, we need to verify the second condition of Opial's lemma, namely, if $\hat{x}$ is a weak sequential cluster point of $(x_n)_{n\geq 0}$, then $\hat{x} \in \Omega$.

Let x^* be an arbitrary element in Ω and let $n \geq 0$ be fixed. Then we have

$$\langle F(x^*), y - x^* \rangle \geq 0, \quad \forall y \in C.$$

Substituting $y := y_n \in C$ into the inequality yields

$$\langle F(x^*), y_n - x^* \rangle \geq 0.$$

From the pseudo-monotonicity of F on C, it follows that

$$\langle F(y_n), y_n - x^* \rangle \geq 0. \tag{5.3}$$

Since $y_n = P_C(x_n - \lambda_n F(y_{n-1}))$ according to Theorem 2.4.4, we get

$$\langle y - y_n, y_n - x_n + \lambda_n F(y_{n-1}) \rangle \geq 0, \quad \forall y \in C, \tag{5.4}$$

which yields for $y = x^* \in C$:

$$\langle x^* - y_n, y_n - x_n + \lambda_n F(y_{n-1}) \rangle \geq 0. \tag{5.5}$$

Multiplying both sides of (5.3) by $\lambda_n > 0$ we have

$$\langle \lambda_n F(y_n), y_n - x^* \rangle \geq 0 \text{ (or } \langle x^* - y_n, -\lambda_n F(y_n) \rangle \geq 0).$$

Adding the above inequality to (5.5) yields

$$\langle x^* - y_n, y_n - x_n + \lambda_n F(y_{n-1}) - \lambda_n F(y_n) \rangle \geq 0,$$

or, equivalently,

$$\langle x^* - y_n, x_{n+1} - x_n \rangle \geq 0.$$

Then, using (5.2) we obtain that

$$\begin{aligned}\langle x_{n+1} - x^*, x_{n+1} - x_n \rangle &\leq \langle x_{n+1} - y_n, x_{n+1} - x_n \rangle \\ &= \|x_{n+1} - x_n\|^2 + \langle x_n - y_n, x_{n+1} - x_n \rangle \\ &\overset{(5.2)}{=} \|x_{n+1} - x_n\|^2 + \langle x_n - y_n, y_n + \lambda_n F(y_{n-1}) - \lambda_n F(y_n) - x_n \rangle \\ &= \|x_{n+1} - x_n\|^2 - \|x_n - y_n\|^2 + \lambda_n \langle x_n - y_n, F(y_{n-1}) - F(y_n) \rangle.\end{aligned} \tag{5.6}$$

On the other hand, we have

$$\|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 + \|x_{n+1} - x_n\|^2 = 2\langle x_{n+1} - x^*, x_{n+1} - x_n \rangle. \tag{5.7}$$

Combining (5.6) and (5.7), we obtain that

$$\begin{aligned}
\|x_{n+1}-x^*\|^2 &\overset{(5.6)}{=} \|x_n-x^*\|^2-\|x_{n+1}-x_n\|^2+2\langle x_{n+1}-x^*, x_{n+1}-x_n\rangle \\
&\overset{(5.7)}{\leq} \|x_n-x^*\|^2-\|x_{n+1}-x_n\|^2 \\
&\quad +2[\|x_{n+1}-x_n\|^2-\|x_n-y_n\|^2+\lambda_n\langle x_n-y_n, F(y_{n-1})-F(y_n)\rangle] \\
&= \|x_n-x^*\|^2+\|x_{n+1}-x_n\|^2-2\|x_n-y_n\|^2 \\
&\quad +2\lambda_n\langle x_n-y_n, F(y_{n-1})-F(y_n)\rangle.
\end{aligned} \tag{5.8}$$

Using the Lipschitz continuity of F and (5.2), we obtain that

$$\begin{aligned}
\|x_{n+1}-x_n\|^2 &\overset{(5.2)}{=} \|y_n+\lambda_n F(y_{n-1})-\lambda_n F(y_n)-x_n\|^2 \\
&= \|y_n-x_n\|^2+2\lambda_n\langle y_n-x_n, F(y_{n-1})-F(y_n)\rangle+\lambda_n^2\|F(y_{n-1})-F(y_n)\|^2 \\
&\leq \|y_n-x_n\|^2+2\lambda_n\langle y_n-x_n, F(y_{n-1})-F(y_n)\rangle+\lambda_n^2L^2\|y_{n-1}-y_n\|^2.
\end{aligned} \tag{5.9}$$

From (5.8) and (5.9), we derive

$$\begin{aligned}
\|x_{n+1}-x^*\|^2 &\leq \|x_n-x^*\|^2+\big[\|y_n-x_n\|^2+2\lambda_n\langle y_n-x_n, F(y_{n-1})-F(y_n)\rangle+\lambda_n^2L^2\|y_{n-1}-y_n\|^2\big] \\
&\quad -2\|x_n-y_n\|^2+2\lambda_n\langle x_n-y_n, F(y_{n-1})-F(y_n)\rangle \\
&= \|x_n-x^*\|^2-\|y_n-x_n\|^2+\lambda_n^2L^2\|y_{n-1}-y_n\|^2.
\end{aligned} \tag{5.10}$$

By Parallelogram identity, we know that $\|y_n-y_{n-1}\|^2+\|(x_n-y_n)+(x_n-y_{n-1})\|^2 = 2\|x_n-y_n\|^2+2\|x_n-y_{n-1}\|^2$, hence together with the Lipschitz property of F we obtain that

$$\begin{aligned}
\|x_n-y_n\|^2 &\geq -\|x_n-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2 \\
&\overset{(5.2)}{=} -\|y_{n-1}+\lambda_{n-1}(F(y_{n-2})-F(y_{n-1}))-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2 \\
&\geq -(\lambda_{n-1}L)^2\|y_{n-2}-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2.
\end{aligned} \tag{5.11}$$

It follows from (5.10) and (5.11) that

$$\begin{aligned}
\|x_{n+1}-x^*\|^2 &\leq \|x_n-x^*\|^2-\big[-(\lambda_{n-1}L)^2\|y_{n-2}-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2\big] \\
&\quad +\lambda_n^2L^2\|y_{n-1}-y_n\|^2.
\end{aligned}$$

Thus, we have

$$\|x_{n+1}-x^*\|^2+\left(\frac{1}{2}-(\lambda_nL)^2\right)\|y_n-y_{n-1}\|^2 \leq \|x_n-x^*\|^2+(\lambda_{n-1}L)^2\|y_{n-2}-y_{n-1}\|^2. \tag{5.12}$$

Since $\limsup_{n\to\infty}\lambda_n < \frac{1}{2L}$, then we have $\limsup_{n\to\infty}\left\{2(\lambda_nL)^2\right\} < \frac{1}{2}$. This means that

$$\liminf_{n\to\infty}\left\{\frac{1}{2}-2(\lambda_nL)^2\right\} > 0,$$

and so there exists $n_0 \in \mathbb{N}$ such that

$$\frac{1}{2} - 2(\lambda_n L)^2 > \eta > 0, \quad \forall n > n_0, \text{ for some } \eta \in \mathbb{R}_{++}. \tag{5.13}$$

Now, from (5.12) we can derive the following inequality, for all $n > n_0$,

$$\begin{aligned} &\|x_{n+1} - x^*\|^2 + \left(\frac{1}{2} - 2(\lambda_n L)^2\right) \|y_n - y_{n-1}\|^2 + (\lambda_n L)^2 \|y_n - y_{n-1}\|^2 \\ &\leq \|x_n - x^*\|^2 + (\lambda_{n-1} L)^2 \|y_{n-2} - y_{n-1}\|^2. \end{aligned} \tag{5.14}$$

By (5.13), (5.14) and Lemma 2.1.3, we obtain that $\|x_n - x^*\|^2 + (\lambda_{n-1}L)^2\|y_{n-2} - y_{n-1}\|^2$ converges (hence it is bounded) and $\sum_{n\in\mathbb{N}} \left(\frac{1}{2} - 2(\lambda_n L)^2\right) \|y_n - y_{n-1}\|^2 < +\infty$. Moreover, from (5.13) we have

$$\sum_{n\in\mathbb{N}} \|y_n - y_{n-1}\|^2 < +\infty, \tag{5.15}$$

which implies $y_n - y_{n-1} \to 0$ as $n \to +\infty$. Recall from (5.10) that $\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - \|y_n - x_n\|^2 + (\lambda_n L)^2 \|y_{n-1} - y_n\|^2$. Since $\limsup_{n\to+\infty} \lambda_n < \frac{1}{2L}$ (and $0 < \liminf_{n\in\mathbb{N}} \lambda_n$) and (5.15) by using Lemma 2.1.3 again, we obtain

$$\sum_{n\in\mathbb{N}} \|y_n - x_n\|^2 < +\infty, \tag{5.16}$$

and $\lim_{n\to+\infty} \|x_n - x^*\|^2$ exists. Furthermore, we also obtain that $\lim_{n\to+\infty} \|y_n - x_n\|^2 = 0$.

In this part of the proof we will show the second part of Opial's Lemma. We proceed similarly as in [21, Theorem 2.1]. Let $\hat{x} \in \mathcal{H}$ be a weak sequential cluster point of x_n as $n \to +\infty$. Since $\lim_{n\to+\infty} \|x_n - y_n\| = 0$, we also have $y_n \rightharpoonup \hat{x}$ as $n \to +\infty$. Furthermore, since F is Lipschitz continuous, $\|F(y_{n-1}) - F(y_n)\| \to 0$ as $n \to +\infty$. We want to show that $\hat{x} \in \Omega$. We assume that $F(\hat{x}) \neq 0$, otherwise the conclusion follows automatically. Since $(y_n)_{n\geq 0} \subseteq C$, and C is weakly closed (because C is a closed convex set), then we have $\hat{x} \in C$. Let $n \in \mathbb{N}, n \geq 0$. From (5.4), for all $y \in C$,

$$\langle y - y_n, x_n - \lambda_n F(y_{n-1}) - y_n \rangle \leq 0,$$

or, equivalently,

$$\frac{1}{\lambda_n} \langle x_n - y_n, y - y_n \rangle \leq \langle F(y_{n-1}) - F(y_n), y - y_n \rangle + \langle F(y_n), y - y_n \rangle. \tag{5.17}$$

Considering the inequality (5.17) and taking into account that $\lim_{n\to+\infty} \|x_n - y_n\| = 0$, $\|F(y_{n-1}) - F(y_n)\| \to 0$ (as $n \to +\infty$), $(y_n)_{n\geq 0}$ is bounded and $\liminf_{n\to+\infty} \lambda_n > 0$, it follows

$$0 \leq \liminf_{n\to+\infty} \langle F(y_n), y - y_n \rangle, \quad \forall y \in C.$$

On the other hand, we have that $(y_n)_{n\geq 0}$ converges weakly to $\hat{x}$ as $n \to +\infty$. Since F is sequentially weak-to-weak continuous, $(F(y_n))_{n\geq 0}$ converges weakly to $F(\hat{x})$ as $n \to +\infty$. Because the norm mapping is convex and continuous, by Theorem 2.2.3 it is weakly sequentially lower semicontinuous. So we have $0 < \|F(\hat{x})\| \leq \liminf_{n\to+\infty} \|F(y_n)\|$. Then there exists $n_{-1} \geq 0$ such that $F(y_n) \neq 0$ for all $n \geq n_{-1}$. Let $(\varepsilon_k)_{k\geq 0}$ be a positive strictly decreasing sequence which converges to 0 as $k \to +\infty$ and $y \in C$. Since $\sup_{N\geq 0} \inf_{n\geq N} \langle F(y_n), y - y_n\rangle = \liminf_{n\to+\infty} \langle F(y_n), y - y_n\rangle > -\varepsilon_0$, there exits $N_0 \geq 0$ such that $\inf_{n\geq N_0} \langle F(y_n), y - y_n\rangle > -\varepsilon_0$. Taking $n_0 > \max\{N_0, n_{-1}\}$, we have $\langle F(y_{n_0}), y - y_{n_0}\rangle + \varepsilon_0 > 0$ and $F(y_{n_0}) \neq 0$. We can continue this construction inductively and assume to this end that $n_0 < n_1 < \cdots < n_k$ are given. Then there exists $N_{k+1} \geq 0$ such that $\inf_{n\geq N_{k+1}} \langle F(y_n), y - y_n\rangle > -\varepsilon_{k+1}$ $(> -\varepsilon_0)$. Taking $n_{k+1} > \max\{N_{k+1}, n_k\}$, we have

$$\langle F(y_{n_{k+1}}), y - y_{n_{k+1}}\rangle + \varepsilon_{k+1} \geq 0 \quad \text{and} \quad F(y_{n_{k+1}}) \neq 0.$$

In this way, we obtain a strictly increasing sequence $(n_k)_{k\geq 0}$ with the property that

$$\langle F(y_{n_k}), y - y_{n_k}\rangle + \varepsilon_k \geq 0 \text{ and } F(y_{n_k}) \neq 0,\ \forall k \geq 0. \tag{5.18}$$

Setting for every $k \geq 0$,

$$z_k := \frac{F(y_{n_k})}{\|F(y_{n_k})\|^2},$$

it holds that $\langle F(y_{n_k}), z_k\rangle = 1$. According to (5.18) we have that

$$\begin{aligned} 0 \leq \langle F(y_{n_k}), y - y_{n_k}\rangle + \varepsilon_k &= \langle F(y_{n_k}), y - y_{n_k}\rangle + \langle F(y_{n_k}), \varepsilon_k z_k\rangle \\ &= \langle F(y_{n_k}), y + \varepsilon_k z_k - y_{n_k}\rangle, \quad \forall k \geq 0\ \forall y \in C. \end{aligned}$$

Since F is pseudo-monotone on $\mathcal{H}$, it yields

$$\langle F(y + \varepsilon_k z_k), y + \varepsilon_k z_k - y_{n_k}\rangle \geq 0, \quad \forall k \geq 0. \tag{5.19}$$

Using that $(F(y_{n_k}))_{n\geq 0}$ is bounded (since $(F(y_n))_{n\geq 0}$ converges weakly to $F(\hat{x})$), we have

$$\lim_{k\to+\infty} \|\varepsilon_k z_k\| = \lim_{k\to+\infty} \frac{\varepsilon_k}{\|F(y_{n_k})\|} = 0.$$

Taking the limit in (5.19) as $k \to +\infty$, we obtain

$$\langle F(y), y - \hat{x}\rangle \geq 0, \quad \forall k \geq 0.$$

As y was arbitrarily chosen in C, if follows from Proposition 5.1.2 that $\hat{x} \in \Omega$. Hence, by Opial's lemma (Lemma 2.1.6), we can conclude that the sequence x_n converges weakly to a point in Ω.

□

Remark 5.2.2.

(i) Following the proof of Theorem 5.2.1, we can demonstrate the first part of Opial's lemma ($\lim_{n\to+\infty} \|x_n - x^*\|$ exists) by using the assumption that F is pseudo-monotone on C (not necessary on $\mathcal{H}$) and $\Omega \neq \emptyset$. Moreover, we also obtain from this part of the proof that $\sum_{n\in\mathbb{N}} \|y_n - y_{n-1}\|^2 < +\infty$ and $\sum_{n\in\mathbb{N}} \|y_n - x_n\|^2 < +\infty$.

(ii) For the second part, the sequentially weak-to-weak continuity and pseudo-monotonicity on $\mathcal{H}$ of F are essential to prove that every weak cluster point of $(x_n)_{n\geq 0}$ belongs to the solution set of $VI(F,C)$.

(iii) Notice from the proof of Theorem 5.2.1 that if we suppose D is an open set of $\mathcal{H}$ containing C, then from $y \in C \subset D$ there exists $\delta > 0$ such that $\mathcal{B}(y,\delta) \subset D$. Since $\varepsilon_k z_k \to 0$, then $y + \varepsilon_k z_k \to y \in C \subset D$ (as $k \to 0$). Hence, there is $k' > 0$ such that $y + \varepsilon_k z_k \in \mathcal{B}(y,\delta) \subset D, \forall k \geq k'$. Therefore, we can relax the assumption of F as a pseudo-monotone on D.

Remark 5.2.3 (adaptive stepsize strategy). On the other hand, when (an upper bound of) the Lipschitz constant of F is not available, we can use in our algorithm the following stepsize strategy, see also [21]:

$$\lambda_{n+1} := \begin{cases} \min\left\{\frac{\mu\|y_{n-1}-y_n\|}{\|F(y_{n-1})-F(y_n)\|}, \lambda_n\right\}, & \text{if } F(y_{n-1}) - F(y_n) \neq 0, \\ \lambda_n, & \text{otherwise,} \end{cases}$$

where $\mu \in (0, \frac{1}{2})$ and $\lambda_0 > 0$. The sequence $(\lambda_n)_{n\geq 0}$ is nonincreasing. If $F(y_{n-1}) - F(y_n) \neq 0$, for $n \geq 0$, then it holds

$$\frac{\mu\|y_{n-1} - y_n\|}{\|F(y_{n-1}) - F(y_n)\|} \geq \frac{\mu\|y_{n-1} - y_n\|}{L\|y_{n-1} - y_n\|} = \frac{\mu}{L},$$

which shows that $(\lambda_n)_{n\geq 0}$ is bounded from below by $\min\{\lambda_0, \frac{\mu}{L}\}$ (this means $\lim_{n\to+\infty} \lambda_n$ exits). Notice that, if $\lambda_0 \leq \frac{\mu}{L}$, then $(\lambda_n)_{n\geq 0}$ is a constant sequence, which leads to a fixed stepsize strategy. Consequently, the $\lim_{n\to+\infty} \lambda_n$ exists and it is a positive real number. We can adapt the proof of Theorem 5.2.1 to the new adaptive stepsize strategy. Indeed, from (5.9), we can write

$$\begin{aligned} \|x_{n+1} - x_n\|^2 &\overset{(5.9)}{=} \|y_n - x_n\|^2 + 2\lambda_n\langle y_n - x_n, F(y_{n-1}) - F(y_n)\rangle + \lambda_n^2\|F(y_{n-1}) - F(y_n)\|^2 \\ &\leq \|y_n - x_n\|^2 + 2\lambda_n\langle y_n - x_n, F(y_{n-1}) - F(y_n)\rangle + \frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2. \end{aligned}$$

Then it follows from (5.8) and the above inequality that

$$\begin{aligned} \|x_{n+1} - x^*\|^2 &\leq \|x_n - x^*\|^2 + \big[\|y_n - x_n\|^2 + 2\lambda_n\langle y_n - x_n, F(y_{n-1}) - F(y_n)\rangle \\ &\quad + \frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2\big] - 2\|x_n - y_n\|^2 + 2\lambda_n\langle x_n - y_n, F(y_{n-1}) - F(y_n)\rangle \\ &= \|x_n - x^*\|^2 - \|x_n - y_n\|^2 + \frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2. \end{aligned} \tag{5.20}$$

By Parallelogram identity, we know that $\|y_n - y_{n-1}\|^2 + \|(x_n - y_n) + (x_n - y_{n-1})\|^2 = 2\|x_n - y_n\|^2 + 2\|x_n - y_{n-1}\|^2$, then

$$\begin{aligned}
\|x_n - y_n\|^2 &\geq -\|x_n - y_{n-1}\|^2 + \frac{1}{2}\|y_n - y_{n-1}\|^2 \\
&\overset{(5.2)}{=} -\|y_{n-1} + \lambda_{n-1}(F(y_{n-2}) - F(y_{n-1})) - y_{n-1}\|^2 + \frac{1}{2}\|y_n - y_{n-1}\|^2 \\
&= -\|\lambda_{n-1}(F(y_{n-2}) - F(y_{n-1}))\|^2 + \frac{1}{2}\|y_n - y_{n-1}\|^2 \\
&\geq -\left(\frac{\lambda_{n-1}\mu}{\lambda_n}\right)^2 \|y_{n-2} - y_{n-1}\|^2 + \frac{1}{2}\|y_n - y_{n-1}\|^2. \qquad (5.21)
\end{aligned}$$

It follows from (5.20) and (5.21) that

$$\begin{aligned}
\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 &- \left[-\left(\frac{\lambda_{n-1}\mu}{\lambda_n}\right)^2 \|y_{n-2} - y_{n-1}\|^2 + \frac{1}{2}\|y_n - y_{n-1}\|^2\right] \\
&+ \frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2.
\end{aligned}$$

Thus

$$\|x_{n+1}-x^*\|^2+\left(\frac{1}{2}-\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2\right)\|y_n-y_{n-1}\|^2 \leq \|x_n-x^*\|^2+\left(\frac{\lambda_{n-1}\mu}{\lambda_n}\right)^2\|y_{n-2}-y_{n-1}\|^2.$$

Hence

$$\begin{aligned}
&\|x_{n+1}-x^*\|^2+\left(\frac{1}{2}-2\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2\right)\|y_n - y_{n-1}\|^2+\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2\|y_n-y_{n-1}\|^2 \\
&\leq \|x_n-x^*\|^2+\left(\frac{\lambda_{n-1}\mu}{\lambda_n}\right)^2\|y_{n-2} - y_{n-1}\|^2.
\end{aligned}$$

Since $\mu \in (0, \frac{1}{2})$, then

$$0 < \frac{1}{2} - 2\mu^2 = \lim_{n\to+\infty}\left(\frac{1}{2} - 2\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2\right).$$

This means that there exists $n_{00} \in \mathbb{N}$ such that $\frac{1}{2} - 2\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2 > \eta > 0$, $\forall n > n_{00}$ for some $\eta > 0$.

The proof of convergence is similar to the proof in Theorem 5.2.1. By using Lemma 2.1.3, we have that the sequence $(x_n)_{n\geq 0}$ is bounded and $\sum\limits_{n=2}^{\infty}\left(\frac{1}{2} - 2\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2\right)\|y_n - y_{n-1}\|^2 < +\infty$ and moreover $\sum_{n\in\mathbb{N}}\|y_n - y_{n-1}\|^2 < +\infty$. Now from (5.20), the fact that $\lim\limits_{n\to+\infty}\lambda_n$ exists and $\sum_{n\in\mathbb{N}}\|y_n - y_{n-1}\|^2 < +\infty$, by using Lemma 2.1.3 again, we obtain that $\sum_{n\in\mathbb{N}}\|y_n - x_n\|^2 < +\infty$. Furthermore, we also get that $\lim\limits_{n\to+\infty}\|x^* - x_n\|^2 = 0$. The rest of the proof is similar to one of Theorem 5.2.1.

Next, we will show that the convergence result in Theorem 5.2.1 holds in finite dimensional spaces under a weaker assumption: F is pseudo-monotone only on $C(\subset \mathcal{H})$ instead of $\mathcal{H}$.

Theorem 5.2.4. *Let $\mathcal{H}$ be a finite-dimensional real Hilbert space. Assume that the solution set Ω is nonempty, F is pseudo-monotone on C and Lipschitz continuous with constant $L > 0$, and $0 < \liminf_{n\to+\infty} \lambda_n \leq \limsup_{n\to+\infty} \lambda_n < \frac{1}{2L}$. Then the sequence $(x_n)_{n\geq 0}$ generated by (5.2) converges to a solution of $VI(F, C)$.*

Proof. Let $x^* \in \Omega$ be fixed. The first part of Opial's lemma follows directly from Remark 5.2.2 (i). Thus we have $\lim_{n\to+\infty} \|x_n - x^*\|$ exists. In addition, we have also that $\sum_{n\in\mathbb{N}} \|y_n - x_n\|^2 < +\infty$, hence $\lim_{n\to+\infty} \|y_n - x_n\| = 0$ (see also (5.16)). Let us prove the second part of Opial's lemma. Let $\hat{x}$ be a cluster point of $(x_n)_{n\geq 0}$. Then there exists a subsequence $(x_{n_k})_{k\geq 0}$ of $(x_n)_{n\geq 0}$, which converges to $\hat{x}$ as $k \to +\infty$. Since $\lim_{n\to+\infty} \|y_n - x_n\| = 0$, then $(y_{n_k})_{k\geq 0}$ also converges to $\hat{x}$ as $k \to +\infty$. Let $y \in C$ be fixed. It follows from (5.4) that

$$\langle y - y_{n_k}, y_{n_k} - x_{n_k} + \lambda_{n_k} F(y_{n_k - 1})\rangle \geq 0, \quad \forall k \geq 0. \tag{5.22}$$

Because the sequence $(\lambda_{n_k})_{k\geq 0}$ is bounded, it has a subsequence which converges to $\tilde{\lambda} > 0$ (since $0 < \liminf_{n\to+\infty} \lambda_n$). Taking the limit along this subsequence in (5.22) and using that F is continuous, we obtain

$$\langle y - \hat{x}, F(\hat{x})\rangle \geq 0. \tag{5.23}$$

Since $y \in C$ was chosen arbitrarily, it follows that $\hat{x}$ is a solution of $VI(F, C)$. Now we can apply the Opial's Lemma (Lemma 2.1.6) to complete the proof. $\square$

5.3 Numerical Experiments for Pseudo-Monotone Variational Inequalities

In this part, we consider a numerical experiment which is carried out in order to compare the classical Tseng's algorithm (FBF) and our algorithm (FBF-EP) for solving pseudo-monotone variational inequalities. We implemented the numerical codes in MATLAB and performed all computations on a Windows desktop with an Intel(R) Core(TM) i5-8250U processor at 1.60 GHz up to 1.8 GHz and RAM of 8 GB. In this experiment, we considered variational inequalities governed by a pseudo-monotone operator, which is not monotone.

Example 5.3.1.

We follow the construction of the experiment in [21] and give our example in a higher dimension (10 and 20 dimensions). It will be shown as below.

We consider the $VI(F, C)$ with

$$C = \left\{ \mathbf{x} \in \mathbb{R}^m \;\middle|\; \sum_{i=1}^{m} x_i \leq 5, 0 \leq x_i \leq 5, \forall i = 1, \ldots, m \right\}$$

and

$$F : \mathbb{R}^m \to \mathbb{R}^m, F(\mathbf{x}) = (e^{-\|\mathbf{x}\|^2} + \alpha)(\mathbf{M}\mathbf{x} + \mathbf{p}),$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^m$, $\alpha = 0.1$, $\mathbf{p}$ is a given vector in $\mathbb{R}^m$, $t \mapsto e^t$ is the exponential function and choose the matrix $\mathbf{M}$ as: for $m = 10$, we pick

$$\mathbf{M_{10}} := \begin{bmatrix} 25 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 \\ -5 & 37 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 \\ 10 & -8 & 14 & -3 & 7 & 1 & 3 & 6 & 14 & 9 \\ 0 & 18 & -3 & 34 & 0 & -2 & 10 & 21 & 2 & 0 \\ 10 & -2 & 7 & 0 & 21 & 6 & 17 & 0 & 7 & 11 \\ 5 & 5 & 1 & -2 & 6 & 5 & 6 & -1 & 1 & 1 \\ 15 & -3 & 3 & 10 & 17 & 6 & 31 & 2 & 7 & 3 \\ 5 & 11 & 6 & 21 & 0 & -1 & 2 & 29 & 15 & 6 \\ 10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 56 & 10 \\ 0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 41 \end{bmatrix},$$

and for $m = 20$, we choose

$$\mathbf{M_{20}}$$

$$:= \begin{bmatrix} 25 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 & 25 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 \\ -5 & 37 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 & -5 & 37 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 \\ 10 & -8 & 14 & -3 & 7 & 1 & 3 & 6 & 14 & 9 & 10 & -8 & 14 & -3 & 7 & 1 & 3 & 6 & 14 & 9 \\ 0 & 18 & -3 & 34 & 0 & -2 & 10 & 21 & 2 & 0 & 0 & 18 & -3 & 34 & 0 & -2 & 10 & 21 & 2 & 0 \\ 10 & -2 & 7 & 0 & 21 & 6 & 17 & 0 & 7 & 11 & 10 & -2 & 7 & 0 & 21 & 6 & 17 & 0 & 7 & 11 \\ 5 & 5 & 1 & -2 & 6 & 5 & 6 & -1 & 1 & 1 & 5 & 5 & 1 & 2 & 6 & 5 & 6 & -1 & 1 & 1 \\ 15 & -3 & 3 & 10 & 17 & 6 & 31 & 2 & 7 & 3 & 15 & -3 & 3 & 10 & 17 & 6 & 31 & 2 & 7 & 3 \\ 5 & 11 & 6 & 21 & 0 & -1 & 2 & 29 & 15 & 6 & 5 & 11 & 6 & 21 & 0 & -1 & 2 & 29 & 15 & 6 \\ 10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 56 & 10 & 10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 56 & 10 \\ 0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 41 & 0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 41 \\ 25 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 & 41 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 \\ -5 & 37 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 & -5 & 46 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 \\ 10 & -8 & 14 & -3 & 7 & 1 & 3 & 6 & 14 & 9 & 10 & -8 & 18 & -3 & 7 & 1 & 3 & 6 & 14 & 9 \\ 0 & 18 & -3 & 34 & 0 & -2 & 10 & 21 & 2 & 0 & 0 & 18 & -3 & 35 & 0 & -2 & 10 & 21 & 2 & 0 \\ 10 & -2 & 7 & 0 & 21 & 6 & 17 & 0 & 7 & 11 & 10 & -2 & 7 & 0 & 25 & 6 & 17 & 0 & 7 & 11 \\ 5 & 5 & 1 & -2 & 6 & 5 & 6 & -1 & 1 & 1 & 5 & 5 & 1 & -2 & 6 & 14 & 6 & -1 & 1 & 1 \\ 15 & -3 & 3 & 10 & 17 & 6 & 31 & 2 & 7 & 3 & 15 & -3 & 3 & 10 & 17 & 6 & 47 & 2 & 7 & 3 \\ 5 & 11 & 6 & 21 & 0 & -1 & 2 & 29 & 15 & 6 & 5 & 11 & 6 & 21 & 0 & -1 & 2 & 54 & 15 & 6 \\ 10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 56 & 10 & 10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 72 & 10 \\ 0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 41 & 0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 50 \end{bmatrix},$$

which are positive definite matrices (the matrices were constructed from the upper triangular matrices in both 10 and 20 dimensions and followed Theorem 8.3.3 in [65]). In general, the operator F is not monotone (see Bianchi et al. [13]). Moreover, we can show that this operator is pseudo-monotone (see Boţ et al. [21]), i.e., for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ such that $\langle F(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \geq 0$ and since $g(\mathbf{x}) := e^{-\|\mathbf{x}\|^2} \geq 0$ then we have $\langle \mathbf{Mx} + \mathbf{p}, \mathbf{y} - \mathbf{x}\rangle \geq 0$ and thus

$$\begin{aligned} \langle \mathbf{F}(\mathbf{y}), \mathbf{y} - \mathbf{x}\rangle &= g(\mathbf{y})\langle \mathbf{My} + \mathbf{p}, \mathbf{y} - \mathbf{x}\rangle \\ &\geq g(\mathbf{y})(\langle \mathbf{My} + \mathbf{p}, \mathbf{y} - \mathbf{x}\rangle - \langle \mathbf{Mx} + \mathbf{p}, \mathbf{y} - \mathbf{x}\rangle) \\ &= g(\mathbf{y})(\langle \mathbf{M}(\mathbf{y} - \mathbf{x}) + \mathbf{p}, \mathbf{y} - \mathbf{x}\rangle) \geq 0. \end{aligned}$$

We computed the unique solution $\mathbf{x}^*$ (x^* is unique since F is strongly pseudo-monotone, see [21, Example 2.1 and Example 2.2]) of the variational inequality $VI(F, C)$ by running 10000 iterations of Tseng's algorithm for all $n \geq 0$ and stepsize $\lambda_n = \frac{0.49}{L}$.

In the first trial, we give $\mathbf{p} = \bar{\mathbf{1}}_m$, a vector in $\mathbb{R}^m$ which all elements are equal to one. We compared the performances of the Tseng's algorithm (FBF) and the Tseng's algorithm with extrapolation (FBF-EP) by considering the random initial points

For 10 dimensions:
$\mathbf{x_0}^{10} = (-4, 1, 8, -9, 0, -1, 8, 3, 10, 2)^T$,
$\mathbf{y}_{-1}^{10} = \hat{\mathbf{y}}_{-1}^{10} := (8, -10, -8, 5, -2, -2, 2, -10, -2, -9)^T$,

For 20 dimensions:
$\mathbf{x_0}^{20} = (-5, 1, 3, -9, 0, -1, 8, -5, 3, -2, -1, 0, 2, -8, 4, 0, -3, -10, 1, 2)^T$,
$\mathbf{y}_{-1}^{20} = \hat{\mathbf{y}}_{-1}^{20} := (11, -2, 10, 7, -8, 4, -6, 7, -1, 10, -17, 9, 13, -1, 0, 3, 12, -8, 9, 15)^T$,

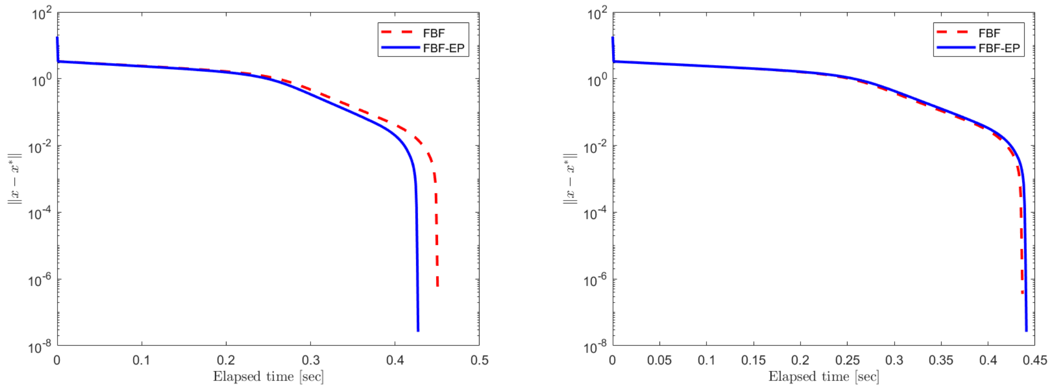
and $\|\mathbf{x_n} - \mathbf{x}^*\| \leq 10^{-6}$ as stopping criterion. The projection on C was computed by using the quadprog function in MATLAB. Figure 5.1, Table 5.1 and Figure 5.2, Table 5.2 show that at the first implementation of this trial for 10 dimensions and 20 dimensions, respectively. We can see that our algorithm (Tseng's algorithm with extrapolation from the past or FBF-EP) spends less time than the classical Tseng's algorithm (FBF), whereas at the second implementation it is not the case. Therefore, one may see that our algorithm is quite sensitive with respect to computer errors or engine computational power.

Table 5.1: [10 dimensions] The table of the performances for 2 attempts of FBF and FBF-EP with $\mathbf{x_0}^{10} = (-4, 1, 8, -9, 0, -1, 8, 3, 10, 2)^T$ and $\mathbf{y}_{-1}^{10} = \hat{\mathbf{y}}_{-1}^{10}$ (when the stepsize is $\lambda_n = \frac{0.49}{L}$).

Attempt	FBF			FBF-EP		
	$\|\mathbf{x_n} - \mathbf{x}^*\|$	No. iter	CPU-time	$\|\mathbf{x_n} - \mathbf{x}^*\|$	No. iter	CPU-time
1	3.567×10^{-7}	464	0.44943	2.6077×10^{-7}	456	0.42628
2	3.567×10^{-7}	464	0.43565	2.6077×10^{-7}	456	0.43983

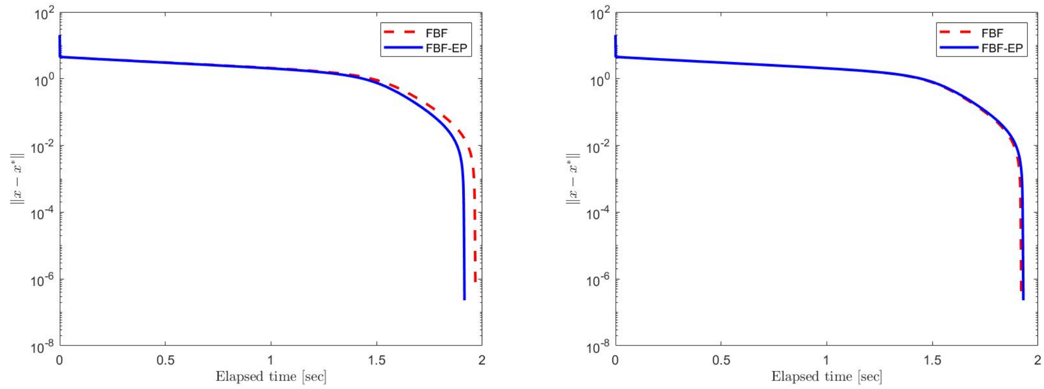
Table 5.2: [20 dimensions] The table of the performances for 2 attempts of FBF and FBF-EP with $\mathbf{x_0}^{20} = (-5, 1, 3, -9, 0, -1, 8, -5, 3, -2, -1, 0, 2, -8, 4, 0, -3, -10, 1, 2)^T$ and $\mathbf{y}_{-1}^{20} = \hat{\mathbf{y}}_{-1}^{20}$ (when the stepsize is $\lambda_n = \frac{0.49}{L}$).

Attempt	FBF			FBF-EP		
	$\lVert\mathbf{x_n} - \mathbf{x}^*\rVert$	No. iter	CPU-time	$\lVert\mathbf{x_n} - \mathbf{x}^*\rVert$	No. iter	CPU-time
1	4.0665×10^{-7}	1976	1.9663	2.295×10^{-7}	1973	1.9147
2	4.0665×10^{-7}	1976	1.9194	2.295×10^{-7}	1973	1.9293



(a) First execution (b) Second execution

Figure 5.1: [10 dimensions] A figure of two graphs for comparison between the classical Tseng's algorithm (FBF) and Tseng's algorithm with extrapolation from the past (FBF-EP) for two different executions with $\mathbf{x_0}^{10} = (-4, 1, 8, -9, 0, -1, 8, 3, 10, 2)^T$ and $\mathbf{y}_{-1}^{10} = \hat{\mathbf{y}}_{-1}^{10}$ (when the stepsize is $\lambda_n = \frac{0.49}{L}$).



(a) First execution

(b) Second execution

Figure 5.2: [20 dimensions] A figure of two graphs for comparison between the classical Tseng's algorithm (FBF) and Tseng's algorithm with extrapolation from the past (FBF-EP) for two different executions with $\mathbf{x_0}^{20} = (-5,1,3,-9,0,-1,8,-5,3,-2,-1,0,2,-8,4,0,-3,-10,1,2)^T$ and $\mathbf{y}_{-1}^{20} = \hat{\mathbf{y}}_{-1}^{20}$ (when the stepsize is $\lambda_n = \frac{0.49}{L}$).

In order to confirm the effectiveness of our method, we designed the experiment as follows. We randomise an initial vector $\mathbf{x_0}$ in which each coordinate is an integer shuffled from the interval $[-10, 10]$, denoted by $\text{rand}_{[-10,10]}$ (in MATLAB). We also select $\mathbf{y}_{-1}$ to be $\mathbf{x_0}$, $\hat{\mathbf{y}}_{-1}$ and $\text{rand}_{[-10,10]}$, and put vector $\mathbf{p}$ (from the definition of F) as following vectors:

For 10 dimensions:

$\mathbf{p_1}^{10} = \bar{\mathbf{1}}_{10}$;

$\mathbf{p_2}^{10} = (1,2,1,0,-1,2,0,1,-1,2)^T$,

$\mathbf{p_3}^{10} = (5,1,0,-2,-1,0,-5,4,-5,-1)^T$,

For 20 dimensions:

$\mathbf{p_1}^{20} = \bar{\mathbf{1}}_{20}$;

$\mathbf{p_2}^{20} = (1,-2,1,2,-1,2,0,1,-1,2,1,-2,1,0,-1,2,0,1,-1,2)^T$;

$\mathbf{p_3}^{20} = (5,1,0,-2,-1,0,-5,4,-5,-1,5,1,0,-2,-1,0,-5,4,-5,-1)^T$.

Then, we run 100 executions for FBF and FBF-EP algorithms. Finally, we collect the achievements when FBF-EP takes less time than FBF and call this value "satisfaction". As a result, we could deduce from Table 5.3 (and Table 5.4) that the FBF-EP algorithm overcomes the FBF algorithm by more than half of 100 executions. In 10 dimensions, the highest satisfaction value is 70 (62 in 20 dimensions), and the lowest is 57 (51 in 20 dimensions) for this experiment.

Table 5.3: The table shows the result of satisfaction after 100 executions in 10 dimensions by choosing $\mathbf{x_0} = \text{rand}_{[-10,10]}$ (for 10 dimensions), $\mathbf{y}_{-1} = \mathbf{x_0}^{10}$, $\hat{\mathbf{y}}^{10}_{-1}$, $\text{rand}_{[-10,10]}$ (for 10 dimensions) and $\mathbf{p} = \mathbf{p_1}^{10}, \mathbf{p_2}^{10}$ and $\mathbf{p_2}^{10}$(when the stepsize is $\lambda_n = \frac{0.49}{L}$)

$\mathbf{x_0}$	$\mathbf{y}_{-1}$	vector $\mathbf{p}$	satisfaction[1]
$\text{rand}_{[-10,10]}$	$\mathbf{x_0}^{10}$		69 out of 100
$\text{rand}_{[-10,10]}$	$\hat{\mathbf{y}}^{10}_{-1}$	$\mathbf{p_1}^{10}$	65 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		66 out of 100
$\text{rand}_{[-10,10]}$	$\mathbf{x_0}^{10}$		66 out of 100
$\text{rand}_{[-10,10]}$	$\hat{\mathbf{y}}^{10}_{-1}$	$\mathbf{p_2}^{10}$	70 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		57 out of 100
$\text{rand}_{[-10,10]}$	$\mathbf{x_0}^{10}$		61 out of 100
$\text{rand}_{[-10,10]}$	$\hat{\mathbf{y}}^{10}_{-1}$	$\mathbf{p_3}^{10}$	60 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		60 out of 100

[1] The number of executions in which the CPU time for Tseng-EP algorithm is less than the Tseng algorithm, within 100 executions.

Table 5.4: The table shows the result of satisfaction after 100 executions in 20 dimensions by choosing $\mathbf{x_0} = \text{rand}_{[-10,10]}$ (for 20 dimensions), $\mathbf{y}_{-1} = \mathbf{x_0}^{20}$, $\hat{\mathbf{y}}^{20}_{-1}$ and $\text{rand}_{[-10,10]}$ (for 20 dimensions), and $\mathbf{p} = \mathbf{p_1}^{20}, \mathbf{p_2}^{20}, \mathbf{p_3}^{20}$ (when the stepsize is $\lambda_n = \frac{0.49}{L}$).

$\mathbf{x_0}$	$\mathbf{y}_{-1}$	vector $\mathbf{p}$	satisfaction[1]
$\text{rand}_{[-10,10]}$	$\mathbf{x_0}^{20}$		62 out of 100
$\text{rand}_{[-10,10]}$	$\hat{\mathbf{y}}^{20}_{-1}$	$\mathbf{p_1}^{20}$	61 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		56 out of 100
$\text{rand}_{[-10,10]}$	$\mathbf{x_0}^{20}$		51 out of 100
$\text{rand}_{[-10,10]}$	$\hat{\mathbf{y}}^{20}_{-1}$	$\mathbf{p_2}^{20}$	53 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		53 out of 100
$\text{rand}_{[-10,10]}$	$\mathbf{x_0}^{20}$		58 out of 100
$\text{rand}_{[-10,10]}$	$\hat{\mathbf{y}}^{20}_{-1}$	$\mathbf{p_3}^{20}$	55 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		51 out of 100

[1] The number of executions in which the CPU time for Tseng-EP algorithm is less than the Tseng algorithm, within 100 executions.

We show next another example in infinite dimensional Hilbert space. We consider our proposed algorithm for solving the variational inequality problem in L^2-space as below.

Example 5.3.2.

Let $\mathcal{H} = L^2([0,1])$ with norm $\|x\| := \left(\int_0^1 |x(t)|^2 dt\right)^{\frac{1}{2}}$ and inner product $\langle x, y\rangle := \int_0^1 x(t)y(t)dt$, $\forall x, y \in \mathcal{H}$. Define an operator $F : \mathcal{H} \to \mathcal{H}$ by (see [86, Example 5.3])

$$(Fx)(t) = \max\{x(t), 0\}, \quad x \in \mathcal{H}, \quad t \in [0,1]. \tag{5.24}$$

It is easy to show that F is monotone (hence, pseudo-monotone) and Lipschitz continuous with Lipschitz constant $L = 1$. We provide the feasible set as the ball $C := \{x \in \mathcal{H} \mid \|x\| \leq 2\}$. To implement, we consider either the FBF or FBF-EP methods, including their adaptive stepsize strategy (see Remark 5.2.3). We represent FBF and FBF-EP with adaptive stepsize strategy by aFBF and aFBF-EP, respectively, and the stopping criterion is $\|x_n - x_{n-1}\| \leq \varepsilon$ with $\varepsilon = 10^5$. Note that the aFBF algorithm is obtained from [21]. The parameter values of the algorithm in Example 5.3.2 are determined as follows.

Case I.

FBF : $\mathbf{x_0} = t^3, \lambda_n = 0.49;$

FBF-EP (our proposed algorithm) : $\mathbf{x_0} = \mathbf{y_{-1}} = t^3, \lambda_n = 0.49;$

aFBF [21] : $\mathbf{x_0} = t^3, \mu = 0.49, \lambda_0 = 0.7;$

aFBF-EP(our proposed algorithm) : $\mathbf{x_0} = \mathbf{y_{-1}} = t^3, \mu = 0.49, \lambda_0 = 0.7.$

Case II.

FBF : $\mathbf{x_0} = \dfrac{\cos(-3t) + \sin(-10)}{200}, \lambda_n = 0.49;$

FBF-EP (our proposed algorithm) : $\mathbf{x_0} = \mathbf{y_{-1}} = \dfrac{\cos(-3t) + \sin(-10)}{200}, \lambda_n = 0.49;$

aFBF [21] : $\mathbf{x_0} = \dfrac{\cos(-3t) + \sin(-10)}{200}, \mu = 0.49, \lambda_0 = 0.7;$

aFBF-EP(our proposed algorithm) : $\mathbf{x_0} = \mathbf{y_{-1}} = \dfrac{\cos(-3t) + \sin(-10)}{200}, \mu = 0.49, \lambda_0 = 0.7.$

Case III.

FBF : $\mathbf{x_0} = (10t^3 - 3t^2)/20, \lambda_n = 0.49;$

FBF-EP (our proposed algorithm) : $\mathbf{x_0} = \mathbf{y_{-1}} = (10t^3 - 3t^2)/20, \lambda_n = 0.49;$

aFBF [21] : $\mathbf{x_0} = (10t^3 - 3t^2)/20, \mu = 0.49, \lambda_0 = 0.7;$

aFBF-EP(our proposed algorithm) : $\mathbf{x_0} = \mathbf{y_{-1}} = (10t^3 - 3t^2)/20, \mu = 0.49, \lambda_0 = 0.7.$

Case IV.
In this case, we use the same values of starting points $(x_0, y{-1})$, μ and λ_n in the methods without adaptive stepsize strategy as in Case III, and we replace instead $\lambda_0 = 0.9$ in the adaptive stepsize scheme.

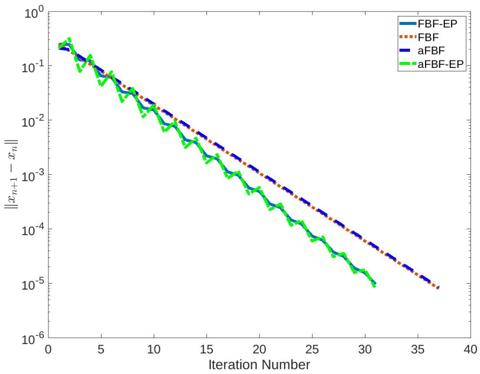
Table 5.5: The result of computation in Example 5.3.2

Algorithms	Case I			Case II		
	$\|\mathbf{x_{n+1}} - \mathbf{x_n}\|$	No. iter	CPU-time	$\|\mathbf{x_{n+1}} - \mathbf{x_n}\|$	No. iter	CPU-time
FBF	7.980980×10^{-6}	38	15.9776	8.472908×10^{-6}	20	79.437
FBF-EP*	9.608509×10^{-6}	32	7.9064	6.706328×10^{-6}	18	53.6874
aFBF	8.405511×10^{-6}	38	18.2491	8.923607×10^{-6}	20	85.596
aFBF-EP**	8.08286×10^{-6}	32	9.7161	9.771265×10^{-6}	16	47.0231
Algorithms	**Case III**			**Case IV**		
	$\|\mathbf{x_{n+1}} - \mathbf{x_n}\|$	No. iter	CPU-time	$\|\mathbf{x_{n+1}} - \mathbf{x_n}\|$	No. iter	CPU-time
FBF	8.823636×10^{-6}	34	14.9177	8.823636×10^{-7}	34	14.7945
FBF-EP*	6.640949×10^{-6}	30	7.59	6.640949×10^{-6}	30	8.0234
aFBF	9.292991×10^{-6}	34	16.8252	8.029509×10^{-6}	35	17.3107
aFBF-EP**	5.520665×10^{-6}	30	9.4287	9.317460×10^{-6}	90	28.3885

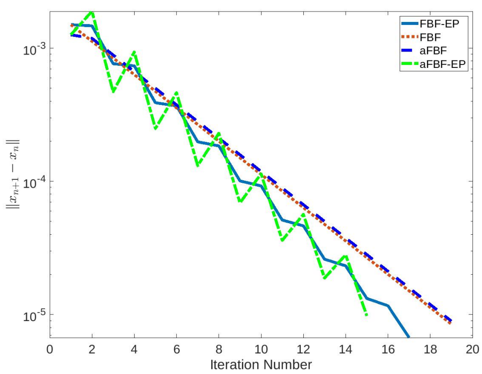
* our proposed algorithm

** our proposed algorithm with adaptive stepsize strategy

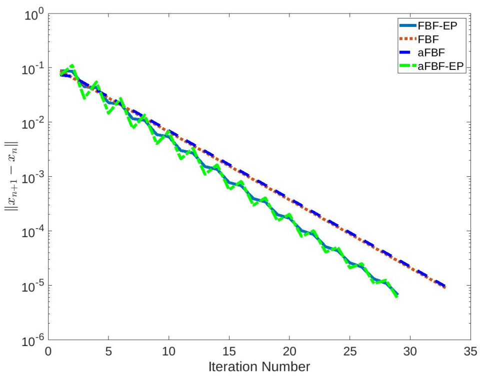
Table 5.5 and Figure 5.3 depict the results of the experiment by showing $\|x_{n+1} - x_n\|$, the number of iteration (No. iter), CPU-time (second) in the table, and plotting between $\|x_{n+1} - x_n\|$ and their iteration numbers in different cases. One can notice that our proposed algorithm, FBF-EP, is always faster than FBF method, which sometimes consumes twice the CPU-time more significantly than the CPU-time of FBF-EP. Further, the number of iterations of FBF-EP is less than the iteration numbers of FBF. For this numerical experiment, we know that the Lipschitz constant L equals to 1 ($L = 1$). Therefore, we have to choose $\lambda_n < \frac{1}{2L}(= \frac{1}{2}$, because $L = 1)$ for the FBF-EP method (see Theorem 5.2.1). Nevertheless, the constant L might not figure out in certain cases. Then, we can use the adaptive stepsize strategy described in Remark 5.2.3 with $\mu = 0.49 < \frac{1}{2L}$ and $\lambda_0 > 0$. We have seen that the starting stepsize λ_0 have a significant effect in both aFBF-EP and aFBF. If the stepsize is a suitable one, it can reduce the CPU-time and turn the method to the fastest one, for example, the FBF-EP in Case II; however, since the time of the process to finding λ_n in the next iteration of the adaptive stepsize strategy has been counted in computation then the method with adaptive stepsize strategy can give the time performance, which is slower than the normal method without the adaptive stepsize strategy. In addition, the adaptive stepsize strategy can provide a small stepsize, which then impacts the number of iterations of the iterative method shown in Case IV.



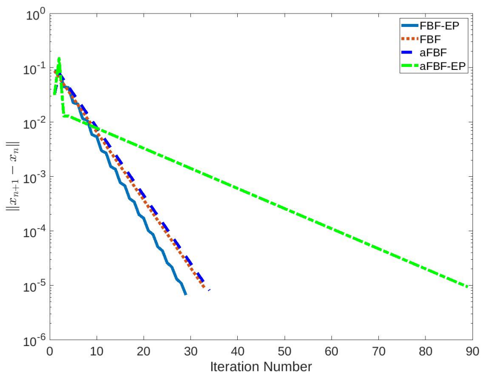
(a) Case I



(b) Case II



(c) Case III



(d) Case IV

Figure 5.3: The graphs plot the value between $\|x_{n+1} - x_n\|$ and the iteration number of the experiments in Case I, Case II, Case III and Case IV.

Bibliography

[1] S. Adly, B. K. Le, *Douglas–Rachford splitting algorithm for solving state-dependent maximal monotone inclusions*. Optimization Letters 15, 2861–2878, DOI: 10.1007/s11590-021-01718-z, 2021.

[2] F. J. Aragón-Artacho, R. I. Boț, D. Torregrosa-Belén, *A primal-dual splitting algorithm for composite monotone inclusions with minimal lifting*, Numerical Algorithms 93, DOI: 10.1007/s11075-022-01405-9, 103–130, 2023.

[3] K. J. Arrow, L. Hurwicz, H. Uzawa, *Studies in Linear and Nonlinear Programming*, Moscow, Foreign Languages Press, 1962.

[4] H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*. Journal of Differential Equations 248, 1315–1344, DOI: 10.1016/j.jde.2009.06.014, 2010.

[5] H. Attouch, M.-O. Czarnecki, J. Peypouquet, *Coupling Forward-Backward with Penalty Schemes and Parallel Splitting for Constrained Variational Inequalities.* SIAM Journal on Optimization 21(4), 1251–1274, DOI: 10.1137/110820300, 2011.

[6] H. Attouch, M.-O. Czarnecki, J. Peypouquet, *Prox-Penalization and Splitting Methods for Constrained Variational Problems*, SIAM Journal on Optimization 21, 149–173, DOI: 10.1137/100789464, 2011.

[7] S. Banert, R. I. Boț, *Backward Penalty Schemes for Monotone Inclusion Problems*, Journal of Optimization Theory and Applications 166, 930–948, DOI: 10.1007/s10957-014-0700-x, 2015.

[8] H. H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, DOI: 10.1007/978-3-319-48311-5, 2011.

[9] H. H. Bauschke, D. A. McLaren, H. S. Sendov, *Fitzpatrick Functions: Inequalities, examples, and remarks on a problem by S. Fitzpatrick*, Journal of Convex Analysis 13, 499–523, 2006.

[10] S. Becker, J. Fadili, *A quasi-Newton proximal splitting method*, Advances in Neural Information Processing Systems 25, 2012.

[11] A. Ben-Tal, T. Margalit, A. Nemirovski A., *The ordered subsets mirror descent optimization method with applications to tomography*, SIAM Journal on Optimization 12(1), 79–108, DOI: 10.1137/S1052623499354564, 2001.

[12] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, DOI: 10.1016/C2013-0-10366-2, 1982.

[13] M. Bianchi, N. Hadjisavvas, S. Schaible, *On pseudomonotone maps t for which -t is also pseudomonotone*, Journal of Convex Analysis 10(1), 149–168, 2003.

[14] A. Böhm, M. Sedlmayer, E. R. Csetnek, R. I. Boţ, *Two Steps at a Time—Taking GAN Training in Stride with Tseng's Method*, SIAM Journal on Mathematics of Data Science 4(2), 750–771, DOI: 10.1137/21M1420939, 2022.

[15] R. I. Boţ, *Conjugate duality in convex optimization*, Lecture Notes in Economics and Mathematical Systems. Springer Berlin Heidelberg, DOI: 10.1007/978-3-642-04900-2, 2010.

[16] R. I. Boţ, E. R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numerical Algorithms 71(3), 519–540, DOI: 10.1007/s11075-015-0007-5, 2016.

[17] R. I. Boţ, E. R. Csetnek, *A Tseng's Type Penalty Scheme for Solving Inclusion Problems Involving Linearly Composed and Parallel-Sum Type Monotone Operators.* Vietnam Journal of Mathematics 42, 451–465, DOI: 10.1007/s10013-013-0050-2, 2013.

[18] R. I. Boţ, E. R. Csetnek, *Forward-Backward and Tseng's Type Penalty Schemes for Monotone Inclusion Problems.* Set-Valued and Variational Analysis 22, 313–331, DOI: 10.1007/s11228-014-0274-7, 2014.

[19] R. I. Boţ, E. R. Csetnek, *On the convergence rate of a forward-backward type primal-dual splitting algorithm for convex optimization problems*, Optimization 64(1), 5–23, DOI: 10.1080/02331934.2014.966306, 2015.

[20] R. I. Boţ, E. R. Csetnek, N. Nimana, *Gradient-type penalty method with inertial effects for solving constrained convex optimization problems with smooth data*, Optimization Letters 12, 17–33, DOI: 10.1007/s11590-017-1158-1, 2017.

[21] R. I. Boţ, E. R. Csetnek, P. T. Vuong, *The forward-backward-forward method from continuous and discrete perspective for pseudo-monotone variational inequalities in Hilbert spaces*, European Journal of Operational Research 287(1), 49–60, DOI: 10.1016/j.ejor.2020.04.035, 2020.

[22] R. I. Boţ, M. Sedlmayer, P. T. Vuong, *A relaxed inertial forward-backward-forward algorithm for solving monotone inclusions with application to GANs*, Journal of Machine Learning Research 24(1), 1–37, 2023.

[23] L. M. Briceño-Arias, P. L. Combettes, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM Journal on Optimization 21(4), 1230–1250, DOI: 10.1137/10081602X, 2011.

[24] J. V. Burke, M. Qian, *A variable metric proximal point algorithm for monotone operators*, SIAM Journal on Control and Optimization 37(2), 353–375, DOI: 10.1137/S0363012992235547, 1999.

[25] J. V. Burke, M. Qian, *On the superlinear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating*, Mathematical Programming 88(1), 157–181, DOI: 10.1007/PL00011373, 2000.

[26] Y. Censor, A. Gibali, S. Reich, *The subgradient extragradient method for solving variational inequalities in Hilbert space*, Journal of Optimization Theory and Applications 148(2), 318–335, DOI: 10.1007/s10957-010-9757-3, 2011.

[27] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical imaging and vision 20(1), 89–97, DOI: 10.1023/B:JMIV.0000011325.36760.1e, 2004.

[28] G. Chantas, N. Galatsanos, A. Likas, M. Saunders, *Variational bayesian image restoration based on a product of t-distributions image prior*, IEEE Transactions on Image Processing 17(10), 1795–1805, DOI: 10.1109/TIP.2008.2002828, 2008.

[29] F. Chorobura, I. Necoara, J.-C. Pesquet, *An adaptive forward-backward-forward splitting algorithm for solving pseudo-monotone inclusions*, arXiv:2503.01070v1, 2025.

[30] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms* Studies in Computational Mathematics 8, 115–52, DOI: 10.1016/S1570-579X(01)80010-0, 2001.

[31] P. L. Combettes, J.-C. Pesquet, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators*, Set-Valued and Variational Analysis Article 20, 307–330, DOI: 10.1007/s11228-011-0191-y, 2012.

[32] P. L. Combettes, B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization 63(9), 1289–1318, DOI: 10.1080/02331934.2012.733883, 2014.

[33] P. L. Combettes, B. C. Vũ, *Variable metric quasi-Fejér monotonicity.* Nonlinear Analysis: Theory, Methods & Applications 78, 17–31, DOI: 10.1016/j.na.2012.09.008, 2013.

[34] R. W. Cottle, J. C. Yao, *Pseudo-monotone complementarity problems in Hilbert space*, Journal of Optimization Theory and Applications 75, 281–295, DOI: 10.1007/BF00941468, 1992.

[35] E. R. Csetnek, Y. Malitsky, M. K. Tam, *Shadow Douglas-Rachford splitting for monotone inclusions*, Applied Mathematics & Optimization 80(3), 665–678, DOI: 10.1007/s00245-019-09597-8, 2019.

[36] C. Daskalakis, A. Ilyas, V. Syrgkanis, H. Zeng, *Training gans with optimism*, arXiv:1711.00141, 2017.

[37] C. Daskalakis, I. Panageas, *The Limit Points of (Optimistic) Gradient Descent in Min-Max Optimization*, Neural Information Processing Systems, 2018.

[38] J. Eckstein, D. P. Bertsekas, *On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming 55, 293–318, DOI: 10.1007/BF01581204, 1992.

[39] J. Eckstein, M. C. Ferris, *Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control*, INFORMS Journal on Computing 10(2), 121–260, DOI: 10.1287/ijoc.10.2.218, 1998.

[40] J. Eckstein, M. Fukushima, *Some reformulations and applications of the alternating direction method of multipliers, in Large Scale Optimization: State of the Art*, W. W. Hager, D. W. Hearn, and P. M. Pardalos, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 115–134, DOI: 10.1007/978-1-4613-3632-7_7, 1994.

[41] F. Facchinei, J.-S. Pang (Eds.), *Finite-dimensional variational inequalities and complementarity problems Volume II*, Springer, New York, DOI: 10.1007/b97543, 2003.

[42] S. Fitzpatrick, *Representing monotone operators by convex functions.* In Workshop/Miniconference on Functional Analysis and Optimization, Australian National University, Mathematical Sciences Institute, 59–66, 1988.

[43] G. Gidel, H. Berard, G. Vignoud, P. Vincent, S. Lacoste-Julien, *A variational inequality perspective on generative adversarial networks*, arXiv:1802.10551, 2018.

[44] M. S. Gowda, M. Teboulle, *A comparison of constraint qualifications in infinite-dimensional convex programming*, SIAM Journal on Control and Optimization 28(4), 925-935, DOI: 10.1137/0328051, 1990.

[45] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM Journal on Control and Optimization 29, 403–419, DOI: 10.1137/0329022, 1991.

[46] S. P. Han and G. Lou, *A parallel algorithm for a class of convex programs*, SIAM Journal on Control and Optimization 26(2), 345–355, DOI: 10.1137/0326019, 1988.

[47] P. Hartman, G. Stampacchia, *On some non-linear elliptic differential-functional equations*, Acta Mathematica 115, 271–310, DOI: 10.1007/BF02392210, 1966.

[48] J. B. Hiriart-Urruty, C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, 1993.

[49] A. N. Iusem, B. F. Svaiter, *A variant of Korpelevich's method for variational inequalities with a new search strategy*, Optimization 42(4), 309–321, DOI: 10.1080/02331939708844365, 1997.

[50] L. Jiao, D. S. Kim, *Robust minimax optimization problems with applications*, RIMS Kyoto University 2114, 96–102, 2019.

[51] S. Karamardian, S. Schaible, *Seven kinds of monotone maps*, Journal of Optimization Theory and Applications 66(1), 37–46, DOI: 10.1007/BF00940531, 1990.

[52] T. Kato, *Perturbation Theory for Linear Operators*, 1966.

[53] D. Kinderlehrer, G. Stampacchia, *An introduction to variational inequalities and their applications*, Society for Industrial and Applied Mathematics, DOI: 10.1137/1.9780898719451, 2000.

[54] K. C. Kiwiel, *Proximal minimization methods with generalized Bregman functions*, SIAM Journal on Control and Optimization 35(4), DOI: 10.1137/S0363012995281742, 1142–1168, 1997.

[55] P. D. Khanh, P. T. Vuong, *Modified projection method for strongly pseudomonotone variational inequalities*, Journal of Global Optimization 58(2), 341–350, DOI: 10.1007/s10898-013-0042-5, 2014.

[56] E. N. Khobotov, *Modification of the extra-gradient method for solving variational inequalities and certain optimization problems*, USSR Computational Mathematics and Mathematical Physics 27(5), 120–127, DOI: 10.1016/0041-5553(87)90058-9, 1987.

[57] G. M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Ekonomika i Matematicheskie Metody 12(4), 747–756, 1976.

[58] B. Lemaire, *The proximal algorithm*, in New Methods in Optimization and Their Industrial Uses, International Series of Numerical Mathematics 87, 73–87, 1989.

[59] P. L. Lions, B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis 16(6), 964–979, DOI: 10.1137/0716071, 1979.

[60] J. Löfberg, *Approximations of closed-loop minimax MPC*. 42nd IEEE International Conference on Decision and Control 2, DOI: 10.1109/CDC.2003.1272813, 2003.

[61] J. Löfberg, *Minimax Approaches to Robust Model Predictive Control*, Dissertations, Linköping University, 2003.

[62] Y. Malitsky, M. K. Tam, *A forward-backward splitting method for monotone inclusions without cocoercivity*, SIAM Journal on Optimization 30(2), 1451–1472, DOI: 10.1137/18M1207260, 2020.

[63] B. Martinet, *Determination approchée d'un point fixe d'une application pseudo-contractante*, Comptes rendus de l'Académie des sciences Paris Series A-B 274, 163–165, 1972.

[64] B. Martinet, *Regularisation d'inéquations variationelles par approximations successives*, Revue française d'informatique et de recherche opérationnelle 4, 154–158, 1970.

[65] W. K. Nicholson, *Linear Algebra with Applications*, 2019.

[66] N. Noun, J. Peypouquet, *Forward-Backward Penalty Scheme for Constrained Convex Minimization Without Inf-Compactness*, Journal of Optimization Theory and Applications 158, 787–795, DOI: 10.1007/s10957-013-0296-6, 2013.

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

[68] L. A. Parente, P. A. Lotito, M. V. Solodov, *A class of inexact variable metric proximal point algorithms*, SIAM Journal on Optimization 19(1), 240–260, DOI: 10.1137/070688146, 2008.

[69] G. B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, Journal of Mathematical Analysis and Applications 72, 383–390, DOI: 10.1016/0022-247X(79)90234-8, 1979.

[70] J. Peypouquet, *Coupling the Gradient Method with a General Exterior Penalization Scheme for Convex Minimization*, Journal of Optimization Theory and Applications 153, 123–138, DOI: 10.1007/s10957-011-9936-x, 2011.

[71] B. T. Polyak, *Gradient methods for the minimisation of functionals*, USSR Computational Mathematics and Mathematical Physics 3(4), 864–878, DOI: 10.1016/0041-5553(63)90382-3, 1963.

[72] B. T. Polyak, *Introduction to Optimization*, Optimization Software, 1987.

[73] L. D. Popov, *A modification of the Arrow-Hurwicz method for search of saddle points.* Mathematical notes of the Academy of Sciences of the USSR 28(5), 845–848, DOI: 10.1007/BF01141092, 1980.

[74] A. Rakhlin, K. Sridharan, *Online learning with predictable sequences*, Proceedings of Machine Learning Research 30, 993–1019, 2013.

[75] S. Rakhlin, K. Sridharan, *Optimization, learning, and games with predictable sequences*, Advances in Neural Information Processing Systems 26, 2013.

[76] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of Operations Research 1(2), 97–116, 1976.

[77] R. T. Rockafellar, *Convex Analysis.* Princeton University Press, New Jersey, 1970.

[78] R. T. Rockafellar, *Duality and stability in extremum problems involving convex functions*, Pacific Journal of Mathematics 21(1), 167–187, 1967.

[79] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization 14(5), 877–898, DOI: 10.1137/0314056, 1976.

[80] R. T. Rockafellar, *On the maximal monotonicity of subdifferential mappings*, Pacific Journal of Mathematics 33(1), 209–216, 1970.

[81] D. Salinger, *A Splitting Algorithm for Multistage Stochastic Programming with Application to Hydropower Scheduling*, Ph.D. thesis, Department of Applied Mathematics, University of Washington, Seattle, WA, 1996.

[82] M. Scarvalone, *Game Theory and the Minimax Theorem*. Universidade de Chicago. `http://www.math.uchicago.edu/~may/VIGRE/VIGRE2008/REUPapers/Scarvalone.pdf`, 2008. Accessed 12 April 2024

[83] S. Shalev-Shwartz, *Online learning and online convex optimization*, Foundations and Trends in Machine Learning 4(2), 107–194, DOI: 10.1561/2200000018, 2012.

[84] M. Sibony, *Méthodes itératives pour les équations et inéquations aux dérivées partielles non-linéares de type monotone*, Calcolo 7, 65–183, 1970.

[85] J. E. Spingarn, *Applications of the method of partial inverses to convex programming: Decomposition*, Mathematical Programming 32, 199–223, DOI: 10.1007/BF01586091, 1985.

[86] Y. Tang, Z. Ji, *Perturbation Resilience of Self-Adaptive Step-Size Algorithms for Solving Split Variational Inclusion Problems and their Applications*, Numerical Functional Analysis and Optimization 44(13), 1343–1370, DOI: 10.1080/01630563.2023.2247615, 2023.

[87] B. Tongnoi, *A Modified Tseng's Algorithm with Extrapolation from the Past for Pseudo-monotone Variational Inequalities.* Taiwanese Journal of Mathematics 28(1), 187–210, DOI: 10.11650/tjm/230906, 2024.

[88] B. Tongnoi, *The forward-backward-forward algorithm with extrapolation from the past and penalty scheme for solving monotone inclusion problems and applications.* Numerical Algorithms, 1–31, DOI: 10.1007/s11075-024-01866-0, 2024.

[89] B. Tongnoi, *Tseng's Algorithm with Extrapolation from the Past Endowed with Variable Metrics and Error Terms.* Numerical Functional Analysis and Optimization 44, 87–123, DOI: 10.1080/01630563.2022.2158196, 2022.

[90] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM Journal on Control and Optimization 38(2), DOI: 10.1137/S0363012998338806, 431–446, 2000.

[91] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Advances in Computational Mathematics 38(3), 667–681, DOI: 10.1007/s10444-011-9254-8, 2013.

[92] B. C. Vũ, *A variable metric extension of the forward-backward-forward algorithm for monotone operators*, Numerical Functional Analysis and Optimization 34(9), 1050–1065, DOI: 10.1080/01630563.2013.763825, 2013.

[93] S. Wang, J. Fadili, P. Ochs, *Inertial Quasi-Newton Methods for Monotone Inclusion: Efficient Resolvent Calculus and Primal-Dual Methods*, arXiv:2209.14019, 2022.

[94] S. Yang, X. Li, G. Lan, *Data-Driven Minimax Optimization with Expectation Constraints*. Operations Research, DOI: 10.1287/opre.2022.0110, 2024.

[95] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, 2002.

[96] G. Zhang, *Understanding minimax optimization in modern machine learning*, Doctoral dissertation, University of Waterloo, UWSpace, `http://hdl.handle.net/10012/17157`, 2021. Accessed 12 April 2024

[97] D. L. Zhu, P. Marcotte, *Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities*, SIAM Journal on Optimization 6(3), DOI: 10.1137/S1052623494250415, 714–726, 1996.