GRADIENT AND FORWARD-BACKWARD INERTIAL METHODS FOR SOLVING CONSTRAINED CONVEX OPTIMIZATION AND MONOTONE INCLUSION PROBLEMS AND THEIR APPLICATIONS PERFORMANCE

NATTHAPHON ARTSAWANG

A Thesis Submitted to the Graduate School of Naresuan University
in Partial Fulfillment of the Requirements
for the Doctor of Philosophy Degree in Mathematics
September 2020

This thesis entitled "Gradient and forward-backward inertial methods for solving constrained convex optimization and monotone inclusion problems and their applications performance"

by Natthaphon Artsawang

has been approved by the Graduate School as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Mathematics of Naresuan University

Oral Defense Committee

.... Chair

(Professor Poom Kumam, Ph.D.)

.... Advisor

(Associate Professor Kasamsuk Ungchittrakool, Ph.D.)

.... Internal Examiner

(Associate Professor Rabian Wangkeeree, Ph.D.)

.... Internal Examiner

(Associate Professor Narin Petrot, Ph.D.)

.... External Examiner

(Associate Professor Prasit Cholamjiak, Ph.D.)

Approved

....

(Professor Paisarn Muneesawang, Ph.D.)

Dean of the Graduate School

22 SEP 2020

Title GRADIENT AND FORWARD-BACKWARD INERTIAL METHODS FOR SOLVING CONSTRAINED CONVEX OPTIMIZATION AND MONOTONE INCLUSION PROBLEMS AND THEIR APPLICATIONS PERFORMANCE

Author Natthaphon Artsawang

Advisor Associate Professor Kasamsuk Ungchittrakool, Ph.D.

Academic Paper Thesis Ph.D. in Mathematics, Naresuan University, 2020.

Keywords Convex optimization problems, gradient algorithm, nonexpansive mapping, forward-backward method, monotone inclusion problems, inertial method, constraint convex optimization problems, Mann-type algorithm.

ABSTRACT

In this thesis, we separate into two parts. First, we study two cases of the constrained convex optimization problems in Hilbert spaces. While the first case concerns the smooth convex objective function over the set of minimizers of a convex differentiable constrained function, the second case deals with the nonsmooth convex objective function over the same constrained set. We also present several iterative methods for approaching solutions of these problems. Second, we present methods for solving the monotone inclusion problems. Moreover, we also present iterative methods for solving fixed point problems which can be applied to solve the monotone inclusion problems. Some numerical examples are provided in order to support the convergence results.

LIST OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure **Page**

CHAPTER I

INTRODUCTION

Constrained convex optimization problem is one class of convex optimization problem concerning minimization of a convex objective function over a convex feasible set. It is preferable to minimize a convex function over a convex set because for a convex function any local minimum must be a global minimum. It is worth noting that the constrained convex optimization has application in many areas such as estimation and signal processing, image processing, communications and networks, electronic circuit design, data analysis and modeling, statistics, and finance [1]. When we try to find common minimizers of two convex functions we recognize that it complicated to consider each function, as a result, we set one of them as the objective function and another convex function becomes constraint. There are several methods for solving the convex optimization problem such as gradient, subgradient, polyhedral approximation, proximal, and interior point methods.

Actually in 2010, Attouch and Czarnecki [2] have represented the starting point of numerical algorithms for solving general constrained convex optimization problems. In 2011, Attouch et al. [3, 4] studied the constrained convex optimization problem in the form of minimization of a convex objective function over a set of minima of another convex function which is also called constrained convex optimization problem or hierarchical-type problem. They also proposed iterative methods for sovling this problem in many cases of the objective function and the penalization function. The convergent results of their iterative methods are presented under the inf-compactness assumption. In order to solving the constrained convex optimization problem without the inf-compactness assumption, in 2012, Peypouquet [5] proposed iterative method combining the gradient method and the penalty method for solving this problem in the case that both the objective function and the penalization function are nonsmooth. After that, in 2013, Noun and Peypouquet [6] proposed an algorithm for solving the constrained convex optimization problem and also proved convergence result without the inf-compactness assumption. We refer the reader to the series of papers [3,4,7–16] for more iterative schemes for solving general constrained convex optimization problems.

To improve the convergence behavior of the iterative methods, one is the inertial concept. Algorithms of inertial type were first introduced by Polyak in [17] and Bertsekas in [18] in the context of the minimization of a differentiable function. Since the works [17,18], one can notice an increasing number of research efforts dedicated to algorithms of inertial type (see [19–32]). For a variety of situations, in particular in the context of solving real-world problems, the presence of inertial terms improves the convergence behavior of the generated sequences. Recently, in 2017, Bot et al. [33] applied the idea of the gradient penalty method and the inertial concept to propose a new algorithm for solving the constrained convex optimization problem in the case that both the objective function and the penalization function are smooth. They also proved a convergence result. After that, Bot et al. [34] introduced an algorithm combining the proximal method and the inertial method for solving the generalized constrained convex optimization problem.

Let us come now to the monotone inclusion problem: find $x \in \mathcal{H}$ such that

$$0 \in Ax + Bx,$$

where $A : \mathcal{H} \to \mathcal{H}$ is a single-valued mapping and $B : \mathcal{H} \to 2^{\mathcal{H}}$ a multi-valued mapping. Many interesting problems can be formed into the monotone inclusion problem, such as convex minization problems, variational inequalities and equilibium problems, image processing problems, etc. Most well known algorithms to approximate the solution of this problem is the forward-backward algorithm **(FB)** [35–37]. In 2001, Alvarez and Attouch [21] introduced a new algorithm by using the idea of the inertial method to solve the monotone inclusion problem consisting one maximal monotone operator. After that, Moudafi and Oliny [32] proposed iterative method which involed the idea of the inertial method for solving the monotone inclusion problem consisting of two maximal monotone operators. Another type of the inertial methods was introduced by Polyak [38], which is a two-step iterative method in which the next iterate is defined by making use of the previous two. The several methods that are in reference to this study are reviewed in the next extensively (see, e.g. [27,28,39–45]). Recently, Kitkuan et al. [46] proposed the viscosity approximation algorithm concerning the inertial forward-backward for finding a solution of the considered problem. In 2019, Kitkuan et al. [47] presented a new method combined Halpern-type method and forward-backward splitting method for solving the monotone inclusion problem.

On the other hand, the monotone inclusion problems can be reformulated to the fixed point problems for nonexpansive mappings. Approximating a fixed point of nonexpansive mappings has been happened to the difference of the iterative methods. The well-known iterative method to solve the fixed point problem for nonexpansive mapping was introduced by Mann [48]. Many researchers have generalized, improved and extended his iterative method for solving various problems. For more details and most recent works on the methods for solving fixed point problems, we refer the reader to [27,28,30,33,34,39,41,42,44,46,47,49–63].

Motivated and inspired by the work mentioned above, we separate into two parts. First, we are going to consider the constrained convex optimization problems where the objective function is a convex function and the the constraint set is a set of minima of another convex functions. Moreover, we also propose the various algorithm for solving this problem of both smooth and nonsmooth cases of the objective functions. Second, we present iterative methods for solving the monotone inclusion problem and generalized monotone inclusion problem. Furthermore, we show the numerical experiments to demonstrate the effectiveness of our algorithms in every part.

This thesis is organized in the following way.

Chapter II. We will include some basic definitions, lemmas and theorems that are useful in the framework of the problems considered in this thesis.

Chapter III. This chapter, firstly, we propose an algorithm for solving constrianed convex optimization problem with smooth objective function. Under the observation of some appropriate choices for the available properties of the considered functions and scalars, we can generate a suitable algorithm that weakly converges to the solution. Further, we also provide a numerical example to compare among the our algorithm, the algorithm introduced by Peypouquet [5] and the algorithm introduced by Bot et al. [33]. Secondly, motivated and inspirated by the recent work, we propose an algorithm for solving constrianed convex optimization problem with nonsmooth objective function. Under suitable choices for the step sizes, the convergent results can be obtained. We also give applications and numerical results for proposed algorithm.

Chapter IV. This chapter, firstly, we propose an algorithm which is a combination of viscosity forward-backward algorithm and inertial extrapolation steps to solve monotone inclustion problem with sum of two monotone operators in a real Hilbert space. By using some suitable control conditions, the strong convergence is obtained. For the virtue of the main theorem, it can be applied to find a solution of the convex minimization problems. As an illustration of the behavior of the proposed algorithm, we compare the convergent behavior of our method and the algorithm was introduced by Kitkuan et al. [46]. Secondly, motivated and inspirated by the recent interest on inertial-type algorithm and the work in [41, 62], we propose a new Mann-type method combining both inertial terms and errors to find a fixed point of a nonexpansive mapping in a real Hilbert space. The strong convergence theorem of the iterate under some appropriate assumptions of parameters sequences are obtained. For the virtue of the main theorem, it can be applied to solve the monotone inclusion problem with sum of three monotone operators. Moreover, we give applications and numerical results for the proposed algorithm.

Chapter IV. We give some conclusions.

CHAPTER II

PRELIMINARIES

This chapter, we summarize some useful notations, definitions, properties, and some results, which are used throughout this thesis.

In this thesis, we denote two specific sets that $\mathbb{R}$ stands for the set of all real numbers and $\mathbb{N}$ the set of all natural numbers.

2.1 Basic results.

Definition 2.1.1. [64] A *linear space* or *vector space* $\mathcal{H}$ over $\mathbb{R}$ is a set $\mathcal{H}$ with binary operation *addition* defined for elements in $\mathcal{H}$ and *scalar multiplication* defined for numbers in $\mathbb{R}$ with elements in $\mathcal{H}$ satisfying the following properties: for all $x, y, z \in \mathcal{H}$ and $\alpha, \beta \in \mathbb{R}$

(V1) $x + y = y + x$.

(V2) $(x + y) + z = x + (y + z)$.

(V3) there exists an element $0 \in \mathcal{H}$ called the zero vector of $\mathcal{H}$ such that $x + 0 = x$ for all $x \in \mathcal{H}$.

(V4) for every element $x \in \mathcal{H}$, there exists an element $-x \in \mathcal{H}$ called the additive inverse or the negative of x such that $x + (-x) = 0$.

(V5) $\alpha(x + y) = \alpha x + \alpha y$.

(V6) $(\alpha + \beta)x = \alpha x + \beta x$.

(V7) $(\alpha\beta)x = \alpha(\beta x)$.

(V8) $1 \cdot x = x$.

The elements of a vector space $\mathcal{H}$ are called *vectors*, and the elements of $\mathbb{R}$ called *scalars*.

Definition 2.1.2. [64] A *normed space* is a vector space $\mathcal{H}$ on which there is defined a real-valued function $\|\cdot\|$ which maps each element x in $\mathcal{H}$ into a real number $\|x\|$ called the *norm* of x. The norm satisfies the following properties:

(N1) $\|x\| \geq 0$ for all $x \in \mathcal{H}$, $\|x\| = 0$ if and only if $x = 0$.

(N2) $\|\alpha x\| = |\alpha|\|x\|$ for all scalars $\alpha \in \mathbb{R}$ and each $x \in \mathcal{H}$.

(N3) $\|x + y\| \leq \|x\| + \|y\|$ for each $x, y \in \mathcal{H}$.

Definition 2.1.3. [64] An *inner product space* is a vector space $\mathcal{H}$ on which there is defined a real-valued function $\langle\cdot,\cdot\rangle$ which maps any pair of elements x and y in $\mathcal{H}$ into a real number $\langle x, y\rangle$ called the *inner product* of x and y. The inner product satisfies the following properties:

(I1) $\langle x, x\rangle \geq 0$ for all $x \in \mathcal{H}$, $\langle x, x\rangle = 0$ if and only if $x = 0$.

(I2) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$ for all scalar $\alpha \in \mathbb{R}$ and each $x, y \in \mathcal{H}$.

(I3) $\langle x, y\rangle = \langle y, x\rangle$ for each $x, y \in \mathcal{H}$.

(I4) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ for each $x, y, z \in \mathcal{H}$.

Let $\mathcal{H}$ be an inner product space. The function $\|\cdot\| : \mathcal{H} \to \mathbb{R}$, defined by

$$\|x\| = \sqrt{\langle x, x\rangle} \qquad \text{for every } x \in \mathcal{H},$$

is a *norm* on $\mathcal{H}$. Indeed, it is clear that $\|x\| \geq 0$ for every $x \in \mathcal{H}$ and $\|x\| = 0 \iff x = 0$. Moreover, for each $\alpha \in \mathbb{R}$ and $x \in \mathcal{H}$, we have $\|\alpha x\|^2 = \langle \alpha x, \alpha x\rangle = \alpha^2\|x\|^2$. It only remains to show that (N3) holds. We need the following inequality which is known as the *Cauchy-Bunyakovsky-Schwarz inequality* (in short, Schwarz inequality).

Theorem 2.1.4 (Schwarz inequality). *Let $\mathcal{H}$ be an inner product space. For each $x, y \in \mathcal{H}$, we have*

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

Proof. See [64, Lemma 3.2-1]. □

We use this inequality to deduce that for each $x, y \in \mathcal{H}$

$$\|x+y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

whence (N3) holds. In this situation, we conclude that the inner product space with the inner product $\langle \cdot, \cdot \rangle$ is a normed space with the norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$.

Definition 2.1.5. [64] A sequence $\{x_k\}_{k\in\mathbb{N}}$ in a normed space $\mathcal{H}$ is said to *converges* (strongly) to an element $x \in \mathcal{H}$ if $\lim_{k\to+\infty} \|x_k - x\| = 0$. We usually write $\lim_{k\to+\infty} x_k = x$ or $x_k \to x$ as $k \to +\infty$ and call the element x the limit of the sequence $\{x_k\}_{k\in\mathbb{N}}$. If a subsequence $\{x_{k_j}\}_{j\in\mathbb{N}}$ of $\{x_k\}_{k\in\mathbb{N}}$ converges to $x \in \mathcal{H}$, then x is called a *cluster point* of the sequence $\{x_k\}_{k\in\mathbb{N}}$.

Definition 2.1.6. [64] A sequence $\{x_k\}_{k\in\mathbb{N}}$ in a normed space $\mathcal{H}$ is said to be *Cauchy* if for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that $\|x_k - l_n\| \leq \varepsilon$ for all $k, l \geq N$.

Definition 2.1.7. [65] A normed space is said to be *complete* if every Cauchy sequence is convergent.

Definition 2.1.8. [65] A *Hilbert space* is a complete inner product space.

Example 2.1.9. *Let us consider the square-summable sequence space* $\ell_2 = \{x := (\xi_1, \xi_2, \ldots) : \sum_{n=1}^{+\infty} |\xi_n|^2 < +\infty\}$ *over* $\mathbb{R}$ *with the inner product* $\langle x, y\rangle = \sum_{n=1}^{+\infty} \xi_n \eta_n$ *where* $x = (\xi_1, \xi_2, \ldots)$ *and* $y = (\eta_1, \eta_2, \ldots)$ *and the associated norm*

$$\|x\| = \left(\sum_{n=1}^{+\infty} |\xi_n|^2\right)^{1/2}.$$

We know that ℓ_2 *is a Hilbert space (see [64, Example 3.1-6]).*

Definition 2.1.10. [64] A sequence $\{x_k\}_{k\in\mathbb{N}}$ in an inner product space $\mathcal{H}$ is said to *converges weakly* to an element $x \in \mathcal{H}$ if for any $y \in \mathcal{H}$, $\lim_{k\to+\infty}\langle x_k - x, y\rangle = 0$. We write $x_k \rightharpoonup x$ as $k \to +\infty$ and call the element x the weak limit of the sequence $\{x_k\}_{k\in\mathbb{N}}$. If a subsequence $\{x_{k_j}\}_{j\in\mathbb{N}}$ of $\{x_k\}_{k\in\mathbb{N}}$ converges weakly to $x \in \mathcal{H}$, then x is called a *weak cluster point* of the sequence $\{x_k\}_{k\in\mathbb{N}}$.

Theorem 2.1.11. *A strong convergent sequence in a Hilbert space is weak convergent with the same limit. In particular, a weakly convergent sequence of a finite dimensional Hilbert space is strong convergent with the same limit.*

Proof. See [64, Theorem 4.8-4]. □

The following example shows that the converse is not generally true.

Example 2.1.12. [64, Example 3.1-6] *Consider the sequence* $\{x_k\}_{k\in\mathbb{N}} \subset \ell_2$ *where* $x_k = (e_{k1}, e_{k2}, \ldots)$, *where*

$$e_{ki} = \begin{cases} 1 & ; i = k, \\ 0 & ; i \neq k. \end{cases}$$

For any $y = (\eta_1, \eta_2, \ldots) \in \ell_2$, *we have* $\langle x_k - 0, y \rangle = \eta_k \to 0$ *as* $k \to +\infty$. *This means that* $x_k \rightharpoonup 0$ *as* $k \to +\infty$. *Note, however, that* $\|x_k - 0\| = 1$ *for every* $k \geq 1$. *Hence,* $\{x_k\}_{k\in\mathbb{N}}$ *does not converge strongly to* 0 *as* $k \to +\infty$.

Definition 2.1.13. [65] A sequence $\{x_k\}_{k\in\mathbb{N}}$ in a normed space $\mathcal{H}$ is said to be *bounded* if there exists a positive number M such that $\|x_k\| \leq M$ for all $k \in \mathbb{N}$.

A Hilbert space has an important property which is expressed in the following theorem.

Theorem 2.1.14. *Every bounded sequence in a Hilbert space possesses a weakly convergent subsequence.*

Proof. See [66, Lemma 2.37]. □

Let $\mathcal{H}$ be a normed space, we denote the set $\mathcal{B}(x;r) := \{z \in \mathcal{H} : \|x - z\| < r\}$ a ball with center $x \in \mathcal{H}$ and radius $r > 0$. Next, we recall some useful sets in a normed space.

Definition 2.1.15. [64] A subset A of a normed space $\mathcal{H}$ is said to be *open* if for each $x \in A$, there exists $r > 0$ such that $\mathcal{B}(x;r) \subset A$. A subset B of $\mathcal{H}$ is said to be *closed* if its complement $\mathcal{H} \setminus B$ is open.

Definition 2.1.16. [64] Let A be a subset of a normed space $\mathcal{H}$ and $x \in \mathcal{H}$. Then, x is said to be an *interior point* of A if there exists $r > 0$ such that $\mathcal{B}(x; r) \subset A$. The *interior* of A is the set of all interior points of A and may be denoted by $\text{int}(A)$.

Definition 2.1.17. [64] Let A be a subset of a normed space $\mathcal{H}$. The *closure* of A is the smallest closed set containing A; it is denoted by $\text{cl}(A)$.

Definition 2.1.18. [64] Let A be a subset of a normed space $\mathcal{H}$. The *boundary* of A is the closure of A without the interior of A; it is denoted by $\text{bd}(A)$.

Definition 2.1.19. [64] Let $\mathcal{H}$ be a Hilbert space. A subset A of $\mathcal{H}$ is said to be *compact* if every sequence A has a convergent subsequence whose limit is an element of A.

Definition 2.1.20. [64] Let $\mathcal{H}$ be a Hilbert space. A subset A of $\mathcal{H}$ is said to be *relatively compact* if $\text{cl}(A)$ is compact.

Let us recall useful facts related to convergence and closedness which will be needed later.

Theorem 2.1.21. *Let A be a subset of a normed space $\mathcal{H}$. Then,*

(1) $x \in \text{cl}(A)$ if and only if there is a sequence $\{x_k\}_{k\in\mathbb{N}} \subset A$ such that $x_k \to x$ as $k \to +\infty$.

(2) A is closed if and only if for any sequence $\{x_k\}_{k\in\mathbb{N}} \subset A$ with $x_k \to x \in \mathcal{H}$ as $k \to +\infty$, we have $x \in A$.

Proof. See [64, Theorem 1.4-6]. □

2.2 Convexity.

Throughout this subsection, we let $\mathcal{H}$ be a Hilbert space. In the following definition we recall the convexity of a real-valued function which goes together with the convexity of a set as we are recalled above.

In practical properties of convexity, then, we denote the extended real number $[-\infty,+\infty] := \mathbb{R} \cup \{-\infty,+\infty\}$.

Definition 2.2.1. [65] A subset $\mathcal{C}$ of $\mathcal{H}$ is said to be *convex* if $\alpha x + (1-\alpha)y \in \mathcal{C}$ for every $x, y \in \mathcal{C}$ and for every $\alpha \in (0,1)$.

Theorem 2.2.2. [66] Let $\{C_j : j \in \mathcal{J}\}$ be an arbitrary collection of convex sets in $\mathcal{H}$. Then, their intersection $\bigcap_{j\in\mathcal{J}} C_j$ is also convex.

Definition 2.2.3. [66] Let $f : \mathcal{H} \to [-\infty,+\infty]$. The *domain* of f is

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

the *graph* of f is

$$\operatorname{gra}(f) = \{(x,\xi) \in \mathcal{H} \times \mathbb{R} \mid f(x) = \xi\},$$

the *epigraph* of f is

$$\operatorname{epi}(f) = \{(x,\xi) \in \mathcal{H} \times \mathbb{R} \mid f(x) \leq \xi\}.$$

The function f is *proper* if $-\infty \notin f(\mathcal{H})$ and $\operatorname{dom}(f) \neq \emptyset$.

Observe that if f is a function from $\mathcal{H}$ into $\mathbb{R}$, then $\operatorname{dom}(f) = \mathcal{H}$ and $-\infty \notin f(\mathcal{H})$.

Definition 2.2.4. [66] Let $f : \mathcal{H} \to [-\infty,+\infty]$. Then f is *convex* if its epigraph is a convex subset of $\mathcal{H} \times \mathbb{R}$.

Proposition 2.2.5. *Let $f : \mathcal{H} \to [-\infty,+\infty]$. Then f is convex if and only if for any $x, y \in \operatorname{dom}(f)$ and for any $\alpha \in (0,1)$, we have*

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

Proof. See [66, Proposition 4.8]. □

Definition 2.2.6. [66] Let $f : \mathcal{H} \to [-\infty,+\infty]$. Then f is *β-strongly convex* with $\beta > 0$ if,

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) + \frac{1}{2}\beta\alpha(1-\alpha)\|x-y\|^2$$

for all $x, y \in \operatorname{dom}(f)$ and for all $\alpha \in (0,1)$.

One more considering with regard to the generalization of the inequality in Proposition 2.2.5 in the case of convex combination of more than two points this so-called Jensen's inequality.

Theorem 2.2.7. *A function* $f : \mathcal{H} \to [-\infty, +\infty]$ *is convex if and only if for any finite families* $\{x_i : i \in \mathcal{I}\} \subset \operatorname{dom}(f)$ *and* $\{\alpha_i : i \in \mathcal{I}\} \subset (0,1)$ *such that* $\sum_{i\in\mathcal{I}} \alpha_i = 1$*, there holds*

$$f\left(\sum_{i\in\mathcal{I}} \alpha_i x_i\right) \leq \sum_{i\in\mathcal{I}} \alpha_i f(x_i).$$

Proof. See [67, Theorem 7.5]. □

Definition 2.2.8. [68] A subset C of H is a *cone* if $\alpha x \in C$ whenever $x \in C$ and $\alpha \in (0, +\infty)$

Definition 2.2.9. [66] Let $\mathcal{C}$ be nonempty convex subset of $\mathcal{H}$ and let $x \in \mathcal{H}$. The *normal cone* to $\mathcal{C}$ at a point x is

$$N_{\mathcal{C}}(x) := \begin{cases} \{\overline{x} \in \mathcal{H} : \langle \overline{x}, c - x\rangle \leq 0 \ \text{ for all } c \in \mathcal{C}\}, & \text{if } x \in \mathcal{C} \\ \emptyset, & \text{otherwise.} \end{cases}$$

Definition 2.2.10. [66] Let $\mathcal{C}$ be nonempty subset of $\mathcal{H}$. The *indicator function* of $\mathcal{C}$ is $\iota_{\mathcal{C}} : \mathcal{H} \to [-\infty, +\infty]$, which is defined by

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

for all $x \in \mathcal{H}$.

Definition 2.2.11. [66] Let $\mathcal{C}$ be nonempty subset of $\mathcal{H}$. The *support function* of $\mathcal{C}$ is $\sigma_{\mathcal{C}} : \mathcal{H} \to [-\infty, +\infty]$, which is defined by

$$\sigma_{\mathcal{C}}(x) := \sup_{c\in\mathcal{C}}\{\langle c, x\rangle\} \text{ for all } x \in \mathcal{H}.$$

We can observe that $\overline{x} \in N_{\mathcal{C}}(x)$ if and only if $\sigma_{\mathcal{C}}(\overline{x}) = \langle \overline{x}, x\rangle$.

Next, we recall some semicontinuities of a function on a Hilbert space.

Definition 2.2.12. [65] A function $f : H \to \mathbb{R} \cup \{+\infty\}$ is said to be *upper semicontinuous* on H if $\{x \in H : f(x) < \lambda\}$ is an open set for all $\lambda \in \mathbb{R}$.

Definition 2.2.13. [65] A function $f : H \to \mathbb{R} \cup \{+\infty\}$ is said to be *lower semicontinuous* on H if $\{x \in H : f(x) \leq \lambda\}$ is a closed set for all $\lambda \in \mathbb{R}$.

Definition 2.2.14. [66] Let D be a subset of $[-\infty, +\infty]$. A number $a \in [-\infty, +\infty]$ is the (necessarily unique) *infimum* (or the greatest lower bound) of D if it is a lower bound of D and if, for every lower bound $\bar{a}$ of D, we have $a \leq \bar{a}$. This number is denoted by $\inf(D)$. The *supremum* (or least upper bound) of D is $\sup(D) := -\inf\{-b : b \in D\}$.

Remark 2.2.15. Note that If D is bounded from above in $\mathbb{R}$, we know from the completeness of $\mathbb{R}$ that there exists the supremum $\sup(D)$ of D in $\mathbb{R}$. If D is not bounded from above in $\mathbb{R}$, in this situation, we have $\sup(D) = +\infty$. Similarly, if D is not bounded from below in $\mathbb{R}$, we have the infimum $\inf(D) = -\infty$. In this viewpoint, the set D always admits an infimum and a supremum in $[-\infty, +\infty]$.

Definition 2.2.16. [49] Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a function. For a sequence $\{x_k\}_{k\in\mathbb{N}} \in \mathcal{H}$, the *limit inferior* of $\{f(x_k)\}_{k\in\mathbb{N}}$ in $[-\infty, +\infty]$ is

$$\liminf_{k\to+\infty} f(x_k) := \sup_{k\geq 1} \inf_{k\leq n} f(x_k)$$

and its *limit superior* in $[-\infty, +\infty]$ is

$$\limsup_{k\to+\infty} f(x_k) := \inf_{k\geq 1} \sup_{k\leq n} f(x_k).$$

With these means the following theorem gives the characterization of lower semicontinuity in the term of limit inferior.

Theorem 2.2.17. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a function. Then, f is lower semicontinuous at $x \in \mathcal{H}$ if and only if, for every sequence $\{x_k\}_{k\in\mathbb{N}}$ in $\mathcal{H}$,*

$$x_k \to x \text{ as } k \to +\infty \implies f(x) \leq \liminf_{k\to+\infty} f(x_k).$$

Proof. See [49, Theorem 1.3.2]. □

It is alike to the upper semicontinuity, we also have the characterization of upper semicontinuity in the term of limit superior.

Theorem 2.2.18. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a function. Then, f is upper semicontinuous at $x \in \mathcal{H}$ if and only if, for every sequence $\{x_k\}_{k\in\mathbb{N}}$ in $\mathcal{H}$,*

$$x_k \to x \text{ as } k \to +\infty \implies \limsup_{k\to+\infty} f(x_k) \leq f(x).$$

Proof. See [49, Problem 1.3(7)]. □

Definition 2.2.19. A function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is said to be *continuous* at $x \in \mathcal{H}$ if, it is lower and upper semicontinuous at x.

The following theorem concerns a sufficient condition for continuity of a convex function.

Theorem 2.2.20. *Assume that $\mathcal{H}$ is finite dimensional. Then a convex function $f : \mathcal{H} \to \mathbb{R}$ is continuous.*

Proof. See [69, Theorem 5.23]. □

Now, we give a definition of inf-compactness function.

Definition 2.2.21. [70] A function $f : \mathcal{H} \to \mathbb{R}$ is said to be *inf-compact* if,

$$\forall r > 0,\ \forall \lambda \in \mathbb{R},\ \{x \in \mathcal{H} : \|x\| \leq r,\ f(x) \leq \lambda\}$$

is relatively compact.

Furthermore, in a practical point of view of Hilbert space, it sometimes concerns with weak convergence. Also, motivated by Theorem 2.2.17 we can consider the semicontinuity relating to weak convergence.

Definition 2.2.22. A function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is said to be *weakly lower semicontinuous* at $x \in \mathcal{H}$ if, for every sequence $\{x_k\}_{k\in\mathbb{N}}$ in $\mathcal{H}$,

$$x_k \rightharpoonup x \text{ as } k \to +\infty \implies f(x) \leq \liminf_{k\to+\infty} f(x_k).$$

Herewith we have got a practical relation of lower semicontinuity and weakly lower semicontinuity.

Theorem 2.2.23. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Then, f is lower semicontinuous if and only if its is weakly lower semicontinuous.*

Proof. See [66, Theorem 9.1]. $\square$

There exist two definitions involving differentiabilities of a function in the setting of Hilbert space.

Definition 2.2.24. [66] Let $f : \mathcal{H} \to \mathbb{R}$ be a function and $x, s \in \mathcal{H}$ be given. The *directional derivative* of f at x in the direction s is

$$f'(x; s) := \lim_{t \to 0} \frac{f(x + ts) - f(x)}{t}$$

whenever this limit exists. The function f is said to be *Gâteaux differentiable* at x if its has directional derivatives $f'(x; s)$ for all $s \in \mathcal{H}$ and

$$f'(x; s) = \langle g, s \rangle$$

holds for some $g \in \mathcal{H}$. The element g is called *Gâteaux derivative* or *Gâteaux gradient* of f at x and is denoted by $\nabla f(x)$.

Definition 2.2.25. [66] Let $f : \mathcal{H} \to \mathbb{R}$ be a function and $x \in \mathcal{H}$ be given. The function f is said to be *Fréchet differentiable* or, shortly, *differentiable* at x if there exists an element $y \in \mathcal{H}$ such that

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

The element y is called *Fréchet derivative* or *gradient* of f at x and is denoted by $Df(x)$.

One of the main points of interest at the relation between these two differentiabilities is advocated by the following theorem.

Theorem 2.2.26. *Let $f : \mathcal{H} \to \mathbb{R}$ be a function and $x \in \mathcal{H}$. If f is Fréchet differentiable at x, then it is Gâteaux differentiable at x and $Df(x) = \nabla f(x)$.*

Proof. See [66, Lemma 2.49]. □

Convexity can be characterized in the term of Gâteaux differentiability as presented in the following theorem.

Theorem 2.2.27. *Let $f : \mathcal{H} \to \mathbb{R}$ be a Gâteaux differentiable function. Then the following are equivalent:*

(i) f *is convex.*

(ii) $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$ *for every* $x, y \in \mathcal{H}$.

(iii) $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0$ *for every* $x, y \in \mathcal{H}$.

Proof. See [66, Proposition 17.10]. □

The next theorem shows not only the necessary condition for a Gâteaux differentiable function to be Fréchet differentiable but also useful property of Fréchet differentiability.

Theorem 2.2.28. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex lower semicontinuous and Gâteaux differentiable function and $x \in \mathcal{H}$. Then, f is Fréchet differentiable at x if and only if ∇f is continuous at x, that is, for every sequence $\{x_k\}_{k\in\mathbb{N}}$, $x_k \to x$ as $k \to +\infty$, we have $\nabla f(x_k) \to \nabla f(x)$ as $k \to +\infty$.*

Proof. See [66, Corollary 17.33]. □

Definition 2.2.29. [66] Let a function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper and $x \in \mathcal{H}$. An element $g \in \mathcal{H}$ is a *subgradient* of f at x if

$$f(y) \geq f(x) + \langle g, y - x \rangle \quad \text{for every } y \in \mathcal{H}.$$

The set of all subgradients of f at x is called *subdifferential* of f at x and may be denoted by $\partial f(x)$. If $\partial f(x) \neq \emptyset$, we say that f is subdifferentiable at x.

In order to guarantee subdifferentiability of a function, the continuity is an important one as follows.

Theorem 2.2.30. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex function. If f is continuous at some element $x_0 \in \mathcal{H}$, then it is subdifferentiable. Furthermore, if f is lower semicontinuous, then it is also subdifferentiable.*

Proof. See [69, Theorem 5.35] and [71, Theorem 2.4.12]. □

The relation between differentiability and subdifferentiability is referred.

Theorem 2.2.31. *Let $f : \mathcal{H} \to \mathbb{R}$ be a convex function and $x \in \mathcal{H}$. If f is Gâteaux differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.*

Proof. See [66, Proposition 17.26]. □

We provide the characterization of minimizers of a proper function in the following Theorem.

Theorem 2.2.32 (Fermats rule). *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper and convex. Then*

$$\arg\min f = \operatorname{zer}(\partial f) := \{x \in \mathcal{H} \mid 0 \in \partial f(x)\},$$

where $\arg\min f = \{x \in \mathcal{H} \mid f(x) \leq f(y)\ \forall y \in \mathcal{H}\}$.

Proof. See [66, Theorem 16.3]. □

We close this subsection by providing the definition of a Fenchel conjugate of a function.

Definition 2.2.33. [66] Let $f : \mathcal{H} \to [-\infty, +\infty]$. The *Fenchel conjugate* of f is $f^* : \mathcal{H} \to [-\infty, +\infty]$, which is defined by

$$f^*(u) = \sup_{x \in \mathcal{H}} \{\langle u, x \rangle - f(x)\} \quad \text{for all } u \in \mathcal{H}.$$

Notice that $\iota_C^* = \sigma_C$.

2.3 Operators.

Throughout this section we also let $\mathcal{C}$ be a nonempty closed convex subset of a Hilbert space $\mathcal{H}$. We denote the subset of fixed points of an operator $T : \mathcal{H} \to \mathcal{H}$ by

$$\mathrm{Fix}(T) := \{x \in \mathcal{H} : Tx = x\}.$$

More than that the crucial basic operators serving as nonlinear operators are presented and with this we normally apply in the later chapters.

Definition 2.3.1. [66] Let $T : \mathcal{H} \to \mathcal{H}$ be an operator.

(i) T is said to be *nonexpansive* if

$$\|Tx - Ty\| \leq \|x - y\| \quad \text{for all } x, y \in \mathcal{H}.$$

(ii) T is said to be *firmly nonexpansive* if

$$\|Tx - Ty\|^2 \leq \langle x - y, Tx - Ty\rangle \quad \text{for all } x, y \in \mathcal{H}.$$

(iii) T is said to be *contraction operator* if there exists a positive real number $\rho \in (0, 1)$ such that

$$\|Tx - Ty\| \leq \rho\|x - y\| \quad \text{for all } \; x, y \in \mathcal{H}.$$

(iv) T is said to be *Lipschitz operator* if there exists a positive real number L such that

$$\|Tx - Ty\| \leq L\|x - y\| \quad \text{for all } \; x, y \in \mathcal{H}.$$

(v) T is said to be *monotone* if

$$0 \leq \langle x - y, Tx - Ty\rangle, \quad \text{for all } x, y \in \mathcal{H}.$$

(vi) T is said to be *β-stongly monotone* with $\beta > 0$ if

$$\beta\|x-y\|^2 \leq \langle x-y, Tx-Ty\rangle \quad \text{for all } x, y \in \mathcal{H}.$$

(vii) T is said to be *β-cocoercive* (or *β-inverse strongly monotone*) with $\beta > 0$ if

$$\beta\|Tx-Ty\|^2 \leq \langle x-y, Tx-Ty\rangle \quad \text{for all } x, y \in \mathcal{H}.$$

(viii) T is said to be *pseudomonotone* if

$$0 \leq \langle Ty, x-y\rangle \Rightarrow 0 \leq \langle Tx, x-y\rangle \quad \text{for all } x, y \in \mathcal{H}.$$

It is easy to check that a β-strongly monotone mapping is monotone and a monotone mapping is pseudomonotone.

In addition, it is viewed in some of work as an essential of the metric projection's definition and its properties should be focus.

Definition 2.3.2. Let C be a nonempty subset of $\mathcal{H}$ and $x \in \mathcal{H}$. If there exists an element $y \in C$ such that

$$\|x-y\| \leq \|x-c\| \quad \text{for all } c \in C,$$

then the element y is called a *metric projection* of x onto C and is denoted by $\text{proj}_C(x)$. Further, if $\text{proj}_C(x)$ exists and uniquely determined for all $x \in \mathcal{H}$, then the operator $\text{proj}_C : \mathcal{H} \to C$ is called the *metric projection* onto C.

We can guarantee the existence and uniqueness of the metric projection by the following theorem.

Theorem 2.3.3. *Let C be a nonempty closed and convex subset of $\mathcal{H}$. Then for any $x \in \mathcal{H}$ there exists a unique metric projection* $\text{proj}_C(x)$.

Proof. See [72, Theorem 1.2.3]. □

Likewise, there exists a useful properties of metric projection as follows.

Theorem 2.3.4. *Let C be a nonempty closed and convex subset of $\mathcal{H}$. Then the operator $\text{proj}_C : \mathcal{H} \to \mathcal{H}$ is firmly nonexpansive and $\text{Fix}(\text{proj}_C) = C$.*

Proof. See [72, Theorem 2.2.21]. □

The correspondence definitions and theorems are provided as follows.

Definition 2.3.5. [66] Let $T : \mathcal{H} \to 2^{\mathcal{H}}$. The *domain* of T is

$$\text{dom}(T) = \{x \in \mathcal{H} \mid Tx \neq \emptyset\},$$

the *range* of T is

$$\text{ran}(T) = T(\mathcal{H}),$$

the *graph* of T is

$$\text{gra}(T) = \{(x, \xi) \in \mathcal{H} \times 2^{\mathcal{H}} \mid \xi \in Tx\}.$$

Definition 2.3.6. Let $T : \mathcal{H} \to 2^{\mathcal{H}}$ be a set-value operator. Then the operator T is called *monotone* if,

$$\langle u - v, x - y \rangle \geq 0$$

for all $(x, u), (y, v) \in \text{gra}(T)$,

the operator T is said to be *maximal monotone* if, it is monotone and there exists no proper monotone extension of the graph of T and

the operator T is said to be *β-strongly monotone* if,

$$\beta\|x - y\|^2 \leq \langle x - y, u - v \rangle,$$

for all $(x, u), (y, v) \in \text{gra}(T)$

A fundamental example of a maximally monotone operator is the subdifferential of a proper, convex, and lower semicontinuous function.

Theorem 2.3.7 (Moreau). *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex, and lower semicontinuous function. Then ∂f is a maximal monotone.*

Proof. See [66, Theorem 20.25]. □

The following definition also plays an important role in convergent analysis.

Definition 2.3.8. Let $T : \mathcal{H} \to 2^{\mathcal{H}}$ be a set-values operator. The *resolvent* of T with parameter $\lambda > 0$, $J_\lambda^T : \mathcal{H} \to 2^{\mathcal{H}}$ defined by $J_\lambda^T := (Id + \lambda T)^{-1}$, where Id is the *identity operator* from $\mathcal{H}$ to $\mathcal{H}$.

If T is maximally monotone, J_λ^T is a single-valued.

The following definition involving proximality of a function in the setting of Hilbert space.

Definition 2.3.9. [66] Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex, and lower semicontinuous function. For $\lambda > 0$, the mapping $\text{prox}_{\lambda f} : \mathcal{H} \to \mathcal{H}$ is given by

$$\text{prox}_{\lambda f}(x) = \arg\min_{y \in \mathcal{H}} \{f(y) + \frac{1}{2\lambda}\|x - y\|^2\}$$

is called *proximal operator* of the function f with scaling parameter λ.

Next, we present the characterization of a proximal operator.

Theorem 2.3.10. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper, convex, and lower semicontinuous. Let $\lambda > 0$, and let x and p be in $\mathcal{H}$. Then*

$$p = \text{prox}_{\lambda f}(x) \Leftrightarrow x - p \in \lambda \partial f.$$

In other words,

$$\text{prox}_{\lambda f} = (Id + \lambda \partial f)^{-1}$$

Proof. See [66, Proposition 16.44]. □

We observe that if $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, convex, and lower semicontinuous, then $\text{prox}_{\lambda f} = J_\lambda^{\partial f}$.

In the following definition, we recall the Fitzpatrick function.

Definition 2.3.11. [66] Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. The *Fitzpatrick function* of A is $F_A : \mathcal{H} \times \mathcal{H} \to [-\infty, +\infty]$, which is defined by

$$F_A(x, u) = \sup_{(y,v) \in \text{gra}(A)} \{\langle y, u\rangle + \langle x, v\rangle - \langle y, v\rangle\} \text{ for all } (x, u) \in \mathcal{H} \times \mathcal{H}.$$

In addition, F_A is convex and lower semicontinuous (see [73, Proposition 4.2]).

Theorem 2.3.12. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximal monotone. Then*

$$F_A(x, u) \geq \langle x, u \rangle \text{ for all } (x, u) \in \mathcal{H} \times \mathcal{H}.$$

Proof. See [66, Proposition 20.58]. □

Theorem 2.3.13. *Let $f : \mathcal{H} \to [-\infty, +\infty]$ be convex, proper, and lower semicontinuous. Then*

$$F_{\partial f}(x, u) \leq f(x) + f^*(u) \text{ for all } (x, u) \in \mathcal{H} \times \mathcal{H}.$$

Proof. See [74, Proposition 2.1]. □

We next state some results in real Hilbert spaces that will be useful in the later chapter.

Lemma 2.3.14. [49] *Let $\mathcal{H}$ be a real Hilbert space. The conditions are verifiable, as follows.*

(i) $2\langle x, y \rangle = \|x\|^2 + \|y\|^2 - \|x - y\|^2$ *for all* $x, y \in \mathcal{H}$,

(ii) $\|x - y\|^2 = \|x\|^2 - \|y\|^2 - 2\langle x - y, y \rangle$ *for all* $x, y \in \mathcal{H}$,

(iii) $\|x + y\|^2 \leq \|x\|^2 + 2\langle x + y, y \rangle$ *for all* $x, y \in \mathcal{H}$,

(iv) $\|rx + (1 - r)y\|^2 = r\|x\|^2 + (1 - r)\|y\|^2 - r(1 - r)\|x - y\|^2$ *for all* $r \in [0, 1]$ *and* $x, y \in \mathcal{H}$.

The following definition is very important and gets along with the cutter.

Definition 2.3.15. An operator $T : \mathcal{H} \to \mathcal{H}$ with $\text{Fix}(T) \neq \emptyset$ is said to satisfy the *demiclosed principle* if for every sequence $\{x_k\}_{k \in \mathbb{N}} \subset \mathcal{H}$ such that $x_k \rightharpoonup x \in \mathcal{H}$ and $Tx_k - x_k \to 0$ as $k \to +\infty$, we have $x \in \text{Fix}(T)$.

The following theorem is due to Opial [75] involves the demiclosed principle of a nonexpansive operator.

Theorem 2.3.16. *If an operator $T : \mathcal{H} \to \mathcal{H}$ is nonexpansive with* $\mathrm{Fix}(T) \neq \emptyset$*, then it is satisfying the demiclosed principle.*

Proof. See [75, Lemma 2]. □

The following definition of operator will play an important role in this thesis. In what follows, we let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces with the inner products $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}_2}$ and the associate norms $\| \cdot \|_{\mathcal{H}_1}$ and $\| \cdot \|_{\mathcal{H}_2}$, respectively. We denote the range of an operator $A : \mathcal{H}_1 \to \mathcal{H}_2$ by $\mathrm{Ran}(A)$, that is,

$$\mathrm{Ran}(A) := \{y \in \mathcal{H}_2 : y = Ax, \text{ for some } x \in \mathcal{H}_1\}.$$

For a subset $D \subset \mathcal{H}_2$, we denote the inverse image of D under A by $A^{-1}(D)$, i.e.,

$$A^{-1}(D) := \{x \in \mathcal{H}_1 : Ax \in D\}.$$

Definition 2.3.17. Let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be an operator.

(i) A is said to be *linear* if

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay \quad \text{for all } x, y \in \mathcal{H}_1 \text{ and for all } \alpha, \beta \in \mathbb{R}.$$

(ii) A is said to be *bounded* if there exists a real number $M > 0$ such that $\|Ax\|_{\mathcal{H}_2} \leq M\|x\|_{\mathcal{H}_1}$ for all $x \in \mathcal{H}_1$.

(ii) A is said to be *continuous* at an element $x \in \mathcal{H}_1$ if for every sequence $\{x_k\}_{k\in\mathbb{N}} \subset \mathcal{H}_1$ such that $x_k \to x \in \mathcal{H}_1$ as $k \to +\infty$, we have the sequence $\{Ax_k\}_{k\in\mathbb{N}} \subset \mathcal{H}_2$ satisfies $Ax_k \to Ax \in \mathcal{H}_2$ as $k \to +\infty$. And, A is said to be *continuous* if it is continuous at every element of $\mathcal{H}_1$.

Definition 2.3.18. Let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded linear operator. The number

$$\|A\| := \sup_{0 \neq x \in \mathcal{H}_1} \left\{ \frac{\|Ax\|_{\mathcal{H}_2}}{\|x\|_{\mathcal{H}_1}} \right\}$$

is called a *norm* of the operator A.

The following theorem gives some useful properties of a linear operator.

Theorem 2.3.19. *Let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a linear operator. Then the following statements are true:*

(i) If A is bounded, then

$$\|Ax\|_{\mathcal{H}_2} \leq \|A\|\|x\|_{\mathcal{H}_1} \quad \text{for every } x \in \mathcal{H}_1.$$

(ii) A is bounded if and only if A is continuous.

Proof. See [64, Theorem 2.7-9]. □

Definition 2.3.20. Let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded linear operator. An operator $A^* : \mathcal{H}_2 \to \mathcal{H}_1$ is called *adjoint operator* of A if

$$\langle Ax, y\rangle_{\mathcal{H}_2} = \langle x, A^*y\rangle_{\mathcal{H}_1} \quad \text{for all } x \in \mathcal{H}_1 \text{ and for all } y \in \mathcal{H}_2.$$

Of course, we can guarantee the well-defined of the adjoint operator of a bounded linear operator by the following theorem.

Theorem 2.3.21. *Let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded linear operator. Then there exists a unique adjoint operator $A^* : \mathcal{H}_2 \to \mathcal{H}_1$ of A. Furthermore, the adjoint operator A^* is bounded linear operator with norm*

$$\|A^*\| = \|A\|.$$

Proof. See [64, Theorem 3.9-2]. □

The following theorem provides a general property of adjoint operator which is used frequently.

Theorem 2.3.22. *Let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded linear operator. Then it holds that*

$$\|A^*A\| = \|AA^*\| = \|A\|^2.$$

Proof. See [64, Theorem 3.9-4]. □

2.4 Further Convergence Tools.

Lemma 2.4.1. [50,76] *Let $\{a_k\}_{k\in\mathbb{N}}$ and $\{\mu_k\}_{k\in\mathbb{N}}$ be sequences of nonnegative real numbers and satisfy the inequality*

$$a_{k+1} \leq (1-\delta_k)a_k + \mu_k + \varepsilon_k \quad \forall k \geq 1,$$

where $0 \leq \delta_k \leq 1$ for all $k \geq 1$. Assume that $\sum_{k\geq 1} \varepsilon_k < +\infty$. Then the following statement hold:

(i) *If $\mu_k \leq c\delta_k$ (where $c \geq 0$), then $\{a_k\}_{k\in\mathbb{N}}$ is bounded.*

(ii) *If $\sum_{k\geq 1} \delta_k = \infty$ and $\limsup_{k\to+\infty} \frac{\mu_k}{\delta_k} \leq 0$, then the sequence $\{a_k\}_{k\in\mathbb{N}}$ converges to 0.*

Lemma 2.4.2. *Let $\{\gamma_k\}_{k\in\mathbb{N}}$, $\{\delta_k\}_{k\in\mathbb{N}}$ and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ be real sequences. Assume that $\{\gamma_k\}_{k\in\mathbb{N}}$ is bounded from below, $\{\delta_k\}_{k\in\mathbb{N}}$ is non-negative and $\sum_{k=1}^{+\infty} \varepsilon_k < +\infty$ such that*

$$\gamma_{k+1} - \gamma_k + \delta_k \leq \varepsilon_k \quad \text{for all} \quad k \geq 1.$$

Then $\lim_{k\to\infty} \gamma_k$ exists and $\sum_{k=1}^{+\infty} \delta_k < +\infty$.

Proof. See [3, Lemma 2] □

Lemma 2.4.3 (Opial Lemma). *Let $\mathcal{H}$ be a real Hilbert space, $\mathcal{C} \subseteq \mathcal{H}$ be nonempty set, $\{x_k\}_{k\in\mathbb{N}}$ be any arbitrary sequence, $\{\lambda_k\}_{k\in\mathbb{N}}$ a sequences of positive real numbers and $\{z_k\}_{k\in\mathbb{N}}$ defined by*

$$z_k = \frac{1}{\tau_k}\sum_{n=1}^{k} \lambda_n x_n, \qquad \text{where} \qquad \tau_k = \sum_{n=1}^{k} \lambda_n$$

such that:

(i) *For every $z \in \mathcal{C}$, $\lim_{k\to+\infty} \|x_k - z\|$ exists;*

(ii) *Every weak cluster point of the sequence $\{x_k\}_{k\in\mathbb{N}}$ (resp., $\{z_k\}_{k\in\mathbb{N}}$) belongs to $\mathcal{C}$.*

Then the sequence $\{x_k\}_{k\in\mathbb{N}}$ (resp., $\{z_k\}_{k\in\mathbb{N}}$) converges weakly to a point in $\mathcal{C}$.

Proof. See [66, Lemma 2.47] □

Lemma 2.4.4. [77] *Let A be an μ-inverse strongly monotone operator from a real Hilbert space $\mathcal{H}$ into itself and $B : \mathcal{H} \to 2^{\mathcal{H}}$ a maximal nomotone operator. Then, the following inequalities hold.*

$$\|\Gamma_\lambda^{A,B}x - \Gamma_\lambda^{A,B}y\|^2 \leq \|x-y\|^2 - \lambda(2\mu-\lambda)\|Ax-Ay\|^2 - \|(Id-J_\lambda^B)(Id-\lambda A)x - (Id-J_\lambda^B)(Id-\lambda A)y\|^2 \tag{2.4.1}$$

for all $x, y \in B_\lambda := \{z \in \mathcal{H} : \|z\| \leq \lambda\}$, where

$$\Gamma_\lambda^{A,B} := J_\lambda^B(Id+\lambda B)^{-1}(Id-\lambda A),\ \lambda \geq 0.$$

Lemma 2.4.5. [50, Lemma 2.5] *Let $\{S_k\}_{k\in\mathbb{N}}$ be a sequence of nonnegative real numbers satisfying the following inequalities*

$$S_{k+1} \leq (1-\rho_k)S_k + \rho_k\sigma_k \quad \forall k \geq 1 \text{ and } S_{k+1} \leq S_k - \eta_k + \pi_k \quad \forall k \geq 1,$$

where $\{\rho_k\}_{k\in\mathbb{N}}$ is a sequence in $(0,1)$, $\{\eta_k\}_{k\in\mathbb{N}}$ is a sequence of nonnegative real number, $\{\sigma_k\}_{k\in\mathbb{N}}$ and $\{\pi_k\}_{k\in\mathbb{N}}$ are real sequences such that

(i) $\sum_{k=1}^{+\infty}\rho_k = \infty$;

(ii) $\lim_{k\to+\infty}\pi_k = 0$;

(iii) $\lim_{i\to+\infty}\eta_{k_i} = 0$ *implies* $\limsup_{i\to+\infty}\sigma_{k_i} \leq 0$ *for any subsequence* $\{\eta_{k_i}\}_{i\in\mathbb{N}}$ *of* $\{\eta_k\}_{k\in\mathbb{N}}$.

Then the sequence $\{S_k\}_{k\in\mathbb{N}}$ converges to 0.

Lemma 2.4.6. [77] *Let A be an μ-inverse strongly monotone operator from a real Hilbert space $\mathcal{H}$ into itself and $B : \mathcal{H} \to 2^{\mathcal{H}}$ a maximal nomotone operator. Then, the following conditions hold.*

(i) *For* $\lambda > 0$, $\mathbf{Fix}(\Gamma_\lambda^{A,B}) = (A+B)^{-1}(0)$;

(ii) *For* $0 < \delta \leq \lambda$ *and* $x \in \mathcal{H}$, $\|x - \Gamma_\delta^{A,B}x\| \leq 2\|x - \Gamma_\lambda^{A,B}x\|$.

CHAPTER III

THE CONSTRAINED CONVEX OPTIMIZATION PROBLEMS

In this chapter, we study and propose iterative methods of solving the constrained convex optimization problem for both smooth and nonsmooth cases.

Mathematically, the constrained convex optimization problem which was introduced by Attouch and Czarnecki [2] deals with

$$\min_{x \in \arg\min g} f(x), \tag{3.0.2}$$

where $f : \mathcal{H} \to \mathbb{R}$ is a convex lower semicontinuous function and $g : \mathcal{H} \to \mathbb{R}$ is a convex and Fréchet differentiable function on a Hilbert space $\mathcal{H}$. We denote $\mathcal{S} := \arg\min\{f(x) : x \in \arg\min g\}$ the set of all solutions of the problem (3.0.2). In order to find a solution of the problem (3.0.2) in nonsmooth objective function, we mention that, in 2011, Attouch et al. [4] applied the forward-backward method to offer the so-called diagonal forward-backward algorithm, which is defined by

$$x_{k+1} = (Id + \lambda_k \partial f)^{-1}(x_k - \lambda_k \beta_k \nabla g(x_k)), \quad \forall k \geq 1,$$

where an initial point $x_1 \in \mathcal{H}$ and $\{\lambda_k\}_{k \in \mathbb{N}}$ and $\{\beta_k\}_{k \in \mathbb{N}}$ are sequences of positive real numbers. Under inf-compactness assumption on the functions f or g and some appropriate assumptions of parameters sequences, the convergence results of the sequence $\{x_k\}_{k \in \mathbb{N}}$ to a solution of the considered constrained convex optimization problem was presented in the paper.

To obtain the convergence results of the sequence $\{x_k\}_{k \in \mathbb{N}}$ without inf-compactness assumption to solve the problem (3.0.2) in smooth objective function, Peypouquet [5] applied the gradient method with a general exterior penalization scheme to offer the so-called *diagonal gradient scheme* (**DGS**), which is defined by

$$x_{k+1} = x_k - \lambda_k \nabla f(x_k) - \lambda_k \beta_k \nabla g(x_k), \quad \forall k \geq 1,$$

where an initial point $x_1 \in \mathcal{H}$ and $\{\lambda_k\}_{k \in \mathbb{N}}$ and $\{\beta_k\}_{k \in \mathbb{N}}$ are sequences of positive real numbers.

In 2017, Bot et al. [33] proposed a new algorithm called *gradient-type penalty* *with inertial effects method* (**GPIM**) for solving the problem (3.0.2) in smooth objective function. For given points $x_0, x_1 \in \mathcal{H}$, generate a sequence $\{x_k\}_{k\in\mathbb{N}}$ by the following iterative scheme

$$x_{k+1} = x_k + \alpha(x_k - x_{k-1}) - \lambda_k \nabla f(x_k) - \lambda_k \beta_k \nabla g(x_k), \quad \forall k \geq 1,$$

where $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{\beta_k\}_{k\in\mathbb{N}}$ are sequences of positive real numbers and $\alpha \in [0, 1)$. He also proved the weak convergence of the sequence produced by the above procedure to a solution of the problem (3.0.2).

Inspired by the research works in this direction, we are interested in the development of the method for finding solutions of the considered problem. While in Section 3.1 we focus on the constrained convex optimization in the case when the objective function is differentiable, in Section 3.2 we consider the case that the objective function is not differentiable.

3.1 Gradient Method for Solving Constrained Convex Optimization Problem with Smooth Objective Function

In this section, we consider constrained convex optimization problem (3.0.2) in the case that f and g are convex Fréchet differentiable and gradient Lipschitz continuous functions with constants L_f and L_g, respectively.

We assume that the solution set S is a nonempty set. Further, without loss of generality, we may assume that $\min g = 0$, that is $g(x) = 0$ for all $x \in \arg\min g$.

We wish to establish the algorithm called *rapid gradient penalty algorithm* (**RGPA**) for solving (3.0.2) which is generated by a controlling sequence of scalars together with the gradient of objective and penalization function. The iterative method for solving Problem (3.0.2) is presented as follows.

Algorithm 1: (RGPA)

Initialization: Given $\{\alpha_k\}_{k\in\mathbb{N}} \subseteq (0,1)$ and two positive sequences $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{\beta_k\}_{k\in\mathbb{N}}$. Choose $x_1 \in \mathcal{H}$ arbitrarily.

Iterative Steps: For a given current iterate $x_k \in \mathcal{H}$, calculate as follows:

Step 1. Compute y_k as

$$y_k = x_k - \lambda_k \nabla f(x_k) - \lambda_k \beta_k \nabla g(x_k).$$

Step 2. Compute

$$x_{k+1} = y_k + \alpha_k(y_k - x_k).$$

Update k := k + 1 and return to Step 1.

For $k \in \mathbb{N}$, we write $\Omega_k := f + \beta_k g$, which is also Fréchet differentiable function. Therefore, $\nabla\Omega_k$ is Lipschitz continuous with constant $L_k := L_f + \beta_k L_g$. In particular, if we setting $\alpha_k = 0$ for all $k \geq 1$, the Algorithm 1 can be reduced to (**DGS**) in Peypouquet [5].

We recall definitions of ℓ_1 and ℓ_2 spaces which are needed in main assumptions as follows:

$$\ell_1 = \{x := (\xi_1, \xi_2, \ldots) \subseteq \mathbb{R} \mid \sum_{n=1}^{+\infty} |\xi_n| < +\infty\}$$

and

$$\ell_2 = \{x := (\xi_1, \xi_2, \ldots) \subseteq \mathbb{R} \mid \sum_{n=1}^{+\infty} |\xi_n|^2 < +\infty\}.$$

In order to analyze the main convergence theorem, we present main assumptions, which is indicating some important properties of the sequences generated by Algorithm 1.

Assumption 3.1.1. (I) The function f is bounded from below;

(II) There exists a positive $K > 0$ such that

$$\beta_{k+1} - \beta_k \leq K\lambda_{k+1}\beta_{k+1},\ \frac{L_k}{2} - \frac{1}{2\lambda_k} \leq -K$$

and $\frac{\alpha_k^2 - 1}{2\lambda_k} + (1+\alpha_k)^2 K < 0$ for all $k \geq 1$;

(III) $\{\alpha_k\}_{k\in\mathbb{N}} \in \ell_2 \setminus \ell_1$, $\sum_{k=1}^{\infty} \lambda_k = +\infty$ and $\liminf_{k\to\infty} \lambda_k \beta_k > 0$;

(IV) For each $p \in \operatorname{ran}(N_{\arg\min g})$, we have

$$\sum_{k=1}^{\infty} \lambda_k \beta_k \left[g^* \left(\frac{p}{\beta_k} \right) - \sigma_{\arg\min g} \left(\frac{p}{\beta_k} \right) \right] < +\infty.$$

Remark 3.1.2. The conditions in Assumption 3.1.1 sparsely extend of the hypotheses in [5]. The differences are given by the second and third inequality in (II), which here involves a sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ which controls the inertial terms, and by $\{\alpha_k\}_{k\in\mathbb{N}} \in l_2 \setminus l_1$.

The following remark, we present some situations where Assumption 3.1.1 is verified.

Remark 3.1.3. Let $K > 0$, $q \in (0,1)$, $\delta > 0$ and $\gamma \in (0, \frac{1}{3L_g})$ be any given. Then we set $\alpha_k := \frac{1}{k+1}$ for all $k \geq 1$, which implies that $\lim_{k\to+\infty} \alpha_k = 0$, $\sum_{k=1}^{+\infty} \alpha_k^2 < +\infty$ and $\alpha_k \leq \frac{1}{2}$ for all $k \geq 1$. We also set

$$\beta_k := \frac{3\gamma[L_f + 2(K+\delta)]}{1 - 3\gamma L_g} + \gamma K k^q \quad \text{and} \quad \lambda_k := \frac{\gamma}{\beta_k} \quad \text{for all } k \geq 1.$$

Since $\beta_k \geq \frac{3\gamma[L_f+2(K+\delta)]}{1-3\gamma L_g}$, we have for each $k \geq 1$

$$\beta_k(1 - 3\gamma L_g) \geq 3\gamma[L_f + 2(K+\delta)].$$

It follows that

$$\frac{1}{3\lambda_k} - \beta_k L_g \geq L_f + 2(K+\delta) \quad \text{for all } k \geq 1,$$

which implies that

$$-(K+\delta) \geq \frac{L_k}{2} - \frac{1}{6\lambda_k} \quad \text{for all } k \geq 1. \tag{3.1.1}$$

According to (3.1.1), we obtain that

$$-K \geq \frac{L_k}{2} - \frac{1}{2\lambda_k} \quad \text{and} \quad \frac{1}{3} > 2\lambda_k K \quad \text{for all } k \geq 1.$$

Let us consider, for each $k \geq 1$

$$\frac{\alpha_k^2 - 1}{2\lambda_k} + (1+\alpha_k)^2 K \leq \frac{-\frac{3}{4} + \frac{9}{4} 2\lambda_k K}{2\lambda_k} < \frac{-\frac{3}{4} + \frac{3}{4}}{2\lambda_k} = 0.$$

On the other hand,

$$\beta_{k+1} - \beta_k = \gamma K[(k+1)^q - k^q] \leq \gamma K = K\lambda_{k+1}\beta_{k+1}.$$

Hence, we can conclude that Assumption 3.1.1 (II) holds.

Since $q \in (0,1)$, we obtain that $\sum_{k=1}^{+\infty} \frac{1}{\beta_k} = +\infty$, so $\sum_{k=1}^{+\infty} \lambda_k = +\infty$. Notice that $\lambda_k\beta_k = \gamma$ for all $k \geq 1$. It follows that $\liminf_{k\to+\infty} \lambda_k\beta_k = \liminf_{k\to+\infty} \gamma > 0$. Thus Assumption 3.1.1 (III) holds.

Finally, since $g^* - \sigma_{\arg\min g} \geq 0$. If $g(x) \geq \frac{\kappa}{2}\operatorname{dist}^2(x, \arg\min g)$ where $\kappa > 0$, then $g^*(x) - \sigma_{\arg\min g}(x) \leq \frac{1}{2\kappa}\|x\|^2$ for all $x \in \mathcal{H}$.

Therefore, for each $p \in \operatorname{ran}(N_{\arg\min g})$, we obtain that

$$\lambda_k\beta_k \left[g^*\left(\frac{p}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{p}{\beta_k}\right)\right] \leq \frac{\lambda_k}{2k\beta_k}\|p\|^2.$$

Thus, $\sum_{k=1}^{+\infty} \lambda_k\beta_k \left[g^*\left(\frac{p}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{p}{\beta_k}\right)\right]$ converges, if $\sum_{k=1}^{+\infty} \frac{\lambda_k}{\beta_k}$ converges, which is equivalently to $\sum_{k=1}^{+\infty} \frac{1}{\beta_k^2}$ converges. This holds for the above choices of $\{\beta_k\}_{k\in\mathbb{N}}$ and $\{\lambda_k\}_{k\in\mathbb{N}}$ when $q \in (\frac{1}{2}, 1)$.

Convergence Analysis

In this part, we present the convergence of the sequence of $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 and of the sequence of objective values $\{f(x_k)\}$.

We start the convergence analysis of this section with technical lemmas.

Lemma 3.1.4. *Let x^* be an arbitrary element in $\mathcal{S}$ and set $p^* := -\nabla f(x^*)$. Then for each $k \geq 1$*

$$\begin{aligned}\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 + (1+\alpha_k)\lambda_k\beta_k g(x_k) &\\ \leq (1+\alpha_k)^2\|x_k - y_k\|^2 &\\ + (1+\alpha_k)\lambda_k\beta_k \left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right]. &\qquad (3.1.2)\end{aligned}$$

Proof. Applying to the first-order optimality condition, we have

$$0 \in \nabla f(x^*) + N_{\arg\min g}(x^*).$$

It follows that

$$p^* = -\nabla f(x^*) \in N_{\arg\min g}(x^*).$$

Note that for each $k \geq 1$, $\frac{x_k - y_k}{\lambda_k} - \beta_k \nabla g(x_k) = \nabla f(x_k)$.

By monotoncity of ∇f, we obtain that

$$\left\langle \frac{x_k - y_k}{\lambda_k} - \beta_k \nabla g(x_k) + p^*, x_k - x^* \right\rangle = \langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle$$
$$\geq 0 \qquad \forall k \geq 1,$$

and hence, for each $k \geq 1$

$$2 \langle x_k - y_k, x_k - x^* \rangle \geq 2\lambda_k \beta_k \langle \nabla g(x_k), x_k - x^* \rangle - 2\lambda_k \langle p^*, x_k - x^* \rangle. \qquad (3.1.3)$$

Since g is convex and differentiable, we have for each $k \geq 1$

$$\langle \nabla g(x_k), x^* - x_k \rangle + g(x_k) \leq g(x^*) = 0,$$

whence

$$2\lambda_k \beta_k g(x_k) \leq 2\lambda_k \beta_k \langle \nabla g(x_k), x_k - x^* \rangle. \qquad (3.1.4)$$

On the other hand,

$$2\langle x_k - y_k, x_k - x^* \rangle = \|x_k - y_k\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2. \qquad (3.1.5)$$

Combining (3.1.3), (3.1.4) and (3.1.5), we get that

$$\|y_k - x^*\|^2 \leq \|x_k - y_k\|^2 + \|x_k - x^*\|^2 - 2\lambda_k \beta_k g(x_k) + 2\lambda_k \langle p^*, x_k - x^* \rangle. \quad (3.1.6)$$

Since $x^* \in \mathcal{S}$ and $p^* \in N_{\arg\min g}(x^*)$, we have

$$\sigma_{\arg\min g}(p^*) = \langle p^*, x^* \rangle.$$

From the inequality (3.1.6), we observe that

$$2\lambda_k \langle p^*, x_k - x^* \rangle - \lambda_k \beta_k g(x_k) = 2\lambda_k \langle p^*, x_k \rangle - \lambda_k \beta_k g(x_k) - 2\lambda_k \langle p^*, x^* \rangle$$

$$= \lambda_k\beta_k\left[\left\langle \frac{2p^*}{\beta_k}, x_k\right\rangle - g(x_k) - \left\langle \frac{2p^*}{\beta_k}, x^*\right\rangle\right]$$
$$\leq \lambda_k\beta_k\left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right]. \tag{3.1.7}$$

Combining (3.1.7) and (3.1.6), we obtain that

$$\|y_k - x^*\|^2 \leq \|x_k - y_k\|^2 + \|x_k - x^*\|^2 - \lambda_k\beta_k g(x_k)$$
$$+ \lambda_k\beta_k\left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right]. \tag{3.1.8}$$

On the other hand, we observe that

$$\|x_{k+1} - x^*\|^2 = \|y_k + \alpha_k(y_k - x_k) - x^*\|^2$$
$$= \|(1+\alpha_k)(y_k - x^*) + \alpha_k(x^* - x_k)\|^2$$
$$= (1+\alpha_k)\|y_k - x^*\|^2 - \alpha_k\|x_k - x^*\|^2$$
$$+ \alpha_k(1+\alpha_k)\|x_k - y_k\|^2. \tag{3.1.9}$$

By (3.1.8) and (3.1.9), we obtain the desired result. □

Lemma 3.1.5. *For all $k \geq 1$, we have*

$$\Omega_{k+1}(x_{k+1}) \leq \Omega_k(x_k) + (\beta_{k+1} - \beta_k)g(x_{k+1}) + \frac{\alpha_k^2 - 1}{2\lambda_k}\|y_k - x_k\|^2$$
$$+ \left[\frac{L_k}{2} - \frac{1}{2\lambda_k}\right]\|x_{k+1} - x_k\|^2.$$

Proof. Since $\nabla\Omega$ is L_k-Lipschitz continuous and by Descent Lemma (see [66, Theorem 18.15]), we obtain that

$$\Omega_k(x_{k+1}) \leq \Omega_k(x_k) + \langle \nabla\Omega_k(x_k), x_{k+1} - x_k\rangle + \frac{L_k}{2}\|x_{k+1} - x_k\|^2.$$

Recall that $-\frac{y_k - x_k}{\lambda_k} = \nabla\Omega_k(x_k)$.

It follows that

$$f(x_{k+1}) + \beta_k g(x_{k+1})$$
$$\leq f(x_k) + \beta_k g(x_k) - \left\langle \frac{y_k - x_k}{\lambda_k}, x_{k+1} - x_k\right\rangle + \frac{L_k}{2}\|x_{k+1} - x_k\|^2$$
$$= f(x_k) + \beta_k g(x_k) - \frac{1}{2\lambda_k}\|y_k - x_k\|^2 - \frac{1}{2\lambda_k}\|x_{k+1} - x_k\|^2$$

$$+\frac{1}{2\lambda_k}\|y_k - x_{k+1}\|^2 + \frac{L_k}{2}\|x_{k+1} - x_k\|^2$$
$$= f(x_k) + \beta_k g(x_k) + \frac{\alpha_k^2 - 1}{2\lambda_k}\|y_k - x_k\|^2 + \left[\frac{L_k}{2} - \frac{1}{2\lambda_k}\right]\|x_{k+1} - x_k\|^2.$$

Adding $\beta_{k+1}g(x_{k+1})$ to both sides, we have

$$f(x_{k+1}) + \beta_{k+1}g(x_{k+1}) \leq f(x_k) + \beta_k g(x_k) + (\beta_{k+1} - \beta_k)g(x_{k+1})$$
$$+ \frac{\alpha_k^2 - 1}{2\lambda_k}\|y_k - x_k\|^2 + \left[\frac{L_k}{2} - \frac{1}{2\lambda_k}\right]\|x_{k+1} - x_k\|^2,$$

which means that

$$\Omega_{k+1}(x_{k+1}) \leq \Omega_k(x_k) + (\beta_{k+1} - \beta_k)g(x_{k+1}) + \frac{\alpha_k^2 - 1}{2\lambda_k}\|y_k - x_k\|^2$$
$$+ \left[\frac{L_k}{2} - \frac{1}{2\lambda_k}\right]\|x_{k+1} - x_k\|^2.$$

□

Let $K > 0$ be such that Assumption 3.1.1 holds. For $k \geq 1$ and $x^* \in \mathcal{S}$, we denote by

$$\Lambda_k := f(x_k) + (1 - (1 + \alpha_k)K\lambda_k)\beta_k g(x_k) + K\|x_k - x^*\|^2$$
$$= \Omega_k(x_k) - (1 + \alpha_k)K\lambda_k\beta_k g(x_k) + K\|x_k - x^*\|^2.$$

Lemma 3.1.6. *Let $x^* \in \mathcal{S}$ and set $p^* := -\nabla f(x^*)$. Assume that Assumption 3.1.1 holds. Then there exists $\theta > 0$ such that for each $k \geq 1$*

$$\Lambda_{k+1} - \Lambda_k + \theta\|y_k - x_k\|^2 \leq (1 + \alpha_k)K\lambda_k\beta_k\left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right].$$

Proof. From Lemma 3.1.5 and Assumption 3.1.1 (II), we obtain that

$$\Omega_{k+1}(x_{k+1}) - \Omega_k(x_k) \leq K\lambda_{k+1}\beta_{k+1}g(x_{k+1}) + \frac{\alpha_k^2 - 1}{2\lambda_k}\|y_k - x_k\|^2$$
$$\leq (1 + \alpha_{k+1})K\lambda_{k+1}\beta_{k+1}g(x_{k+1})$$
$$+ \frac{\alpha_k^2 - 1}{2\lambda_k}\|y_k - x_k\|^2. \tag{3.1.10}$$

On the other hand, multiplying (3.1.2) by K, we have

$$K\|x_{k+1} - x^*\|^2 - K\|x_k - x^*\|^2 + (1 + \alpha_k)K\lambda_k\beta_k g(x_k)$$

$$\leq (1+\alpha_k)^2 K\|x_k - y_k\|^2 + (1+\alpha_k)K\lambda_k\beta_k\left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right]. \tag{3.1.11}$$

Combining (3.1.10) and (3.1.11), we have

$$\Lambda_{k+1} - \Lambda_k \leq \left[\frac{\alpha_k^2 - 1}{2\lambda_k} + (1+\alpha_k)^2 K\right]\|y_k - x_k\|^2 + (1+\alpha_k)K\lambda_k\beta_k\left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right]. \tag{3.1.12}$$

For each $k \geq 1$, $\frac{\alpha_k^2-1}{2\lambda_k} + (1+\alpha_k)^2 K < 0$, we have there exists $\theta > 0$ such that

$$\frac{\alpha_k^2 - 1}{2\lambda_k} + (1+\alpha_k)^2 K < -\theta.$$

From (3.1.12), we have

$$\Lambda_{k+1} - \Lambda_k + \theta\|y_k - x_k\|^2 \leq (1+\alpha_k)K\lambda_k\beta_k\left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right].$$

This completes the proof. □

Lemma 3.1.7. *Assume that Assumption 3.1.1 holds. Then the following statements hold:*

(i) *The sequence $\{\Lambda_k\}$ is bounded from below and $\lim_{k\to+\infty} \Lambda_k$ exists;*

(ii) $\sum_{k=1}^{+\infty} \|y_k - x_k\|^2 < +\infty$;

(iii) *Let $x^* \in \mathcal{S}$. Then $\lim_{k\to+\infty} \|x_k - x^*\|^2$ exists and $\sum_{k=1}^{+\infty} \lambda_k\beta_k g(x_k) < +\infty$;*

(iv) $\lim_{k\to+\infty} \Omega_k(x_k)$ *exists;*

(v) $\lim_{k\to+\infty} g(x_k) = 0$ *and every weak cluster point of the sequence $\{x_k\}_{k\in\mathbb{N}}$ belongs to* $\arg\min g$.

Proof. We set $p^* := -\nabla f(x^*)$.

(i). From Assumption 3.1.1 (II) implies $1-(1+\alpha_k)K\lambda_k \geq 0$. Since f is convex and differentiable, we have for each $k \geq 1$

$$\begin{aligned}
\Lambda_k &= f(x_k) + (1-(1+\alpha_k)K\lambda_k)\beta_k g(x_k) + K\|x_k - x^*\|^2 \\
&\geq f(x_k) + K\|x_k - x^*\|^2 \\
&\geq f(x^*) + \langle \nabla f(x^*), x_k - x^* \rangle + K\|x_k - x^*\|^2 \\
&= f(x^*) - \left\langle \frac{p^*}{\sqrt{2K}}, \sqrt{2K}(x_k - x^*) \right\rangle + K\|x_k - x^*\|^2 \\
&\geq f(x^*) - \frac{\|p^*\|^2}{4K} - K\|x_k - x^*\|^2 + K\|x_k - x^*\|^2 \\
&= f(x^*) - \frac{\|p^*\|^2}{4K}.
\end{aligned}$$

Therefore, $\{\Lambda_k\}$ is bounded from below.

Next, we set $\gamma_n = \Lambda_k$, $\delta_k = \theta\|y_k - x_k\|^2$ and

$$\varepsilon_k = (1+\alpha_k)K\lambda_k\beta_k \left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right].$$

Recall that $\min g = 0$. Thus $g \leq \delta_{\arg\min g}$. Therefore $\sigma_{\arg\min g} = (\delta_{\arg\min g})^* \leq g^*$ and hence, $g^* - \sigma_{\arg\min g} \geq 0$. It follows that

$$\begin{aligned}
\varepsilon_k &= (1+\alpha_k)K\lambda_k\beta_k \left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right] \\
&\leq 2K\lambda_k\beta_k \left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right].
\end{aligned}$$

By using Assumption 3.1.1 (IV) and $p^* \in N_{\arg\min g}(x^*)$, we have $\sum_{k=1}^{+\infty} \varepsilon_k < +\infty$. Applying Lemma 3.1.6 and Lemma 2.4.2, we obtain that $\lim_{k\to+\infty} \Lambda_k$ exists.

(ii). Follows immediately from Lemmas 3.1.6 and 2.4.2.

(iii). We set $\gamma_k = \|x_k - x^*\|^2$, $\delta_k = (1+\alpha_k)\lambda_k\beta_k g(x_k)$ and

$$\varepsilon_k = (1+\alpha_k)^2\|y_k - x_k\|^2 + (1+\alpha_k)\lambda_k\beta_k \left[g^*\left(\frac{2p^*}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{2p^*}{\beta_k}\right)\right].$$

From statement (ii), Lemma 2.4.2 and Lemma 3.1.4, we get that

$$\lim_{k\to+\infty} \|x_k - x^*\| \ \text{ exists and } \sum_{k=1}^{+\infty} \lambda_k\beta_k g(x_k) < +\infty.$$

For (iv) since for each $k \geq 1$ $\Omega_k(x_k) = \Lambda_k + (1+\alpha_k)K\lambda_k\beta_k g(x_k) - K\|x_k - x^*\|^2$, by using (i), (iii) and $\lim_{k\to+\infty} \alpha_k = 0$, we have $\lim_{k\to+\infty} \Omega_k(x_k)$ exists.

(v). It follows from Assumption 3.1.1 (III) that $\liminf_{k\to+\infty} \lambda_k\beta_k > 0$. According to statement (iii) implies $\lim_{k\to+\infty} g(x_k) = 0$. Let $\overline{x}$ be any weak cluster point of the sequence $\{x_k\}_{k\in\mathbb{N}}$. Therefore, there exists subsequence $\{x_{n_k}\}$ of $\{x_k\}_{k\in\mathbb{N}}$ converges weakly to $\overline{x}$ as $k \to +\infty$. By the weak lower semicontinuity of g, we get that

$$g(\overline{x}) \leq \liminf_{k\to+\infty} g(x_{n_k}) \leq \lim_{k\to+\infty} g(x_k) = 0,$$

which means that $\overline{x} \in \arg\min g$. This completes the proof. □

Lemma 3.1.8. *Let $x^* \in S$. Assume that Assumption 3.1.1 holds. Then*

$$\sum_{k=1}^{+\infty} \lambda_k \left[\Omega_k(x_k) - f(x^*)\right] < +\infty.$$

Proof. Since f is differentiable and convex, we obtain that for each $k \geq 1$

$$f(x^*) \geq f(x_k) + \langle \nabla f(x_k), x^* - x_k \rangle.$$

Since g is differentiable, convex and $\min g = 0$, we obtain that for each $k \geq 1$

$$0 = g(x^*) \geq g(x_k) + \langle \nabla g(x_k), x^* - x_k \rangle,$$

which implies that

$$0 \geq \beta_k g(x_k) + \langle \beta_k \nabla g(x_k), x^* - x_k \rangle, \quad \text{for all } k \geq 1.$$

Therefore, we can conclude that

$$\begin{aligned} f(x^*) &\geq \Omega_k(x_k) + \langle \nabla \Omega_k(x_k), x^* - x_k \rangle \\ &= \Omega_k(x_k) + \left\langle \frac{x_k - y_k}{\lambda_k}, x^* - x_k \right\rangle. \end{aligned} \tag{3.1.13}$$

From (3.1.13), we obtain that

$$\begin{aligned} 2\lambda_k \left[\Omega_k(x_k) - f(x^*)\right] &\leq 2\langle x_k - y_k, x_k - x^* \rangle \\ &= \|x_k - y_k\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2. \end{aligned} \tag{3.1.14}$$

On the other hand, for each $k \geq 1$

$$\begin{aligned}\|y_k - x^*\|^2 &= \|x_{k+1} - \alpha_k(y_k - x_k) - x^*\|^2 \\ &= \|x_{k+1} - x^*\|^2 + \alpha_k^2\|y_k - x_k\|^2 - 2\langle \alpha_k(x_{k+1} - x^*), y_k - x_k\rangle \\ &= \|x_{k+1} - x^*\|^2 + \alpha_k^2\|y_k - x_k\|^2 - \alpha_k^2\|x_{k+1} - x^*\|^2 - \|y_k - x_k\|^2 \\ &\quad + \|\alpha_k(x_{k+1} - x^*) - (y_k - x_k)\|^2 \\ &\geq \|x_{k+1} - x^*\|^2 + \alpha_k^2\|y_k - x_k\|^2 - \alpha_k^2\|x_{k+1} - x^*\|^2 - \|y_k - x_k\|^2,\end{aligned}$$

which implies that

$$\begin{aligned}-\|y_k - x^*\|^2 \leq -\|x_{k+1} - x^*\|^2 - \alpha_k^2\|y_k - x_k\|^2 + \alpha_k^2\|x_{k+1} - x^*\|^2 \\ + \|y_k - x_k\|^2. \qquad (3.1.15)\end{aligned}$$

Combining (3.1.14) and (3.1.15), we have for all $k \geq 1$

$$\begin{aligned}2\lambda_k\left[\Omega_k(x_k) - f(x^*)\right] &\leq (2 - \alpha_k^2)\|x_k - y_k\|^2 + \|x_k - x^*\|^2 \\ &\quad - \|x_{k+1} - x^*\|^2 + \alpha_k^2\|x_{k+1} - x^*\|^2 \\ &\leq 2\|x_k - y_k\|^2 + \|x_k - x^*\|^2 \\ &\quad - \|x_{k+1} - x^*\|^2 + \alpha_k^2\|x_{k+1} - x^*\|^2.\end{aligned}$$

Therefore, according Lemma 3.1.7 (iii), we get that the sequence $\{\|x_k - x^*\|\}$ is bounded, which means that there exists $M > 0$ such that $\|x_k - x^*\| \leq M$ for all $k \geq 1$.

By Assumption 3.1.1 (III) and Lemma 3.1.7, we obtain that

$$2\sum_{k=1}^{+\infty}\lambda_k[\Omega_k(x_k) - f(x^*)] \leq 2\sum_{k=1}^{+\infty}\|y_k - x_k\|^2 + \|x_1 - x^*\|^2 + M^2\sum_{k=1}^{+\infty}\alpha_k^2 < +\infty.$$

□

We perform our main convergence theorem as follows.

Theorem 3.1.9. *Let $\{x_k\}_{k\in\mathbb{N}}$ be define by Algorithm 1. Assume that Assumption 3.1.1 holds. Then $\{x_k\}_{k\in\mathbb{N}}$ converges weakly to a point in $\mathcal{S}$. Moreover, the sequence $\{f(x_k)\}_{k\in\mathbb{N}}$ converges to the optimal objective value of the problem* (3.0.2).

Proof. From Lemma 3.1.7 (iii), $\lim_{k\to+\infty}\|x_k - x^*\|$ exists for all $x^* \in \mathcal{S}$. Let $\overline{x}$ be any weak cluster point of $\{x_k\}_{k\in\mathbb{N}}$. Then there exists a subsequence $\{x_{k_i}\}_{i\in\mathbb{N}}$ of $\{x_k\}_{k\in\mathbb{N}}$ such that $\{x_{k_i}\}_{i\in\mathbb{N}}$ converges weakly to $\overline{x}$ as $k \to +\infty$. According to Lemma 3.1.7 (v) implies $\overline{x} \in \arg\min g$. It suffices to show that $f(\overline{x}) \leq f(x)$ for all $x \in \arg\min g$. Since $\sum_{k=1}^{+\infty}\lambda_k = +\infty$, and by Lemma 3.1.8 and Lemma 3.1.7 (iv), we have

$$\lim_{k\to+\infty}\Omega_k(x_k) - f(x^*) \leq 0 \quad \text{for all } x^* \in \mathcal{S}.$$

Therefore, $f(\overline{x}) \leq \liminf_{i\to+\infty} f(x_{k_i}) \leq \lim_{k\to+\infty}\Omega_k(x_k) \leq f(x^*),\ \ \forall x^* \in \mathcal{S}$, which implies that $\overline{x} \in \mathcal{S}$. Applying Lemma 2.4.3 (Opial Lemma), we obtain that $\{x_k\}_{k\in\mathbb{N}}$ converges weakly to a point in $\mathcal{S}$. The last statement follows immediately from the above. □

Numerical Experiments

In this section, we present the performance of the Algorithm 1 for the minimization problem with linear equality constraints. We consider the following problem, say, the minimization problem with linear equality constraints.

$$\begin{aligned}&\text{minimize } \frac{1}{2}\|x\|^2\\ &\text{subject to } \mathbf{A}x = \mathbf{b},\end{aligned} \tag{3.1.16}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$ with $n > m$.
The problem (3.1.16) is equivalenct to the following problem.

$$\begin{aligned}&\text{minimize } \frac{1}{2}\|x\|^2\\ &\text{subject to } x \in \arg\min \frac{1}{2}\|\mathbf{A}(\cdot) - \mathbf{b}\|^2,\end{aligned} \tag{3.1.17}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$ with $n > m$.

We set $f(x) := \frac{1}{2}\|x\|^2$ and $g(x) := \frac{1}{2}\|\mathbf{A}(x) - \mathbf{b}\|^2$. In this setting, we have $\nabla f(x) = x$ and notice that ∇f is 1-Lipschitz continuous. Furthermore, we get that $\nabla g(x) = \mathbf{A}^\top(\mathbf{A}x - \mathbf{b})$ and notice that ∇g is $\|\mathbf{A}\|^2$-Lipschitz continuous.

We begin with the problem by random matrix $\mathbf{A}$ in $\mathbb{R}^{m\times n}$, vector $x_1 \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$ with $m = 1000$ and $n = 4000$ generated by using MATLAB, which the entries of $\mathbf{A}$, x_1 and $\mathbf{b}$ are integer in [-10,10]. Next, we are going to compare a performance of three algorithms consisting of Algorithms 1, (**DGS**) and (**GPIM**). The choice of the parameters for the computational experiment in both Algorithms 1 and (**DGS**) is based on Remark 3.1.3. We choose $\gamma = \frac{1}{4\|\mathbf{A}\|^2}$ and $\delta = 1$. We consider different choices of the parameters $K \in (0,1]$ and $q \in (\frac{1}{2}, 1)$. The choice of the parameters for (**GPIM**) is based on Remark 12 in [33] with $\alpha = 0.001$. We obtain the CPU times (seconds) and the number of iterations by using the stopping criteria :

$$\|x_k - x_{k-1}\| \leq 10^{-6}.$$

Table 1: Comparison of the convergence of Algorithm 1 and (**DGS**) for the parameters $K = 0.001$ and $q \in (\frac{1}{2}, 1)$.

q	Algorithm 1		(DGS)	
	Time (sec)	#(Iters)	Time (sec)	#(Iters)
0.6	2.38	566	10.23	2221
0.7	2.31	568	107.78	25336
0.8	2.46	581	384.00	90636
0.9	44.96	11458	447.11	103487

In Table 1 we present a comparison of the convergence between two algorithms including Algorithm 1 and (**DGS**) for the parameters $K = 0.001$ and different choices for the parameters $q \in (\frac{1}{2}, 1)$. We observe that when $q = 0.6$ leads to lowest computation time for Algorithm 1 and (**DGS**) with 2.38 second and 10.23 second, respectively. Furthermore, we also observe that (**DGS**) hit a big number of iterations than Algorithm 1 for all choices of parameter q.

Table 2: Comparison of the convergence of Algorithm 1 and **(DGS)** for the parameters $q = 0.6$ and $K \in (0, 1]$.

K	Algorithm 1		**(DGS)**	
	Time (sec)	#(Iters)	Time (sec)	#(Iters)
0.001	2.38	566	10.23	2221
0.005	2.40	585	171.46	40888
0.01	6.63	1612	254.93	64469
0.05	83.22	20480	288.39	65722
0.1	107.41	26257	212.02	52464
0.5	79.95	18606	100.33	24419
1	51.46	13414	67.20	16616

In Table 2 we present a comparison of the convergence of Algorithm 1 and **(DGS)** for the parameters $q = 0.6$ and $K \in (0, 1]$. We observe that the number of iterations and computation time for Algorithm 1 smaller than the number of iterations for **(DGS)** for each choices of parameters K. Furthermore, Algorithm 1 need tiny computation time to reach the optimality tolerance than **(DGS)** for each choices of parameter K.

Finally, we give the comparison of convergence for Algorithm 1, **(DGS)** and **(GPIM)** in Figure 1.

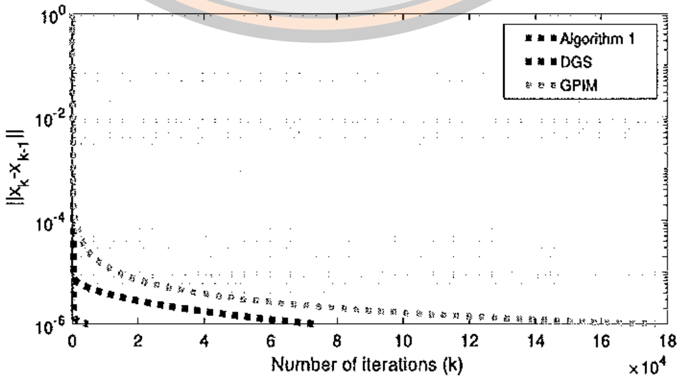


Figure 1: Illustration of the behavior of $\|x_k - x_{k-1}\|$ for Algorithm 1, **(DGS)** and **(GPIM)** when $q = 0.9$ and $(m, n) = (100, 400)$.

We observe that the our algorithm performs an adventage behavior when comparing with algorithm (**DGS**) for all different choices of parameters. Note that the number of iterations for (**RGPA**) smaller than the number of iterations for (**DGS**). Furthermore, Algorithm 1 need tiny computation time to reach the optimality tolerance than (**DGS**) for each different choices of parameters. Furthermore, the our algorithm performs an adventage behavior when comparing with (**DGS**) and (**GPIM**).

3.2 Forward-Backward Method for Solving Constrained Convex Optimization Problem with Nonsmooth Objective Function

In this section, we consider constrained convex optimization problem (3.0.2) in the case that $f : \mathcal{H} \to [-\infty, +\infty]$ is proper convex lower semicontinuous and $g : \mathcal{H} \to \mathbb{R}$ is Fréchet differentiable on the space $\mathcal{H}$ and gradient Lipschitz continuos functions with constants L_g. Assume that the solution set of the problem (3.0.2) is nonempty and some qualifications in [66, Proposition 27.8] hold. Then, problem (3.0.2) is equivalent to the following problem: find $x \in \mathcal{H}$ such that

$$0 \in \partial f(x) + N_{\arg\min g}(x). \tag{3.2.1}$$

In order to solve the problem (3.0.2), we are going to consider the following problem.

Find $x \in \mathcal{H}$ such that

$$0 \in A(x) + N_{(\mathrm{zer}(B))}(x), \tag{3.2.2}$$

where, $A : \mathcal{H} \to 2^{\mathcal{H}}$ is a maximally monotone operator and $B : \mathcal{H} \to \mathcal{H}$ is a cocoercive operator with parameter $\omega > 0$.

We recall the set of all zeros of the operator B denoted by $\mathrm{zer}(B) := \{z \in \mathcal{H} : 0 = B(z)\}$.

Note that if $A = \partial f$ and $B = \nabla g$, then the problem (3.2.1) is a special case of MIP (3.2.2).

The aim of this work is to employ the forward-backward penalty method to solve (3.2.2) from [3] with a new inertial effect. Inspired by the research works mentioned above, we wish to develop the algorithm called a new forward-backward penalty algorithm (**NFBP**) for solving (3.2.2) as follows:

Algorithm 2: (NFBP)

Initialization: Given three positive sequences $\{\alpha_k\}_{k\in\mathbb{N}}$, $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{\beta_k\}_{k\in\mathbb{N}}$. Choose $x_1 \in \mathcal{H}$ arbitrarily.

Iterative Steps: For a given current iterate $x_k \in \mathcal{H}$, calculate as follows:

Step 1. Compute y_k as

$$y_k = J^A_{\lambda_k}(x_k - \lambda_k\beta_k B(x_k)).$$

Step 2. Compute

$$x_{k+1} = y_k + \alpha_k(y_k - x_k).$$

Update k := k + 1 and return to Step 1.

The proposed numerical scheme can be reduced to the algorithm investigated in [4] which is called *forward-backward* method (**FB**) when $\alpha_k = 0$, $\forall k \geq 1$.

Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 2, $\{\lambda_k\}_{k\in\mathbb{N}}$ a sequence of positive real numbers and $\{z_k\}_{k\in\mathbb{N}}$ the sequence of weighted averages

$$z_k = \tfrac{1}{\tau_k}\sum_{n=1}^{k}\lambda_n x_n, \quad \text{where} \quad \tau_k = \sum_{n=1}^{k}\lambda_n. \tag{3.2.3}$$

We carry out the convergence analysis for new gradient penalty algorithm (**NFBP**) which is settled by the following hypotheses.

Assumption 3.2.1. (I) The qualification condition $\text{zer}(B) \cap \text{int dom}(A) \neq \emptyset$ holds.

(II) $\{\lambda_k\} \in \ell_2 \setminus \ell_1$, $\lim_{k\to+\infty}\alpha_k = 0$ and $0 < \liminf_{k\to+\infty}\lambda_k\beta_k \leq \limsup_{k\to+\infty}\lambda_k\beta_k < \omega$.

(III) For each $p \in \text{ran}(N_{\text{zer}(B)})$, we have

$$\sum_{k=1}^{+\infty}\lambda_k\beta_k\left[\sup_{x^*\in\text{zer}(B)} F_B\left(\tfrac{p}{\beta_k}, x^*\right) - \sigma_{\text{zer}(B)}\left(\tfrac{p}{\beta_k}\right)\right] < +\infty.$$

We present some siuations that satisfy the Assumption 3.2.1 as the following remark.

Remark 3.2.2. (i) Since A and $N_{\text{zer}(B)}$ are maximal monotone and Assumption 3.2.1 (I) holds, we obtain that $A+N_{\text{zer}(B)}$ is maximal monotone operator (see [66, Example 20.26 and Corollary 25.5]).

(ii) There are some examples satisfying Assumption 3.2.1 (II) e.g. sequences $\lambda_k \sim \frac{1}{k}$, $\beta_k \sim \overline{\omega}k$ for some $\overline{\omega} \in (0,\omega)$ and $\alpha_k \sim \frac{1}{k}$ for all $k \in \mathbb{N}$.

(iii) Assumption 3.2.1 (III) has already been used in [9] in order to show the convergence of the proposed algorithm (see [9, Assumption (H_{fitz})]). They also pointed out that for each $p \in \text{ran}(N_{\text{zer}(B)})$ and any $k \in \mathbb{N}$ one has $\sup_{x^* \in \text{zer}(B)} F_B\left(\frac{p}{\beta_k}, x^*\right) - \sigma_{\text{zer}(B)}\left(\frac{p}{\beta_k}\right) \geq 0$. Some examples of the operator B satisfying Assumption 3.2.1 (III) can be found in [78, Section 5].

Let us denote an arbitrary sequence verifying Algorithm 2 by $\{x_k\}_{k\in\mathbb{N}}$ and provide some estimations.

Lemma 3.2.3. *Let $x^* \in \text{zer}(B) \cap \text{dom}(A)$ and $v \in A(x^*)$. Then the following inequality holds for each $k \in \mathbb{N}$ and $\varepsilon \geq 0$*

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 &+ (1+\alpha_k)\left(\tfrac{2\varepsilon}{1+\varepsilon}\right)\lambda_k\beta_k\langle B(x_k), x_k - x^*\rangle \\
&+ (1+\alpha_k)\left(\tfrac{\varepsilon}{1+\varepsilon} - \alpha_k\right)\|y_k - x_k\|^2 \\
&\leq (1+\alpha_k)\lambda_k\beta_k\left((1+\varepsilon)\lambda_k\beta_k - \tfrac{2\omega}{1+\varepsilon}\right)\|B(x_k)\|^2 \\
&\quad + 2(1+\alpha_k)\lambda_k\langle v, x^* - y_k\rangle.
\end{aligned}
\tag{3.2.4}
$$

Proof. It is not hard to verify from 2 that for each $k \in \mathbb{N}$, $\frac{x_k - y_k}{\lambda_k} - \beta_k B(x_k) \in A(y_k)$. By the monotonicity of A and $v \in A(x^*)$,

$$\left\langle \tfrac{x_k - y_k}{\lambda_k} - \beta_k B(x_k) - v, y_k - x^* \right\rangle \geq 0.$$

It follows that

$$\langle x_k - y_k, x^* - y_k\rangle \leq \lambda_k\langle \beta_k B(x_k) + v, x^* - y_k\rangle, \text{ for all } k \in \mathbb{N}.$$

From Lemma 2.3.14 (i), we obtain that for each $k \in \mathbb{N}$

$$\|x_k - y_k\|^2 + \|x^* - y_k\|^2 - \|x_k - x^*\|^2 \leq 2\lambda_k\langle \beta_k B(x_k) + v, x^* - y_k\rangle, \tag{3.2.5}$$

which mean that

$$\|y_k - x^*\|^2 - \|x_k - x^*\|^2 + \|x_k - y_k\|^2 \leq 2\lambda_k\langle v, x^* - y_k\rangle$$
$$+ 2\lambda_k\beta_k\langle B(x_k), x^* - x_k\rangle$$
$$+ 2\lambda_k\beta_k\langle B(x_k), x_k - y_k\rangle. \qquad (3.2.6)$$

Note that B is ω-cocoercive and $B(x^*) = 0$, we have

$$2\lambda_k\beta_k\langle B(x_k), x^* - x_k\rangle \leq -2\omega\lambda_k\beta_k\|B(x_k)\|^2 \qquad (3.2.7)$$

for all $k \in \mathbb{N}$. From (3.2.7), we observe that

$$2\lambda_k\beta_k\langle B(x_k), x^* - x_k\rangle = \tfrac{1}{1+\varepsilon}2\lambda_k\beta_k\langle B(x_k), x^* - x_k\rangle$$
$$+ \tfrac{\varepsilon}{1+\varepsilon}2\lambda_k\beta_k\langle B(x_k), x^* - x_k\rangle$$
$$\leq -\tfrac{2\omega}{1+\varepsilon}\lambda_k\beta_k\|B(x_k)\|^2$$
$$+ \tfrac{2\varepsilon}{1+\varepsilon}\lambda_k\beta_k\langle B(x_k), x^* - x_k\rangle. \qquad (3.2.8)$$

For each $k \in \mathbb{N}$, let us consider

$$0 \leq \tfrac{1}{1+\varepsilon}\|y_k - x_k + (1+\varepsilon)\lambda_k\beta_k B(x_k)\|^2$$
$$= \tfrac{1}{1+\varepsilon}\|y_k - x_k\|^2 + (1+\varepsilon)\lambda_k^2\beta_k^2\|B(x_k)\|^2 + 2\lambda_k\beta_k\langle B(x_k), y_k - x_k\rangle,$$

which implies that

$$2\lambda_k\beta_k\langle B(x_k), x_k - y_k\rangle \leq \tfrac{1}{1+\varepsilon}\|y_k - x_k\|^2 + (1+\varepsilon)\lambda_k^2\beta_k^2\|B(x_k)\|^2. \qquad (3.2.9)$$

Joining (3.2.8) and (3.2.9) to (3.2.6) together with some simple calculations, we have that

$$\|y_k - x^*\|^2 \leq 2\lambda_k\langle v, x^* - y_k\rangle + \tfrac{2\varepsilon}{1+\varepsilon}\lambda_k\beta_k\langle B(x_k), x^* - x_k\rangle$$
$$+ \lambda_k\beta_k\left((1+\varepsilon)\lambda_k\beta_k - \tfrac{2\omega}{1+\varepsilon}\right)\|B(x_k)\|^2 + \|x_k - x^*\|^2$$
$$- \tfrac{\varepsilon}{1+\varepsilon}\|y_k - x_k\|^2. \qquad (3.2.10)$$

On the other hand, by using Lemma 2.3.14 (iv), we have the following equation

$$\|x_{k+1} - x^*\|^2 = \|y_k + \alpha_k(y_k - x_k) - x^*\|^2$$
$$= \|(1+\alpha_k)(y_k - x^*) - \alpha_k(x_k - x^*)\|^2$$

$$= (1+\alpha_k)\|y_k - x^*\|^2 - \alpha_k\|x_k - x^*\|^2$$
$$+ \alpha_k(1+\alpha_k)\|x_k - y_k\|^2. \tag{3.2.11}$$

Multiplying both sides of (3.2.10) by $(1+\alpha_k)$ and then connecting to (3.2.11), it yields that

$$\begin{aligned}\|x_{k+1} - x^*\|^2 \leq\ & 2(1+\alpha_k)\lambda_k\langle v, x^* - y_k\rangle - (1+\alpha_k)\left(\tfrac{2\varepsilon}{1+\varepsilon}\right)\lambda_k\beta_k\langle B(x_k), x_k - x^*\rangle \\ & + (1+\alpha_k)\lambda_k\beta_k\left((1+\varepsilon)\lambda_k\beta_k - \tfrac{2\omega}{1+\varepsilon}\right)\|B(x_k)\|^2 \\ & + (1+\alpha_k)\|x_k - x^*\|^2 - (1+\alpha_k)\left(\tfrac{\varepsilon}{1+\varepsilon}\right)\|y_k - x_k\|^2 \\ & - \alpha_k\|x_k - x^*\|^2 + \alpha_k(1+\alpha_k)\|x_k - y_k\|^2\end{aligned}$$

for all $k \in \mathbb{N}$. This completes the proof. $\square$

Lemma 3.2.4. *Let $(x^*, w) \in \operatorname{gra}(A + N_{\mathbf{zer}(B)})$, $v \in A(x^*)$ and $p \in N_{\mathbf{zer}(B)}(x^*)$ be such that $w = v + p$. Suppose that $\limsup_{k\to+\infty} \lambda_k\beta_k < \omega$. Then there exist $\overline{k} \in \mathbb{N}$, $\varepsilon_0 > 0$ and $K > 0$ such that for each $k \geq \overline{k}$,*

$$\begin{aligned}&\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 + \left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_k - x_k\|^2 \\ &+ \left(\tfrac{\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\right)\langle B(x_k), x_k - x^*\rangle + \left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_k\beta_k\|B(x_k)\|^2 \\ &\leq \tfrac{(1+K)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2p(1+\varepsilon_0)}{\varepsilon_0\beta_k}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{2p(1+\varepsilon_0)}{\varepsilon_0\beta_k}\right)\right] \\ &\quad + 2(1+K)\lambda_k\langle w, x^* - x_k\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_k^2\|v\|^2.\end{aligned}$$

Proof. Since $\limsup_{k\to+\infty} \lambda_k\beta_k < \omega$, there exists $N_0 \in \mathbb{N}$ such that $\lambda_k\beta_k < \omega$ for all $k \geq N_0$. So, we can find $\varepsilon_0 \in \left(0, \sqrt{\frac{\omega}{\limsup_{k\to+\infty}\lambda_k\beta_k}} - 1\right)$ and hence, $(1+\varepsilon_0)\lambda_k\beta_k < \frac{\omega}{1+\varepsilon_0}$ for all $k \geq N_0$. Note that $\alpha_k \to 0$ as $k \to +\infty$, there exists $N_1 \in \mathbb{N}$ such that $\alpha_k < \frac{\varepsilon_0}{4(1+\varepsilon_0)}$ for all $k \geq N_1$. Choose $\overline{k} := \max\{N_0, N_1\}$. For each $k \in \mathbb{N}$, by applying Lemma 2.3.14 (i), the following inequality holds:

$$\begin{aligned}2(1+\alpha_k)\lambda_k\langle v, x^* - y_k\rangle &= 2(1+\alpha_k)\lambda_k\langle v, x^* - x_k\rangle + 2(1+\alpha_k)\langle\lambda_k v, x_k - y_k\rangle \\ &\leq 2(1+\alpha_k)\lambda_k\langle v, x^* - x_k\rangle \\ &\quad + 2(1+\alpha_k)\langle\sqrt{\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}}\lambda_k v, \sqrt{\tfrac{\varepsilon_0}{2(1+\varepsilon_0)}}(x^* - x_k)\rangle \\ &\leq 2(1+\alpha_k)\lambda_k\langle v, x^* - x_k\rangle\end{aligned}$$

$$+ (1+\alpha_k)\left(\tfrac{\varepsilon_0}{2(1+\varepsilon_0)}\right)\|y_k - x_k\|^2$$
$$+ (1+\alpha_k)\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)\lambda_k^2\|v\|^2. \tag{3.2.12}$$

Combining (3.2.12) to (3.2.4), we obtain that for each $k \geq \overline{k}$,

$$\begin{aligned}
&\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 + (1+\alpha_k)\left(\tfrac{\varepsilon_0}{2(1+\varepsilon_0)} - \alpha_k\right)\|y_k - x_k\|^2 \\
&\quad + (1+\alpha_k)\left(\tfrac{2\varepsilon_0}{1+\varepsilon_0}\right)\lambda_k\beta_k\langle B(x_k), x_k - x^*\rangle \\
&\quad + (1+\alpha_k)\left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_k\beta_k\|B(x_k)\|^2 \\
&\quad \leq (1+\alpha_k)\lambda_k\beta_k\left((1+\varepsilon_0)\lambda_k\beta_k - \tfrac{\omega}{1+\varepsilon_0}\right)\|B(x_k)\|^2 \\
&\quad\quad + (1+\alpha_k)\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)\lambda_k^2\|v\|^2 + 2(1+\alpha_k)\lambda_k\langle v, x^* - x_k\rangle.
\end{aligned} \tag{3.2.13}$$

From (3.2.13), we get that for each $k \geq \overline{k}$,

$$\begin{aligned}
&\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 + (1+\alpha_k)\left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_k - x_k\|^2 \\
&\quad + (1+\alpha_k)\left(\tfrac{2\varepsilon_0}{1+\varepsilon_0}\right)\lambda_k\beta_k\langle B(x_k), x_k - x^*\rangle \\
&\quad + (1+\alpha_k)\left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_k\beta_k\|B(x_k)\|^2 \\
&\quad \leq 2(1+\alpha_k)\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)\lambda_k^2\|v\|^2 + 2(1+\alpha_k)\lambda_k\langle v, x^* - x_k\rangle.
\end{aligned} \tag{3.2.14}$$

Next, for each $k \geq \overline{k}$, we focus on the following terms of (3.2.14)

$$\begin{aligned}
&2(1+\alpha_k)\lambda_k\langle v, x^* - x_k\rangle - \tfrac{(1+\alpha_k)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\langle B(x_k), x_k - x^*\rangle \\
&\quad = 2(1+\alpha_k)\lambda_k\langle w - p, x^* - x_k\rangle - \tfrac{(1+\alpha_k)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\langle B(x_k), x_k - x^*\rangle \\
&\quad = 2(1+\alpha_k)\lambda_k\langle w, x^* - x_k\rangle + 2(1+\alpha_k)\lambda_k\langle p, x_k\rangle \\
&\quad\quad - \tfrac{(1+\alpha_k)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\langle B(x_k), x_k - x^*\rangle - 2(1+\alpha_k)\lambda_k\langle p, x^*\rangle \\
&\quad = \tfrac{(1+\alpha_k)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\left[\left\langle \tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x_k\right\rangle + \langle B(x_k), x^*\rangle\right. \\
&\quad\quad \left. - \langle B(x_k), x_k\rangle - \left\langle \tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right\rangle\right] + (1+\alpha_k)2\lambda_k\langle w, x^* - x_k\rangle \\
&\quad \leq \tfrac{(1+\alpha_k)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\left[\sup_{x^* \in \operatorname{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right) - \left\langle \tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right\rangle\right] \\
&\quad\quad + 2(1+\alpha_k)\lambda_k\langle w, x^* - x_k\rangle.
\end{aligned} \tag{3.2.15}$$

Since $p \in N_{\operatorname{zer}(B)}(x^*)$, we have $\frac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p \in N_{\operatorname{zer}(B)}(x^*)$ for all $k \in \mathbb{N}$. It is equivalent to say that $\sigma_{\operatorname{zer}(B)}\left(\frac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p\right) = \left\langle \frac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right\rangle$ for all $k \in \mathbb{N}$. It follows from (3.2.15) that

$$2(1+\alpha_k)\lambda_k\langle v, x^*-x_k\rangle \le \tfrac{(1+\alpha_k)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\langle B(x_k), x_k-x^*\rangle + \tfrac{(1+\alpha_k)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\left[\sup_{x^*\in\operatorname{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p\right)\right] + 2(1+\alpha_k)\lambda_k\langle w, x^*-x_k\rangle. \tag{3.2.16}$$

Combining (3.2.16) to (3.2.13), it appears the result that for each $k \ge \overline{k}$,

$$\begin{aligned}&\|x_{k+1}-x^*\|^2 - \|x_k-x^*\|^2 + (1+\alpha_k)\left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_k-x_k\|^2 \\ &+ \tfrac{(1+\alpha_k)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\langle B(x_k), x_k-x^*\rangle + (1+\alpha_k)\left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_k\beta_k\|B(x_k)\|^2 \\ &\le \tfrac{(1+\alpha_k)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\left[\sup_{x^*\in\operatorname{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p\right)\right] \\ &\quad + 2(1+\alpha_k)\lambda_k\langle w, x^*-x_k\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+\alpha_k)\lambda_k^2\|v\|^2.\end{aligned}$$

Note that the positive sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ is bounded, there exists $K>0$ such that $\alpha_k \le K$ for all $k\in\mathbb{N}$. Since $\langle B(x_k), x_k-x^*\rangle$ is nonnegative for all $k\in\mathbb{N}$, we obtain that

$$\begin{aligned}&\|x_{k+1}-x^*\|^2 - \|x_k-x^*\|^2 + \left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_k-x_k\|^2 \\ &+ \left(\tfrac{\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\right)\langle B(x_k), x_k-x^*\rangle + \left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_k\beta_k\|B(x_k)\|^2 \\ &\le \tfrac{(1+K)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\left[\sup_{x^*\in\operatorname{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p\right)\right] \\ &\quad + 2(1+K)\lambda_k\langle w, x^*-x_k\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_k^2\|v\|^2, \quad \forall k \ge \overline{k}.\end{aligned}$$

This completes the proof. □

Convergence Results for Monotone Inclusion Problem

In this section, some convergence results for Algorithm 2 are demonstrated. Before going into the main results, it is useful to know the following propositions.

Proposition 3.2.5. *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 2. If all assumptions in Assumption 3.2.1 hold, then the following hold:*

(i) *For each* $x^* \in \operatorname{zer}(A + N_{\operatorname{zer}(B)})$, $\lim_{k\to+\infty} \|x_k - x^*\|$ *exists.*

(ii) *The series* $\sum_{k=1}^{+\infty} \|y_k - x_k\|^2$, $\sum_{k=1}^{+\infty} \lambda_k\beta_k\|B(x_k)\|^2$, *and* $\sum_{k=1}^{+\infty} \lambda_k\beta_k\langle B(x_k), x_k - x^*\rangle$ *are convergent.*

(iii) $\lim_{k\to+\infty} \|y_k - x_k\|^2 = \lim_{k\to+\infty} \|B(x_k)\| = \lim_{k\to+\infty} \langle B(x_k), x_k - x^*\rangle = 0.$

Proof. Let $x^* \in \operatorname{zer}(A + N_{\operatorname{zer}(B)})$. Taking $w = 0$ in Lemma 3.1.5, we get that

$$\begin{aligned}
&\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 + \left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_k - x_k\|^2 \\
&\quad + \left(\tfrac{\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\right)\langle B(x_k), x_k - x^*\rangle + \left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_k\beta_k\|B(x_k)\|^2 \\
&\quad \le \tfrac{(1+K)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\left[\sup_{x^*\in\operatorname{zer}(B)} F_B\left(\tfrac{2p(1+\varepsilon_0)}{\varepsilon_0\beta_k}, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{2p(1+\varepsilon_0)}{\varepsilon_0\beta_k}\right)\right] \\
&\quad + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_k^2\|v\|^2, \quad \forall k \ge \overline{k}.
\end{aligned}$$

By Assumption 3.2.1, the conclusion in (i) and (ii) follows from Lemma 2.4.2.

(iii) From (ii), we have $\lim_{n\to+\infty} \|y_k - x_k\|^2 = 0$. Since $\liminf_{n\to+\infty} \lambda_k\beta_k > 0$, we obtain that $\lim_{k\to+\infty} \|B(x_k)\| = \lim_{k\to+\infty} \langle B(x_k), x_k - x^*\rangle = 0$. □

Theorem 3.2.6. *Let* $\{x_k\}_{k\in\mathbb{N}}$ *be a sequence generated by Algorithm 2 and* $\{z_k\}_{k\in\mathbb{N}}$ *be a sequence of weighted averages as* (3.2.3). *Suppose that all assumptions in Assumption 3.2.1 hold. Then the sequence* $\{z_k\}_{k\in\mathbb{N}}$ *converges weakly to an element in* $\operatorname{zer}(A + N_{\operatorname{zer}(B)})$.

Proof. Let z be a weak cluster point of $\{z_k\}_{k\in\mathbb{N}}$. Then there exists a subsequence $\{z_{k_i}\}_{i\in\mathbb{N}}$ of $\{z_k\}_{k\in\mathbb{N}}$ such that $z_{k_i} \rightharpoonup z$ as $i \to +\infty$. We will show that $z \in \operatorname{zer}(A + N_{\operatorname{zer}(B)})$. Since $A + N_{\operatorname{zer}(B)}$ is a maximal monotone operator, it suffices to show that $\langle w, x^* - z\rangle \ge 0$ for all $(x^*, w) \in \operatorname{gra}(A + N_{\operatorname{zer}(B)})$.

Let $(x^*, w) \in \operatorname{gra}(A + N_{\operatorname{zer}(B)})$, $v \in A(x^*)$ and $p \in N_{\operatorname{zer}(B)}(x^*)$ be such that $w = v + p$. Recall from Lemma 3.1.5 that

$$\begin{aligned}
&\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 + \left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_k - x_k\|^2 \\
&\quad + \left(\tfrac{\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\right)\langle B(x_k), x_k - x^*\rangle + \left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_k\beta_k\|B(x_k)\|^2
\end{aligned}$$

$$\leq \tfrac{(1+K)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\left[\sup_{x^*\in\operatorname{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p\right)\right]$$
$$+ 2(1+K)\lambda_k\langle w, x^* - x_k\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_k^2\|v\|^2, \quad \forall k \geq \overline{k}.$$

Discarding nonnegative terms $\langle B(x_k), x_k - x^*\rangle$, $\|B(x_k)\|^2$ and $\|y_k - x_k\|^2$, we deduce to

$$\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2$$
$$\leq \tfrac{(1+K)\varepsilon_0\lambda_k\beta_k}{1+\varepsilon_0}\left[\sup_{x^*\in\operatorname{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p\right)\right]$$
$$+ 2(1+K)\lambda_k\langle w, x^* - x_k\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_k^2\|v\|^2, \quad \forall k \geq \overline{k}.$$

Summing up for $k = \overline{k}, \overline{k}+1, ..., k_i$ in the above inequality, we have

$$\|x_{k_i+1} - x^*\|^2 - \|x_{\overline{k}} - x^*\|^2$$
$$\leq 2(1+K)\left\langle w, \sum_{k=\overline{k}}^{k_i}\lambda_k x^* - \sum_{k=\overline{k}}^{k_i}\lambda_k x_k\right\rangle + L_1$$
$$= 2(1+K)\left\langle w, \sum_{k=1}^{k_i}\lambda_k x^* - \sum_{k=1}^{\overline{k}-1}\lambda_k x^* - \sum_{k=1}^{k_i}\lambda_k x_k + \sum_{k=1}^{\overline{k}-1}\lambda_k x_k\right\rangle + L_1$$

where

$$L_1 := \tfrac{(1+K)}{2}\sum_{k=\overline{k}}^{k_i}\lambda_k\beta_k\left[\sup_{x^*\in\operatorname{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_k}p\right)\right]$$
$$+ 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\sum_{k=\overline{k}}^{k_i}\lambda_k^2\|v\|^2.$$

Discarding the nonnegative term $\|x_{k_i+1} - x^*\|^2$ and dividing inequality above by $2(1+K)\tau_{k_i}$, we obtain

$$-\tfrac{\|x_{\overline{k}} - x^*\|^2}{2(1+K)\tau_{k_i}} \leq \langle w, x^* - z_{k_i}\rangle + \tfrac{L_2}{2(1+K)\tau_{k_i}}, \tag{3.2.17}$$

where $L_2 := L_1 + 2(1+K)\left\langle w, -\sum_{k=1}^{\overline{k}-1}\lambda_k x^* + \sum_{k=1}^{\overline{k}-1}\lambda_k x_k\right\rangle$. Note that L_2 is a finite real number. Taking $i \to +\infty$ (so that $\lim_{i\to+\infty}\tau_{k_i} = +\infty$) on both sides of (3.2.17), we get that

$$0 \leq \langle w, x^* - z\rangle.$$

Since $(x^*, w) \in \text{gra}(A + N_{\text{zer}(B)})$ is arbitrary, we have $z \in \text{zer}(A + N_{\text{zer}(B)})$. By Lemma 2.4.3, we can conclude that the sequence $\{z_k\}_{k\in\mathbb{N}}$ converges weakly to an element in $\text{zer}(A + N_{\text{zer}(B)})$. □

Next, we will prove the strong convergence of the sequence $\{x_k\}_{k\in\mathbb{N}}$.

Theorem 3.2.7. *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 2 and the operator A be a γ-strongly monotone with $\gamma > 0$. If all assumptions in Assumption 3.2.1 hold, then the sequence $\{x_k\}_{k\in\mathbb{N}}$ converges in norm to the unique $x^* \in \text{zer}(A+ N_{\text{zer}(B)})$.*

Proof. Let x^* be the unique element in $\text{zer}(A + N_{\text{zer}(B)})$. Then there exists $v \in A(x^*)$ and $p \in N_{\text{zer}(B)}(x^*)$ such that $0 = v + p$. Since $\frac{x_k - y_k}{\lambda_k} - \beta_k B(x_k) \in A(y_k)$ and $v \in A(x^*)$, the strong monotonicity of A implies

$$\lambda_k \gamma \|y_k - x^*\|^2 \leq \langle x_k - y_k - \lambda_k(\beta_k B(x_k) + v), y_k - x^*\rangle$$

for all $k \in \mathbb{N}$. It follows that

$$\lambda_k \gamma \|y_k - x^*\|^2 + \langle x_k - y_k, x^* - y_k\rangle \leq \lambda_k \langle \beta_k B(x_k) + v, x^* - y_k\rangle \tag{3.2.18}$$

for all $k \in \mathbb{N}$. By applying Lemma 2.3.14 (i), we have

$$\begin{aligned} 2\lambda_k \gamma \|y_k - x^*\|^2 + \|y_k - x^*\|^2 - \|x_k - x^*\|^2 \leq\ & 2\lambda_k \langle \beta_k B(x_k) + v, x^* - y_k\rangle \\ & - \|x_k - y_k\|^2 \end{aligned} \tag{3.2.19}$$

for all $k \in \mathbb{N}$. Focusing on the right hand side of (3.2.19), we see that

$$\begin{aligned} & 2\lambda_k \langle \beta_k B(x_k) + v, x^* - y_k\rangle - \|x_k - y_k\|^2 \\ & \quad = 2\lambda_k \langle \beta_k B(x_k) + v, x^* - x_k\rangle + 2\lambda_k \langle \beta_k B(x_k) + v, x_k - y_k\rangle - \|x_k - y_k\|^2 \\ & \quad \leq 2\lambda_k \langle \beta_k B(x_k) + v, x^* - x_k\rangle + \lambda_k^2 \|\beta_k B(x_k) + v\|^2 \\ & \quad \leq 2\lambda_k \langle \beta_k B(x_k) + v, x^* - x_k\rangle + 2\lambda_k^2 \beta_k^2 \|B(x_k)\|^2 + 2\lambda_k^2 \|v\|^2, \quad \forall k \in \mathbb{N}. \end{aligned} \tag{3.2.20}$$

Next, we consider the first term on right hand side of (3.2.20),

$$2\lambda_k \langle \beta_k B(x_k) + v, x^* - x_k\rangle$$

$$
\begin{aligned}
&= 2\lambda_k\langle \beta_k B(x_k), x^* - x_k\rangle + 2\lambda_k\langle v, x^* - x_k\rangle \\
&= 2\lambda_k\langle \beta_k B(x_k), x^* - x_k\rangle + 2\lambda_k\langle p, x_k\rangle - 2\lambda_k\langle p, x^*\rangle \\
&= 2\lambda_k\beta_k \left[\left\langle \tfrac{p}{\beta_k}, x_k\right\rangle + \langle B(x_k), x^*\rangle - \langle B(x_k), x_k\rangle - \left\langle \tfrac{p}{\beta_k}, x^*\right\rangle\right] \\
&\leq 2\lambda_k\beta_k \left[\sup_{x^*\in \operatorname{zer}(B)} F_B\left(\tfrac{p}{\beta_k}, x^*\right) - \left\langle \tfrac{p}{\beta_k}, x^*\right\rangle\right] \\
&= 2\lambda_k\beta_k \left[\sup_{x^*\in \operatorname{zer}(B)} F_B\left(\tfrac{p}{\beta_k}, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{p}{\beta_k}, x^*\right)\right], \quad \forall k \in \mathbb{N}.
\end{aligned} \tag{3.2.21}
$$

Combining (3.2.19), (3.2.20) and (3.2.21), we have

$$
\begin{aligned}
&2\lambda_k\gamma\|y_k - x^*\|^2 + \|y_k - x^*\|^2 - \|x_k - x^*\|^2 \\
&\qquad \leq 2\lambda_k\beta_k \left[\sup_{x^*\in \operatorname{zer}(B)} F_B\left(\tfrac{p}{\beta_k}, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{p}{\beta_k}, x^*\right)\right] \\
&\qquad\quad + 2\lambda_k^2\beta_k^2\|B(x_k)\|^2 + 2\lambda_k^2\|v\|^2, \quad \forall k \in \mathbb{N}.
\end{aligned} \tag{3.2.22}
$$

By simple calculation using (3.2.22), we get the result that

$$
\begin{aligned}
\|y_k - x^*\|^2 &\leq \tfrac{2\lambda_k\beta_k}{2\lambda_k\gamma+1}\left[\sup_{x^*\in \operatorname{zer}(B)} F_B\left(\tfrac{p}{\beta_k}, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{p}{\beta_k}, x^*\right)\right] \\
&\quad + \tfrac{1}{2\lambda_k\gamma+1}\|x_k - x^*\|^2 + \tfrac{2\lambda_k^2\beta_k^2}{2\lambda_k\gamma+1}\|B(x_k)\|^2 + \tfrac{2\lambda_k^2}{2\lambda_k\gamma+1}\|v\|^2.
\end{aligned} \tag{3.2.23}
$$

Combining (3.2.23) to (3.2.11), we have the following inequality

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 &\leq (1+\alpha_k)\tfrac{2\lambda_k\beta_k}{2\lambda_k\gamma+1}\left[\sup_{x^*\in \operatorname{zer}(B)} F_B\left(\tfrac{p}{\beta_k}, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{p}{\beta_k}, x^*\right)\right] \\
&\quad + (1+\alpha_k)\left[\tfrac{1}{2\lambda_k\gamma+1}\|x_k - x^*\|^2 + \tfrac{2\lambda_k^2\beta_k^2}{2\lambda_k\gamma+1}\|B(x_k)\|^2\right] \\
&\quad + \left(\tfrac{1+\alpha_k}{2\lambda_k\gamma+1}\right)2\lambda_k\|v\|^2 - \tfrac{\alpha_k}{2\lambda_k\gamma+1}\|x_k - x^*\|^2 \\
&\quad + \alpha_k(1+\alpha_k)\|x_k - y_k\|^2.
\end{aligned} \tag{3.2.24}
$$

It is not hard to verify from (3.2.24) and it yields that

$$
\begin{aligned}
&2\lambda_k\gamma\|x_{k+1} - x^*\|^2 + \|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 \\
&\qquad \leq (1+\alpha_k)2\lambda_k\beta_k\left[\sup_{x^*\in \operatorname{zer}(B)} F_B\left(\tfrac{p}{\beta_k}, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{p}{\beta_k}, x^*\right)\right] \\
&\qquad\quad + (1+\alpha_k)2\lambda_k^2\beta_k^2\|B(x_k)\|^2
\end{aligned}
$$

$$+ (1+\alpha_k)\, 2\lambda_k^2\|v\|^2 + \alpha_k(1+\alpha_k)(2\lambda_k\gamma+1)\|x_k - y_k\|^2.$$

Since nonnegative sequences $\{\lambda_k\}_{k\in\mathbb{N}}$, $\{\lambda_k\beta_k\}_{k\in\mathbb{N}}$ and $\{\alpha_k\}_{k\in\mathbb{N}}$ are bounded, there exists positive numbers M, c and K such that $\lambda_k \le M$, $\lambda_k\beta_k \le c$, and $\alpha_k \le K$ for all $k \in \mathbb{N}$.

Hence,

$$\begin{aligned} 2\lambda_k\gamma\|x_{k+1}-x^*\|^2 + \|x_{k+1}-x^*\|^2 - \|x_k - x^*\|^2 &\le (1+K)2\lambda_k\beta_k\left[\sup_{x^*\in \operatorname{zer}(B)} F_B\left(\tfrac{p}{\beta_k}, x^*\right) - \sigma_{\operatorname{zer}(B)}\left(\tfrac{p}{\beta_k}, x^*\right)\right] \\ &\quad + (1+K)2c\lambda_k\beta_k\|B(x_k)\|^2 + (1+K)\,2\lambda_k^2\|v\|^2 \\ &\quad + K(1+K)(2M\gamma+1)\|x_k - y_k\|^2, \end{aligned} \tag{3.2.25}$$

and then

$$\begin{aligned} 2\gamma\sum_{k=1}^{+\infty}\lambda_k\|x_{k+1}-x^*\|^2 &\le \|x_1 - x^*\|^2 \\ &\quad + (1+K)\left[2c\sum_{k=1}^{+\infty}\lambda_k\beta_k\|B(x_k)\|^2 + 2\sum_{k=1}^{+\infty}\lambda_k^2\|v\|^2\right] \\ &\quad + K(1+K)(2M\gamma+1)\sum_{k=1}^{+\infty}\|x_k - y_k\|^2 \\ &\quad + (1+K)2\sum_{k=1}^{+\infty}\lambda_k\beta_k\left[\sup_{x^*\in\operatorname{zer}(B)} F_B\left(\tfrac{p}{\beta_k}, x^*\right)\right. \\ &\quad \left. -\sigma_{\operatorname{zer}(B)}\left(\tfrac{p}{\beta_k}, x^*\right)\right]. \end{aligned}$$

By all assumptions in Assumption 3.2.1 and Proposition 3.2.5, we have

$$2\gamma\sum_{k=1}^{+\infty}\lambda_k\|x_{k+1}-x^*\|^2 < +\infty.$$

From (3.2.25) and Lemma 2.4.2, we obtain that $\lim_{k\to+\infty}\|x_k - x^*\|$ exists.

Since $\sum_{k=1}^{+\infty}\lambda_k = +\infty$, we have $\lim_{k\to+\infty}\|x_k - x^*\| = 0$. This completes the proof.

□

Applications to Constrained Convex Optimization Problem

In this part, we will apply the results obtained in the previous part to solve constrained convex optimization problems (3.0.2). Furthermore, we may assume without loss of generality that $\min g = 0$. We assume that the solution set of the problem (3.0.2) $\mathcal{S}$ is a nonempty set. Notice that f is proper convex lower semicontinuous, we have that the subdifferential ∂f is maximally monotone. Moreover, since the function g is convex differentiable, by using the Theorem of Baillon-Haddad (see [66, Corollary 18.17]), ∇g is $\frac{1}{L_g}$-cocoercive and $\arg\min g = \operatorname{zer}(\nabla g)$.

By using this and Algorithm 2, we will consider the following algorithm.

Algorithm 3:

Initialization: Given three positive sequences $\{\alpha_k\}_{k\in\mathbb{N}}$, $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{\beta_k\}_{k\in\mathbb{N}}$. Choose $x_1 \in \mathcal{H}$ arbitrarily.

Iterative Steps: For a given current iterate $x_k \in \mathcal{H}$, calculate as follows:

Step 1. Compute y_k as

$$y_k = \operatorname{prox}_{\lambda_k f}(x_k - \lambda_k\beta_k\nabla g(x_k)).$$

Step 2. Compute

$$x_{k+1} = y_k + \alpha_k(y_k - x_k).$$

Update k := k + 1 and return to Step 1.

In order to obtain the convergence of the sequence generated by Algorithm 3, we have to assume the following assumption.

Assumption 3.2.8. (a) The qualification condition $\arg\min g \cap \operatorname{int}\operatorname{dom}(f) \neq \emptyset$ holds.

(b) $\{\lambda_k\} \in \ell_2 \setminus \ell_1$, $\lim\limits_{k\to+\infty} \alpha_k = 0$ and $0 < \liminf\limits_{k\to+\infty} \lambda_k\beta_k \leq \limsup\limits_{k\to+\infty} \lambda_k\beta_k < \frac{1}{L_g}$.

(c) For each $p \in \operatorname{ran}(N_{\arg\min g})$, we have

$$\sum_{k=1}^{+\infty} \lambda_k\beta_k \left[g^*\left(\frac{p}{\beta_k}\right) - \sigma_{\arg\min g}\left(\frac{p}{\beta_k}\right)\right] < +\infty.$$

Note that $F_{\nabla g}\left(x^*, \frac{p}{\beta_k}\right) \leq g(x^*) + g^*\left(\frac{p}{\beta_k}\right) = g^*\left(\frac{p}{\beta_k}\right)$ for all $x^* \in \arg\min g$, we have $\sup_{x^* \in \arg\min g} F_{\nabla g}\left(x^*, \frac{p}{\beta_k}\right) \leq g^*\left(\frac{p}{\beta_k}\right)$. Hence, conditions (a)-(c) in Assumption 3.2.8 imply hypotheses (I)-(III) in Assumption 3.2.1.

Corollary 3.2.9. *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 3 and $\{z_k\}_{k\in\mathbb{N}}$ be a sequence of weighted averages as* (3.2.3)*. Suppose that all assumptions in Assumption 3.2.8 hold. Then the sequence $\{z_k\}_{k\in\mathbb{N}}$ converges weakly to an element in $\mathcal{S}$.*

If we assume that the function f is strongly convex, then its subdifferential ∂f is strongly monotone.

Corollary 3.2.10. *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 3 and the function f be a γ-strongly convex with $\gamma > 0$. If all assumptions in Assumption 3.2.8 hold, then the sequence $\{x_k\}_{k\in\mathbb{N}}$ converges strongly to the unique element in $\mathcal{S}$.*

Numerical Experiments

In this section, we present an example of numerical set for testing the purposed algorithm. Some comparisons of our algorithm Algorithm 3 with the algorithm (**FB**) introduced by Attouch [4] are also reported.

We consider the problem with equality constraints:

$$\begin{aligned} &\text{minimize } \|x\|_1 \\ &\text{subject to } \mathbf{A}x = \mathbf{b}, \end{aligned} \tag{3.2.26}$$

where $\mathbf{A} \in \mathbb{R}^{l\times s}$, $\mathbf{b} \in \mathbb{R}^l$. In addition, we assume that $s > l$. The problem (3.2.26) can be written in the form of the problem (3.0.2) as :

$$\begin{aligned} &\text{minimize } f(x) := \|x\|_1 \\ &\text{subject to } x \in \arg\min g, \end{aligned}$$

where $g(z) := \frac{1}{2}\|\mathbf{A}z - \mathbf{b}\|^2$, for all $z \in \mathbb{R}^s$.

In this setting, we have $\nabla g(z) = \mathbf{A}^T(\mathbf{A}x - \mathbf{b})$ and notice that ∇g is $\|\mathbf{A}\|^2$-Lipschitz continuous. We also get that
$\text{prox}_{\lambda_k f}(x) = \left(\max\left(0, 1 - \frac{\lambda_k}{|x_1|}\right) x_1, \max\left(0, 1 - \frac{\lambda_k}{|x_2|}\right) x_2, ..., \max\left(0, 1 - \frac{\lambda_k}{|x_k|}\right) x_s\right).$

We begin with the problem by random vectors $x_1 \in \mathbb{R}^s$, $\mathbf{b} \in \mathbb{R}^l$ and matrix $\mathbf{A} \in \mathbb{R}^{l\times s}$. Next, we compare the performance of the Algorithm 3 with the algorithm (**FB**). The used of parameters in two algorithms are chosen as follows:

$\beta_k = \frac{k}{(\|\mathbf{A}\|^2)+1}$, $\lambda_k = \frac{1}{k}$, $\forall k \geq 1$. We obtain the CPU times (seconds) and the number of iterations by using the stopping criteria : $\|x_k - x_{k-1}\| \leq 10^{-6}$.

Table 3: Comparison of number of iterations and CPU computation time between Algorithm 3 and (**FB**) with difference of parameter sequences $\{\alpha_k\}_{k\in\mathbb{N}}$.

Algorithm 3	CPU times (s)	Iterations
(**FB**) ($\alpha_k = 0$)	180.44	38352
Algorithm 3 ($\alpha_k = 1/\sqrt{k+1}$)	140.36	35649
Algorithm 3 ($\alpha_k = 1/(k+1)$)	155.79	35589
Algorithm 3 ($\alpha_k = 1/(k+1)^2$)	136.01	33841
Algorithm 3 ($\alpha_k = 1/(k+1)^4$)	150.45	37164
Algorithm 3 ($\alpha_k = 1/(k+1)^{10}$)	154.40	38344

We compare the performance of the Algorithm 3 and (**FB**) for case $s = 4000$, $l = 1000$ with difference of parameter sequences $\{\alpha_k\}_{k\in\mathbb{N}}$. The results are reported in table 3. We observe that (**FB**) spends more CPU computation time than Algorithm 3. We can see that when $\alpha_k = \frac{1}{(k+1)^2}$, it leads to the lowest CPU computation time and number of iterations for Algorithm 3 of 136.01 seconds and 33841 times, respectively. We also observe that our algorithm requires less iterations than (**FB**) for all choice of parameter sequences $\{\alpha_k\}_{k\in\mathbb{N}}$.

Table 4: The comparison of two algorithms with different sizes of matrix **A**.

(l,s)	Algorithm 3		(FB)	
	CPU time (s)	Iterations	CPU time (s)	Iterations
(20,1000)	1.99	34160	5.13	83860
(50,1000)	2.32	32986	5.72	77435
(100,1000)	2.92	30352	7.38	79054
(200,1000)	3.94	30546	7.35	56337
(300,1000)	5.17	26191	6.70	33513
(20,2000)	4.14	37505	11.14	98780
(50,2000)	6.14	45289	10.55	78691
(100,2000)	5.00	27642	10.42	58504
(200,2000)	8.33	24207	23.45	67317
(300,2000)	15.71	27088	28.22	48109
(20,5000)	10.17	40463	25.04	96251
(50,5000)	7.09	22812	21.79	68287
(100,5000)	18.70	29416	42.56	66194
(200,5000)	40.66	33008	92.56	77144
(300,5000)	51.59	27193	123.60	58909

In table 4, we present a comparison between the numerical results of Algorithm 3 and (**FB**) cases for $\alpha_k = \frac{1}{\sqrt{k+1}}, \forall k \geq 1$ and different sizes of matrix **A**. We can see that the number of iterations of Algorithm 3 are smaller than of (**FB**) for all different sizes of matrix **A**. Furthermore, Algorithm 3 requires less CPU computation time to reach the optimality tolerance than (**FB**) for all cases.

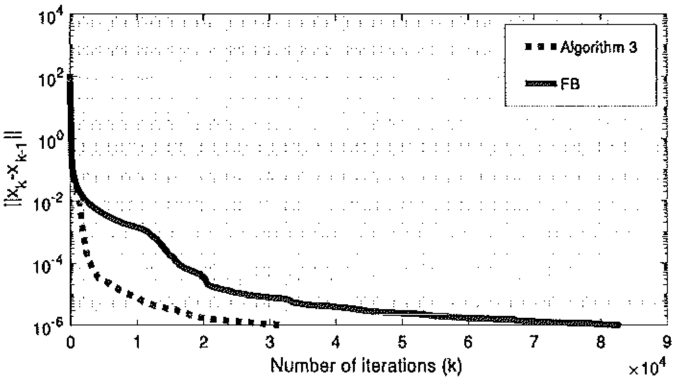


Figure 2: Illustration of the behavior of $\|x_k - x_{k-1}\|$ for Algorithm 3 and (**FB**) methods when $\alpha_k = \frac{1}{\sqrt{k+1}}$ and $(l, s) = (100, 3000)$.

Figure 2 shows the behavior of $\|x_k - x_{k-1}\|$ for Algorithm 3 and (**FB**) methods when $\alpha_k = \frac{1}{\sqrt{k+1}}$ and $(l, s) = (100, 3000)$. We can observe that by using Algorithm 3 the behavior of the red line Algorithm 3 performs better than the blue line (**FB**).

CHAPTER IV

THE MONOTONE INCLUSION PROBLEMS

In this chapter, we propose iterative methods for solving monotone inclusion problem. We also propose iterative methods for solving fixed point problem of nonexpansive mapping to apply the solution of generalized monotone inclusion problem. We have divided into two sections as the following:

4.1 Generalized Viscosity Forward-Backward Splitting Scheme with Inertial Terms for Solving Monotone Inclusion Problems

The purpose of this section is to consider the monotone inclusion problem: find $x \in \mathcal{H}$ such that

$$0 \in Ax + Bx, \tag{4.1.1}$$

where $A : \mathcal{H} \to \mathcal{H}$ is a single-valued mapping and $B : \mathcal{H} \to 2^{\mathcal{H}}$ a multi-valued mapping. We denote the set of all solutions of the problem (4.1.1) by $(A+B)^{-1}(0)$. Most well known algorithms to approximate the solution of the problem (4.1.1) is the forward-backward algorithm (**FB**) [35–37]. The algorithm (**FB**) was first introduced by Passty [36] that was defined by a sequence $\{x_k\}_{k\in\mathbb{N}}$ as follows:

$$x_{k+1} = J_\lambda^B(x_k - \lambda A x_k), \quad \text{for all } k \geq 1, \tag{4.1.2}$$

where $J_\lambda^B = (Id + \lambda B)^{-1}$ is the resolvent of the operator B and $\lambda > 0$ and Id is an identity mapping. This method involve with the proximal point algorithm [79–83] and the gradient method.

In 2001, Alvarez and Attouch [21] introduced a new algorithm by using the idea of the inertial method in [17, 38]. This method is written as follows:

$$\begin{cases} w_k = x_k + \theta_k(x_k - x_{k-1}), \\ x_{k+1} = (Id + \lambda_k B)^{-1} w_k, \quad \text{for all } k \geq 1. \end{cases} \tag{4.1.3}$$

They proved that the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by algorithm (4.1.3) converges weakly to a zero point of the operator B under the following conditions $\{\theta_k\}_{k\in\mathbb{N}} \subseteq [0,1]$ and $\{\lambda_k\}_{k\in\mathbb{N}}$ is non-decreasing with

$$\sum_{k=1}^{+\infty} \theta_k \|x_k - x_{k-1}\|^2 < +\infty. \tag{4.1.4}$$

Moudafi and Oliny [32] proposed iterative method which involed the idea of the inertial method to solve the problem (4.1.1). They also proved weakly convergence of the iterate under the following conditions:

(i) the condition (4.1.4) holds;

(ii) $\lambda_k < 2/L$ with L the Lipschitz constant of A.

Their algorithm is defined by

$$\begin{cases} w_k = x_k + \theta_k(x_k - x_{k-1}), \\ x_{k+1} = (Id + \lambda_k B)^{-1}(w_k - \lambda_k A x_k), \quad \text{for all } k \geq 1, \end{cases} \tag{4.1.5}$$

where $A : \mathcal{H} \to \mathcal{H}$ and $B : \mathcal{H} \to 2^{\mathcal{H}}$.

The several methods that are in reference to this study are reviewed in the next extensively (see, e.g. [27, 28, 39–45])

Recently, Kitkuan et al. [46] proposed the viscosity approximation algorithm concerning the inertial forward-backward for finding a solution of the problem (4.1.1) as follows:

$$\begin{cases} w_k = x_k + \theta_k(x_k - x_{k-1}), \\ x_{k+1} = \gamma_k(f(x_k)) + (1 - \gamma_k) J^B_{\lambda_k}(w_k - \lambda_k A w_k), \quad \text{for all } k \geq 1, \end{cases} \tag{4.1.6}$$

where $A : \mathcal{H} \to \mathcal{H}$ is μ-inverse strongly monotone operator with $\mu > 0$, $B : \mathcal{H} \to 2^{\mathcal{H}}$ is a maximal monotone operator and $f : \mathcal{H} \to \mathcal{H}$ is a contraction with constrant $c \in (0, 1)$. They also proved the strong convergence of their proposed method under some appropriate conditions imposed on the parameters.

On the other hand, in 2019, Kitkuan et al. [47] presented a new method combined Halpern-type method and forward-backward splitting method for solving the monotone inclusion problem (4.1.1) as follows:

$$\begin{cases} u, x_1 \in \mathcal{H}, \\ z_k = \alpha_k x_k + (1-\alpha_k) J^B_{\lambda_k}(x_k - \lambda_k A x_k), \\ y_k = \beta_k x_k + (1-\beta_k) J^B_{\lambda_k}(z_k - \lambda_k A z_k), \\ x_{k+1} = \gamma_k u + (1-\gamma_k) y_k, \quad \text{for all } k \geq 1, \end{cases} \tag{4.1.7}$$

where $J^B_{\lambda_k} = (Id + \lambda_k B)^{-1}$ is the resolvent of B and $\alpha_k, \beta_k, \gamma_k \in (0,1)$. Strong convergence results are obtained under some appropriate conditions.

By employing the *inertial viscosity forward-backward splitting algorithm* (**IVF-BSA**) motivated by the works of Kikuan et al. [46, 47], we propose the following algorithm.

Algorithm 4: (IVFBSA)

Initialization: Given $\{\theta_k\}_{k\in\mathbb{N}} \subseteq [0,\theta]$ with $\theta \in [0,1)$ and three sequences $\{\alpha_k\}_{k\in\mathbb{N}}$, $\{\beta_k\}_{k\in\mathbb{N}}$ and $\{\gamma_k\}_{k\in\mathbb{N}}$ in $[0,1]$. Choose $x_0, x_1 \in \mathcal{H}$ arbitrarily.

Iterative Steps: For a given current iterate $x_{k-1}, x_k \in \mathcal{H}$, calculate as follows:

Step 1. Compute

$$\begin{cases} w_k = x_k + \theta_k(x_k - x_{k-1}), \\ z_k = \alpha_k w_k + (1-\alpha_k) J^B_{\lambda_k}(w_k - \lambda_k A w_k), \\ y_k = \beta_k w_k + (1-\beta_k) J^B_{\lambda_k}(z_k - \lambda_k A z_k) \end{cases}$$

Step 2. Compute

$$x_{k+1} = \gamma_k f(x_k) + (1-\gamma_k) y_k.$$

Update k := k + 1 and return to Step 1.

Remark 4.1.1. If $\alpha_k = 1$ in Algorithm 4, we have the inertial viscosity forward-backward splitting algorithm (4.1.6).

If $\theta_k = 0$ and setting $f(x_k) = u$ in Algorithm 4, we have generalized Halpern-type forward-backward splitting method (4.1.7).

The convergence behavior between the algorithm that obtained from the Algorithm 4 and the algorithm (4.1.6) are illustrated by some numerical experiments.

We present the convergence analysis of our main results as follows.

Theorem 4.1.2. *Let $A : \mathcal{H} \to \mathcal{H}$ be a μ-inverse strongly monotone operator on a real Hilbert space $\mathcal{H}$ with $\mu > 0$ and $B : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximal monotone operator such that $(A+B)^{-1}(0) \neq \emptyset$. Let $f : \mathcal{H} \to \mathcal{H}$ be a contraction mapping with a constant $c \in (0,1)$. Let $\{x_k\}_{k\in\mathbb{N}}$ be generated by Algorithm 4. Assume that the following conditions hold:*

(i) $\lim_{k\to\infty} \gamma_k = 0$ *and* $\sum_{k\geq 1} \gamma_k = +\infty$;

(ii) $\lim_{k\to\infty} \frac{\theta_k}{\gamma_k}\|x_k - x_{k-1}\| = 0$;

(iii) $0 < \liminf_{k\to+\infty} \lambda_k \leq \limsup_{k\to+\infty} \lambda_k < 2\mu$;

(iv) $\liminf_{k\to+\infty}(1-\alpha_k)(1-\beta_k) > 0$.

Then, the sequence $\{x_k\}_{k\in\mathbb{N}}$ converges strongly to $\overline{x} := \mathrm{proj}_{(A+B)^{-1}(0)}(f(\overline{x}))$.

Proof. Let $\Gamma_k = J^B_{\lambda_k}(Id - \lambda_k A)$. By Lemma, we have for each $k \in \mathbb{N}$ the mapping Γ_k is nonexpansive. Next, we claim that $(A+B)^{-1}(0) = \mathrm{Fix}(\Gamma_k)$.

Let us consider,

$$\begin{aligned} \overline{x} = \Gamma_k(\overline{x}) \iff & \overline{x} = J^B_{\lambda_k}(Id - \lambda_k A)\overline{x} \\ \iff & \overline{x} - \lambda_k A\overline{x} \in \overline{x} + \lambda_k B\overline{x} \\ \iff & 0 \in A\overline{x} + B\overline{x}. \end{aligned} \tag{4.1.8}$$

Hence, $(A+B)^{-1}(0) = \mathrm{Fix}(\Gamma_k)$.

We expect that $\{x_k\}_{k\in\mathbb{N}}$ is bounded. Since f is contraction mapping and $\mathrm{proj}_{(A+B)^{-1}(0)}(\cdot)$ is nonexpansive, we have $\mathrm{proj}_{(A+B)^{-1}(0)}(f(\cdot))$ is contraction mapping. Then, there exists the unique fixed point $\overline{x} \in (A+B)^{-1}(0)$ such that $\overline{x} = \mathrm{proj}_{(A+B)^{-1}(0)}(f(\overline{x}))$. Thus $\overline{x} \in \mathrm{Fix}(\Gamma_k)$. It follows that

$$\|z_k - \overline{x}\| = \|\alpha_k w_k + (1-\alpha_k)\Gamma_k w_k - \overline{x}\|$$

$$\leq \alpha_k\|w_k - \overline{x}\| + (1-\alpha_k)\|\Gamma_k w_k - \overline{x}\|$$
$$\leq \|w_k - \overline{x}\|. \tag{4.1.9}$$

and

$$\begin{aligned}\|y_k - \overline{x}\| &= \|\beta_k w_k + (1-\beta_k)\Gamma_k z_k - \overline{x}\| \\ &\leq \beta_k\|w_k - \overline{x}\| + (1-\beta_k)\|\Gamma_k z_k - \overline{x}\| \\ &\leq \beta_k\|w_k - \overline{x}\| + (1-\beta_k)\|z_k - \overline{x}\|. \end{aligned}\tag{4.1.10}$$

On the other hand, we consider

$$\begin{aligned}\|w_k - \overline{x}\| &= \|x_k + \theta_k(x_k - x_{k-1} - \overline{x})\| \\ &\leq \|x_k - \overline{x}\| + \theta_k\|x_k - x_{k-1}\|. \end{aligned}\tag{4.1.11}$$

Combining (4.1.9), (4.1.10) and (4.1.11), we obtain that

$$\begin{aligned}\|x_{k+1} - \overline{x}\| &= \|\gamma_k f(x_k) + (1-\gamma_k)y_k - \overline{x}\| \\ &\leq \gamma_k\|f(x_k) - \overline{x}\| + (1-\gamma_k)\|y_k - \overline{x}\| \\ &\leq \gamma_k\|f(x_k) - f(\overline{x})\| + \gamma_k\|f(\overline{x}) - \overline{x}\| + (1-\gamma_k)\|w_k - \overline{x}\| \\ &\leq \gamma_k c\|x_k - \overline{x}\| + \gamma_k\|f(\overline{x}) - \overline{x}\| + (1-\gamma_k)\|x_k - \overline{x}\| \\ &\quad + (1-\gamma_k)\theta_k\|x_k - x_{k-1}\| \\ &\leq (1-\gamma_k(1-c))\|x_k - \overline{x}\| + \gamma_k\|f(\overline{x}) - \overline{x}\| \\ &\quad + (1-\gamma_k)\theta_k\|x_k - x_{k-1}\| \\ &\leq (1-\gamma_k(1-c))\|x_k - \overline{x}\| + \gamma_k\|f(\overline{x}) - \overline{x}\| \\ &\quad + (1-\gamma_k(1-c))\theta_k\|x_k - x_{k-1}\|. \end{aligned}\tag{4.1.12}$$

Since $\lim_{k\to\infty}\frac{\theta_k}{\gamma_k}\|x_k - x_{k-1}\| = 0$, there exists $M > 0$ such that

$$\frac{(1-\gamma_k(1-c))\theta_k}{\gamma_k}\|x_k - x_{k-1}\| \leq M \text{ for all } k \in \mathbb{N}.$$

Let $M_1 := \frac{2}{1-\mu}\max\{\|f(\overline{x}) - \overline{x}\|, M\}$.

From (4.1.12) and by using mathematical induction, we get that

$$\|x_{k+1} - \overline{x}\| \leq (1-\gamma_k(1-c))\|x_k - \overline{x}\| + \gamma_k(1-c)M_1$$

$$\leq \max\{\|x_k - \overline{x}\|, \|f(\overline{x}) - \overline{x}\|\}$$
$$\vdots$$
$$\leq \max\{\|x_1 - \overline{x}\|, \|f(\overline{x}) - \overline{x}\|\}. \tag{4.1.13}$$

Therefore, $\{x_k\}_{k\in\mathbb{N}}$ is bounded. So $\{w_k\}_{k\in\mathbb{N}}$, $\{z_k\}_{k\in\mathbb{N}}$ and $\{y_k\}_{k\in\mathbb{N}}$ are also bounded. By using the condition (iv) in Lemma 2.3.14 and the definition of $\{x_k\}_{k\in\mathbb{N}}$ and $\{y_k\}_{k\in\mathbb{N}}$, we get that

$$\begin{aligned}\|z_k - \overline{x}\|^2 &= \|\alpha_k w_k + (1-\alpha_k)\Gamma_k w_k - \overline{x}\|^2 \\ &\leq \alpha_k\|w_k - \overline{x}\|^2 + (1-\alpha_k)\|\Gamma_k w_k - \overline{x}\|^2\end{aligned} \tag{4.1.14}$$

and

$$\begin{aligned}\|y_k - \overline{x}\|^2 &= \|\beta_k w_k + (1-\beta_k)\Gamma_k z_k - \overline{x}\|^2 \\ &\leq \beta_k\|w_k - \overline{x}\|^2 + (1-\beta_k)\|\Gamma_k z_k - \overline{x}\|^2.\end{aligned} \tag{4.1.15}$$

Now, consider terms $\|\Gamma_k w_k - \overline{x}\|^2$ and $\|\Gamma_k z_k - \overline{x}\|^2$ by using Lemma 2.4.4, we have

$$\begin{aligned}\|\Gamma_k w_k - \overline{x}\|^2 &= \|\Gamma_k w_k - \Gamma_k\overline{x}\|^2 \\ &\leq \|w_k - \overline{x}\|^2 - \lambda_k(2\mu - \lambda_k)\|Aw_k - A\overline{x}\|^2 \\ &\quad - \|w_k - \lambda_k Aw_k - \Gamma_k w_k + \lambda_k A\overline{x}\|^2.\end{aligned} \tag{4.1.16}$$

and

$$\begin{aligned}\|\Gamma_k z_k - \overline{x}\|^2 &= \|\Gamma_k z_k - \Gamma_k\overline{x}\|^2 \\ &\leq \|z_k - \overline{x}\|^2 - \lambda_k(2\mu - \lambda_k)\|Az_k - A\overline{x}\|^2 \\ &\quad - \|z_k - \lambda_k Az_k - \Gamma_k z_k + \lambda_k A\overline{x}\|^2.\end{aligned} \tag{4.1.17}$$

Substituting (4.1.16) into (4.1.14), we have

$$\begin{aligned}\|z_k - \overline{x}\|^2 &\leq \|w_k - \overline{x}\|^2 - (1-\alpha_k)\lambda_k(2\mu - \lambda_k)\|Aw_k - A\overline{x}\|^2 \\ &\quad - (1-\alpha_k)\|w_k - \lambda_k Aw_k - \Gamma_k w_k + \lambda_k A\overline{x}\|^2.\end{aligned} \tag{4.1.18}$$

Substituting (4.1.17) into (4.1.15), we have

$$\|y_k - \overline{x}\|^2 \leq \beta_k\|w_k - \overline{x}\|^2 + (1-\beta_k)\|z_k - \overline{x}\|^2$$

$$
\begin{aligned}
&- (1-\beta_k)\lambda_k(2\mu-\lambda_k)\|Az_k - A\overline{x}\|^2 \\
&- (1-\beta_k)\|z_k - \lambda_k Az_k - \Gamma_k z_k + \lambda_k A\overline{x}\|^2. \qquad (4.1.19)
\end{aligned}
$$

Combining (4.1.18) and (4.1.19), we can imply that

$$
\begin{aligned}
\|y_k - \overline{x}\|^2 \leq \|w_k - \overline{x}\|^2 &+ (1-\beta_k)(1-\alpha_k)\lambda_k(2\mu-\lambda_k)\|Aw_k - A\overline{x}\|^2 \\
&- (1-\beta_k)(1-\alpha_k)\|w_k - \lambda_k Aw_k - \Gamma_k w_k + \lambda_k A\overline{x}\|^2 \\
&- (1-\beta_k)\lambda_k(2\mu-\lambda_k)\|Az_k - A\overline{x}\|^2 \\
&- (1-\beta_k)\|z_k - \lambda_k Az_k - \Gamma_k z_k + \lambda_k A\overline{x}\|^2. \qquad (4.1.20)
\end{aligned}
$$

From (4.1.20), we obtain

$$
\begin{aligned}
\|x_{k+1} - \overline{x}\|^2 &= \langle \gamma_k f(x_k) + (1-\gamma_k)y_k - \overline{x}, x_{k+1} - \overline{x}\rangle \\
&= \langle \gamma_k (f(x_k) - \overline{x}), x_{k+1} - \overline{x}\rangle + \langle (1-\gamma_k)(y_k - \overline{x}), x_{k+1} - \overline{x}\rangle \\
&= \gamma_k \langle f(x_k) - f(\overline{x}), x_{k+1} - \overline{x}\rangle + \gamma_k\langle f(\overline{x}) - \overline{x}, x_{k+1} - \overline{x}\rangle \\
&\quad + (1-\gamma_k)\langle y_k - \overline{x}, x_{k+1} - \overline{x}\rangle \\
&\leq \gamma_k \|f(x_k) - f(\overline{x})\|\|x_{k+1} - \overline{x}\| + \gamma_k\langle f(\overline{x}) - \overline{x}, x_{k+1} - \overline{x}\rangle \\
&\quad + (1-\gamma_k)\|y_k - \overline{x}\|\|x_{k+1} - \overline{x}\| \\
&\leq \frac{\gamma_k}{2}\left(\|f(x_k) - f(\overline{x})\|^2 + \|x_{k+1} - \overline{x}\|^2\right) + \gamma_k\langle f(\overline{x}) - \overline{x}, x_{k+1} - \overline{x}\rangle \\
&\quad + \frac{(1-\gamma_k)}{2}\left(\|y_k - \overline{x}\|^2 + \|x_{k+1} - \overline{x}\|^2\right) \\
&\leq \frac{\gamma_k c^2}{2}\|x_k - \overline{x}\|^2 + \frac{\gamma_k}{2}\|x_{k+1} - \overline{x}\|^2 + \gamma_k\langle f(\overline{x}) - \overline{x}, x_{k+1} - \overline{x}\rangle \\
&\quad + \frac{(1-\gamma_k)}{2}\|w_k - \overline{x}\|^2 \\
&\quad - \frac{(1-\gamma_k)(1-\beta_k)(1-\alpha_k)}{2}\lambda_k(2\mu-\lambda_k)\|Aw_k - A\overline{x}\|^2 \\
&\quad - \frac{(1-\gamma_k)(1-\beta_k)(1-\alpha_k)}{2}\|w_k - \lambda_k Aw_k - \Gamma_k w_k + \lambda_k A\overline{x}\|^2 \\
&\quad - \frac{(1-\gamma_k)(1-\beta_k)\lambda_k(2\mu-\lambda_k)}{2}\|Az_k - A\overline{x}\|^2 \\
&\quad - \frac{(1-\gamma_k)(1-\beta_k)}{2}\|z_k - \lambda_k Az_k - \Gamma_k z_k + \lambda_k A\overline{x}\|^2 \\
&\quad + \frac{(1-\gamma_k)}{2}\|x_{k+1} - \overline{x}\|^2 \\
&\leq \frac{\gamma_k c^2}{2}\|x_k - \overline{x}\|^2 + \frac{1}{2}\|x_{k+1} - \overline{x}\|^2 + \gamma_k\langle f(\overline{x}) - \overline{x}, x_{k+1} - \overline{x}\rangle \\
&\quad + \frac{(1-\gamma_k)}{2}\left(\|x_k - \overline{x}\|^2 + 2\theta_k\langle x_k - x_{k-1}, w_k - \overline{x}\rangle\right)
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{(1-\gamma_k)(1-\beta_k)(1-\alpha_k)}{2}\lambda_k(2\mu-\lambda_k)\|Aw_k-A\overline{x}\|^2\\
&-\frac{(1-\gamma_k)(1-\beta_k)(1-\alpha_k)}{2}\|w_k-\lambda_kAw_k-\Gamma_kw_k+\lambda_kA\overline{x}\|^2\\
&-\frac{(1-\gamma_k)(1-\beta_k)\lambda_k(2\mu-\lambda_k)}{2}\|Az_k-A\overline{x}\|^2\\
&-\frac{(1-\gamma_k)(1-\beta_k)}{2}\|z_k-\lambda_kAz_k-\Gamma_kz_k+\lambda_kA\overline{x}\|^2\\
\leq&\ \frac{(1-\gamma_k(1-c^2))}{2}\|x_k-\overline{x}\|^2+\frac{1}{2}\|x_{k+1}-\overline{x}\|^2\\
&+\gamma_k\langle f(\overline{x})-\overline{x},x_{k+1}-\overline{x}\rangle+(1-\gamma_k)\theta_k\langle x_k-x_{k-1},w_k-\overline{x}\rangle\\
&-\frac{(1-\gamma_k)(1-\beta_k)(1-\alpha_k)}{2}\lambda_k(2\mu-\lambda_k)\|Aw_k-A\overline{x}\|^2\\
&-\frac{(1-\gamma_k)(1-\beta_k)(1-\alpha_k)}{2}\|w_k-\lambda_kAw_k-\Gamma_kw_k+\lambda_kA\overline{x}\|^2\\
&-\frac{(1-\gamma_k)(1-\beta_k)\lambda_k(2\mu-\lambda_k)}{2}\|Az_k-A\overline{x}\|^2\\
&-\frac{(1-\gamma_k)(1-\beta_k)}{2}\|z_k-\lambda_kAz_k-\Gamma_kz_k+\lambda_kA\overline{x}\|^2. \qquad (4.1.21)
\end{aligned}
$$

Then (4.1.21) reduces to the following:

$$
\begin{aligned}
\|x_{k+1}-\overline{x}\|^2\leq&\ (1-\gamma_k(1-c^2))\|x_k-\overline{x}\|^2+2\gamma_k\langle f(\overline{x})-\overline{x},x_{k+1}-\overline{x}\rangle\\
&+2(1-\gamma_k)\theta_k\langle x_k-x_{k-1},w_k-\overline{x}\rangle\\
&-(1-\gamma_k)(1-\beta_k)(1-\alpha_k)\lambda_k(2\mu-\lambda_k)\|Aw_k-A\overline{x}\|^2\\
&-(1-\gamma_k)(1-\beta_k)(1-\alpha_k)\|w_k-\lambda_kAw_k-\Gamma_kw_k+\lambda_kA\overline{x}\|^2\\
&-(1-\gamma_k)(1-\beta_k)\lambda_k(2\mu-\lambda_k)\|Az_k-A\overline{x}\|^2\\
&-(1-\gamma_k)(1-\beta_k)\|z_k-\lambda_kAz_k-\Gamma_kz_k+\lambda_kA\overline{x}\|^2. \qquad (4.1.22)
\end{aligned}
$$

For each $k\in\mathbb{N}$, we set

$S_k=\|x_{k+1}-\overline{x}\|^2,$

$\rho_k=\gamma_k(1-c^2),\ \pi_k=\rho_k\sigma_k,$

$\sigma_k=\frac{2}{(1-c^2)}\langle f(\overline{x})-\overline{x},x_{k+1}-\overline{x}\rangle+\frac{2(1-\gamma_k)\theta_k}{\gamma_k(1-c^2)}\langle x_k-x_{k-1},w_k-\overline{x}\rangle$ and

$$
\begin{aligned}
\eta_k=&\ (1-\gamma_k)(1-\beta_k)(1-\alpha_k)\lambda_k(2\mu-\lambda_k)\|Aw_k-A\overline{x}\|^2\\
&+(1-\gamma_k)(1-\beta_k)(1-\alpha_k)\|w_k-\lambda_kAw_k-\Gamma_kw_k+\lambda_kA\overline{x}\|^2
\end{aligned}
$$

$$+ (1-\gamma_k)(1-\beta_k)\lambda_k(2\mu-\lambda_k)\|Az_k - A\overline{x}\|^2$$
$$+ (1-\gamma_k)(1-\beta_k)\|z_k - \lambda_k Az_k - \Gamma_k z_k + \lambda_k A\overline{x}\|^2.$$

As a result, inequality (4.1.22) reduces to the following:

$$S_{k+1} \leq (1-\rho_k)S_k + \rho_k\sigma_k \text{ and } S_{k+1} \leq S_k - \eta_k + \pi_k.$$

By the conditions (i) and (ii), we get that $\sum_{k=1}^{+\infty}\rho_k = +\infty$ and $\lim_{k\to+\infty}\pi_k = 0$. In order to complete proof, by applying Lemma 2.4.1, it is sufficient to show that $\lim_{k\to+\infty}\eta_{k_i} = 0$ implies $\limsup_{i\to+\infty}\sigma_{k_i} \leq 0$ for any subsequence $\{\eta_{k_i}\}_{i\in\mathbb{N}}$ of $\{\eta_k\}_{k\in\mathbb{N}}$.

Let $\{\eta_{k_i}\}_{i\in\mathbb{N}}$ be a subsequence of $\{\eta_k\}_{k\in\mathbb{N}}$ such that $\lim_{i\to+\infty}\eta_{k_i} = 0$. Therefore, by the assumptions of Lemma 2.4.1, we can conclude that

$$\lim_{i\to+\infty}\|Aw_{k_i} - A\overline{x}\| = 0;$$
$$\lim_{i\to+\infty}\|Az_{k_i} - A\overline{x}\| = 0;$$
$$\lim_{i\to+\infty}\|w_{k_i} - \lambda_{k_i}Aw_{k_i} - \Gamma_{k_i}w_{k_i} + \lambda_{k_i}A\overline{x}\| = 0;$$
$$\lim_{i\to+\infty}\|z_{k_i} - \lambda_{k_i}Az_{k_i} - \Gamma_{k_i}z_{k_i} + \lambda_{k_i}A\overline{x}\| = 0.$$

This implies that

$$\lim_{i\to+\infty}\|\Gamma_{k_i}w_{k_i} - w_{k_i}\| = 0; \tag{4.1.23}$$
$$\lim_{i\to+\infty}\|\Gamma_{k_i}z_{k_i} - z_{k_i}\| = 0. \tag{4.1.24}$$

From (ii), we have

$$\|w_{k_i} - x_{k_i}\| = \theta_{k_i}\|x_{k_i} - x_{k_{i-1}}\| \to 0 \ (i\to+\infty). \tag{4.1.25}$$

On the other hand, we get

$$\begin{aligned}\|\Gamma_{k_i}z_{k_i} - w_{k_i}\| &\leq \|\Gamma_{k_i}z_{k_i} - z_{k_i}\| + \|z_{k_i} - w_{k_i}\| \\ &= \|\Gamma_{k_i}z_{k_i} - z_{k_i}\| + (1-\alpha_{k_i})\|\Gamma_{k_i}w_{k_i} - w_{k_i}\|.\end{aligned} \tag{4.1.26}$$

From (4.1.23) and (4.1.24), we otain that

$$\lim_{i\to+\infty}\|\Gamma_{k_i}z_{k_i} - w_{k_i}\| = 0. \tag{4.1.27}$$

Since $\liminf_{k\to+\infty} \lambda_k > 0$, there exists $\lambda > 0$ such that $\lambda_k \geq \lambda$ for all $k \in \mathbb{N}$. In particular, $\lambda_{k_i} \geq \lambda$ for all $i \in \mathbb{N}$. By the condition (2.4.6) of Lemma 2.4.6, one has

$$\|\Gamma_\lambda^{A,B} w_{k_i} - w_{k_i}\| \leq 2\|\Gamma_{k_i} w_{k_i} - w_{k_i}\|. \tag{4.1.28}$$

From (4.1.28), we can get that

$$\lim_{i\to+\infty} \|\Gamma_\lambda^{A,B} w_{k_i} - w_{k_i}\| = 0. \tag{4.1.29}$$

Let

$$z_t = tf(\overline{x}) + (1-t)\Gamma_\lambda^{A,B} z_t,\ t \in (0,1). \tag{4.1.30}$$

Applying Theorem 2.3.16, z_t converges strongly to the unique fixed point $\overline{x} = \text{proj}_{(A+B)^{-1}(0)}(f(\overline{x}))$ as $t \to 0$. Therefore, we have

$$\begin{aligned}
\|z_t - w_{k_i}\|^2 &= \|t(f(\overline{x}) - w_{k_i}) + (1-t)(\Gamma_\lambda^{A,B} z_t - w_{k_i})\|^2 \\
&\leq (1-t)^2\|\Gamma_\lambda^{A,B} z_t - w_{k_i}\|^2 + 2t\langle f(\overline{x}) - z_t, z_t - w_{k_i}\rangle \\
&\quad + 2t\langle z_t - w_{k_i}, z_t - w_{k_i}\rangle \\
&\leq (1-t)^2(\|\Gamma_\lambda^{A,B} z_t - \Gamma_\lambda^{A,B} w_{k_i}\| + \|\Gamma_\lambda^{A,B} w_{k_i} - w_{k_i}\|)^2 \\
&\quad + 2t\langle f(\overline{x}) - z_t, z_t - w_{k_i}\rangle + 2t\|z_t - w_{k_i}\|^2 \\
&\leq (1-t)^2(\|z_t - w_{k_i}\| + \|\Gamma_\lambda^{A,B} w_{k_i} - w_{k_i}\|)^2 \\
&\quad + 2t\langle f(\overline{x}) - z_t, z_t - w_{k_i}\rangle + 2t\|z_t - w_{k_i}\|^2.
\end{aligned} \tag{4.1.31}$$

The inequality (4.1.31) reduces to the following:

$$\begin{aligned}
&\langle z_t - f(\overline{x}), z_t - w_{k_i}\rangle \\
&\quad \leq \frac{(1-t)^2}{2t}(\|z_t - w_{k_i}\| + \|\Gamma_\lambda^{A,B} w_{k_i} - w_{k_i}\|)^2 + \frac{(2t-1)}{2t}\|z_t - w_{k_i}\|^2.
\end{aligned} \tag{4.1.32}$$

Combining (4.1.27) and (4.1.32), we get that

$$\limsup_{i\to+\infty} \langle z_t - f(\overline{x}), z_t - w_{k_i}\rangle \leq \frac{1}{2t}[(1-t)^2 + (2t-1)]M_0^2, \tag{4.1.33}$$

where $M_0 = \sup_{i\in\mathbb{N}, t\in(0,1)} \|z_t - w_{k_i}\|$. By taking $t \to 0$ in (4.1.33), we obtain that

$$\limsup_{i\to+\infty} \langle \overline{x} - f(\overline{x}), \overline{x} - w_{k_i}\rangle \leq 0. \tag{4.1.34}$$

Let us consider,

$$\begin{aligned}\langle z - f(\overline{x}), z - x_{k_i}\rangle &= \langle z - f(\overline{x}), z - w_{k_i}\rangle + \theta_{k_i}\langle z - f(\overline{x}), x_{k_i} - x_{k_{i-1}}\rangle \\ &\leq \langle z - f(\overline{x}), z - w_{k_i}\rangle + \theta_{k_i}\|z - f(\overline{x})\|\|x_{k_i} - x_{k_{i-1}}\|\end{aligned} \tag{4.1.35}$$

From (4.1.35), one has

$$\limsup_{i\to+\infty}\langle \overline{x} - f(\overline{x}), \overline{x} - x_{k_i}\rangle \leq 0. \tag{4.1.36}$$

Next, we claim that $\lim_{i\to+\infty}\|x_{k_i+1} - x_{k_i}\| = 0$. By Algorithm 4, we have the following estimates:

$$\begin{aligned}\|x_{k_{i+1}} - x_{k_i}\| &\leq \gamma_{k_i}\|f(\overline{x}) - x_{k_i}\| + (1-\gamma_{k_i})\|y_{k_i} - x_{k_i}\| \\ &\leq \gamma_{k_i}\|f(\overline{x}) - x_{k_i}\| + (1-\gamma_{k_i})(\|y_{k_i} - w_{k_i}\| + \|w_{k_i} - x_{k_i}\|) \\ &\leq \gamma_{k_i}\|f(\overline{x}) - x_{k_i}\| + (1-\gamma_{k_i})\|w_{k_i} - x_{k_i}\| \\ &\quad + (1-\gamma_{k_i})(1-\beta_{k_i})\|\Gamma_{k_i}z_{k_i} - w_{k_i}\|.\end{aligned} \tag{4.1.37}$$

From (4.1.37), by using the boundedness of $\{x_k\}_{k\in\mathbb{N}}$, the condition i, and (4.1.25) and (4.1.27), we obtain that

$$\lim_{i\to+\infty}\|x_{k_i+1} - x_{k_i}\| = 0. \tag{4.1.38}$$

Combining (4.1.38) and (4.1.36), we infer that

$$\limsup_{i\to+\infty}\langle \overline{x} - f(\overline{x}), \overline{x} - x_{k_{i+1}}\rangle \leq 0.$$

Hence, $\limsup_{i\to+\infty}\sigma_{k_i} \leq 0$. By Lemma 2.4.1, we observe that $\lim_{k\to+\infty}S_k = 0$, that is $x_k \to \overline{x}$ as $k \to +\infty$. We thus complete the proof. □

Remark 4.1.3. The condition (ii) in Theorem 4.1.2 is verified, if we choose θ_k such that $0 \leq \theta_k \leq \overline{\theta}_k$, where

$$\overline{\theta}_k = \begin{cases} \min\left\{\theta, \frac{\varepsilon_k}{\|x_k - x_{k-1}\|}\right\}, & \text{if } x_k \neq x_{k-1}, \\ \theta, & \text{otherwise,}\end{cases}$$

and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ is a positive sequence such that $\lim_{k\to+\infty}\frac{\varepsilon_k}{\gamma_k} = 0$.

4.1.1 Applications

This subsection, we present the applications of the Algorithm 4 in the previous part in convex minimization problems and image restoration problems.

Convex Minimization Problems

Let $h : \mathcal{H} \to \mathbb{R}$ be convex and differentiable function and $g : \mathcal{H} \to \mathbb{R}$ be a convex lower-semicontinuous function. We consider the convex minimization problem as follows: find $\overline{x} \in \mathcal{H}$ such that

$$h(\overline{x}) + g(\overline{x}) = \min_{x \in \mathcal{H}} \{h(x) + g(x)\}. \tag{4.1.39}$$

By using Fermats rule, the problem (4.1.39) can be written in the form of the following problem as: find $\overline{x} \in \mathcal{H}$ such that

$$0 \in \nabla h(\overline{x}) + \partial g(\overline{x}),$$

where ∇h is a gradient of h and ∂g is a subdifferential of g.

Remark 4.1.4. [84] If a function $K : \mathcal{H} \to \mathcal{H}$ is $(1/L)$-Lipschitz continuous, then K is L-inverse strongly monotone.

By applying Theorem 4.1.2 and set $A = \nabla h$ and $B = \partial g$, we can obtain the following result.

Theorem 4.1.5. *Let $\mathcal{H}$ be a real Hilbert space. Let $h : \mathcal{H} \to \mathbb{R}$ be convex differentiable function with a $(1/L)$-Lipschitz continuous gradient ∇h and $g : \mathcal{H} \to \mathbb{R}$ be a convex lower-semicontinuous such that $(\nabla h + \partial g)^{-1}(0) \neq \emptyset$. Let $f : \mathcal{H} \to \mathcal{H}$ be a contraction mapping with constant $c \in (0,1)$. Let $\{x_k\}_{k \in \mathbb{N}}$ be generated by $x_0, x_1 \in \mathcal{H}$*

$$\begin{cases} w_k = x_k + \theta_k(x_k - x_{k-1}), \\ z_k = \alpha_k w_k + (1 - \alpha_k) J_{\lambda_k}^{\partial g}(w_k - \lambda_k \nabla h w_k), \\ y_k = \beta_k w_k + (1 - \beta_k) J_{\lambda_k}^{\partial g}(z_k - \lambda_k \nabla h z_k), \\ x_{k+1} = \gamma_k f(x_k) + (1 - \gamma_k) y_k, \qquad \text{for all } k \geq 1. \end{cases} \tag{4.1.40}$$

Assume that the following conditions hold:

(i) $\lim_{k\to+\infty}\gamma_k = 0$ *and* $\sum_{k\geq 1}\gamma_k = +\infty$;

(ii) $\lim_{k\to+\infty}\frac{\theta_k}{\gamma_k}\|x_k - x_{k-1}\| = 0$;

(iii) $0 < \liminf_{k\to+\infty}\lambda_k \leq \limsup_{k\to+\infty}\lambda_k < 2L$;

(iv) $\liminf_{k\to+\infty}(1-\alpha_k)(1-\beta_k) > 0$.

Then, the sequence $\{x_k\}_{k\in\mathbb{N}}$ *converges strongly to* $\overline{x} := \mathrm{proj}_{(\nabla h+\partial g)^{-1}(0)}(f(\overline{x}))$.

Next, we present some comparisons between our method and Kitkuan et al.s algorithm in Equation (4.1.6).

Example 4.1.6. Let $\mathbf{K} \in \mathbb{R}^{l\times s}$ and $\mathrm{b} \in \mathbb{R}^l$ with $s > l$. Let $g : \mathbb{R}^s \to \mathbb{R}$ be defined by $g(x) = \|x\|_1$ for all $x \in \mathbb{R}^s$, and $h : \mathbb{R}^s \to \mathbb{R}$ be defined by $h(x) = \frac{1}{2}\|\mathbf{K}x - \mathbf{b}\|_2^2$ for all $x \in \mathbb{R}^s$. To find the solution of the minimization problem as follows:

$$\begin{aligned}&\text{minimize } \frac{1}{2}\|\mathbf{K}x - \mathbf{b}\|_2^2 + \|x\|_1,\\ &\text{subject to } x \in \mathbb{R}^s.\end{aligned} \tag{4.1.41}$$

By setting this, we obtain that for each $x = (x^1, x^2, ..., x^s) \in \mathbb{R}^s$

$$J_{\lambda_k}^{\partial g}(x) = (\max\{0, 1 - \frac{\lambda_k}{|x^1|}\}x_1, \max\{0, 1 - \frac{\lambda_k}{|x^2|}\}x_2, ..., \max\{0, 1 - \frac{\lambda_k}{|x^s|}\}x_s),$$

$\nabla h(x) = \mathbf{K}^T(\mathbf{K}x - \mathbf{b})$ and ∇h is $\|\mathbf{K}\|^2$-Lipschitz continuous, where $\mathbf{K}^T$ is a transpose of $\mathbf{K}$.

Firstly, we random vectors $x_0, x_1 \in \mathbb{R}^s$ and $\mathbf{b} \in \mathbb{R}^l$ and matrix $\mathbf{K} \in \mathbb{R}^{l\times s}$. After that, we compare the performance between our algorithm and Kitkuan et al.'s algorithm (4.1.6). We set $f(x) = \frac{x}{6}$ for all $x \in \mathbb{R}^s$. We choose the parameters in this example as follows: $\alpha_k = \frac{1}{100k+1}$, $\beta_k = \frac{1}{k+1}$, $\gamma_k = \frac{1}{100k+1}$, $\lambda_k = \frac{1}{\|\mathbf{K}\|^2+1}$ and

$$\theta_k = \begin{cases} \min\left\{\frac{1}{2}, \frac{1}{(k+1)^2\|x_k - x_{k-1}\|}\right\}, & \text{if } x_k \neq x_{k-1},\\ \frac{1}{2}, & \text{otherwise.}\end{cases} \tag{4.1.42}$$

We perform two algorithms and obtain the number of iterations (k) and the elapsed times (seconds) by using the stopping criteria : $\|x_k - x_{k-1}\| \leq 10^{-6}$.

Table 5 shows the behaviors of two algorithms for the problem (4.1.41) with different sizes of matrix $\mathbf{K}$. In Table 5, we see that our algorithm requires the least elapsed time to reach the optimality tolerance for all cases. Furthermore, we can observe that for each size of matrix $\mathbf{K}$, the number of iterations of our algorithm are smaller than the number of iterations of Kitkuan et al.'s algorithm (4.1.6).

Table 5: The comparison of two algorithms with different sizes of matrix $\mathbf{K}$

(l, s)	Our algorithm		Kitkuan et al.'s algorithm (4.1.6)	
	Elapsed time (s)	Iterations	Elapsed time (s)	Iterations
(20,500)	0.3558	4901	1.1314	18517
(50,500)	0.5561	6566	0.7707	15316
(300,500)	0.5716	3158	0.8861	8902
(20,1000)	0.6193	8016	0.9908	19346
(50,1000)	0.9604	8111	1.3478	20877
(300,1000)	2.0653	4871	2.4289	11560
(500,1000)	3.5339	4590	5.8424	12641
(20,2000)	1.1691	8789	1.4671	18546
(50,2000)	0.7983	2869	0.8291	5411
(300,2000)	2.4663	2385	4.2916	6248
(500,2000)	10.2722	4590	11.7226	10107
(1000,2000)	13.9273	3052	20.0587	8617

Image Restoration Problems

In this subsection, we demonstrate the effectiveness of the proposed algorithm by applying to solve the image restoration problems, which involves deblurring and denoising images. The image restoration problem can be formulated by the inversion of the following degradation model:

$$y = \mathbf{H}x + w, \tag{4.1.43}$$

where y, $\mathbf{H}$, x and w represent the degraded image, degradation operator or blurring operator, original image and noise operator, respectively.

To approximate the reconstructed image is obtained by solving the following regularized least-squares minimization problem

$$\min_{x}\left\{\frac{1}{2}\|\mathbf{H}x-y\|_2^2+\mu\phi(x)\right\}, \tag{4.1.44}$$

where $\mu > 0$ is the regularization parameter and $\phi(\cdot)$ is the regularizer. The l_1 norm is a regularization functional, which is well-known that it is used to remove noise in the restoration problem. This is called Tikhonov regularization [85]. The problem (4.1.44) can be formulated by the following problem as:

$$\text{find } x \in \arg\min_{x\in\mathbb{R}^s}\left\{\frac{1}{2}\|\mathbf{H}x-y\|_2^2+\mu\|x\|_1\right\}, \tag{4.1.45}$$

where y is the degraded image and $\mathbf{H}$ is a bounded linear operator. We can see that problem (4.1.45) can be formed in the problem (4.1.1) by setting $B = \partial\|\cdot\|_1$, $\mu = 0.001$ and $A = \nabla L(\cdot)$ where $L(x) = \frac{1}{2}\|\mathbf{H}x - y\|_2^2$. By using this we observe that $A(x) = \nabla L(x) = \mathbf{H}^T(\mathbf{H}x - y)$. Firstly, we degrade image by adding random noise and different types of blurring. Next, solving the problem (4.1.45) by using our algorithm in Theorem 4.1.5 and putting $f(x) = \frac{x}{2}$ for all $x \in \mathbb{R}^s$, $\alpha_k = \frac{1}{k+1}$, $\beta_k = \frac{1}{k+1}$, $\gamma_k = \frac{1}{100k+1}$, $\lambda_k = 0.7$ and θ_k is defined as (4.1.42).

The comparisons of the performance between our proposed algorithm and Kitkuan et al.'s algorithm in Equation (4.1.6) that was introduced by Kitkuan et al. [46] are presented. The quality of the reconstructed image is measured by means of the signal to noise ratio (SNR), that is,

$$\mathbf{SNR}(k) = 20\log_{10}\frac{\|x\|_2^2}{\|x-x_k\|_2^2},$$

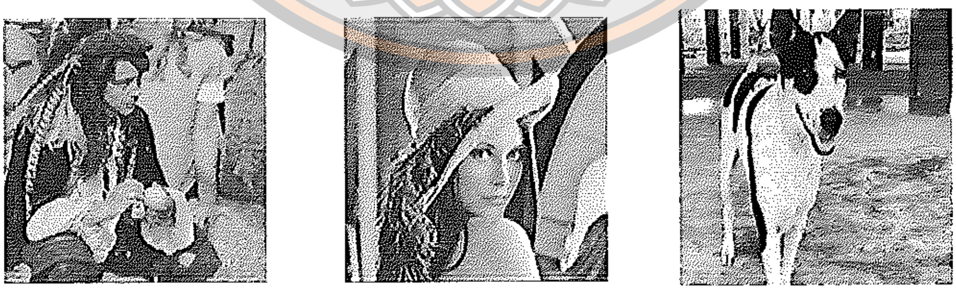
where x and x_k denote the original and the restored image at iteration k, respectively.

The comparisons between our proposed algorithm in Theorem 4.1.5 and Kitkuan et al.'s algorithm (4.1.6) in image restoration problems are presented in Figure 3-4.

(a)

(b)

(c)

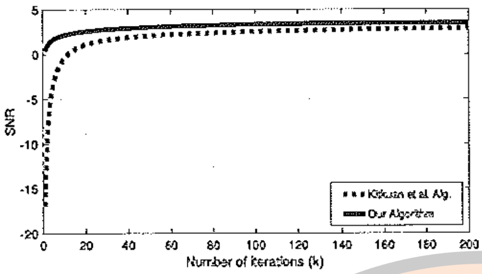


(d)

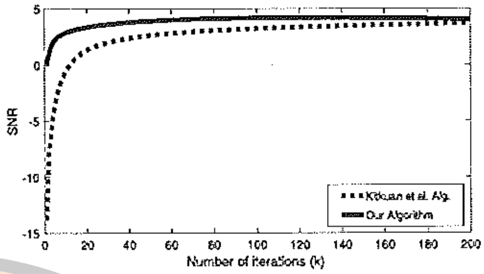
Figure 3: The degraded and reconstructed images with different techniques

Figure 3: **(a)** shows the original image 'Pirate', 'Lena', and 'Dog' image, respectively; **(b)** shows the images degraded by Gaussian blur and random noise,

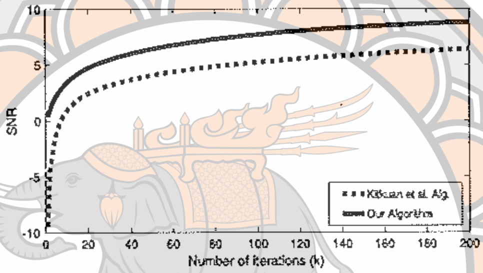
average blur and random noise, and motion blur and random noise, respectively; (**c**) shows the reconstructed images by using Kitkuan et al.'s algorithm; and (**d**) shows the reconstructed images by using our algorithm in Equation (4.1.40).



(**a**) Pirate



(**b**) Lena



(**c**) Dog

Figure 4: Illustration of the behavior of SNR for our algorithm and Kitkuan et al.'s algorithm (4.1.6) in Figure 3c,d

Figure 4: (**a**) The behavior of **SNR** for two algorithms of 'Pirate' image in Figure 3c,d; (**b**) The behavior of **SNR** for two algorithms of 'Lena' image in Figure 3c,d;and (**c**) The behavior of **SNR** for two algorithms of 'Dog' image in Figure 3c,d.

4.2 Inertial Mann-type Algorithm for a Nonexpansive Mapping to Solve Monotone Inclusion Problems

In this section, we propose iterative method to solve the following monotone inclusion problem.

$$\text{Find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx + Cx, \tag{4.2.1}$$

where A, B, C are maximal monotone operators on a Hilbert space $\mathcal{H}$ and C is δ-cocoercive with parameter $\delta > 0$. The problem (4.2.1) was considered by Davis and Yin [86] and it can be reformulated to the fixed point problem for nonexpansive mappings. Therefore, it is interesting to study the fixed point problem in order to apply for solving the problem (4.2.1).

Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping.
Problem: the fixed point problem for the mapping T generally denote as,

$$\text{find } x \in \mathcal{H} \text{ such that } x = Tx.$$

A solution of the fixed point problem for nonexpansive mappings was approximated by the iterative method which was introduced by Mann [48]. In addition, the "Mann Iteration" stated that

$$x_{k+1} = \alpha_k x_k + (1 - \alpha_k)Tx_k, \quad \forall k \geq 1, \tag{4.2.2}$$

where $x_1 \in \mathcal{H}$ and $\{\alpha_k\}_{k \in \mathbb{N}}$ is a real sequence in $[0, 1]$. The weak convergent result of the iterative sequence $\{x_k\}_{k \in \mathbb{N}}$ was obtained under control condition that $\sum_{k \geq 1} \alpha_k(1 - \alpha_k) = +\infty$, see [87, 88]. In order to obtain the strong convergence for the fixed point solutions of nonexpansive mappings, one of the most important methods to solve the fixed point problem for a nonexpansive mapping was introduced by Halpern [89]:

$$x_{k+1} = \alpha_k u + (1 - \alpha_k)Tx_k, \quad \forall k \geq 1, \tag{4.2.3}$$

where $x_1, u \in \mathcal{H}$ and $(\alpha_k)_{k \geq 1}$ is a real sequence in $[0, 1]$. In direction to study and improve this algorithm (4.2.3), many results have been presented (see [44, 50–56]). In 2000, Moudafi [57] proposed iterative method which involved the concept of viscosity to solve strong convergence of the iterate. Moreover, many authors were interested in studying and developing Moudafi's algorithm. The several methods that are in reference to this study are reviewed in the next extensively (see, for example, [49, 55, 58–61]). Recently, Bot et al. [62] proposed a new Mann-type algorithm (**MTA**) to solve the fixed point problem for a nonexpansive mapping and proved strong convergence of the iterate without using viscosity and projection method under some control conditions of parameters sequences. Their algorithm was defined by

$$\textbf{(MTA)} \quad x_{k+1} = (1-\alpha_k)\delta_k x_k + \alpha_k T\delta_k x_k, \quad \forall k \geq 1, \tag{4.2.4}$$

where $x_1 \in \mathcal{H}$ and $\{\alpha_k\}_{k\in\mathbb{N}}$, $\{\delta_k\}_{k\in\mathbb{N}}$ are sequences in $(0,1]$.

In 2015, Combettes and Yamada [63] presented a new Mann algorithm combining error term for solving a common fixed point of averaged nonexpansive mappings in a Hilbert space. By using the concept of the inertial method, the technique of Halpern method and error terms, Shehu et al. [41] introduced an algorithm for solving a fixed point of a nonexpansive mapping which was defined as follows:

$$\begin{cases} x_0, x_1 \in \mathcal{H}, \\ y_k = x_k + \theta_k(x_k - x_{k-1}), \\ x_{k+1} = \alpha_k x_0 + \delta_k y_k + \gamma_k T y_k + e_k, \end{cases} \tag{4.2.5}$$

for all $k \geq 1$, where $\{\theta_k\}_{k\in\mathbb{N}} \subseteq [0,\theta]$ with $\theta \in [0,1)$, $\{\alpha_k\}_{k\in\mathbb{N}}$, $\{\beta_k\}_{k\in\mathbb{N}}$ and $\{\gamma_k\}_{k\in\mathbb{N}}$ are sequences in $(0,1]$ and $\{e_k\}_{k\in\mathbb{N}}$ is a sequence in $\mathcal{H}$.

Being motivated by the above facts, we intend to accelerate the speed of convergence by avoiding the viscosity concept, hence, we propose a Mann-type method combining both inertial terms and errors for finding a fixed point of a nonexpansive mapping in a Hilbert space.

Let a nonexpansive mapping T from $\mathcal{H}$ into itself be such that $\text{Fix}(T) \neq \emptyset$. We propose the following algorithm.

Algorithm 5:

Initialization: Given $\{\theta_k\}_{k\in\mathbb{N}} \subseteq [0,\theta]$ with $\theta \in [0,1)$ two sequences $\{\alpha_k\}_{k\in\mathbb{N}}$ and $\{\beta_k\}_{k\in\mathbb{N}}$ in $(0,1]$ and a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ in $\mathcal{H}$. Choose $x_0, x_1 \in \mathcal{H}$ arbitrarily.

Iterative Steps: For a given current iterate $x_{k-1}, x_k \in \mathcal{H}$, calculate as follows:

Step 1. Compute

$$y_k = x_k + \theta_k(x_k - x_{k-1}).$$

Step 2. Compute

$$x_{k+1} = \delta_k y_k + \alpha_k(T\delta_k y_k - \delta_k y_k) + \varepsilon_k.$$

Update k := k + 1 and return to Step 1.

We first state the assumptions that we will assume to hold through the rest of this part.

Assumption 4.2.1. Let $\{\alpha_k\}_{k\in\mathbb{N}}$ and $\{\delta_k\}_{k\in\mathbb{N}}$ be sequences in $(0,1]$ and let $\{\varepsilon_k\}_{k\in\mathbb{N}}$ be a sequence in $\mathcal{H}$. Assume the conditions are verifiable, as follows.

(i) $\liminf_{k\to+\infty} \alpha_k > 0$ and $\sum_{k=1}^{+\infty} |\alpha_k - \alpha_{k-1}| < +\infty$,

(ii) $\lim_{k\to+\infty} \delta_k = 1$, $\sum_{k=1}^{+\infty}(1-\delta_k) = +\infty$ and $\sum_{k=1}^{+\infty} |\delta_k - \delta_{k-1}| < +\infty$,

(iii) $\sum_{k=1}^{+\infty} \|\varepsilon_k\| < +\infty$.

We have verified Assumption 4.2.1 as shown in the following remark.

Remark 4.2.2. Let $z \in \mathcal{H}$. We set $\delta_k = 1 - \frac{1}{k+2}$, $\alpha_k = \frac{1}{4} - \frac{1}{(k+3)^2}$ and $\varepsilon_k = \frac{z}{(k+1)^3}$ for all $k \geq 1$. It's easy to see that the Assumption 4.2.1 is satisfied.

We discuss the convergence analysis of the proposed algorithm. Beginning with given boundedness of our algorithm as in the following lemma.

Lemma 4.2.3. *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping such that* $\mathrm{Fix}(T) \neq \emptyset$ *and let $\{x_k\}_{k\in\mathbb{N}}$ be generated by Algorithm 5. Let $\{\theta_k\}_{k\in\mathbb{N}}$ be a sequence in $[0,\theta]$ with $\theta \in [0,1)$ such that $\sum_{k=1}^{+\infty} \theta_k\|x_k - x_{k-1}\| < +\infty$. Suppose Assumption 4.2.1 holds. Then $\{x_k\}_{k\in\mathbb{N}}$ is bounded.*

Proof. Let $k \in \mathbb{N}$ and a sequence $\{z_k\}_{k\in\mathbb{N}}$ be defined by

$$z_{k+1} = \delta_k z_k + \alpha_k(T\delta_k z_k - \delta_k z_k) + \varepsilon_k.$$

By nonexpansiveness of T, we have

$$\begin{aligned} \|x_{k+1} - z_{k+1}\| &= \|(1-\alpha_k)\delta_k(y_k - z_k) + \alpha_k(T\delta_k y_k - T\delta_k z_k)\| \\ &\leq (1-\alpha_k)\delta_k\|y_k - z_k\| + \alpha_k\delta_k\|y_k - z_k\| \\ &= \delta_k\|y_k - z_k\| \\ &= \delta_k\|x_k - z_k + \theta_k(x_k - x_{k-1})\| \\ &\leq \delta_k\|x_k - z_k\| + \delta_k\theta_k\|x_k - x_{k-1}\| \\ &\leq \delta_k\|x_k - z_k\| + \theta_k\|x_k - x_{k-1}\|. \qquad (4.2.6) \end{aligned}$$

By applying Lemma 2.4.1, we have $\lim\limits_{k\to+\infty} \|x_k - z_k\| = 0$.

Next, we expect that $\{z_k\}_{k\in\mathbb{N}}$ is bounded. Let $x^* \in \mathrm{Fix}(T)$. It follows that

$$\begin{aligned} \|z_{k+1} - x^*\| &\leq \|\delta_k z_k + \alpha_k(T\delta_k z_k - \delta_k z_k + \varepsilon_k - x^*)\| \\ &\leq (1-\alpha_k)\|\delta_k z_k - x^*\| + \alpha_k\|T\delta_k z_k - x^*\| + \|\varepsilon_k\| \\ &\leq \|\delta_k z_k - x^*\| + \|\varepsilon_k\| \\ &= \|\delta_k(z_k - x^*) + (\delta_k - 1)x^*\| + \|\varepsilon_k\| \\ &\leq \delta_k\|z_k - x^*\| + (1-\delta_k)\|x^*\| + \|\varepsilon_k\|. \qquad (4.2.7) \end{aligned}$$

Notice that $\sum_{k=1}^{+\infty} \varepsilon_k < +\infty$, we can apply Lemma 2.4.1 to obtain that $\{z_k\}_{k\in\mathbb{N}}$ is bounded. Seeing that $\lim\limits_{k\to+\infty} \|x_k - z_k\| = 0$ and $\{z_k\}_{k\in\mathbb{N}}$ is bounded, we get that $\{x_k\}_{k\in\mathbb{N}}$ is bounded. □

Theorem 4.2.4. *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping such that $\mathrm{Fix}(T) \neq \emptyset$ and let $\{x_k\}_{k\in\mathbb{N}}$ be generated by Algorithm 5. Let $\{\theta_k\}_{k\in\mathbb{N}}$ be a sequence in $[0, \theta]$ with $\theta \in [0, 1)$ such that $\sum_{k=1}^{+\infty} \theta_k\|x_k - x_{k-1}\| < +\infty$. Suppose Assumption 4.2.1 holds. Then, the sequence $\{x_k\}_{k\in\mathbb{N}}$ strongly converges to $x^* := \mathrm{proj}_{\mathrm{Fix}(T)}(0)$.*

Proof. From Lemma 4.2.3, we have $\{x_k\}_{k\in\mathbb{N}}$ is bounded. Moreover, $\{y_k\}_{k\in\mathbb{N}}$ is also bounded. Let $x^* := \mathrm{proj}_{\mathrm{Fix}(T)}(0)$. Then $x^* \in \mathrm{Fix}(T)$. By using Lemma 2.3.14 (iii), we get that

$$\|\delta_k y_k - x^*\|^2 = \|\delta_k(y_k - x^*) + (\delta_k - 1)x^*\|^2$$

$$= \delta_k^2\|y_k - x^*\|^2 + 2\delta_k(1-\delta_k)\langle -x^*, y_k - x^*\rangle + (1-\delta_k)^2\|x^*\|^2$$
$$\leq \delta_k\|x_k - x^* + \theta_k(x_k - x_{k-1})\|^2 + (1-\delta_k)(2\delta_k\langle -x^*, y_k - x^*\rangle$$
$$+ (1-\delta_k)\|x^*\|^2)$$
$$\leq \delta_k\|x_k - x^*\|^2 + 2\delta_k\langle \theta_k(x_k - x_{k-1}), y_k - x^*\rangle$$
$$+ (1-\delta_k)\left(2\delta_k\langle -x^*, y_k - x^*\rangle + (1-\delta_k)\|x^*\|^2\right). \tag{4.2.8}$$

By using Lemma 2.3.14 and the nonexpansiveness of T, we have

$$\|x_{k+1} - x^*\|^2 = \|\delta_k y_k + \alpha_k(T\delta_k y_k - \delta_k y_k) + \varepsilon_k - x^*\|^2$$
$$= \|(1-\alpha_k)(\delta_k y_k - x^*) + \alpha_k(T\delta_k y_k - x^*) + \varepsilon_k\|^2$$
$$\leq \|(1-\alpha_k)(\delta_k y_k - x^*) + \alpha_k(T\delta_k y_k - x^*)\|^2 + 2\langle \varepsilon_k, x_{k+1} - x^*\rangle$$
$$= (1-\alpha_k)\|\delta_k y_k - x^*\|^2 + \alpha_k\|T\delta_k y_k - x^*\|^2$$
$$- \alpha_k(1-\alpha_k)\|T\delta_k y_k - \delta_k y_k\|^2 + 2\langle \varepsilon_k, x_{k+1} - x^*\rangle$$
$$\leq \|\delta_k y_k - x^*\|^2 + 2\langle \varepsilon_k, x_{k+1} - x^*\rangle. \tag{4.2.9}$$

Combining (4.2.8) and (4.2.9), we obtain that

$$\|x_{k+1} - x^*\|^2 \leq \delta_k\|x_k - x^*\|^2 + (1-\delta_k)\left(2\delta_k\langle -x^*, y_k - x^*\rangle + (1-\delta_k)\|x^*\|^2\right)$$
$$+ 2\delta_k\langle \theta_k(x_k - x_{k-1}), y_k - x^*\rangle + 2\langle \varepsilon_k, x_{k+1} - x^*\rangle$$
$$\leq \delta_k\|x_k - x^*\|^2 + (1-\delta_k)\left(2\delta_k\langle -x^*, y_k - x^*\rangle + (1-\delta_k)\|x^*\|^2\right)$$
$$+ 2\delta_k\|y_k - x^*\|\left(\theta_k\|(x_k - x_{k-1})\|\right) + 2\|x_{k+1} - x^*\|(\|\varepsilon_k\|). \tag{4.2.10}$$

Next, we claim that $\|x_{k+1} - x_k\| \to 0$ as $n \to +\infty$. By the boundedness of a sequence $(y_k)_{k\geq 1}$ and the nonexpansiveness of T, we have

$$\|x_{k+1} - x_k\| = \|\delta_k y_k + \alpha_k(T\delta_k y_k - \delta_k y_k) + \varepsilon_k - (x_k)\|$$
$$\leq \|(1-\alpha_k)(\delta_k y_k - \delta_{k-1}y_{k-1}) + (\alpha_k - \alpha_{k-1})\delta_{k-1}y_{k-1}\|$$
$$+ \|\alpha_k(T\delta_k y_k - T\delta_{k-1}y_{k-1}) + (\alpha_k - \alpha_{k-1})T\delta_{k-1}y_{k-1}\|$$
$$+ \|\varepsilon_k - \varepsilon_{k-1}\|$$
$$\leq \|\delta_k y_k - \delta_{k-1}y_{k-1}\| + |\alpha_k - \alpha_{k-1}|(\|\delta_{k-1}y_{k-1}\| + \|T\delta_{k-1}y_{k-1}\|)$$
$$+ \|\varepsilon_k - \varepsilon_{k-1}\|$$
$$\leq \|\delta_k y_k - \delta_{k-1}y_{k-1}\| + |\alpha_k - \alpha_{k-1}|C_1 + \|\varepsilon_k - \varepsilon_{k-1}\|, \tag{4.2.11}$$

where $C_1 > 0$. After that we will consider the term $\|\delta_k y_k - \delta_{k-1}y_{k-1}\|$ in the inequality (4.2.11).

Let us consider,

$$\begin{aligned}\|\delta_k y_k - \delta_{k-1}y_{k-1}\| &= \|\delta_k(y_k - y_{k-1}) + (\delta_k - \delta_{k-1})y_{k-1}\| \\ &\le \delta_k\|y_k - y_{k-1}\| + |\delta_k - \delta_{k-1}|(\|y_{k-1}\|) \\ &\le \delta_k\|x_k - x_{k-1}\| + \delta_k\theta_k\|x_k - x_{k-1}\| + \delta_n\theta_{k-1}\|x_{k-1} - x_{n-2}\| \\ &\quad + |\delta_k - \delta_{k-1}|(\|y_{k-1}\|) \\ &\le \delta_k\|x_k - x_{k-1}\| + \theta_k\|x_k - x_{k-1}\| + \theta_{k-1}\|x_{k-1} - x_{n-2}\| \\ &\quad + |\delta_k - \delta_{k-1}|C_2, \qquad (4.2.12)\end{aligned}$$

where $C_2 > 0$. Combining (4.2.11) and (4.2.12), we get that

$$\begin{aligned}\|x_{k+1} - x_k\| \le \delta_k\|x_k - x_{k-1}\| + \theta_k\|x_k - x_{k-1}\| + \theta_{k-1}\|x_{k-1} - x_{n-2}\| \\ + |\alpha_k - \alpha_{n-1}|C_1 + |\delta_k - \delta_{k-1}|C_2 + \|\varepsilon_k - \varepsilon_{k-1}\|. \qquad (4.2.13)\end{aligned}$$

By applying Lemma 2.4.1 and the Assumption 4.2.1, we can conclude that

$\|x_{k+1} - x_k\| \to 0$ as $k \to +\infty$.

In the following, we prove that $\|T\delta_k y_k - \delta_k y_k\| \to 0$ as $k \to +\infty$. We observe that

$$\begin{aligned}\|T\delta_k y_k - \delta_k y_k\| &= \|T\delta_k y_k - x_{k+1} + x_{k+1} - \delta_k y_k\| \\ &\le \|T\delta_k y_k - x_{k+1}\| + \|x_{k+1} - \delta_k y_k\| \\ &= \|(1-\alpha_k)(T\delta_k y_k - \delta_k y_k) - \varepsilon_k\| \\ &\quad + \|(1-\delta_k)x_{k+1} + \delta_k x_{k+1} - \delta_k y_k\| \\ &\le (1-\alpha_k)\|T\delta_k y_k - \delta_k y_k\| + \|\varepsilon_k\| + (1-\delta_k)\|x_{k+1}\| \\ &\quad + \delta_k\|x_{k+1} - y_k\| \\ &= (1-\alpha_k)\|T\delta_k y_k - \delta_k y_k\| + \|\varepsilon_k\| + (1-\delta_k)\|x_{k+1}\| \\ &\quad + \delta_k\|x_{k+1} - x_k\| + \delta_k\theta_k\|x_k - x_{k-1}\|. \qquad (4.2.14)\end{aligned}$$

It follows that

$$\|T\delta_k y_k - \delta_k y_k\| \le \frac{1}{\alpha_k}(\|\varepsilon_k\| + (1-\delta_k)\|x_{k+1}\| + \delta_k\|x_{k+1} - x_k\|$$

$$+ \delta_k\theta_k\|x_k - x_{k-1}\|). \tag{4.2.15}$$

Since $\lim_{k\to+\infty} \|x_{k+1} - x_k\| = 0$ and the properties of the sequences involved, we can conclude that $\lim_{k\to+\infty} \|T\delta_k y_k - \delta_k y_k\| = 0$.

In order to show that the sequence $\{x_k\}_{k\in\mathbb{N}}$ strongly converges to x^*, it is sufficient to prove that

$$\limsup_{n\to+\infty}\langle -x^*, y_k - x^*\rangle \leq 0. \tag{4.2.16}$$

On the other hand, assume that the inequality (4.2.16) does not hold then there exists a real number $k > 0$ and a subsequence $(y_{k_i})_{i\in\mathbb{N}}$ such that

$$\langle -x^*, y_{k_i} - x^*\rangle \geq k > 0 \quad \forall i \geq 1.$$

For $\{y_k\}_{k\in\mathbb{N}}$ is bounded on a Hilbert space $\mathcal{H}$, we can find a subsequence of $\{y_k\}_{k\in\mathbb{N}}$ weakly converges to a point $y \in \mathcal{H}$. Without loss of generality, we can assume that $y_{k_i} \rightharpoonup y$ as $i \to +\infty$. Therefore,

$$0 < k \leq \lim_{i\to+\infty} \langle -x^*, y_{k_i} - x^*\rangle = \langle -x^*, y - x^*\rangle. \tag{4.2.17}$$

Notice that $\lim_{k\to+\infty} \delta_k = 1$, we get $\delta_{k_i} y_{k_i} \rightharpoonup y$ as $i \to +\infty$. Applying Lemma 2.4.4, we obtain that $y \in \operatorname{Fix}(T)$. With this, we have $\langle -x^*, y - x^*\rangle \leq 0$, which is a contradiction. Hence, the inequality (4.2.16) is verifyed. It follows that

$$\limsup_{k\to+\infty} \left(2\delta_k\langle -x^*, y_k - x^*\rangle + (1-\delta_k)\|x^*\|^2\right) \leq 0.$$

Using Lemma 2.4.1 and (4.2.10), we can conclude that $\lim_{k\to+\infty} x_k = x^*$. Based on what is described earlier, the proof is complete. □

Remark 4.2.5. The assumption of the seqence $\{\theta_k\}_{k\in\mathbb{N}}$ in Theorem 4.2.4 is verified, if we choose θ_k such that $0 \leq \theta_k \leq \bar{\theta}_k$, where

$$\bar{\theta}_k = \begin{cases} \min\left\{\theta, \frac{c_k}{\|x_k - x_{k-1}\|}\right\}, & \text{if } x_k \neq x_{k-1}, \\ \theta, & \text{otherwise,} \end{cases}$$

and $\sum_{k=1}^{+\infty} c_k < +\infty$.

4.2.1 Applications

This subsection is devoted to discussing the applications of the Algoruthm 5 in the monotone inclusion problems (4.2.1). We assume that $\operatorname{zer}(A+B+C) \neq \emptyset$. We propose the following algorithm for solving the problem (4.2.1).

Algorithm 6:

Initialization: Given $\{\theta_k\}_{k\in\mathbb{N}} \subseteq [0,\theta]$ with $\theta \in [0,1)$, $\mu \in (0,2\delta)$, two sequences $\{\alpha_k\}_{k\in\mathbb{N}}$ and $\{\beta_k\}_{k\in\mathbb{N}}$ in $(0,1]$ and a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ in $\mathcal{H}$. Choose $x_0, x_1 \in \mathcal{H}$ arbitrarily.

Iterative Steps: For a given current iterate $x_{k-1}, x_k \in \mathcal{H}$, calculate as follows:

Step 1. Compute

$$\begin{cases} a_k = x_k + \theta_k(x_k - x_{k-1}), \\ y_k = J_\mu^B(\delta_k a_k), \\ z_k = J_\mu^A(2y_k - \delta_k a_k - \mu C y_k). \end{cases}$$

Step 2. Compute

$$x_{k+1} = \delta_k a_k + \alpha_k(z_k - y_k) + \varepsilon_k.$$

Update k := k + 1 and return to Step 1.

The above algorithm can be rewritten as

$$\begin{aligned} x_{k+1} &= \delta_k a_k + \alpha_k[J_\mu^A \circ (2J_\mu^B - \mathrm{Id} - \mu C \circ J_\mu^B) + Id - J_{\mu_k}^B](\delta_k a_k) + \varepsilon_k \\ &= \delta_k a_k + \alpha_k(T\delta_k a_k - \delta_k a_k) + \varepsilon_k \end{aligned}$$

where $x_0, x_1 \in \mathcal{H}$, $a_k := x_k + \theta_k(x_k - x_{k-1})$ and

$$T := J_\mu^A \circ (2J_\mu^B - Id - \mu C \circ J_\mu^B) + Id - J_\mu^B. \tag{4.2.18}$$

The following proposition is the important tool for verifying the convergence of Algorithm 6 (see [86, Proposition 2.1])

Proposition 4.2.6. *Let $T_1, T_2 : \mathcal{H} \to \mathcal{H}$ be two firmly nonexpansive operators and C be a δ-cocoercive operator with $\delta > 0$. Let $\mu \in (0,2\delta)$. Then operator $T := Id - T_2 + T_1 \circ (2T_2 - Id - \mu C \circ T_2)$ is α-averaged with coefficient $\alpha := \frac{2\delta}{4\delta - \mu} < 1$.*

In particular, the following inequality holds for all $z, w \in \mathcal{H}$

$$\|Tz - Tw\|^2 \leq \|z-w\|^2 - \frac{(1-\alpha)}{\alpha}\|(Id-T)z - (Id-T)w\|^2.$$

The following lemma is a characterization of $\mathbf{zer}(A+B+C)$.

Lemma 4.2.7. [86, Lemma 2.2] *Let* $A : \mathcal{H} \to 2^{\mathcal{H}}$ *and* $B : \mathcal{H} \to 2^{\mathcal{H}}$ *be maximal monotone operators and* $C : \mathcal{H} \to \mathcal{H}$ *be a* δ*-cocoercive operator with* $\delta > 0$. *Suppose that* $\operatorname{zer}(A+B+C) \neq \emptyset$. *Then*

$$\operatorname{zer}(A+B+C) = J_\mu^B(\operatorname{Fix}(T)),$$

where $T := J_\mu^A \circ (2J_\mu^B - Id - \mu C \circ J_\mu^B) + Id - J_\mu^B$ *with* $\mu > 0$.

Remark 4.2.8.

(i) If we set $Cx = 0$ for all $x \in \mathcal{H}$ in Lemma 4.2.7, $\operatorname{zer}(A+B) = J_\mu^B(\operatorname{Fix}(T))$, where $T := J_\mu^A \circ (2J_\mu^B - Id) + Id - J_\mu^B$ with $\mu > 0$.

(ii) If we set $Bx = 0$ for all $x \in \mathcal{H}$ in Lemma 4.2.7, $\operatorname{zer}(A+C) = \operatorname{Fix}(T)$, where $T := J_\mu^A \circ (Id - \mu C)$ with $\mu > 0$.

Theorem 4.2.9. *Let* $A, B : \mathcal{H} \to 2^{\mathcal{H}}$ *be two maximal monotone operators and* $C : \mathcal{H} \to \mathcal{H}$ *be* δ*-cocoercive with* $\delta > 0$. *Suppose that* $\operatorname{zer}(A+B+C) \neq \emptyset$. *Let* $\{\theta_k\}_{k\in\mathbb{N}}$ *be a sequence in* $[0, \theta]$ *with* $\theta \in [0,1)$ *and* $\mu \in (0, 2\delta)$. *Let* $\{x_k\}_{k\in\mathbb{N}}$, $\{y_k\}_{k\in\mathbb{N}}$ *and* $\{z_k\}_{k\in\mathbb{N}}$ *be generated by Algorithm 6. Assume that the Assumption 4.2.1 holds and* $\sum_{k=1}^{+\infty}\theta_k\|x_k - x_{k-1}\| < +\infty$. *Then the following statements are true:*

(a) $\{x_k\}_{k\in\mathbb{N}}$ *strongly converges to* $x^* := \operatorname{proj}_{\operatorname{Fix}(T)}(0)$, *where* $T := J_\mu^A \circ (2J_\mu^B - Id - \mu C \circ J_\mu^B) + Id - J_\mu^B$ *for some* $\mu > 0$.

(b) $\{y_k\}_{k\in\mathbb{N}}$ *and* $\{z_k\}_{k\in\mathbb{N}}$ *strongly converge to* $J_\mu^B(x^*) \in \operatorname{zer}(A+B+C)$.

Proof. (a): Let $\{x_k\}_{k\in\mathbb{N}}$ be generated by Algorithm 6. Then the iterative method can be rewritten as

$$x_{k+1} = \delta_k a_k + \alpha_k(T\delta_k a_k - \delta_k a_k)$$

where $x_0, x_1 \in \mathcal{H}$, $a_k := x_k + \theta_k(x_k - x_{k-1})$ and

$$T := J_\mu^A \circ (2J_\mu^B - Id - \mu C \circ J_\mu^B) + Id - J_\mu^B.$$

By applying Proposition 4.2.6, we get T is nonexpansive.

On the other hand, by Lemma 4.2.7, we obtain that

$$J_\mu^B(\text{Fix}(T)) = \text{zer}(A + B + C) \neq \emptyset.$$

It means that $\text{Fix}(T) \neq \emptyset$. By applying Theorem 4.2.4, we have the sequence $\{x_k\}_{k\in\mathbb{N}}$ strongly converges to $x^* := \text{proj}_{\text{Fix}(T)}(0)$ as $k \to +\infty$.
(b): The sequences $\{a_k\}_{k\in\mathbb{N}}$ as Algorithm 6, we obtain that $a_k \to x^*$ as $k \to +\infty$. Since J_μ^B is continuous, we have $y_k \to J_\mu^B(x^*) \in \text{zer}(A + B + C)$. From the last line of Algorithm 6, we get that $\lim_{k\to+\infty} \|z_k - y_k\| = 0$. This proof is complete. □

Using similar arguments as in Theorem 4.2.9 and set $Cx = 0$ for all $x \in \mathcal{H}$, we can prove the following results.

Corollary 4.2.10. *Let $A, B : \mathcal{H} \to 2^{\mathcal{H}}$ be two maximal monotone operators and $\text{zer}(A + B)$ be a nonempty set. We consider the following algorithm:*

$$(\forall k \geq 1) \begin{cases} a_k = x_k + \theta_k(x_k - x_{k-1}), \\ y_k = J_\mu^B(\delta_k a_k), \\ z_k = J_\mu^A(2y_k - \delta_k a_k), \\ x_{k+1} = \delta_k a_k + \alpha_k(z_k - y_k) + \varepsilon_k, \end{cases}$$

where $x_0, x_1 \in \mathcal{H}$, $\mu \in (0, 2\delta)$, $\{\theta_k\}_{k\in\mathbb{N}} \subseteq [0, \theta]$ with $\theta \in [0, 1)$, and $\{\alpha_k\}_{k\in\mathbb{N}}$ and $\{\delta_k\}_{k\in\mathbb{N}}$ are sequences in $(0, 1]$ and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ is a sequence in $\mathcal{H}$. Assume that the Assumption 4.2.1 holds and $\sum_{k=1}^{+\infty} \theta_k \|x_k - x_{k-1}\| < +\infty$. Then the following statements hold:

(a) *$\{x_k\}_{k\in\mathbb{N}}$ strongly converges to $x^* := \text{proj}_{\text{Fix}(J_\mu^A \circ (2J_\mu^B - Id) + Id - J_\mu^B)}(0)$ for some $\mu > 0$.*

(b) *$\{y_k\}_{k\in\mathbb{N}}$ and $\{z_k\}_{k\in\mathbb{N}}$ strongly converge to $J_\mu^B(x^*) \in \text{zer}(A + B)$.*

Proof. It follows from the proof of Theorem 4.2.9. ☐

Using similar arguments as in Theorem 4.2.9 and set $Bx = 0$ for all $x \in \mathcal{H}$, we can prove the following results.

Corollary 4.2.11. *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximal monotone operator and $C : \mathcal{H} \to \mathcal{H}$ a δ-cocoercive operator with $\delta > 0$ and $\operatorname{zer}(A+C) \neq \emptyset$. Let $\mu \in (0, 2\delta)$ and $\{x_k\}_{k\in\mathbb{N}}$ be generated by the following iterative scheme*

$$\begin{cases} x_0, x_1 \in \mathcal{H}, \\ y_k = x_k + \theta_k(x_k - x_{k-1}), \\ x_{k+1} = (1-\alpha_k)\delta_k y_k + \alpha_k J_\mu^A(\delta_k y_k - \mu C \delta_k y_k) + \varepsilon_k, \end{cases} \tag{4.2.19}$$

for all $k \geq 1$, where $\{\theta_k\}_{k\in\mathbb{N}} \subseteq [0, \theta]$ with $\theta \in [0, 1)$, and $\{\alpha_k\}_{k\in\mathbb{N}}$ and $\{\delta_k\}_{k\in\mathbb{N}}$ are sequences in $(0, 1]$ and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ is a sequence in $\mathcal{H}$. Assume that the Assumption 4.2.1 holds and $\sum_{k=1}^{+\infty} \theta_k \|x_k - x_{k-1}\| < +\infty$.

Then, the sequence $\{x_k\}_{k\in\mathbb{N}}$ strongly converges to a point $\operatorname{proj}_{\mathbf{zer}(A+C)}(0)$.

4.2.2 Numerical experiments

To illustrate the behavior of the proposed iterative method, we provide a numerical example in a convex minimization problem and compare the convergence performance of the proposed algorithm with some algorithms in the literature. Moreover, we also employ our algorithm in the context of image restoration problems. All the experiments are implemented in MATLAB R2016b running on a MacBook Air 13-inch, Early 2017 with a 1.8 GHz Intel Core i5 processor and 8 GB 1600 MHz DDR3 memory.

Convex Minimization Problems

In this subsection, we present some comparisons among Algorithm 6, **MTA**, and Shehu et al. algorithm (4.2.5) ([41, Algorothm 3.1]) in convex minimization problem.

Example 4.2.12. Let $f : \mathbb{R}^s \to \mathbb{R}$ be defied by $f(x) = \|x\|_1$ for all $x \in \mathbb{R}^s$, $g : \mathbb{R}^s \to \mathbb{R}$ be defined by indicator function $g(x) = \delta_W(x)$ with $W := \{x : \mathbf{A}x = \mathrm{b}\}$ for all $x \in \mathbb{R}^s$, where $\mathbf{A} : \mathbb{R}^s \to \mathbb{R}^l$ is a non-zero linear transformation, $\mathrm{b} \in \mathbb{R}^l$ and $s > l$ and $h : \mathbb{R}^s \to \mathbb{R}$ be defied by $h(x) = \frac{1}{2}\|x\|_2^2$ for all $x \in \mathbb{R}^s$. Since $s > l$, we get that $\mathbf{A}$ is an asymmetric transformation. Finding the solution of the following problem:

$$\begin{aligned} &\text{minimize } \|x\|_1 + \delta_W(x) + \frac{1}{2}\|x\|_2^2 \\ &\text{subject to } x \in \mathbb{R}^s. \end{aligned} \tag{4.2.20}$$

The problem (4.2.20) can be written in the form of the problem (4.2.1) as:

$$\text{find } x \in \mathbb{R}^s \text{ such that } 0 \in \partial\|x\|_1 + \partial\delta_W(x) + \nabla h(x), \tag{4.2.21}$$

where $A = \partial\|\cdot\|_1$, $B = \partial\delta_W(\cdot)$ and $C = \nabla h(\cdot)$.

In this setting, we have $J_\mu^{\partial\delta_W}(x) = x + \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathrm{b} - \mathbf{A}x)$,

$$J_\mu^{\partial\|\cdot\|_1}(x) = (\max\{0, 1 - \frac{\mu}{|x^1|}\}x_1, \max\{0, 1 - \frac{\mu}{|x^2|}\}x_2, ..., \max\{0, 1 - \frac{\mu}{|x^s|}\}x_s),$$

and $\nabla h(x) = x$, where $x = (x^1, x^2, ..., x^s) \in \mathbb{R}^s$.

We begin with the problem by random vectors $z, x_0, x_1 \in \mathbb{R}^s$ and $\mathrm{b} \in \mathbb{R}^l$ and matrix $\mathbf{A} \in \mathbb{R}^{l\times s}$. Next, we compare the Algorithm 6 performance with two remained performance. The parameters that are used in our algorithm are chosen as follows: $\alpha_k = 1 - \frac{1}{(k+2)^2}$, $\delta_k = 1 - \frac{1}{k+2}$, $\varepsilon_k = \frac{z}{(100k)^2}$, and

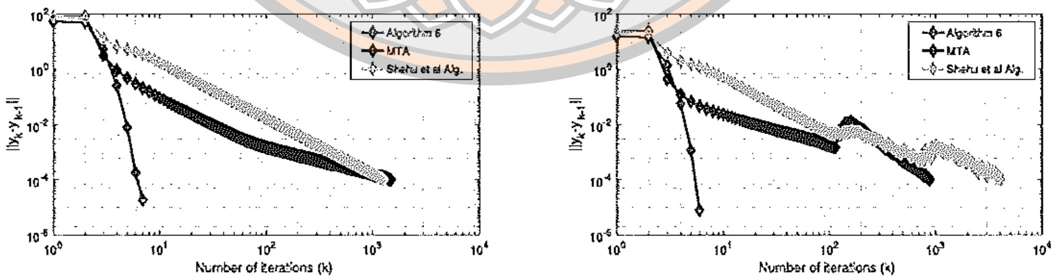
$$\theta_k = \begin{cases} \min\left\{\frac{1}{2}, \frac{1}{(k+1)^2\|x_k - x_{k-1}\|}\right\}, & \text{if } x_k \neq x_{k-1}, \\ \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{4.2.22}$$

We choose $\alpha_k = \frac{1}{k+1}$, $\delta_k = \gamma_k = \frac{1}{2(k+1)}$ and $e_k = \varepsilon_k$ for the algorithm of Shehu et al. (4.2.5) in [41]. We obtain the CPU times (seconds) and the number of iterations by using the stopping criteria : $\|y_k - y_{k-1}\| \leq 10^{-4}$.

Table 6: Comparison: Algorithm 6, **MTA** and Shehu et al. Alg. (4.2.5)

(l,s)	Algorithm 6		**MTA**		**Shehu et al. Alg.** (4.2.5)	
	CPU Time (s)	Iterations	CPU Time (s)	Iterations	CPU Time (s)	Iterations
(20,700)	0.0218	7	0.0428	278	0.0756	626
(20,800)	0.0189	7	0.0914	350	0.1745	796
(20,7000)	0.0302	7	1.7751	1273	0.0977	53
(20,8000)	0.0308	6	1.2419	1290	0.0671	54
(200,7000)	0.0365	8	1.9452	858	4.6538	2028
(200,8000)	0.0406	7	2.5115	977	0.1425	53
(500,7000)	0.0403	7	4.1647	892	8.3620	1956
(500,8000)	0.0548	8	4.3239	813	9.0929	1835
(1000,7000)	0.0703	7	6.7954	786	14.1693	1751
(1000,8000)	0.0728	7	7.8302	825	16.3752	1784
(3000,7000)	0.1597	7	18.0559	779	44.8129	1940
(3000,8000)	0.1763	7	22.3514	841	49.6872	1891
(100,80000)	0.1376	8	26.6863	1489	1.5926	94
(1000,80000)	0.6949	8	344.7048	3289	9.4181	93

In table 6 we present a comparison among the numerical results of Algorithm 6, **MTA**, and Shehu et al. Algorithm (4.2.5) in different sizes of matrix **A**. The smallest number of iterations is generated by Algorithm 6 for all different sizes of matrix **A**. Moreover, Algorithm 6 requires the least CPU computation time to reach the optimality tolerance for all cases.



(a) Case: $(l,s) = (100, 80000)$ (b) Case: $(l,s) = (500, 7000)$

Figure 5: Illustration the behavior of $\|y_k - y_{k-1}\|$ for Algorithm 6, **MTA**, and Shehu et al. Alg. (4.2.5)

Figure 5 shows the behavior of $\|y_k - y_{k-1}\|$ for Algorithm 6, **MTA**, and Shehu et al. Algorithm (4.2.5) in two different choices of (l,s). We can observe that by using our algorithm the behavior of the red line Algorithm 6 is the best performance.

Image Restoration Problems

In this subsection, we apply the proposed algorithm, image restoration problems, which involves deblurring and denoising images. We recall the following problem as:

$$\text{find } x \in \arg\min_{x \in \mathbb{R}^s} \left\{ \frac{1}{2}\|Hx - y\|_2^2 + \mu\|x\|_1 \right\}, \tag{4.2.23}$$

where y is the degraded image and H is a bounded linear operator. Note that problem (4.2.23) is a spacial case of problem (4.2.1) by setting $A = \partial f(\cdot)$, $B = 0$, and $C = \nabla L(\cdot)$ where $f(x) = \|x\|_1$ and $L(x) = \frac{1}{2}\|Hx - y\|_2^2$. This setting we have that $C(x) = \nabla L(x) = H^*(Hx - y)$, where H^* is a transpose of H. We begin the problem by choosing images and degrade them by random noise and different types of blurring. The random noise in this study is provided by Gaussian white noise of zero mean and 0.001 variance. We solve the problem (4.2.23) by using our algorithm in Corollary 4.2.11. We set $\alpha_k = 1 - \frac{1}{(k+1)^2}$, $\delta_k = 1 - \frac{1}{100k+1}$, $\mu = 0.001$, $\varepsilon_k = 0$ and θ_k is defined as (4.2.22).

We compare our proposed algorithm with the inertial Mann-type algorithm that was introduced by Kitkuan et al. [46]. In Kitkuan et al. Algorithm ([46, Algorithm in Theorem 3.1]), we choose $\varsigma_k = \theta_k$, $\alpha_k = \frac{1}{k+1}$, $\lambda_k = 0.001$ and $h(x) = \frac{1}{12}\|x\|_2^2$. We assess the quality of the reconstructed image by using the signal to noise ratio (**SNR**) for monochrome images which is defined by

$$\mathbf{SNR}(k) = 20 \log_{10} \frac{\|x\|_2^2}{\|x - x_k\|_2^2},$$

where x and x_k denote the original and the restored image at iteration k, respectively.

For colour images, we estimate the quality of the reconstructed image by using the normalized colour difference (**NCD**) [90] which is defined by

$$\mathbf{NCD}(k) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\sqrt{(L_{i,j}^o - L_{i,j}(k))^2 + (u_{i,j}^o - u_{i,j}(k))^2 + (v_{i,j}^o - v_{i,j}(k))^2}}{\sum_{i=1}^{N}\sum_{j=1}^{M}\sqrt{(L_{i,j}^o)^2 + (u_{i,j}^o)^2 + (v_{i,j}^o)^2}},$$

where i, j are indices of the sample position, N, M characterize an image size and $L^o_{i,j}$, $u^o_{i,j}$, $v^o_{i,j}$ and $L_{i,j}(k)$, $u_{i,j}(k)$, $v_{i,j}(k)$ are values of the perceived lightness and two representatives of chrominance related to the original and the restored image at iteration k, respectively. We generated the noised model in order to obviously see the differences between degraded and original figure as follows. Figure 6 firstly shows the original image. Secondly, the degrated image was corrupted by average blur (size 20 by 20) and Gaussian noise (zero mean and 0.001 variance). We randomly selected parameters which visibly showed the differences sharpness level and. Lastly, reconstructed images are shown. Figure 7 firstly shows the original image. Secondly, the degrated image was corrupted by Gaussian blur (size 20 by 20 with the standard deviation 20) and Gaussian noise (zero mean and 0.001 variance). With this point, we found that any adjustment of the standard deviation as much as small might not shown the difference between degraded and original figure. Lastly, reconstructed images are shown. Figure 8 firstly shows the original image. Secondly, the degrated image was corrupted by motion blur (the linear motion of a camera by 30 pixels with an angle of 60 degrees) and Gaussian noise (zero mean and 0.001 variance). We randomly selected parameters which visibly showed the differences sharpness level. Lastly, reconstructed images are shown. The comparisons between our proposed algorithm (4.2.19) and Kitkuan et al. Algorithm ([46, Algorithm in Theorem 3.1]) in image restoration problems are presented in Figure 4 and Table 7. Furthermore, we also present the comparison Kitkuan et al. Algorithm ([46, Algorithm in Theorem 3.1]), our algorithm, and the well-known technique for image restoration which is Weiner filtering (**WF**) [91, 92]. In Figure 5 present the comparative results of two degradation images 'Artsawang' and 'Mandril' corrupted by motion blur and different salt & pepper noise from 0% to 10%.

(a) camera man

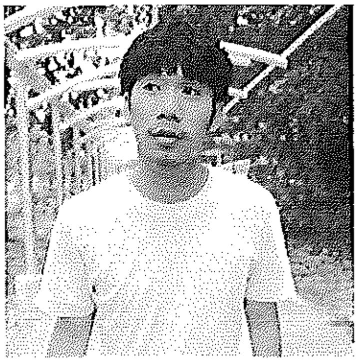
(b) average blur & random noise

(c) Weiner Filtering

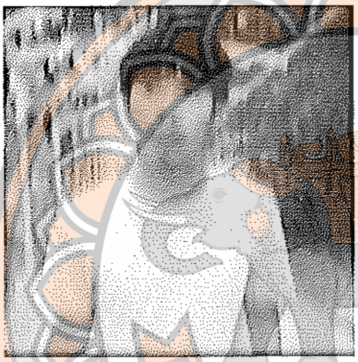
(d) Kitkuan et al. Alg.

(e) our algorithm

Figure 6: The degraded and reconstructed 'camera man' images with different techniques



(a) Artsawang



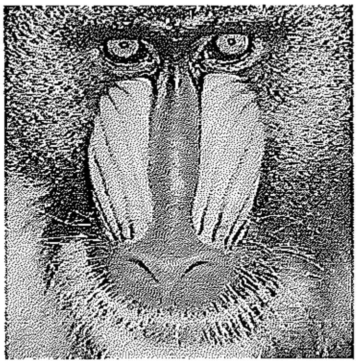
(b) Gaussian blur & random noise

(c) Weiner Filtering

(d) Kitkuan et al. Alg.

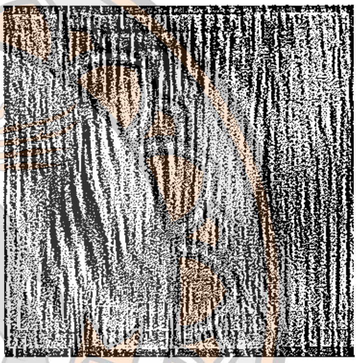
(e) our algorithm

Figure 7: The degraded and reconstructed 'Atrsawang' images with different techniques

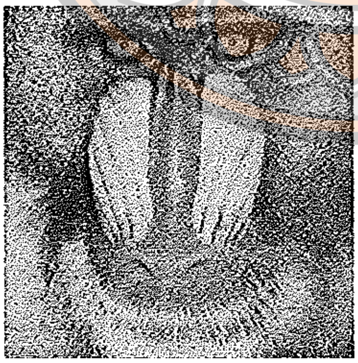


(a) Mandril

(b) motion blur & random noise



(c) Weiner Filtering



(d) Kitkuan et al. Alg.

(e) our algorithm

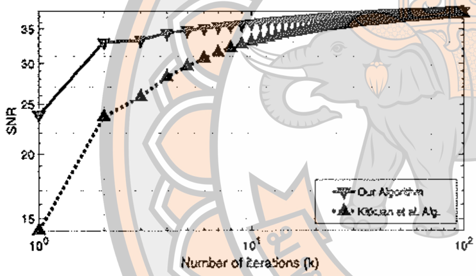
Figure 8: The degraded and reconstructed 'Mandril' images with different techniques

Figure 6: **(a)** shows the original image 'camera man', figure **(b)** shows the

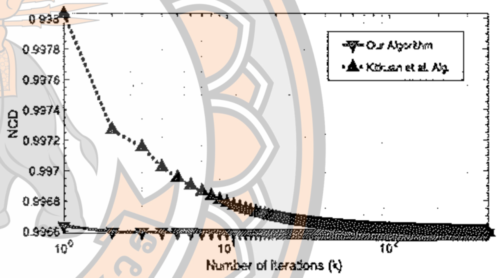
images degraded by average blur and random noise (Gaussian noise) and figure **(c)**, **(d)**, **(e)** show the reconstructed image by using Weiner filter, Kitkuan et al. algorithm, and our algorithm (4.2.19)., respectively.

Figure 7: **(a)** shows the original image 'Artsawang', figure **(b)** shows the images degraded by Gaussian blur and random noise (Gaussian noise) and figure **(c)**, **(d)**, **(e)** show the reconstructed image by using Weiner filter, Kitkuan et al. algorithm, and our algorithm (4.2.19)., respectively.

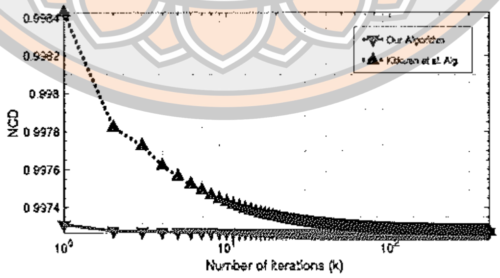
Figure 8: **(a)** shows the original image 'Mandril', figure **(b)** shows the images degraded by motion blur and random noise (Gaussian noise) and figure **(c)**, **(d)**, **(e)** show the reconstructed image by using Weiner filter, Kitkuan et al. algorithm, and our algorithm (4.2.19)., respectively.



(a) camera man



(b) Artsawang



(c) Mandril

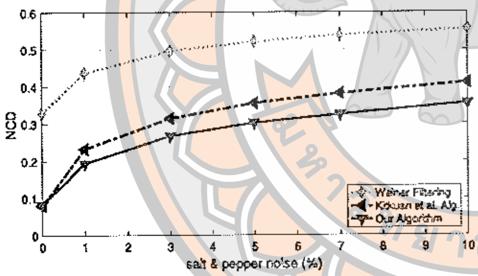
Figure 9: Illustration of the behavior of **SNR** and **NCD** for our algorithm and Kitkuan et al.'s algorithm in Figure 6, 7, and 8.

Figure 9: **(a)** shows the behavior of **SNR** for two algorithms in figure **(d)**, **(e)** of figure 6, figure **(b)** shows the behavior of **NCD** for two algorithms in figure **(d)**, **(e)** of figure 7 and figure **(c)** shows the behavior of **NCD** for two algorithms

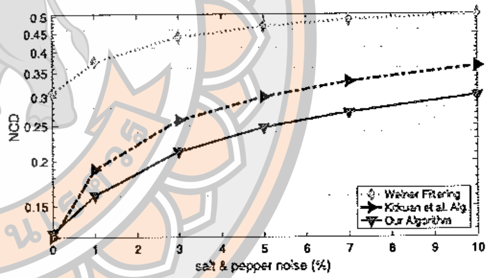
in figure (d), (e) of figure 8.

Table 7: The performance of the normalized colour difference (**NCD**) in two images.

The normalized colour difference (**NCD**)				
	Kitkuan et al. Alg.		Our algorithm (4.2.19)	
k	Artsawang image	Mandril image	Artsawang image	Mandril image
1	0.99803	0.99842	0.99663	0.99731
50	0.99660	0.99730	0.99659	0.99727
100	0.99661	0.99729	0.99658	0.99726
200	0.99660	0.99728	0.99658	0.99726
300	0.99659	0.99727	0.99658	0.99726
400	0.99659	0.99727	0.99658	0.99726



(a) Artsawang



(b) Mandril

Figure 10: Illustration of the behavior of **NCD** in motion blur and different different salt & pepper noise from 0% to 10%.

CHAPTER V

CONCLUSION

In this thesis, we presented a number of contributions in the context of solving constrained convex optinmization problems and monotone inclusion problems in Hilbert spaces by means of optimization iterative algorithms in two parts.

The first part of the thesis addresses the constrained convex optinmization problem, which is to minimize a smooth convex objective function subject to the set of minima of another differentiable convex function (3.0.2). In order to investigate the convergence properties of this problem, we proposed iterative algorithm combines the gradient method with a penalization technique, which is called rapid gadient penalty algorithm (in short **RGPA**). Moreover, the iterative mathod of our algorithm are developed by using the new computation of gradient mehod. The convergence results will not go on whenever the key assumptions, Assumption 3.1.1, has not been verified. We also presented a numerical example to illustrate the convergence behavior of the iterate and compared the performance of the algorithm (**RGPA**), the algorithm introduced by Peypouquet (**DGS**) [5] and the algorithm introduced by Bot et al. (**GPIM**) [33]. It has been showed that our algorithm (**RGPA**) performs better behavior when comparing with other algorithms.

Subsequently, we investigated the constrained convex optinmization problem, which is to minimize a nonsmooth convex objective function subject to the set of minima of another differentiable convex function (3.0.2). In a similar fashion with smooth convex objective conterpart, we proposed the so-called new forward–backward penalty method Algorithm, which combines the proximal method with a penalization technique. Under some appropriate assumptions of parameters, the main convergence result for the sequence generated by this method was presented in Theorem 3.2.7. We also discussed a numerical example to illustrate the convergence behavior of the iterate.

In the second part of the thesis we approached the solving of the monotone inclusion problem (4.1.1). This problem involved with the sum of two maximal monotone operators. The monotone inclusion problem can be considered to be a generalization of many existing mathematical problems such as fixed point problems [41] and constrained convex minimization problems [27, 28, 39–44, 46]. We proposed the so-called inertial viscosity forward-backward splitting algorithm (IVFBSA) for solving the problem (4.1.1). By using some suitable control conditions, the strong convergence was obtained in Theorem 4.1.2. For the virtue of the main theorem, it can be applied to find a solution of the convex minimization problems. As an illustration of the behavior of the proposed algorithm, we compared the convergent behavior of our method and the algorithm introduced by Kitkuan et al. [46].

Finally, we investigated the iterative method combining both inertial terms and errors to find a fixed point of a nonexpansive mapping the strong convergence of the iterate under some appropriate assumptions was presented in Theorem 4.2.4. For the virtue of the Theorem 4.2.4, it can be applied to an approximately zero point of the sum of three monotone operators, that was presented in Theorem 4.2.9. We also illustrated the functionality of the algorithm through numerical experiments addressing image restoration problems.

REFERENCES

REFERENCES

1. Boyd S, Vandenberghe L. Convex optimization. UK: Cambridge university press; 2004.
2. Attouch H, Czarnecki M-O. Asymptotic behavior of coupled dynamical systems with multiscale aspects. J. Differ. Equat. 2010;248:1315–1344.
3. Attouch H, Czarnecki M-O, Peypouquet J. Prox-penalization and splitting methods for constrained variational problems. SIAM J. Optim. 2011;21:149–173.
4. Attouch H, Czarnecki M-O, Peypouquet J. Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities. SIAM J. Optim. 2011;21:1251–1274.
5. Peypouquet J. Coupling the gradient method with a general exterior penalization scheme for convex minimization. J. Optim. Theory Appl. 2012;153:123–138.
6. Noun N, Peypouquet J. Forward-backward penalty scheme for constrained convex minimization without inf-compactness. J. Optim. Theory Appl. 2013;158:787–795.
7. Attouch H, Cabot A, Czarnecki M-O. Asymptotic behavior of nonautonomous monotone and subgradient evolution equations. Trans. Am. Math. Soc. 2016. https://doi.org/10.1090/tran/6965,arXiv:1601.00767(2016) .
8. Attouch H, Czarnecki M-O. Asymptotic behavior of gradient-like dynamical systems involving inertia and multiscale aspects. J. Differ. Equ. 2017;262:2745–2770.
9. Bot RI, Csetnek ER. Forward-backward and Tseng's type penalty schemes for monotone inclusion problems. Set-Valued Var. Anal. 2014;22:313–331.
10. Bot RI, Csetnek ER. Approaching the solving of constrained variational inequalities via penalty term-based dynamical systems. J. Math. Anal. Appl. 2016;435:1688–1700.
11. Bot RI, Csetnek ER. Levenberg-Marquardt dynamics associated to variational inequalities. Set-Valued Var. Anal. 2017.

12. Bot RI, Csetnek ER. A Tsengs type penalty scheme for solving inclusion problems involving linearly composed and parallel-sum type monotone operators. Vietnam J. Math. 2014;42:451–465.

13. Bot RI, Csetnek ER. Penalty schemes with inertial effects for monotone inclusion problems. Optimization. 2017;66:965–982.

14. Bot RI, Csetnek ER. Second-order dynamical systems associated to variational inequalities. Appl. Anal. 2017;96:799–809.

15. Bot RI, Csetnek ER. A second order dynamical system with Hessian-driven damping and penalty term associated to variational inequalities. Optimization. 2019;68(7):1265–1277.

16. Czarnecki M-O, Noun N, Peypouquet J. Splitting forward-backward penalty scheme for constrained variational problems. J. Convex Anal. 2016;23:531–565.

17. Polyak BT. Introduction to Optimization. Optimization Software New York. 1987.

18. Bertsekas DP. Nonlinear Programming, 2nd ed. Cambridge: Athena Scientific; 1999.

19. Alvarez F. On the minimizing property of a second order dissipative system in Hilbert spaces. SIAM J. Control. Optim. 2000;38:1102–1119.

20. Alvarez F. Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space. SIAM J. Optim. 2004;14:773–782.

21. Alvarez F, Attouch H. An inertial proximal method for monotone operators via discretization of a nonlinear oscillator with damping. Set Valued Anal. 2001;9:3-11.

22. Attouch H, Peypouquet J, Redont P. A dynamical approach to an inertial forward-backward algorithm for convex minimization. SIAM J. Optim. 2014;24:232–256.

23. Bot RI, Csetnek ER. A hybrid proximal-extragradient algorithm with inertial effects. Numer. Funct. Anal. Optim. 2015;36:951–963.

24. Bot RI, Csetnek ER. An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems. Numer. Algorithms. 2016;71:519–540.

25. Bot RI, Csetnek ER. An inertial alternating direction method of multipliers. Minimax Theory Appl. 2016;1:29–49.

26. Bot RI, Csetnek ER. An inertial Tsengs type proximal algorithm for nonsmooth and nonconvex optimization problems. J. Optim. Theory Appl. 2016;171:600–616.

27. Bot RI, Csetnek ER, Hendrich C. Inertial DouglasRachford splitting for monotone inclusion problems. Appl. Math. Comput. 2015;256:472–487.

28. Bot RI, Csetnek ER, László S. An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions. EURO J. Comput. Optim. 2016;4:3–25.

29. Attouch H, Peypouquet J, Redont P. A dynamical approach to an inertial forward-backward algorithm for convex minimization. SIAM Journal on Optimization. 2014;24(1):232–256.

30. Maingé PE. Convergence theorems for inertial KM-type algorithms. J. Comput. Appl. Math. 2008;219:223–236.

31. Maingé PE, Moudafi A. Convergence of new inertial proximal methods for DC programming. SIAM Journal on Optimization. 2008;19(1):397–413.

32. Moudafi A, Oliny M. Convergence of a splitting inertial proximal method for monotone operators. J. Comput. Appl. Math. 2003;155:447-454.

33. Bot RI, Csetnek ER, Nimana N. Gradient-type penalty method with inertial effects for solving constrained convex optimization problems with smooth data. Optim. Lett. 2017. https://doi.org/10.1007/s11590-017-1158-1;(2017)

34. Bot RI, Csetnek ER, Nimana N. An inertial proximal-gradient penalization scheme for constrained convex optimization problems. Vietnam J. Math. 2017.

35. Lions P-L, Mercier B. Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. 1979;16:964–979.

36. Passty GB. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J. Math. Anal. Appl. 1979;72:383–390.

37. Tseng P. A modified forward–backward splitting method for maximal monotone mappings. SIAM J. Control. Optim. 2000;38:431–446.

38. Polyak BT. Some methods of speeding up the convergence of iterative methods. Zh. Vychisl. Mat. Mat. Fiz. 1964;4:1–17.

39. Cholamjiak W, Cholamjiak P, Suantai S. An inertial forward–backward splitting method for solving inclusion problems in Hilbert spaces. J. Fixed Point Theory Appl. 2018;20(42). doi:10.1007/s11784-018-0526-5

40. Artsawang N, Ungchittrakool K. Inertial Mann-type algorithm for a nonexpansive mapping to solve monotone inclusion and image restoration problems. Symmetry. 2020;12(5):750.

41. Shehu Y, Iyiola OS, Ogbuisi FU. Iterative method with inertial terms for nonexpansive mappings: applications to compressed sensing. Numer. Algor. 2019.

42. Cholamjiak P, Shehu Y. Inertial forward–backward splitting method in Banach spaces with application to compressed sensing. Appl. Math. 2019;64:409–435.

43. Shehu Y, Cai G. Strong convergence result of forward-backward splitting methods for accretive operators in banach spaces with applications. RACSAM. 2016.

44. Cholamjiak P. A generalized forward-backward splitting method for solving quasi inclusion problems in Banach spaces. Numer. Alg. 2016;71(4):915–932.

45. Lorenz DA, Pock T. An inertial forward-backward algorithm for monotone inclusions. Journal of Mathematical Imaging and Vision. 2015;51(2):311–325.

46. Kitkuan D, Kumam P. Martínez-Moreno J, Sitthithakerngkiet K. Inertial viscosity forward-backward splitting algorithm for monotone inclusions and its application to image restoration problems. Int. J. Comput. Math. 2019;1–19.

47. Kitkuan D, Kumam P, Martínez-Moreno J. Generalized Halpern-type forward–backward splitting methods for convex minimization problems with application to image restoration problems. Optimization. 2019. 10.1080/02331934.2019.1646742.

48. Mann WR. Mean value methods in iteration. Proc. Amer. Math. Soci. 1953;4:506–510.

49. Takahashi W. Nonlinear Functional Analysis Fixed Point Theory and its Applications. Yokohama: Yokohama Publishers; 2000.

50. Xu H-K. Iterative algorithms for nonlinear operators. J. London Math. Soc. 2002;66(1):240–256.

51. Chidume CE, Chidume CO. Iterative approximation of fixed points of nonexpansive mappings. J. Math. Anal. Appl. 2006;318:288–295.

52. Lions P-L. Approximation de points fixes de contractions. C. R. Acad. Sci. Paris Sér. 1977;284:A1357–A1359.

53. Reich S. Strong convergence theorems for resolvents of accretive operators in Banach spaces. J. Math. Anal. Appl. 1980;75:287–292.

54. Reich S. Some problems and results in fixed point theory. Contemp. Math. 1983;21:179–187.

55. Shioji N, Takahashi W. Strong convergence of approximated sequences for nonexpansive mapping in Banach spaces. Proc. Amer. Math. Soc. 1997;125:3641–3645.

56. Wittmann R. Approximation of fixed points of nonexpansive mappings. Arch. Math. 1992;58:486–491.

57. Moudafi A. Viscosity approximation methods for fixed points problems. J. Math. Anal. Appl. 2000;241:46–55.

58. Dong QL, Lu YY. A new hybrid algorithm for a nonexpansive mapping. Fixed Point Theroy Appl. 2015;2015:37.

59. Dong QL, Yuan HB. Accelerated Mann and CQ algorithms for finding a fixed point of a nonexpansive mapping. Fixed Point Theroy Appl. 2015;2015:125.

60. Kanzow C, Shehu Y. Generalized Krasnoselskii-Mann-type iterations for nonexpansive mappings in Hilbert spaces. Comput. Optim. Appl. 201767(3):595–620.

61. Kim TH, Xu HK. Strong convergence of modified Mann iterations. Non-linear Anal. 2005;61:51–60.

62. Bot RI, Csetnek ER, Meier D. Inducing strong convergence into the asymptotic behaviour of proximal splitting algorithms in Hilbert spaces. Optim. Methods Software. 2019;34(3):489–514.

63. Combettes PL, Yamada I. Compositions and convex combinations of averaged nonexpansive operators. J. Math. Anal. Appl. 2015;425(1):55–70.

64. Kreyszig E. Introductory functional analysis with applications. Singapore: John Wiley & Sons; 1978.

65. Takahashi W. Introduction to Nonlinear Analysis. Yokohama: Yokohama Publishers; 2009.

66. Bauschke HH, Combettes PL. Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics. New York: Springer; 2011.

67. Beck A. Introduction to Nonlinear Optimization Theory, Algorithms, and Applications with MATLAB. Philadelphia: SIAM; 2014.

68. Dhara A, Joydeep D. Optimality conditions in convex optimization. Taylor & Francis Group; 2012.

69. Van TJ. Convex Analysis, An Introductory Text. Chichester: Wiley; 1984.

70. Hernández-Lerma O, Lasserre JB. Markov chains and invariant probabilities (Vol. 211). Birkhäuser; 2012.

71. Zălinescu C. Convex Analysis in General Vector Spaces. Singapore: World Scientific; 2002.

72. Cegielski A. Iterative Methods for Fixed Point Problems in Hilbert Spaces. Heidelberg: Springer; 2012.

73. Fitzpatrick S. Representing monotone operators by convex functions. Workshop/Miniconference on Functional Analysis and Optimization. Centre for Mathematics and its Applications, Mathematical Sciences Institute, The Australian National University; 1988.

74. Bauschke H, McLaren D, Sendov H. Fitzpatrick functions: Inequalities, examples, and remarks on a problem by S. Fitzpatrick. Journal of Convex Analysis. 2005;13.

75. Opial Z. Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Am. Math. Soc. 1967;73:591–597.

76. Mainge PE. Approximation methods for common fixed points of nonexpansive mappings in Hilbert spaces. J. Math. Anal. Appl. 2007;325:469–479.

77. López G, Martín-Márquez V, Wang F, et al. Forward-backward splitting methods for accretive operators in Banach spaces. Abstr. Appl. Anal. 2012;2012(25).

78. Banert S, Bot RI. Backward penalty schemes for monotone inclusion problems. J. Optim. Theory Appl. 2015;166:930–948.

79. Bauschke HH, Matoušková E, Reich S. Projection and proximal point methods:convergence results and counterexamples. Nonlinear Anal. 2004;56:715-738.

80. Bruck RE, Reich S. Nonexpansive projections and resolvents of accretive operators in Banach spaces. Houston J. Math. 1977;3:459-470.

81. Güler O. On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. 1991;29:403-419.

82. Chen GHG, Rockafellar RT. Convergence rates in forward backward splitting. SIAM J. Optim. 1997;7:421-444.

83. Rockafellar RT. Monotone operators and the proximal point algorithm. SIAM J. Control Optim. 1976;14:877-898.

84. Baillon JB, Haddad G. Quelques proprietes des operateurs anglebornes et cycliquement monotones. Israel J. Math. 1977;26:137-150.

85. Tikhonov AN, Arsenin VY. Solutions of Ill—Posed Problems. SIAM Rev. 1979;21:266–267.

86. Davis D, Yin W. A Three-Operator Splitting Scheme and its Optimization Applications. Set-Valued Var. Anal. 2017;25:829–858.

87. Bauschke HH, Combettes PL. A weak-to-strong convergence principle for fejer-monotone methods in Hilbert spaces. Math. Oper. Res. 2001;26:248–264.

88. Reich S. Weak convergence theorems for nonexpansive mappings in Banach spaces. J. Math. Anal. Appl. 1979;67:274–276.

89. Halpern B. Fixed points of nonexpansive maps. Bull. Amer. Math. Soc. 1967;73:957961.

90. Ma Z, Wu HR. Partition based vector filtering technique for color suppression of noise in digital color images. IEEE Trans. Image Process. 2006;15(8):2324–2342.

91. Gonzalez RC, Woods RE. Digital Image Processing. 4th ed. Prentice-Hall; 2007.

92. Lim JS. Two-Dimensional Signal Processing. Englewood Cliffs, NJ: Prentice-Hall; 1990.