

Article

A Novel Two-Step Inertial Viscosity Algorithm for Bilevel Optimization Problems Applied to Image Recovery

Rattanakorn Wattanataweekul ¹, Kobkoon Janngam ² and Suthep Suantai ^{3,*}

¹ Department of Mathematics, Statistics and Computer, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand; rattanakorn.w@ubu.ac.th

² Graduate Ph.D. Degree Program in Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand; kobkoon_jan@cmu.ac.th

³ Research Center in Optimization and Computational Intelligence for Big Data Prediction, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

* Correspondence: suthep.s@cmu.ac.th

Abstract: This paper introduces a novel two-step inertial algorithm for locating a common fixed point of a countable family of nonexpansive mappings. We establish strong convergence properties of the proposed method under mild conditions and employ it to solve convex bilevel optimization problems. The method is further applied to the image recovery problem. Our numerical experiments show that the proposed method achieves faster convergence than other related methods in the literature.

Keywords: convex bilevel optimization; forward–backward algorithms; image restoration problems; two-step inertial; viscosity approximation.

MSC: 47H09; 90C25; 65K10



Citation: Wattanataweekul, R.; Janngam, K.; Suantai, S. A Novel Two-Step Inertial Viscosity Algorithm for Bilevel Optimization Problems Applied to Image Recovery. *Mathematics* **2023**, *11*, 3518. <https://doi.org/10.3390/math11163518>

Academic Editor: Timilehin Opeyemi Alakoya

Received: 16 July 2023

Revised: 31 July 2023

Accepted: 10 August 2023

Published: 15 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bilevel optimization has received significant attention in recent years, having arisen as a powerful tool for many machine learning applications such as hyperparameter optimization [1,2], signal processing [3,4], and reinforcement learning [5]. It is defined as a mathematical program in which an optimization problem contains another optimization problem as a constraint. In this paper, we consider the bilevel optimization problem in which the following minima are sought:

$$\min_{x \in S_*} \omega(x), \quad (1)$$

where $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ is assumed to be strongly convex and differentiable, while S_* is a nonempty set of inner level optimizers satisfying

$$\min_{x \in \mathbb{R}^n} \{\psi_1(x) + \psi_2(x)\}, \quad (2)$$

where $\psi_1 : \mathbb{R}^n \rightarrow \mathbb{R}$ is a differentiable and convex function such that $\nabla \psi_1$ is L -Lipschitz continuous and $\psi_2 : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is a convex, proper, and lower semi-continuous function. We let Λ be the solution set of (1).

Observe that this bilevel optimization model contains the inner level minimization problem (2) as a constraint to the outer level optimization problem (1). It is a well-known form (1) that

$$x^* \in \Lambda \text{ if and only if } \langle \nabla \omega(x^*), x - x^* \rangle \geq 0 \text{ for all } x \in S_*.$$

Many researchers have proposed algorithms for solving problem (2); see [6–10]. The basic algorithm is the proximal forward–backward technique, or proximal gradient method, defined by the iterative equation

$$x_{n+1} = \text{prox}_{\alpha_n \psi_2}(I - \alpha_n \nabla \psi_1)(x_n), \quad n \in \mathbb{N}, \tag{3}$$

where $\alpha_n > 0$ is the step-size, prox_{ψ_2} is the proximity operator of ψ_2 , and $\nabla \psi_1$ is the gradient of ψ_1 [6,11]. Equation (3) is referred to in the literature as the forward–backward splitting algorithm (FBSA). The FBSA can be used to solve the inner level optimization problem if ψ_1 is L -Lipschitz continuous [7].

The proximal gradient method can also be viewed as a fixed-point algorithm, where the iterated mapping is given by

$$T := \text{prog}_{\alpha \psi_2}(I - \alpha \nabla \psi_1) \tag{4}$$

and is called the forward–backward mapping [12]. The forward–backward mapping, T , is nonexpansive if $0 < \alpha < 2/L$, where L is a Lipschitz constant of $\nabla \psi_1$ and, in that case, $\text{Fix}(T) = \text{argmin}\{\psi_1(x) + \psi_2(x)\}$. It is noted that implementation of the forward–backward operator can be simplified by first changing the inner level optimization problem into a zero-point problem of the sum of two monotone operators, and then, after analysis, translating back into the fixed-point problem. Exemplifying the fixed-point approach, Sabach et al. [13] proposed the bilevel gradient sequential averaging method (BiG-SAM) for solving problems (1) and (2). The iterative process can be defined as

$$\begin{cases} u_n = \text{prox}_{c g}(x_{n-1} - c \nabla f(x_{n-1})), \\ v_n = x_{n-1} - \lambda \nabla \omega(x_{n-1}), \\ x_{n+1} = \gamma_n v_n + (1 - \gamma_n) u_n, \quad n \geq 1 \end{cases} \tag{5}$$

where $c \in (0, \frac{2}{L_f})$, $\lambda \in (0, \frac{2}{L_\omega + \sigma})$, ω is strongly convex with parameter σ , and where L_f and L_ω are Lipschitz constants for the gradients of f and ω . The authors analyzed the convergence behavior of BiG-SAM using an existing fixed-point algorithm and discussed its rate of convergence.

In optimization problems like those presented above, mathematicians frequently employ a technique known as inertial-type extrapolation [14,15] to accelerate the convergence of the iterative equations. This approach involves utilizing a term $\theta_n(x_n - x_{n-1})$, where θ_n denotes an inertial parameter, to govern the momentum $x_n - x_{n-1}$. One such algorithm that has enjoyed immense popularity was developed by Nesterov [14]. He used an inertial or extrapolation technique to solve convex optimization problems of the form of (2), where $F := \psi_1 + \psi_2$ is a convex, smooth function. Nesterov’s algorithm takes the following form:

$$\begin{cases} z_n = x_n + \theta_n(x_n - x_{n-1}), \\ x_{n+1} = z_n + c \nabla F(z_n), \quad n \in \mathbb{N}, \end{cases} \tag{6}$$

where the inertial parameter $\theta_n \in (0, 1)$ for all n and $c > 0$ is the step size depending on the Lipschitz continuity modulus of ∇F . Nesterov proved that Equation (6) has a faster convergence rate than the general gradient algorithm by selecting $\{\theta_n\}$ such that $\sup_n \theta_n = 1$. Similarly, in 2009, Beck et al. [16] introduced the fast iterative shrinkage-thresholding algorithm (FISTA) for solving linear inverse problems. Their result combined the proximity algorithm with the inertial technique, again resulting in the algorithm’s convergence rate being considerably accelerated.

In 2019, Shehu et al. [17] presented an inertial forward–backward algorithm, called the inertial bilevel gradient sequential averaging method (iBiG-SAM) for solving

problems (1) and (2). Their method was subsequently improved by Sabach et al. [13], using the following iterative algorithm:

$$\begin{cases} s_n = x_n + \theta_n(x_n - x_{n-1}), \\ u_n = \text{prox}_{cg}(I - c\nabla f)(s_n), \\ v_n = s_n - \lambda \nabla \omega(s_n), \\ x_{n+1} = \gamma_n v_n + (1 - \gamma_n)u_n, \quad n \geq 1. \end{cases} \tag{7}$$

The authors transformed the bilevel optimization problem into a fixed-point problem for a nonexpansive mapping in an infinite dimensional Hilbert space and then proved strong convergence.

As the above suggests, research on fixed-point problems for nonexpansive mappings has become crucial for developing optimization methods. The Mann iterative process is a well-known method for approximating fixed points of nonexpansive mappings on Hilbert spaces. However, Mann’s process provides only weak convergence. Many authors have demonstrated fixed-point problems exhibiting strong convergence for nonexpansive mappings on Hilbert spaces using the viscosity approximation method, expressed by the equation

$$x_{n+1} = \beta_n S(x_n) + (1 - \beta_n)Tx_n, \quad n \geq 1, \tag{8}$$

where $\{\beta_n\} \in (0, 1)$, S is a contraction on Hilbert spaces H and $x_1 \in H$; see [18,19].

In 2009, Takahashi [20] modified the viscosity approximation method, selecting a particular fixed point of the nonexpansive self-mapping of Moudafi [18]. The iterative process is given by

$$x_{n+1} = \beta_n S(x_n) + (1 - \beta_n)T_n x_n, \quad n \geq 1, \tag{9}$$

where $\{\beta_n\} \in (0, 1)$, S is a contraction of C into itself, $\{T_n\}$ is a countable family of nonexpansive of C into itself, C is subset of a Banach space, and $x_1 \in C$. Takahashi proved the strong convergence of (9) to a common fixed point of T_n .

Jailoka et al. [21] introduced a fast viscosity forward–backward algorithm (FVFBA) with the inertial technique for finding a common fixed point of a countable family of nonexpansive mappings. They proved a strong convergence result and applied it to solving a convex minimization problem of the sum of two convex functions. The iterative process can be formulated by

$$\begin{cases} u_n = x_n + \theta_n(x_n - x_{n-1}), \\ v_n = (1 - \alpha_n)T_n u_n + \alpha_n S(u_n), \\ x_{n+1} = (1 - \beta_n)T_n u_n + \beta_n T_n v_n, \quad n \geq 1, \end{cases} \tag{10}$$

where $\{\alpha_n\}, \{\beta_n\} \in (0, 1)$, S is a contraction on Hilbert spaces H and $x_1 \in H$.

Recently, Janngam et al. [22] presented an inertial viscosity modified SP algorithm (IVMSPA). The authors proved a strong convergence of their algorithm and applied it to solving the convex bilevel optimization problems (problems 1 and 2). Their algorithm was given by

$$\begin{aligned} y_n &= x_n + \theta_n(x_n - x_{n-1}), \\ z_n &= (1 - \alpha_n)y_n + \alpha_n S(y_n), \\ w_n &= (1 - \beta_n)z_n + \beta_n T_n z_n, \\ x_{n+1} &= (1 - \gamma_n)w_n + \gamma_n T_n w_n, \quad n \geq 1, \end{aligned} \tag{11}$$

where $\{\alpha_n\}, \{\beta_n\}, \{\gamma_n\} \in (0, 1)$, S is contraction mapping on Hilbert spaces H and $x_1 \in H$.

The above authors all employ a single inertial parameter to accelerate the convergence of their algorithms. However, it has been noted that the incorporation of two inertial parameters enhances motion modeling, improves stability and robustness, increases redundancy and fault tolerance, expands the range of applications, and offers flexibility and adaptability in algorithm design. In [23], it was illustrated through an example that the one-step inertial extrapolation, expressed as $w_n = x_n + \theta_n(x_n - x_{n-1})$ with $\theta_n \in [0, 1)$, may not produce acceleration. Additionally, Ref. [24] mentioned that incorporating more than two points, such as x_n and x_{n-1} , in the inertial process could lead to acceleration. For instance, consider the following two-step inertial extrapolation:

$$y_n = x_n + \theta(x_n - x_{n-1}) + \delta(x_{n-1} - x_{n-2}) \tag{12}$$

where $\theta > 0$ and $\delta < 0$ can provide acceleration. The limitations of employing one-step inertial acceleration in the alternating direction method of multipliers (ADMM) were discussed in [25], which led to the proposal of adaptive acceleration as an alternative solution. In addition, Polyak [26] discussed the potential for multi-step inertial methods to enhance the speed of optimization techniques despite the absence of established convergence or rate results in [26]. Recent research conducted in [27] has further explored and examined various aspects of multi-step inertial methods.

Based on the information provided above, our aim in this paper is to solve the convex bilevel optimization problem by introducing a new accelerated viscosity algorithm with the two-point inertial technique, which we then apply to image recovery. The remainder of the paper is organized as follows. In Section 2, we recall some basic definitions and results that are crucial in the paper. The proposed algorithm and the analysis of its convergence are presented in Section 3. The performance of deblurring images using our algorithm is analyzed and illustrated in Section 4. Finally, we give conclusions and discuss directions for future work in Section 5.

2. Preliminaries

In this section, we present some preliminary material that will be needed for the main theorems.

Let C be a nonempty subset of a real Hilbert space H with norm $\|\cdot\|$, \mathbb{R} denote the set of real numbers, \mathbb{R}_+ denote the non-negative real numbers, $\mathbb{R}_{>0}$ denote the positive real numbers, \mathbb{N} denote the set of positive integers, and let I denote the identity mapping on H .

Definition 1. The mapping $T : C \rightarrow C$ is said to be L -Lipschitz with $L \geq 0$, if

$$\|Tu - Tv\| \leq L\|u - v\|$$

for all $u, v \in C$. Furthermore, if $L \in [0, 1)$ then T is called a contraction mapping, and it is nonexpansive if $L = 1$.

When $\{x_n\}$ is a sequence in C , we denote the strong convergence of x_n to $x \in C$ by $x_n \rightarrow x$, and $Fix(T)$ will symbolize the set of all fixed points of T .

Let $T : C \rightarrow C$ be a nonexpansive mapping and $\{T_n\}$ be a family of nonexpansive mappings of C into itself such that $\emptyset \neq Fix(T) \subset \Gamma := \bigcap_{n=1}^{\infty} Fix(T_n)$. The sequence $\{T_n\}$ is said to satisfy the NST-condition (I) with T [28], if for each bounded sequence $\{x_n\} \subset C$,

$$\lim_{n \rightarrow \infty} \|x_n - T_n x_n\| = 0 \text{ implies } \lim_{n \rightarrow \infty} \|x_n - T x_n\| = 0.$$

The following condition is an essential condition for proving our convergence theorem.

Definition 2 ([29,30]). A sequence $\{T_n\}$ with $\bigcap_{n=1}^\infty \text{Fix}(T_n) \neq \emptyset$ is said to satisfy the condition (Z) if for every bounded sequence $\{u_n\}$ in C such that

$$\lim_{n \rightarrow \infty} \|u_n - T_n u_n\| = 0,$$

then, every weak cluster point of $\{u_n\}$ belongs to $\bigcap_{n=1}^\infty \text{Fix}(T_n)$.

Recall that for a nonempty closed convex subset C of H , the metric projection on C is a mapping $P_C : H \rightarrow C$, defined by

$$P_C x = \operatorname{argmin}\{\|x - y\| : y \in C\}$$

for all $x \in H$. Note that $v = P_C x$ if and only if $\langle x - v, y - v \rangle \leq 0$ for all $y \in C$.

The definition and properties of a proximity operator are presented below.

Definition 3 ([31,32]). Let $g : H \rightarrow \mathbb{R} \cup \{\infty\}$ be a function that is convex, proper, and lower semi-continuous. The function prox_g , known as the proximity operator of g , is defined as follows:

$$\operatorname{prox}_g(x) := \min_{y \in H} \left(g(y) + \frac{1}{2} \|x - y\|^2 \right).$$

Alternatively, it can be expressed as:

$$\operatorname{prox}_g = (I + \partial g)^{-1},$$

where ∂g represents the subdifferential of g defined by:

$$\partial g(x) := \{v \in H : g(x) + \langle v, u - x \rangle \leq g(u) \text{ for all } u \in H\}$$

for any $x \in H$. Additionally, for $\rho > 0$, we know that $\operatorname{prox}_{\rho g}$ is firmly nonexpansive and

$$\operatorname{Fix}(\operatorname{prox}_{\rho g}) = \operatorname{Argmin}(g) := \{v \in H : g(v) \leq g(u) \text{ for all } u \in H\},$$

where $\operatorname{Fix}(\operatorname{prox}_{\rho g})$ is the set of fixed points of $\operatorname{prox}_{\rho g}$.

The following lemmas will be used for proving the convergence of our proposed algorithm.

Lemma 1 ([33]). Let $g : H \rightarrow \mathbb{R} \cup \{\infty\}$ be a convex, proper, and lower semi-continuous function and let $f : H \rightarrow \mathbb{R}$ be a differentiable and convex function such that ∇f is L -Lipschitz continuous. Let

$$T_n := \operatorname{prox}_{\rho_n g}(I - \rho_n \nabla f) \quad \text{and} \quad T := \operatorname{prox}_{\rho g}(I - \rho \nabla f),$$

where $\rho_n, \rho \in (0, 2/L)$ with $\rho_n \rightarrow \rho$ as $n \rightarrow \infty$. Then $\{T_n\}$ satisfies the NST-condition (I) with T .

Lemma 2 ([34]). Let $x_1, x_2 \in H$ and $t \in [0, 1]$. Then, the following properties are true:

- (i) $\|x_1 \pm x_2\|^2 = \|x_1\|^2 \pm 2\langle x_1, x_2 \rangle + \|x_2\|^2$;
- (ii) $\|x_1 + x_2\|^2 \leq \|x_1\|^2 + 2\langle x_2, x_1 + x_2 \rangle$;
- (iii) $\|tx_1 + (1 - t)x_2\|^2 = t\|x_1\|^2 + (1 - t)\|x_2\|^2 - t(1 - t)\|x_1 - x_2\|^2$.

Lemma 3 ([35]). Let $\{a_n\}, \{b_n\} \subset \mathbb{R}_+$ and $\{t_n\} \subset (0, 1)$ such that $\sum_{n=1}^\infty t_n = \infty$. Assume that

$$a_{n+1} \leq (1 - t_n)a_n + t_n b_n$$

for all $n \in \mathbb{N}$. If $\limsup_{i \rightarrow \infty} b_{n_i} \leq 0$ for every subsequence $\{a_{n_i}\}$ of $\{a_n\}$ satisfying

$$\liminf_{i \rightarrow \infty} (a_{n_i+1} - a_{n_i}) \geq 0,$$

then $\lim_{n \rightarrow \infty} a_n = 0$.

3. Main Results

Throughout this section, we let C be closed convex with $\emptyset \neq C \subset H$ and a mapping $F : C \rightarrow C$ be a k -contraction where $0 < k < 1$. Let $\{T_n\}$ is a family of nonexpansive mappings of C into itself satisfying the condition (Z) such that $\Gamma := \bigcap_{n=1}^{\infty} \text{Fix}(T_n) \neq \emptyset$.

For the first of our main results, we draw upon the ideas of Jailoka et al. [21] and Liang [24] and introduce a modified two-step inertial viscosity algorithm (MTIVA) for finding a common fixed point of a family of nonexpansive mappings $\{T_n\}$, as follows:

In Theorem 1, we show that Algorithm 1 converges strongly.

Algorithm 1 Modified Two-Step Inertial Viscosity Algorithm (MTIVA)

Initialization: Let $\{\beta_n\}, \{\gamma_n\} \subset [0, 1], \{\tau_n\} \subset \mathbb{R}_+$ and let $\{\mu_n\}, \{\rho_n\} \subset \mathbb{R}_{>0}$ be bounded sequences. Take $x_{-1}, x_0, x_1 \in H$ arbitrarily. For $n \in \mathbb{N}$.

Step 1. Compute the inertial step:

$$\vartheta_n = \begin{cases} \min\left\{\mu_n, \frac{\tau_n}{\|x_n - x_{n-1}\|}\right\} & \text{if } x_n \neq x_{n-1}, \\ \mu_n & \text{otherwise,} \end{cases} \tag{13}$$

and

$$\delta_n = \begin{cases} \max\left\{-\rho_n, \frac{-\tau_n}{\|x_{n-1} - x_{n-2}\|}\right\} & \text{if } x_{n-1} \neq x_{n-2}, \\ -\rho_n & \text{otherwise,} \end{cases} \tag{14}$$

$$w_n = x_n + \vartheta_n(x_n - x_{n-1}) + \delta_n(x_{n-1} - x_{n-2}). \tag{15}$$

Step 2. Compute the viscosity step:

$$z_n = (1 - \gamma_n)T_n w_n + \gamma_n F(w_n). \tag{16}$$

Step 3. Compute x_{n+1} :

$$x_{n+1} = (1 - \beta_n)T_n w_n + \beta_n T_n z_n. \tag{17}$$

Theorem 1. Let a sequence $\{x_n\}$ be generated by Algorithm 1. Suppose the conditions (C1–C3) hold for the sequences $\{\tau_n\}, \{\gamma_n\}$, and $\{\beta_n\}$. Then, $x_n \rightarrow \check{p} \in \Gamma$, where $\check{p} = P_{\Gamma}F(\check{p})$.

(C1) $\lim_{n \rightarrow \infty} \frac{\tau_n}{\gamma_n} = 0$;

(C2) $0 < \epsilon_1 \leq \beta_n \leq \epsilon_2 < 1$ for some $\epsilon_1, \epsilon_2 \in \mathbb{R}$;

(C3) $0 < \gamma_n < 1$, $\lim_{n \rightarrow \infty} \gamma_n = 0$ and $\sum_{n=1}^{\infty} \gamma_n = \infty$.

Proof. Let $\check{p} = P_{\Gamma}F(\check{p})$. By the definition of z_n , we obtain

$$\begin{aligned} \|z_n - \check{p}\| &= \|(1 - \gamma_n)T_n w_n + \gamma_n F(w_n) - \check{p}\| \\ &\leq (1 - \gamma_n)\|T_n w_n - \check{p}\| + \gamma_n\|F(w_n) - F(\check{p})\| + \gamma_n\|F(\check{p}) - \check{p}\| \\ &\leq (1 - \gamma_n(1 - k))\|w_n - \check{p}\| + \gamma_n\|F(\check{p}) - \check{p}\|. \end{aligned} \tag{18}$$

By the definition of w_n , we obtain

$$\begin{aligned} \|w_n - \check{p}\| &= \|x_n + \vartheta_n(x_n - x_{n-1}) + \delta_n(x_{n-1} - x_{n-2}) - \check{p}\| \\ &\leq \|x_n - \check{p}\| + \vartheta_n\|x_n - x_{n-1}\| + \delta_n\|x_{n-1} - x_{n-2}\|. \end{aligned} \tag{19}$$

Using (18) and (19), we obtain

$$\begin{aligned} \|x_{n+1} - \check{p}\| &\leq (1 - \beta_n)\|T_n w_n - \check{p}\| + \beta_n\|T_n z_n - \check{p}\| \\ &\leq (1 - \beta_n)\|w_n - \check{p}\| + \beta_n\|z_n - \check{p}\| \\ &\leq (1 - \gamma_n \beta_n(1 - k))\|w_n - \check{p}\| + \beta_n \gamma_n \|F(\check{p}) - \check{p}\| \\ &\leq (1 - \gamma_n \beta_n(1 - k))(\|x_n - \check{p}\| + \vartheta_n \|x_n - x_{n-1}\| + \delta_n \|x_{n-1} - x_{n-2}\|) \\ &\quad + \beta_n \gamma_n \|F(\check{p}) - \check{p}\| \\ &\leq (1 - \gamma_n \beta_n(1 - k))\|x_n - \check{p}\| + \beta_n \gamma_n \left(\frac{\vartheta_n}{\beta_n \gamma_n} \|x_n - x_{n-1}\| \right. \\ &\quad \left. + \frac{\delta_n}{\beta_n \gamma_n} \|x_{n-1} - x_{n-2}\| + \|F(\check{p}) - \check{p}\| \right). \end{aligned}$$

By (13), (14) and (C1), we have $\frac{\vartheta_n}{\beta_n \gamma_n} \|x_n - x_{n-1}\| \rightarrow 0$ as $n \rightarrow \infty$ and $\frac{\delta_n}{\beta_n \gamma_n} \|x_{n-1} - x_{n-2}\| \rightarrow 0$ as $n \rightarrow \infty$, and then $M_1, M_2 > 0$ exist such that

$$\frac{\vartheta_n}{\beta_n \gamma_n} \|x_n - x_{n-1}\| \leq M_1 \quad \text{and} \quad \frac{\delta_n}{\beta_n \gamma_n} \|x_{n-1} - x_{n-2}\| \leq M_2$$

for all $n \geq 1$. Then,

$$\begin{aligned} \|x_{n+1} - \check{p}\| &\leq (1 - \gamma_n \beta_n(1 - k))\|x_n - \check{p}\| + \beta_n \gamma_n(1 - k) \left(\frac{M_1 + M_2 + \|F(\check{p}) - \check{p}\|}{1 - k} \right) \\ &\leq \max \left\{ \|x_n - \check{p}\|, \frac{M + \|F(\check{p}) - \check{p}\|}{1 - k} \right\}, \end{aligned}$$

where $M = M_1 + M_2 > 0$. Thus, by mathematical induction, we deduce that

$$\|x_n - \check{p}\| \leq \max \left\{ \|x_1 - \check{p}\|, \frac{M + \|F(\check{p}) - \check{p}\|}{1 - k} \right\}$$

for all $n \geq 1$. Hence, the sequence $\{x_n\}$ is bounded and so are the sequences $\{F(w_n)\}$, $\{T_n w_n\}$, $\{z_n\}$. Now, by Lemma 2, we obtain

$$\begin{aligned} \|z_n - \check{p}\|^2 &= \|(1 - \gamma_n)(T_n w_n - \check{p}) + \gamma_n(F(w_n) - F(\check{p})) + \gamma_n(F(\check{p}) - \check{p})\|^2 \\ &\leq \|\gamma_n(F(w_n) - F(\check{p})) + (1 - \gamma_n)(T_n w_n - \check{p})\|^2 + 2\gamma_n \langle F(\check{p}) - \check{p}, z_n - \check{p} \rangle \\ &\leq \gamma_n \|F(w_n) - F(\check{p})\|^2 + (1 - \gamma_n)\|T_n w_n - \check{p}\|^2 + 2\gamma_n \langle F(\check{p}) - \check{p}, z_n - \check{p} \rangle \\ &\leq (1 - \gamma_n(1 - k))\|w_n - \check{p}\|^2 + 2\gamma_n \langle F(\check{p}) - \check{p}, z_n - \check{p} \rangle \end{aligned} \tag{20}$$

and

$$\begin{aligned} \|w_n - \check{p}\|^2 &= \|x_n - \check{p}\|^2 + 2\langle x_n - \check{p}, \vartheta_n(x_n - x_{n-1}) + \delta_n(x_{n-1} - x_{n-2}) \rangle \\ &\quad + \|\vartheta_n(x_n - x_{n-1}) + \delta_n(x_{n-1} - x_{n-2})\|^2 \\ &\leq \|x_n - \check{p}\|^2 + 2\vartheta_n \|x_n - \check{p}\| \|x_{n-1} - x_n\| + 2|\delta_n| \|x_n - \check{p}\| \|x_{n-1} - x_{n-2}\| \\ &\quad + \vartheta_n^2 \|x_{n-1} - x_n\|^2 + 2\vartheta_n |\delta_n| \|x_{n-1} - x_n\| \|x_{n-1} - x_{n-2}\| \\ &\quad + \delta_n^2 \|x_{n-1} - x_{n-2}\|^2. \end{aligned} \tag{21}$$

Also, from Lemma 2 (iii), (20) and (21), we obtain

$$\begin{aligned}
 \|x_{n+1} - \check{p}\|^2 &= (1 - \beta_n)\|T_n w_n - \check{p}\|^2 + \beta_n\|T_n z_n - \check{p}\|^2 - \beta_n(1 - \beta_n)\|T_n w_n - T_n z_n\|^2 \\
 &\leq (1 - \beta_n)\|w_n - \check{p}\|^2 + \beta_n\|z_n - \check{p}\|^2 - \beta_n(1 - \beta_n)\|T_n w_n - T_n z_n\|^2 \\
 &\leq (1 - \beta_n\gamma_n(1 - k))\|w_n - \check{p}\|^2 + 2\gamma_n\beta_n\langle F(\check{p}) - \check{p}, z_n - \check{p} \rangle \\
 &\quad - \beta_n(1 - \beta_n)\|T_n w_n - T_n z_n\|^2 \\
 &\leq (1 - \beta_n\gamma_n(1 - k))\|x_n - \check{p}\|^2 + 2\vartheta_n\|x_n - \check{p}\|\|x_{n-1} - x_n\| \\
 &\quad + 2|\delta_n|\|x_n - \check{p}\|\|x_{n-1} - x_{n-2}\| + \vartheta_n^2\|x_{n-1} - x_n\|^2 \\
 &\quad + 2\vartheta_n|\delta_n|\|x_{n-1} - x_n\|\|x_{n-1} - x_{n-2}\| + \delta_n^2\|x_{n-1} - x_{n-2}\|^2 \\
 &\quad + 2\gamma_n\beta_n\langle F(\check{p}) - \check{p}, z_n - \check{p} \rangle - \beta_n(1 - \beta_n)\|T_n w_n - T_n z_n\|^2 \\
 &= (1 - \beta_n\gamma_n(1 - k))\|x_n - \check{p}\|^2 - \beta_n(1 - \beta_n)\|T_n w_n - T_n z_n\|^2 \\
 &\quad + \beta_n\gamma_n(1 - k)b_n,
 \end{aligned} \tag{22}$$

where

$$\begin{aligned}
 b_n &= \frac{1}{1 - k} \left(\frac{2\vartheta_n}{\beta_n\gamma_n}\|x_n - \check{p}\|\|x_{n-1} - x_n\| + \frac{2|\delta_n|}{\beta_n\gamma_n}\|x_n - \check{p}\|\|x_{n-1} - x_{n-2}\| \right. \\
 &\quad \left. + \frac{2\vartheta_n|\delta_n|}{\beta_n\gamma_n}\|x_{n-1} - x_n\|\|x_{n-1} - x_{n-2}\| + \frac{\delta_n^2}{\beta_n\gamma_n}\|x_{n-1} - x_{n-2}\|^2 \right. \\
 &\quad \left. + 2\langle F(\check{p}) - \check{p}, z_n - \check{p} \rangle \right).
 \end{aligned}$$

It follows that

$$\beta_n(1 - \beta_n)\|T_n w_n - T_n z_n\|^2 \leq \|x_n - \check{p}\|^2 - \|x_{n+1} - \check{p}\|^2 + \beta_n\gamma_n(1 - k)M', \tag{23}$$

where $M' = \sup\{b_n : n \in \mathbb{N}\}$.

Next, we shall show that the sequence $\{x_n\}$ converges strongly to \check{p} . Take $a_n := \|x_n - \check{p}\|^2$ and $t_n = \beta_n\gamma_n(1 - k)$. From (22), we have

$$a_{n+1} \leq (1 - t_n)a_n + t_nb_n$$

for all $n \in \mathbb{N}$. To apply Lemma 3, we have to show that $\limsup_{i \rightarrow \infty} b_{n_i} \leq 0$ whenever a subsequence $\{a_{n_i}\}$ of $\{a_n\}$ satisfies

$$\liminf_{i \rightarrow \infty} (a_{n_i+1} - a_{n_i}) \geq 0. \tag{24}$$

Suppose that $\{a_{n_i}\}$ is a subsequence of $\{a_n\}$ satisfying (24). It follows from (23) and (C3) that

$$\begin{aligned}
 \limsup_{i \rightarrow \infty} \beta_{n_i}(1 - \beta_{n_i})\|T_{n_i} w_{n_i} - T_{n_i} z_{n_i}\|^2 &\leq \limsup_{i \rightarrow \infty} (a_{n_i} - a_{n_i+1} + \beta_{n_i}\gamma_{n_i}(1 - k)M') \\
 &\leq \limsup_{i \rightarrow \infty} (a_{n_i} - a_{n_i+1}) + (1 - k)M' \lim_{i \rightarrow \infty} \beta_{n_i}\gamma_{n_i} \\
 &= -\liminf_{i \rightarrow \infty} (a_{n_i+1} - a_{n_i}) \\
 &\leq 0.
 \end{aligned}$$

The condition (C2) and above inequality lead to

$$\lim_{i \rightarrow \infty} \|T_{n_i} w_{n_i} - T_{n_i} z_{n_i}\| = 0. \tag{25}$$

Using (C2) and (C3), and since

$$\beta_{n_i} \|z_{n_i} - T_{n_i} w_{n_i}\| = \beta_{n_i} \gamma_{n_i} \|F(w_{n_i}) - T_{n_i} w_{n_i}\|,$$

we obtain

$$\lim_{i \rightarrow \infty} \|z_{n_i} - T_{n_i} w_{n_i}\| = 0. \tag{26}$$

From (25) and (26), we obtain

$$\|z_{n_i} - T_{n_i} z_{n_i}\| \leq \|z_{n_i} - T_{n_i} w_{n_i}\| + \|T_{n_i} w_{n_i} - T_{n_i} z_{n_i}\| \rightarrow 0 \tag{27}$$

as $i \rightarrow \infty$. In order to prove that $\limsup_{i \rightarrow \infty} b_{n_i} \leq 0$, it suffices to show that

$$\limsup_{i \rightarrow \infty} \langle F(\check{p}) - \check{p}, z_{n_i} - \check{p} \rangle \leq 0. \tag{28}$$

Since $\{z_{n_i}\}$ is bounded, a subsequence $\{z_{n_{i_j}}\}$ of $\{z_{n_i}\}$ and $y \in H$ exists such that $\{z_{n_{i_j}}\} \rightarrow y$ as $j \rightarrow \infty$ and

$$\begin{aligned} \limsup_{i \rightarrow \infty} \langle F(\check{p}) - \check{p}, z_{n_i} - \check{p} \rangle &= \lim_{j \rightarrow \infty} \langle F(\check{p}) - \check{p}, z_{n_{i_j}} - \check{p} \rangle \\ &= \langle F(\check{p}) - \check{p}, y - \check{p} \rangle. \end{aligned}$$

Since $\{T_n\}$ satisfies the condition (Z) and (27), we obtain $y \in \Gamma$. From $\check{p} = P_{\Gamma} F(\check{p})$, we obtain

$$\langle F(\check{p}) - \check{p}, z - \check{p} \rangle \leq 0$$

For all $z \in \Gamma$. In particular, we have

$$\langle F(\check{p}) - \check{p}, y - \check{p} \rangle \leq 0.$$

Hence, we obtain (28). Thus, in view of Lemma 3, $\{x_n\}$ converges to \check{p} , as required. \square

In what follows, we impose the assumptions on the mappings ψ_1, ψ_2 , and ω associated with the convex bilevel optimization problems (1) and (2).

(A1) $\psi_1 : H \rightarrow \mathbb{R}$ is a convex and differentiable function such that $\nabla \psi_1$ is Lipschitz continuous with constant $L_{\psi_1} > 0$ and $\psi_2 : H \rightarrow (-\infty, \infty]$ are proper lower semi-continuous and convex functions;

(A2) $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ is strongly convex with parameter σ such that $\nabla \omega$ is L_{ω} -Lipschitz continuous and $s \in (0, \frac{2}{L_{\omega} + \sigma})$.

With the above assumptions in place, we propose the following algorithm, called the two-step inertial forward-backward bilevel gradient method (TIFB-BiGM), for solving problems (1) and (2).

The proposition below is attributable to Sabach and Shtern [13] and is critical to our next result.

Proposition 1. *Suppose that $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ is strongly convex with $\sigma > 0$ and $\nabla \omega$ is Lipschitz continuous with constant L_{ω} . Hence, it follows that for all $s \in (0, \frac{2}{\sigma + L_{\omega}})$, the mapping $S_s = I - s \nabla \omega$ is a contraction such that*

$$\|x - s \nabla \omega(u) - (v - s \nabla \omega(v))\| \leq \sqrt{1 - \frac{2s\sigma L_{\omega}}{\sigma + L_{\omega}}} \|u - v\|$$

for all $u, v \in \mathbb{R}^n$.

Theorem 2. *The sequence $\{x_n\}$ generated by Algorithm 2 converges strongly to $\check{p} \in \Lambda$, where Λ is the set of all solutions of (1) and $\check{p} = P_{S_*}(I - s\nabla\omega)(\check{p})$, provided that all conditions as in Theorem 1 hold.*

Algorithm 2 Two-Step Inertial Forward–Backward Bilevel Gradient Method (TIFB-BiGM)

Initialization: Let $\{\beta_n\}, \{\gamma_n\} \subset [0, 1]$, $\{\tau_n\} \subset \mathbb{R}_+$, and let $\{\mu_n\}, \{\rho_n\} \subset \mathbb{R}_{>0}$ be bounded sequences. Take $x_{-1}, x_0, x_1 \in H$ arbitrarily. Let $\{c_n\} \subset (0, \frac{2}{L_{\psi_1}})$ with $c_n \rightarrow c$ as $n \rightarrow \infty$, where $c \in (0, \frac{2}{L_{\psi_1}})$. For $n \in \mathbb{N}$.

Step 1. Compute the inertial step:

$$\vartheta_n = \begin{cases} \min\left\{\mu_n, \frac{\tau_n}{\|x_n - x_{n-1}\|}\right\} & \text{if } x_n \neq x_{n-1}, \\ \mu_n & \text{otherwise,} \end{cases} \tag{29}$$

and

$$\delta_n = \begin{cases} \max\left\{-\rho_n, \frac{-\tau_n}{\|x_{n-1} - x_{n-2}\|}\right\} & \text{if } x_{n-1} \neq x_{n-2}, \\ -\rho_n & \text{otherwise,} \end{cases} \tag{30}$$

$$w_n = x_n + \vartheta_n(x_n - x_{n-1}) + \delta_n(x_{n-1} - x_{n-2}). \tag{31}$$

Step 2. Compute:

$$z_n = (1 - \gamma_n)prox_{c_n\psi_2}(I - c_n\nabla\psi_1)w_n + \gamma_n(I - s\nabla\omega)(w_n), \tag{32}$$

$$x_{n+1} = (1 - \beta_n)prox_{c_n\psi_2}(I - c_n\nabla\psi_1)w_n + \beta_nprox_{c_n\psi_2}(I - c_n\nabla\psi_1)z_n. \tag{33}$$

Proof. Put $F = I - s\nabla\omega$ and $T_n = prox_{c_n\psi_2}(I - c_n\nabla\psi_1)$, where $c_n \in (0, \frac{2}{L_{\psi_1}})$. Then, by Proposition 1, F is a contraction mapping. We also know that T_n is nonexpansive. Using Theorem 1, we conclude that $x_n \rightarrow \check{p} \in \Gamma$, where $\check{p} = P_{\Gamma}F(\check{p})$. It is noted that, $\Gamma = \bigcap_{n=1}^{\infty} Fix(T_n) = S_*$. Then, for all $x \in S_*$, we have

$$0 \geq \langle F(\check{p}) - \check{p}, x - \check{p} \rangle = \langle \check{p} - s\nabla\omega(\check{p}) - \check{p}, x - \check{p} \rangle = \langle -s\nabla\omega(\check{p}), x - \check{p} \rangle.$$

Dividing above inequalities by $-s$, we obtain

$$\langle \nabla\omega(\check{p}), x - \check{p} \rangle \geq 0$$

for all $x \in S_*$. Hence, $\check{p} \in \Lambda$, so $x_n \rightarrow \check{p} \in \Lambda$. This completes the proof. \square

4. Application to Image Recovery

Algorithm 2 will now be applied to the problem of image restoration. The algorithm’s performance will be compared to that of several existing methods, such as IVMSPA, FVFBA, BiG-SAM, and iBiG-SAM. Image restoration, also known as image deblurring or image deconvolution, is the process of removing or minimizing degradations (blur) in an image. Efforts along these lines began in the 1950s, and applications have been found in a number of areas, including consumer photography, scientific exploration, and image/video decoding; see [36,37]. Mathematically, image restoration can be modeled with the equation

$$v = Ax + \check{b}, \tag{34}$$

where $v \in \mathbb{R}^m$ is the observed image, $A \in \mathbb{R}^{m \times n}$ is the blurring matrix, $x \in \mathbb{R}^n$ is an original image, and \check{b} is an additive noise. The objective is to recover the original image $\check{x} \in \mathbb{R}^n$ that satisfies (34) by minimizing the value of \check{b} using the least squares method as shown in

Equation (35). This method aims to minimize the squared difference between v and $\mathcal{A}x$ defined as follows:

$$\min_x \|v - \mathcal{A}x\|_2^2, \tag{35}$$

where $\|\cdot\|_2$ is the Euclidean norm. Many iterations, such as the Richardson iteration, see [38], can be used to estimate the solution of (35). The problem stated in Equation (35) is considered ill-posed because there are more unknown variables than observations, resulting in a norm result that is too large to be meaningful. This issue is discussed in references [39,40]. To address this problem, various regularization methods have been introduced to improve the least squares problem. One commonly used method is Tikhonov regularization, which was proposed by Tikhonov and involves minimizing a specific equation.

$$\min_x \left\{ \|v - \mathcal{A}x\|_2^2 + \zeta \|Lx\|_2 \right\}, \tag{36}$$

where ζ is a positive parameter known as a regularization parameter, $\|\cdot\|_1$ is the l_1 -norm and $\|\cdot\|_2$ is the Euclidean norm, and $L \in \mathbb{R}^{m \times n}$ is called the Tikhonov matrix. L is set to be the identity in the standard form. A well-known model for solving problem (34) is the least absolute shrinkage and selection operator (LASSO) [41], which is defined by the expression

$$\min_x \left\{ \|v - \mathcal{A}x\|_2^2 + \zeta \|x\|_1 \right\}. \tag{37}$$

The restoration of RGB images presents a challenge for the model (36) due to the significant size of the matrix A , as well as its associated elements, which can make computing the multiplication Ax and $\|x\|_1$ quite expensive. To address this, researchers in this field commonly implement a 2-D fast Fourier transform to transform the images, resulting in a modified version of the model (36) that overcomes this issue.

$$\min_x \left\{ \|v - \mathcal{A}x\|_2^2 + \zeta \|Wx\|_1 \right\}. \tag{38}$$

The blurring operation A , commonly selected as $A = RW$, plays a crucial role in the problem (34). R represents the blurring matrix, while W denotes the two-dimensional fast Fourier transform. The observed image $v \in \mathbb{R}^{m \times n}$ is affected by both blurring and noise, with its dimensions being $m \times n$.

Now, let S_* be the set of all solutions of (38). Among the solutions in S_* , we would also like to select a solution $x^* \in S_*$ in such a way that x^* is a minimizer of

$$\min_{x^* \in S_*} \frac{1}{2} \|x^*\|^2. \tag{39}$$

We consider 2 RGB images (Wat Chedi Luang [42] and Matsue Castle) with the size of 256×256 as the original images (see Figure 1). The pictures we used in this experiment were created by the third author. In order to simulate blurring, we convolved the images using a Gaussian blur filter with a size of 9×9 and a standard deviation of $\sigma = 4$ with noise 10^{-4} .

Peak signal-to-noise ratio (PSNR) [43] and signal-to-noise ratio (SNR) [44] were used as the metrics for evaluating the performance of each algorithm. The PSNR and SNR at x_n are given by

$$PSNR(x_n) = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right), \tag{40}$$

$$SNR(x_n) = 10 \log_{10} \left(\frac{\|x - \bar{x}\|_2}{\|x_n - \bar{x}\|_2} \right), \tag{41}$$

where MAX is the maximum pixel value (usually 255 in 8-bit grayscale images) and $MSE = \frac{1}{256^2} \|x_n - x\|_2^2$ is the mean squared error between the original and the distorted image. Both and SNR are expressed in decibels (dB) as a logarithmic measure of the signal-to-noise or signal-to-error ratio.

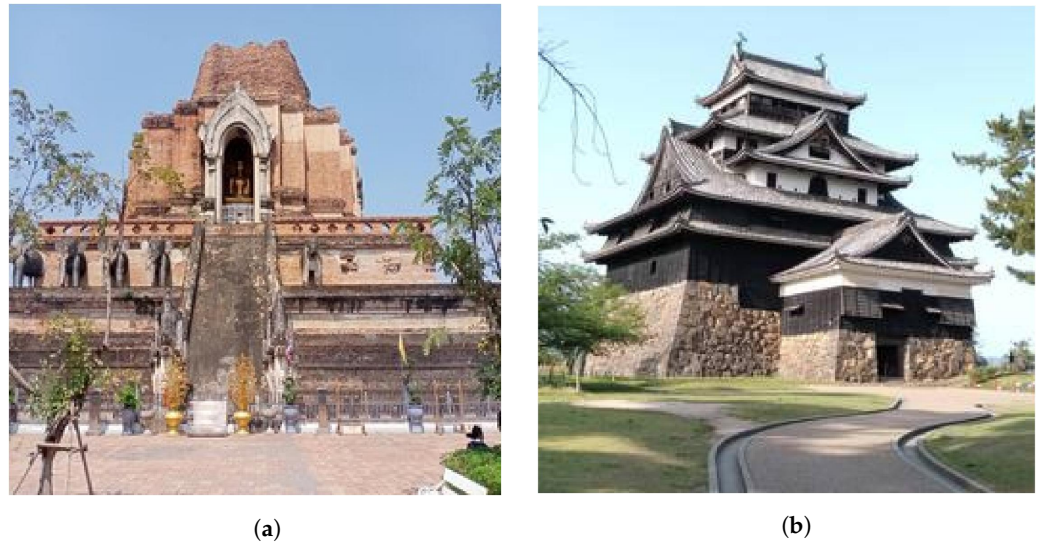


Figure 1. Original images: (a) Wat Chedi Luang, (b) Matsue Castle.

In image restoration, both PSNR and SNR are commonly used as metrics to assess the performance of deblurring results. However, it is important to note that these metrics provide different types of information.

PSNR measures the quality of a deblurred image by comparing it to the original image and evaluating the amount of noise introduced during the restoration process. It calculates the ratio between the peak signal power (the maximum possible value for the pixel) and the mean squared error (MSE) between the original and deblurred images. Higher PSNR values indicate better restoration quality as they indicate a lower level of distortion or noise.

On the other hand, SNR measures the ratio between the signal power and the noise power in the deblurred image. It quantifies the preservation of the original signal after the restoration process. Higher SNR values indicate less noise in the deblurred image.

While both PSNR and SNR are useful metrics, they focus on different aspects of image restoration. PSNR primarily considers the visual quality and fidelity of the deblurred image compared to the original, while SNR focuses more on the amount of noise present in the deblurred image.

To comprehensively evaluate the performance of your deblurring algorithm, it is recommended to consider both PSNR and SNR. They provide complementary information about the restoration quality.

We now employ our proposed algorithm (TIFB-BiGM) in Theorem 2 to solve the convex bilevel optimization problems (38) and (39). In our experiments, the algorithm developed in this paper (TIFB-BiGM) as well as the others are discussed and applied to solve the convex bilevel optimization problems (38) and (39), where $\omega(x) = \frac{1}{2} \|x\|_2^2$, $\psi_1(x) = \|v - Ax\|^2$, $\psi_2(x) = \zeta \|Wx\|_1$ and $\zeta = 5 \times 10^{-5}$. The observed images are blurred images. We compute the Lipschitz constant L_{ψ_1} by using the maximum eigenvalues of the matrix $A^T A$.

For the first experiment, the parameters of the TIFB-BiGM are chosen as follows: $\beta_n = \frac{0.99n}{n+1}$, $\gamma_n = \frac{1}{50n}$, $c_n = \frac{1}{L_{\psi_1}}$, $\tau_n = \frac{10^{14}}{n^2}$ and $s = 0.01$. Now, the experiments for recovering the “Wat Chedi Luang” image with size of 256×256 using TIFB-BiGM with different inertial parameters are shown in Tables 1 and 2. We also observe from Tables 1 and 2 that μ_n tends to 1 and ρ_n tends to 0

$$\mu_n = \frac{0.99n}{n + 0.001} \text{ and } \rho_n = \frac{1}{n^2}$$

gives the highest values of PSNR and SNR for our method.

Table 1. PSNR values for restoration of “Wat Chedi Luang” image by TIFB-BiGM after 300 iterations for different choices of parameters μ_n and ρ_n .

$\mu_n \rightarrow$ $\rho_n \downarrow$	0.1	0.3	0.5	0.9	$\frac{0.99n}{n+0.001}$	1
0.1	22.9755	23.2143	23.5185	24.6769	25.3398	25.4489
0.3	22.7791	22.9764	23.2154	23.9454	24.2129	24.2479
0.5	22.6116	22.7799	22.9773	23.5215	23.6923	23.7133
0.9	22.3362	22.4662	22.6129	22.9789	23.0805	23.0924
$\frac{1}{n^2}$	23.0847	23.3513	23.7038	25.4271	26.2116	24.9267

Table 2. SNR values for restoration of “Wat Chedi Luang” image by TIFB-BiGM after 300 iterations for different choices of parameters μ_n and ρ_n .

$\mu_n \rightarrow$ $\rho_n \downarrow$	0.1	0.3	0.5	0.9	$\frac{0.99n}{n+0.001}$	1
0.1	18.9503	19.1890	19.4932	20.6516	21.3144	21.4236
0.3	18.7539	18.9510	19.1901	19.9200	20.1876	20.2225
0.5	18.5864	18.7545	18.9519	19.4961	19.6670	19.6879
0.9	18.3110	18.4408	18.5875	18.9536	19.0551	19.0670
$\frac{1}{n^2}$	19.0595	18.3260	19.6784	21.4018	22.1913	20.9014

The parameter values for each algorithm were chosen for optimum performance, based on the published literature. The value for γ_n in Table 3 is the best choice for BiG-SAM considered in [13]. For iBiG-SAM, $\alpha = 3$ is the best choice over other values considered in [17], and the same authors found, based on their numerical experiments, $\mu_n = \frac{n}{n+1}$ to be the best choice for FVFBA.

Table 3. Parameters selection of TIFB-BiGM, IVMSPA, FVFBA, BiG-SAM, and iBiG-SAM.

Methods	Setting
TIFB-BiGM	$s = 0.01, c_n = \frac{1}{L_{\Psi_1}}, \beta_n = \frac{0.99n}{n+1}, \gamma_n = \frac{1}{50n},$ $\tau_n = \frac{10^{18}}{n^2}, \mu_n = \frac{0.99n}{n+0.001}, \rho_n = \frac{1}{n^2}$
IVMSPA	$s = 0.01, c_n = \frac{1}{L_f}, \alpha_n = \frac{1}{50n}, \beta_n = \gamma_n = 0.5,$ $\tau_n = \frac{10^{20}}{n}$ $\theta_n = \begin{cases} \min \left\{ \frac{p_n-1}{p_{n+1}}, \frac{\alpha_n \tau_n}{\ x_n - x_{n-1}\ } \right\} & \text{if } x_n \neq x_{n-1} \\ \frac{p_n-1}{p_{n+1}} & \text{otherwise} \end{cases}$ <p style="text-align: center;">where $p_1 = 1$ and $p_{n+1} = \frac{1 + \sqrt{1 + 4p_n^2}}{2}$</p>
FVFBA	$c_n = \frac{n}{n+1}, \beta_n = \frac{0.99n}{n+1}, \gamma_n = \frac{1}{50n}, \tau_n = \frac{10^{15}}{n^2}$ $\theta_n = \begin{cases} \min \left\{ \frac{n}{n+1}, \frac{\tau_n}{\ x_n - x_{n-1}\ } \right\} & \text{if } x_n \neq x_{n-1} \\ \frac{n}{n+1} & \text{otherwise} \end{cases}$

Table 3. Cont.

Methods	Setting
BiG-SAM	$\lambda = 0.01, c = \frac{1}{L_{\psi_1}}, \gamma_n = \frac{2(0.1)}{1-n^{\frac{2+cL_{\psi_1}}{4}}}$
iBiG-SAM	$\lambda = 0.01, c = \frac{1}{L_{\psi_1}}, \gamma_n = \frac{2(0.1)}{1-\frac{2+cL_{\psi_1}}{4}}, \beta_n = \frac{\gamma_n}{n^{0.01}}$
	$\theta_n = \begin{cases} \min\left\{\frac{n}{n+\alpha-1}, \frac{\beta_n}{\ x_n-x_{n-1}\ }\right\} & \text{if } x_n \neq x_{n-1}, \\ \frac{n}{n+\alpha-1} & \text{otherwise} \end{cases}$

The following experiments demonstrate Algorithm 2’s efficiency for image restoration in comparison to IVMSPA, FVFBA, BiG-SAM, and iBiG-SAM using PSNR and SNR as measurements.

The efficiency of restoring images using various algorithms under different iterations are illustrated in Figures 2–7. The results indicate that TIFB-BiGM achieves higher PSNR and SNR values than IVMSPA, FVFBA, BiG-SAM, and iBiG-SAM. Therefore, our algorithm demonstrates superior convergence behavior compared to the aforementioned methods.

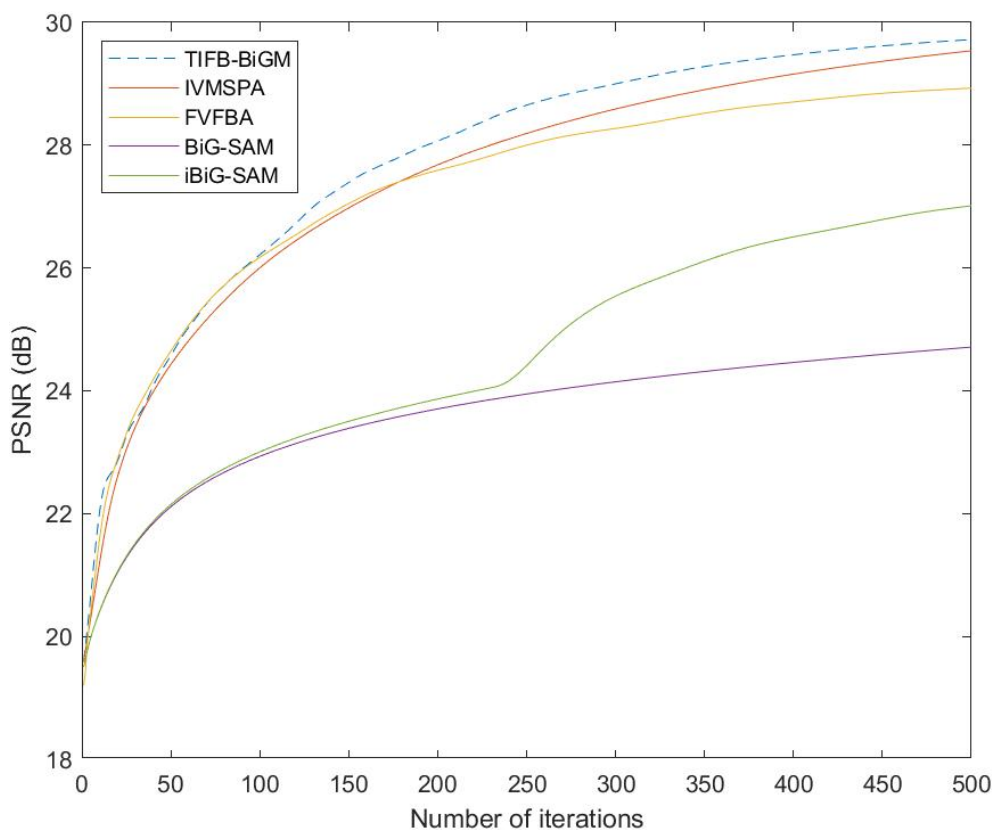


Figure 2. The graphs of PSNR of each algorithm for Wat Chedi Luang.

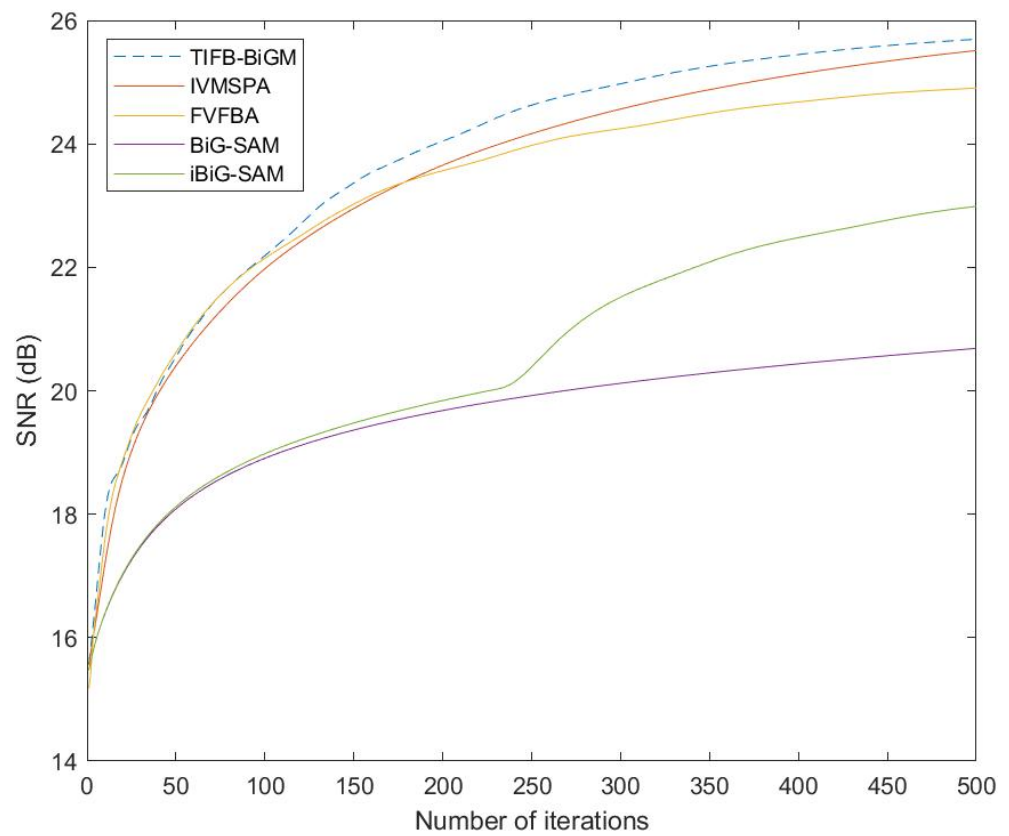


Figure 3. The graphs of SNR of each algorithm for Wat Chedi Luang.

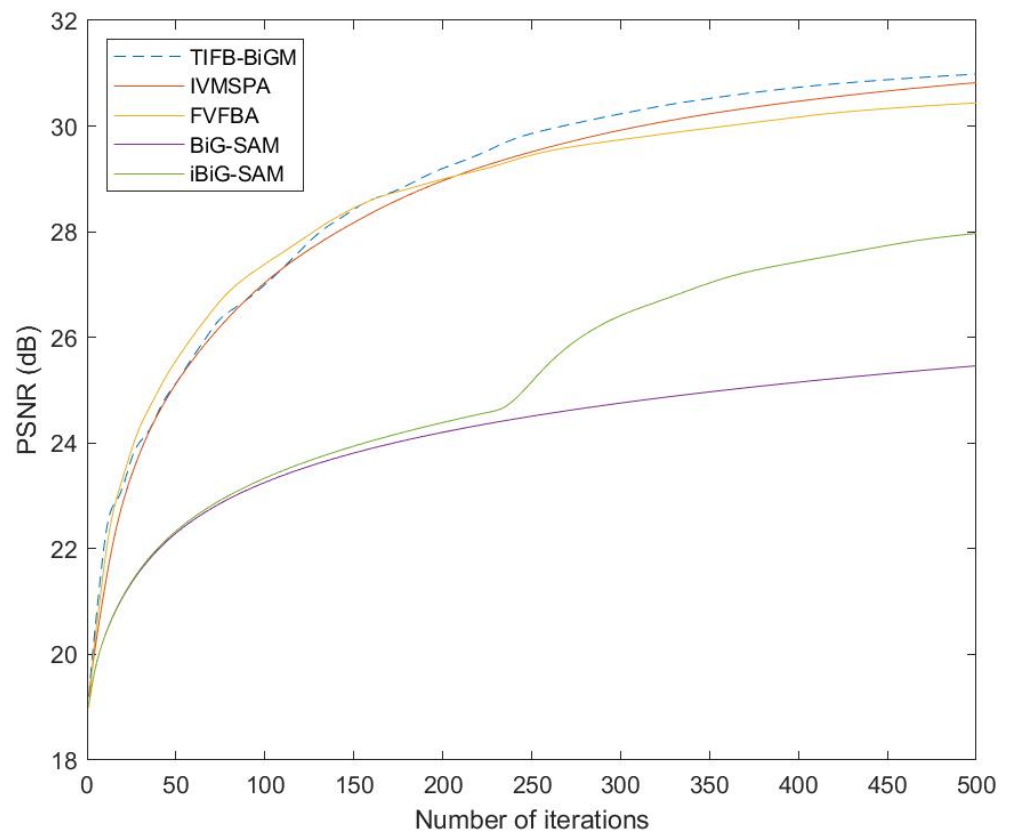


Figure 4. The graphs of PSNR of each algorithm for Matsue Castle.

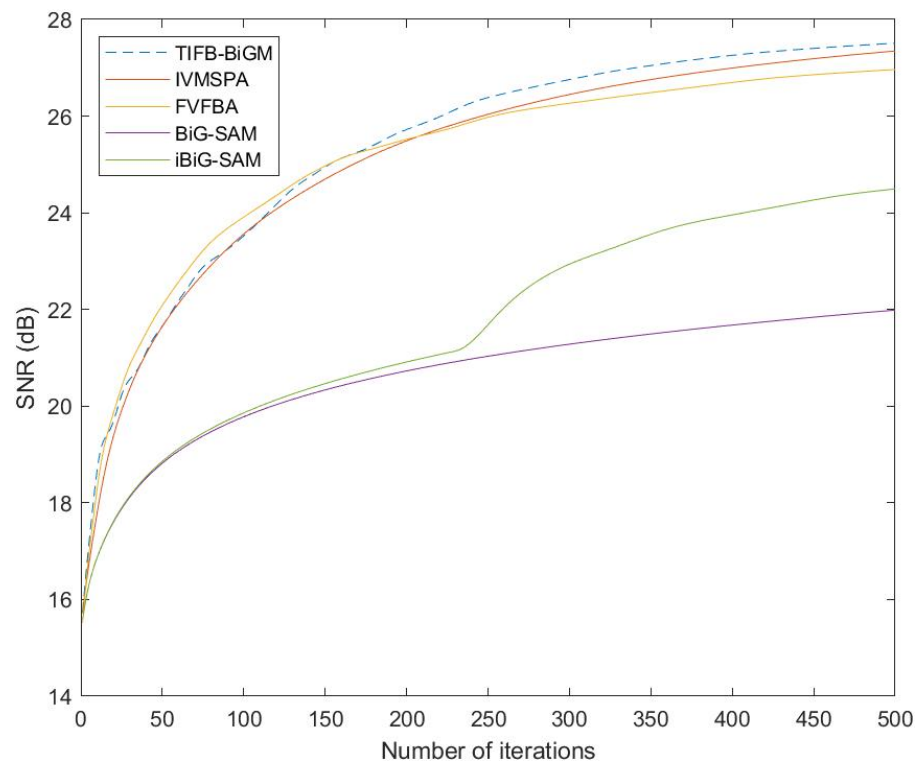


Figure 5. The graphs of SNR of each algorithm for Matsue Castle.



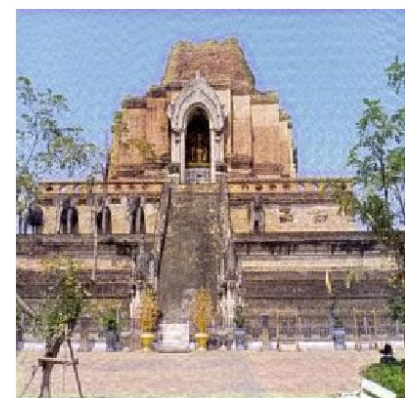
(a)



(b)



(c)

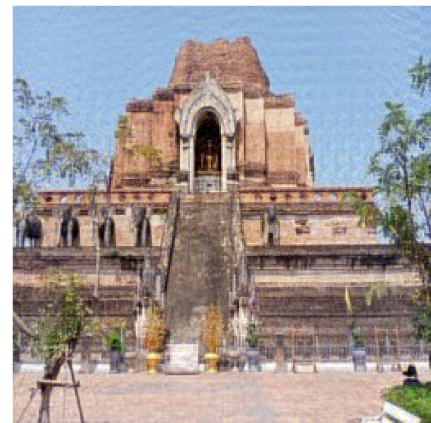


(d)

Figure 6. Cont.



(e)



(f)

Figure 6. Results for deblurring “Wat Chedi Luang” image using various algorithms at the 500th iteration. (a) Gaussian blurred image, (b) TIFB-BiGM (PSNR = 29.7216, SNR = 25.6962), (c) IVMSPA (PSNR = 29.5375, SNR = 25.5121), (d) FVFBA (PSNR = 28.9243, SNR = 24.8989), (e) BiG-SAM (PSNR = 24.7118, SNR = 20.6864), and (f) iBiG-SAM (PSNR = 27.0172, SNR = 22.9918).



(a)



(b)



(c)



(d)

Figure 7. Cont.



Figure 7. Results for deblurring “Matsue Castle” image using various algorithms at the 500th iteration. (a) Gaussian blurred image, (b) TIFB-BiGM (PSNR = 30.9830, SNR = 27.5075), (c) IVMSPA (PSNR = 30.8212, SNR = 27.3457), (d) FVFBA (PSNR = 30.43625, SNR = 26.9636), (e) BiG-SAM (PSNR = 25.4625, SNR = 21.9870), and (f) iBiG-SAM (PSNR = 27.9712, SNR = 24.4957).

5. Conclusions

In this paper, algorithmic solutions to a family of convex bilevel optimization problems are developed and applied to image processing. An interesting connection between minimization problems and fixed-point methods is observed. We first present a modified two-step inertial viscosity algorithm (MTIVA) for finding a common fixed point of a family of nonexpansive operators in a Hilbert space and prove strong convergence under relatively mild conditions. This is then applied to the solution of a convex bilevel optimization problem by introducing a novel two-step inertial forward–backward bilevel gradient method (TIFB-BiGM). The main results are then employed in the solution of an image restoration problem. Through careful comparative analysis, we demonstrate that our algorithm outperforms several existing algorithms such as IVMSPA, FVFBA, BiG-SAM, and iBiG-SAM, in terms of image recovery efficiency, as verified through numerical experiments conducted under specific parameter settings.

There are several potential avenues for future research. Firstly, investigating the adaptability and performance of the proposed algorithm in different image processing tasks could provide valuable insights. Additionally, one might explore the algorithm’s scalability to large-scale image datasets or investigate the incorporation of parallel computing techniques that could enhance the algorithm’s computational efficiency. Moreover, conducting comparative studies with other state-of-the-art image restoration algorithms would provide a comprehensive evaluation of the algorithm’s strengths and limitations. Finally, exploring the applicability of the proposed algorithm to other domains beyond image processing, such as computer vision or signal processing, would broaden its potential impact.

Author Contributions: Conceptualization, S.S.; formal analysis, R.W. and S.S.; investigation, R.W. and K.J.; methodology, R.W. and S.S.; software, K.J.; supervision, S.S.; validation, R.W. and S.S.; writing—original draft, R.W. and K.J.; and writing—review and editing, R.W. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: NSRF via the program Management Unit for Human Resources & Institutional Development, Research, and Innovation (grant number B05F640183).

Data Availability Statement: Not applicable.

Acknowledgments: This research has received funding support from the NSRF via the Program Management Unit for Human Resources and Institutional Development, Research, and Innovation (grant number B05F640183), and it was also partially supported by Chiang Mai University and Ubon Ratchathani University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Franceschi, L.; Frasconi, P.; Salzo, S.; Grazi, R.; Pontil, M. Bilevel programming for hyperparameter optimization and meta-learning. In Proceedings of the International Conference on Machine Learning (ICML), Stockholm, Sweden, 10–15 July 2018; pp. 1568–1577.
2. Shaban, A.; Cheng, C.-A.; Hatch, N.; Boots, B. Truncated back-propagation for bilevel optimization. In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS), Okinawa, Japan, 16–18 April 2019; pp. 1723–1732.
3. Kunapuli, G.; Bennett, K.P.; Hu, J.; Pang, J.-S. Classification model selection via bilevel programming. *Optim. Methods Softw.* **2008**, *23*, 475–489. [\[CrossRef\]](#)
4. Flamary, R.; Rakotomamonjy, A.; Gasso, G. Learning constrained task similarities in graph regularized multitask learning. In *Regularization, Optimization, Kernels, and Support Vector Machines*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2014; Volume 103, ISBN 978-0367658984.
5. Konda, V.R.; Tsitsiklis, J.N. Actor-critic algorithms. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Denver, CO, USA, 30 November–2 December 1999; pp. 1008–1014.
6. Bruck, R.E., Jr. On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space. *J. Math. Anal. Appl.* **1977**, *61*, 159–164. [\[CrossRef\]](#)
7. Lions, P.L.; Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **1979**, *16*, 964–979. [\[CrossRef\]](#)
8. Janngam, K.; Suantai, S. An inertial modified S-Algorithm for convex minimization problems with directed graphs and their applications in classification problems. *Mathematics* **2022**, *10*, 4442. [\[CrossRef\]](#)
9. Cabot A. Proximal point algorithm controlled by a slowly vanishing term: Applications to hierarchial minimization. *SIAM J. Optim.* **2005**, *15*, 555–572. [\[CrossRef\]](#)
10. Xu, H.K. Averaged mappings and the gradient-projection algorithm. *J. Optim. Theory Appl.* **2011**, *150*, 360–378. [\[CrossRef\]](#)
11. Passty, G.B. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.* **1979**, *72*, 383–390. [\[CrossRef\]](#)
12. Beck, A.; Sabach, S. A first order method for finding minimal norm-like solutions of convex optimization problems. *Math. Program.* **2014**, *147*, 25–46. [\[CrossRef\]](#)
13. Sabach, S.; Shtern, S. A first order method for solving convex bilevel optimization problems. *SIAM J. Optim.* **2017**, *27*, 640–660. [\[CrossRef\]](#)
14. Nesterov, Y.E. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Sov. Math. Dokl.* **1983**, *27*, 372–376.
15. Polyak, B.T. Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **1964**, *4*, 1–17. [\[CrossRef\]](#)
16. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [\[CrossRef\]](#)
17. Shehu, Y.; Vuong, P.T.; Zemkoho, A. An inertial extrapolation method for convex simple bilevel optimization. *Optim Methods Softw.* **2019**, *36*, 1–19. [\[CrossRef\]](#)
18. Moudafi, A. Viscosity approximation method for fixed-points problems. *J. Math. Anal. Appl.* **2000**, *241*, 46–55. [\[CrossRef\]](#)
19. Xu, H.K. Viscosity approximation methods for nonexpansive mappings. *J. Math. Anal. Appl.* **2004**, *298*, 279–291. [\[CrossRef\]](#)
20. Takahashi, W. Viscosity approximation methods for countable families of nonexpansive mappings in Banach spaces. *Nonlinear Anal.* **2009**, *70*, 719–734. [\[CrossRef\]](#)
21. Jailoka, P.; Suantai, S. and Hanjing, A. A fast viscosity forward–backward algorithm for convex minimization problems with an application in image recovery. *Carpathian J. Math.* **2021**, *37*, 449–461. [\[CrossRef\]](#)
22. Janngam, K.; Suantai, S.; Cho, Y.J.; Kaewkhao, A.; Wattanataweekul, R. A Novel Inertial Viscosity Algorithm for Bilevel Optimization Problems Applied to Classification Problems. *Mathematics* **2023**, *11*, 3241. [\[CrossRef\]](#)
23. Poon, C.; Liang, J. Geometry of First-order Methods and Adaptive Acceleration. *arXiv* **2020**, arXiv:2003.03910.
24. Liang, J. Convergence Rates of First-Order Operator Splitting Methods. Ph.D. Thesis, Normandie Universit’e, Normandie, France, 2016.
25. Poon, C.; Liang, J. Trajectory of Alternating Direction Method of Multiplier and Adaptive Acceleration. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019.
26. Polyak, B.T. *Introduction to Optimization*; Optimization Software, Publication Division: New York, NY, USA, 1987.
27. Combettes, P.L.; Glaudin, L. Quasi-Nonexpansive Iterations on the Affine Hull of Orbits: From Mann’s Mean Value Algorithm to Inertial Methods. *SIAM J. Optim.* **2017**, *27*, 2356–2380. [\[CrossRef\]](#)
28. Nakajo, K.; Shimoji, K.; Takahashi, W. On strong convergence by the hybrid method for families of mappings in Hilbert spaces. *Nonlinear Anal.* **2009**, *71*, 112–119. [\[CrossRef\]](#)
29. Aoyama, K.; Kimura, Y. Strong convergence theorems for strongly nonexpansive sequences. *Appl. Math. Comput.* **2011**, *217*, 7537–7545. [\[CrossRef\]](#)

30. Aoyama, K.; Kohsaka, F.; Takahashi, W. Strong convergence theorems by shrinking and hybrid projection methods for relatively nonexpansive mappings in Banach spaces. *Nonlinear Anal. Convex Anal.* **2009**, *10*, 7–26.
31. Moreau, J.J. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes Rendus Acad. Sci. Paris Ser. A Math.* **1962**, *255*, 2897–2899.
32. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*; Springer: New York, NY, USA, 2011.
33. Bussaban, L.; Suantai, S.; Kaewkhao, A. A parallel inertial S-iteration forward–backward algorithm for regression and classification problems. *Carpathian J. Math.* **2020**, *36*, 35–44. [[CrossRef](#)]
34. Takahashi, W. *Introduction to Nonlinear and Convex Analysis*; Yokohama Publishers: Yokohama, Japan, 2009.
35. Saejung, S.; Yotkaew, P. Approximation of zeros of inverse strongly monotone operators in Banach spaces. *Nonlinear Anal.* **2012**, *75*, 724–750. [[CrossRef](#)]
36. Maurya, A.; Tiwari, R. A Novel Method of Image Restoration by using Different Types of Filtering Techniques. *Int. J. Eng. Sci. Innov. Technol.* **2014**, *3*, 124–129.
37. Suseela, G.; Basha, S.A.; Babu, K.P. Image Restoration Using Lucy Richardson Algorithm For X-Ray Images. *IJISSET Int. J. Innov.Sci. Eng. Technol.* **2016**, *3*, 280–285.
38. Vogel, C.R. *Computational Methods for Inverse Problems*; SIAM: Philadelphia, PA, USA, 2002.
39. Eldén, L. Algorithms for the Regularization of Ill-Conditioned Least Squares Problems. *BIT Numer. Math.* **1977**, *17*, 134–145. [[CrossRef](#)]
40. Hansen, P.C.; Nagy, J.G.; O’Leary, D.P. *Deblurring Images: Matrices, Spectra, and Filtering (Fundamentals of Algorithms 3) (Fundamentals of Algorithms)*; SIAM: Philadelphia, PA, USA, 2006.
41. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
42. Yatakoat, P.; Suantai, S.; Hanjing, A. On Some Accelerated Optimization Algorithms Based on Fixed Point and Linesearch Techniques for Convex Minimization Problems with Applications. *Adv. Cont. Discr. Mod.* **2022**, *2022*, 43:1–43:13. [[CrossRef](#)]
43. Thung, K.; Raveendran, P. A survey of image quality measures. In Proceedings of the 2009 International Conference for Technical Postgraduates (TECHPOS), Kuala Lumpur, Malaysia, 14–15 December 2009; pp. 1–4.
44. Chen, D.Q.; Zhang, H.; Cheng, L.Z. A fast fixed-point algorithmfixed-point algorithmfixed-point algorithmfixed-point algorithm for total variation deblurring and segmentation. *J. Math. Imaging Vis.* **2012**, *43*, 167–179. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.