

https://dspace.org/

Inertial Stochastic Reflected Forward Backward Method with Applications to Traffic Network Problems

Chinedu Izuchukwu; T Alakoya; S Moutari; S Zeng

SPRINGER/PLENUM PUBLISHERS
https://hdl.handle.net/10539/47582
WOS

Downloaded from a DSpace repository. https://dspace.org/

Journal of Optimization Theory and Applications (2025) 207:23
https://doi.org/10.1007/s10957-025-02779-1

Inertial Stochastic Reflected Forward Backward Method with Applications to Traffic Network Problems

Chinedu Izuchukwu[1] · Timilehin Opeyemi Alakoya[2] · Salissou Moutari[2] · Shengda Zeng[3]

Received: 3 March 2025 / Accepted: 30 June 2025
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

Abstract
This paper introduces a new inertial stochastic reflected-forward-backward splitting method aimed at addressing monotone inclusion problems, specifically involving a maximal monotone set-valued operator and a single-valued Lipschitz continuous and monotone operator within a real separable Hilbert space. Distinct from many existing inertial splitting approaches, this algorithm uniquely depends on one unbiased estimate of the monotone Lipschitz continuous operator and a single backward computation of the maximal monotone operator per iteration. We establish a convergence rate of $\mathcal{O}(\log(i)/(i))$ in expectation for a case of strong monotonicity, and almost sure convergence for a general monotone scenario. Furthermore, we examine its application to traffic flow networks.

Keywords Reflected-forward-backward algorithm · Stochastic optimization · Inertial method · Monotone inclusion · Traffic flow network

Dedicated to Prof. Q.H. Ansari on the occasion of his 65th birthday.

Communicated by Jen-Chih Yao.

✉ Shengda Zeng
zengshengda@163.com

Chinedu Izuchukwu
chinedu.izuchukwu@wits.ac.za ; izuchukwu_c@yahoo.com

Timilehin Opeyemi Alakoya
t.alakoya@qub.ac.uk ; timimaths@gmail.com

Salissou Moutari
s.moutari@qub.ac.uk

1 School of Mathematics, University of the Witwatersrand, Private Bag 3, Johannesburg 2050, South Africa

2 School of Mathematics and Physics, Queen's University Belfast, Belfast, United Kingdom

3 Guangxi Colleges and Universities Key Laboratory of Complex System Optimization and Big Data Processing, Yulin Normal University, Yulin 537000, Guangxi, P.R. China

Published online: 21 July 2025

Mathematics Subject Classification 58E35 · 47N10 · 49J53 · 90C25 · 90C33

1 Introduction

Consider a real separable Hilbert space denoted by $\mathcal{V}$, with two monotone operators $\mathcal{J} : \mathcal{V} \to 2^{\mathcal{V}}$ and $\mathcal{K} : \mathcal{V} \to \mathcal{V}$. The task of searching for a zero of the sum of these operators is termed monotone inclusion problem, which can be expressed as follows:

$$\text{search for } e \in \mathcal{V} \text{ such that } 0 \in (\mathcal{J} + \mathcal{K})e. \tag{1}$$

We use $(\mathcal{J} + \mathcal{K})^{-1}(0)$ to represent the set of solutions to problem (1). This problem is crucial in optimization, signal processing, image recovery, machine learning, obstacle problems, and traffic network problems [27, 47, 48]. In this study, our focus lies on traffic network issues. Traffic congestion remains a significant challenge in many cities around the globe. Data-driven trends indicate that congestion in major cities is worsening year by year, as evidenced by navigation system-based statistics ('https://www.tomtom.com/traffic-index/ranking/, accessed on 03 July 2024, Copyright © 2024 TomTom International BV, all rights reserved'). Given the growing problem of urban congestion mentioned above, cities urgently need effective tools for forecasting and simulating traffic in order to devise solutions. Urban planners and transportation engineers develop various traffic management strategies to address these challenges. One approach to assessing potential projects and analyzing existing road network usage is through mathematical modeling. Traffic modeling plays a crucial role in helping road network authorities better manage infrastructure by directing investments toward new projects, streamlining traffic movement, alleviating congestion, and enabling the quick detection and response to accidents or other incidents. As noted by Ortúzar and Willumsen [35], transportation planning tools include a wide range of mathematical models that are created for tackling complex challenges. This study specifically focuses on traffic allocations, also known as the traffic equilibrium problem. Traffic assignment is particularly significant in traffic engineering, as it plays a critical role in the simulation of traffic and the prediction of transportation demand. The traffic allocation model captures the travel behavior of road users, specifically predicting the route choices each user will make over a given period. The continuous expansion of road networks and the demand for highly accurate solutions have driven researchers to propose numerous methods for solving this problem. We will be applying our results to study and solve traffic equilibrium problems.

Splitting algorithms often address problem (1) by separately addressing each operator ($\mathcal{J}$ and $\mathcal{K}$) through computing the single-valued operator in the forward step and the set-valued operator in the backward step, instead of their sum. Extensive research has aimed at developing faster, more computationally efficient, and broadly applicable methods. One such method is the forward-backward splitting algorithm [22, 36]:

$$m_{i+1} = (I_{\mathcal{V}} + \sigma \mathcal{J})^{-1}(m_i - \sigma \mathcal{K} m_i), \ i \geq 1, \tag{2}$$

where $I_{\mathcal{V}}$ represents the identity operator on $\mathcal{V}$, and $\sigma > 0$. Algorithm (2) performs one forward calculation involving $\mathcal{K}$ and one backward computation involving $\mathcal{J}$ in each iteration. It exhibits weak convergence to an element of $(\mathcal{J}+\mathcal{K})^{-1}(0)$, provided that $\mathcal{K}$ is $\mathcal{L}^{-1}$-cocoercive (a strict condition), $\sigma \in (0, 2\mathcal{L}^{-1})$, $\mathcal{J}$ is a maximal monotone operator, and $(\mathcal{J}+\mathcal{K})^{-1}(0) \neq \emptyset$.
The strict cocoercivity requirement in Algorithm (2) was removed in subsequent research, leading to the following algorithm, named the forward-backward-forward splitting algorithm, alternatively, Tseng's splitting algorithm [43]:

$$\begin{aligned} n_i &= (I_{\mathcal{V}} + \sigma\mathcal{J})^{-1}(m_i - \sigma\mathcal{K}m_i), \\ m_{i+1} &= n_i - \sigma\mathcal{K}n_i + \sigma\mathcal{K}m_i, \ i \geq 1. \end{aligned} \tag{3}$$

Variants of Algorithm (3), such as inexact and variable metric versions, were also studied in [5] and [45], respectively. This algorithm was shown to exhibit weak convergence to a point in $(\mathcal{J}+\mathcal{K})^{-1}(0)$ under a lenient condition that $\mathcal{K}$ is monotone and $\mathcal{L}$-Lipschitz continuous, with $\sigma \in (0, \mathcal{L}^{-1})$, $\mathcal{J}$ being maximal monotone, and the set $(\mathcal{J}+\mathcal{K})^{-1}(0)$ being nonempty. The primary limitation is the need for two forward calculations of $\mathcal{K}$, which can reduce the algorithm's efficacy, specifically in applications where operator computations are costly [24].

To address this drawback, a forward-reflected-backward splitting algorithm was introduced [28]:

$$m_{i+1} = (I_{\mathcal{V}} + \sigma\mathcal{J})^{-1}(m_i - 2\sigma\mathcal{K}m_i + \sigma\mathcal{K}m_{i-1}), \ i \geq 1, \tag{4}$$

where $\sigma \in (0, \frac{1}{2}\mathcal{L}^{-1})$. This algorithm's primary advantages are the requirement of only a single forward calculation of $\mathcal{K}$, while also assuming that $\mathcal{K}$ is monotone Lipschitz continuous. For the reflexive Banach space version of this algorithm, refer to [18]. If $\mathcal{K}$ is linear, Algorithm (4) and the following algorithm discussed in [7], named the reflected-forward-backward splitting algorithm, have the same structure:

$$\begin{aligned} n_i &= 2m_i - m_{i-1} \\ m_{i+1} &= (I_{\mathcal{V}} + \sigma\mathcal{J})^{-1}(m_i - \sigma\mathcal{K}n_i), \ i \geq 1. \end{aligned} \tag{5}$$

Algorithm (5) was proved to converge weakly to an element of $(\mathcal{J}+\mathcal{K})^{-1}(0)$ provided $\mathcal{K}$ is monotone and $\mathcal{L}$-Lipschitz continuous, with σ chosen from the range $(0, \mathcal{L}^{-1}(\sqrt{2}-1))$, $\mathcal{J}$ is a maximal monotone operator, and $(\mathcal{J}+\mathcal{K})^{-1}(0) \neq \emptyset$. Algorithm (5) has also been specifically analyzed in cases where $\mathcal{J}$ in (1) is the normal cone associated with a nonempty, closed, and convex subset of $\mathcal{V}$ (refer to [26]). Furthermore, when $\mathcal{J}$ represents the subdifferential of a convex function, which is also proper and lower semicontinuous, a line-search variant of Algorithm (5) was investigated in [25] (see also [6] for a recent modification in this setting).
It is noteworthy that the forward-backward splitting algorithm (2), the forward-backward-forward splitting algorithm (3), the forward-reflected-backward splitting algorithm (4), and the reflected-forward-backward splitting algorithm (5), along

with their respective variants, are all constrained to deterministic frameworks. The stochastic versions of the forward-backward splitting algorithm were analyzed in [9, 10, 39]. However, they are limited to applications involving cocoercive operators. Also, the stochastic variant of the forward-backward-forward splitting algorithm was explored by the author in [44] and more recently in [11], both requiring two stochastic approximations of $\mathcal{K}$, which are costly. A stochastic improvement of the forward-reflected-backward splitting algorithm was examined in [28] for cases where $\mathcal{K}$ is a finite sum. The drawback here is evaluating $\mathcal{K}m_i$ instead of $\mathcal{K}_{j_i}m_i$ (where j_i is chosen uniformly at random from $\{1, ..., k\}$, see [28, Section 6] for details), making its computational cost similar to that of (4). Recently, a stochastic variant of the reflected-forward-backward splitting algorithm was presented and analyzed by the authors in [34], where convergence rate and almost sure convergence were obtained.

In recent years, fast converging optimization problem-solving algorithms have garnered significant interest. The inertial extrapolation method, inspired by a discrete analogy to a second-order dissipative dynamical system, has proven to be effective in accelerating the convergence of optimization problem solving algorithms (see [17, 32, 37, 42]). Inertial algorithms provide distinct advantages over traditional optimization methods, particularly for high-dimensional and large-scale problems [32, 37]. For instance, Nesterov's Accelerated Gradient (NAG) method often outperforms standard gradient descent in deep learning tasks, yielding faster training and improved model performance [14, 32]. A key strength of inertial algorithms is their capacity to accelerate convergence, thereby reducing iteration counts, which is crucial when resources or time are limited. Their scalability makes them especially effective in various fields, such as machine learning [40], data science [14], and signal processing, where handling massive datasets and complex models efficiently is essential [14]. For notable results on inertial algorithms for addressing problem (1), see [2, 3, 8, 12, 23, 30] and the sources cited therein.

Building on the discussion above, this paper introduces and investigates an inertial stochastic modification of the reflected-forward-backward splitting algorithm (5) within a real separable Hilbert space. Specifically, we propose and analyze an inertial stochastic reflected-forward-backward splitting algorithm, and achieve a convergence rate of $\mathcal{O}(\log(i)/i)$ in expectation for the strongly monotone case, and almost sure convergence for the general monotone case. We bypass the limitations of [11, 28, 44], by considering an inertial stochastic improvement of (5) that uses a single unbiased estimation of $\mathcal{K}(2m_i - m_{i-1})$, aligning with the framework of the standard stochastic forward–backward splitting algorithms in [9, 10, 39], but capable of addressing a broader range of challenges associated with non-cocoercive operators; a unique feature among inertial splitting algorithms. Furthermore, we explore challenges in traffic flow networks and show how our algorithm can be utilized to find solutions to these problems.

2 Preliminaries

We use $\langle \cdot, \cdot \rangle$ to represent the inner product and $\| \cdot \|$ to represent the norm in a separable real Hilbert space $\mathcal{V}$. Also, for a sequence $\{m_i\}$ in $\mathcal{V}$ and a point e in $\mathcal{V}$, we represent

the strong convergence of $\{m_i\}$ to e by $m_i \to e$ and the weak convergence of $\{m_i\}$ to e by $m_i \rightharpoonup e$.
The operator $\mathcal{K} : \mathcal{V} \to \mathcal{V}$ is referred to as $\mathcal{L}$-cocoercive if there is a constant $\mathcal{L} > 0$ such that

$$\langle \mathcal{K}e - \mathcal{K}c, e - c\rangle \geq \mathcal{L}\|\mathcal{K}e - \mathcal{K}c\|^2 \ \forall e, c \in \mathcal{V},$$

and monotone if

$$\langle \mathcal{K}e - \mathcal{K}c, e - c\rangle \geq 0 \ \ \forall e, c \in \mathcal{V}.$$

The operator $\mathcal{K}$ is referred to as an $\mathcal{L}$-Lipschitz continuous operator if there is $\mathcal{L} > 0$ such that

$$\|\mathcal{K}e - \mathcal{K}c\| \leq \mathcal{L}\|e - c\| \ \forall e, c \in \mathcal{V}.$$

Let $\mathcal{J}$ be a set-valued operator $\mathcal{J} : \mathcal{V} \to 2^{\mathcal{V}}$, $\mathcal{J}$ is referred to as a monotone operator if

$$\langle f - g, e - c\rangle \geq 0 \ \ \forall e, c \in \mathcal{V}, \ \ f \in \mathcal{J}e, \ \ g \in \mathcal{J}c,$$

and strongly monotone if there is $\epsilon > 0$ such that

$$\langle f - g, e - c\rangle \geq \epsilon\|e - c\|^2 \ \ \forall e, c \in \mathcal{V}, \ \ g \in \mathcal{J}e, \ \ g \in \mathcal{J}c.$$

$\mathcal{J}$ is termed a maximal monotone operator if the graph of $\mathcal{J}$, $\mathcal{G}(\mathcal{J})$ given as

$$\mathcal{G}(\mathcal{J}) := \{(e, f) \in \mathcal{V} \times \mathcal{V} : f \in \mathcal{J}e\},$$

is not a proper subset of the graph of any other monotone operator. Specifically, $\mathcal{J}$ is termed maximal monotone if, for $(e, f) \in \mathcal{V} \times \mathcal{V}$, we have that $\langle f - g, e - c\rangle \geq 0$ for all $(c, g) \in \mathcal{G}(\mathcal{J})$ implies $f \in \mathcal{J}e$. Associated with a set-valued operator $\mathcal{J}$ is the resolvent, which is defined as the mapping $\text{Res}_{\sigma}^{\mathcal{J}} : \mathcal{V} \to 2^{\mathcal{V}}$ given by

$$\text{Res}_{\sigma}^{\mathcal{J}}(e) := (I_{\mathcal{V}} + \sigma\mathcal{J})^{-1}(e), \ \ e \in \mathcal{V}, \ \sigma > 0.$$

When the monotone operator $\mathcal{J}$ is maximal and $\mathcal{K}$ is single-valued, both $\text{Res}_{\sigma}^{\mathcal{J}}$ and $\text{Res}_{\sigma}^{\mathcal{J}}(I_{\mathcal{V}} - \sigma\mathcal{K})$ are guaranteed to be single-valued and are defined everywhere in the entire space $\mathcal{V}$.
As described in [21], let $(\Omega, \mathcal{F}, P)$ represent a probability space. A $\mathcal{V}$-valued random variable taking values in the Hilbert space $\mathcal{V}$ is defined as a measurable function $\mathcal{X} : \Omega \to \mathcal{V}$, with $\mathcal{V}$ equipped with the Borel $\widehat{\sigma}$-algebra. The $\widehat{\sigma}$-field generated by $\mathcal{X}$ is denoted by $\widehat{\sigma}(\mathcal{X})$. The expectation of a random variable $\mathcal{X}$ is represented by $\mathbb{E}[\mathcal{X}]$, and the conditional expectation of $\mathcal{X}$ denoted as a $\widehat{\sigma}$-field $\mathcal{A} \subset \mathcal{F}$, is represented by $\mathbb{E}[\mathcal{X} \mid \mathcal{A}]$. A random process in $\mathcal{V}$ is described as a sequence $\{m_i\}$ of $\mathcal{V}$-valued random variables. The term a.s. is used as an abbreviation for 'almost surely'. A random variable (or stochastic process) $\mathcal{X}$ taking values in a Hilbert space $\mathcal{V}$ is $\mathcal{V}$-measurable

if it is measurable as a function $\mathcal{X} : (\Omega, \mathcal{F}) \to (\mathcal{V}, \mathcal{B}(\mathcal{V}))$, i.e., for every Borel set $\mathcal{B} \subset \mathcal{V}$, the event $\omega : \mathcal{X}(\omega) \in \mathcal{B}$ lies in the $\widehat{\sigma}$-algebra $\mathcal{V}$. For an integrable random variable $\mathcal{X}$ and a sub-$\widehat{\sigma}$-algebra, the conditional expectation $\mathbb{E}[\mathcal{X} \mid \mathcal{F}]$ is the unique $\mathcal{F}$-measurable random variable satisfying $\mathbb{E}[\mathbb{E}[\mathcal{X} \mid \mathcal{F}]\mathbf{1}_{\mathcal{A}}] = \mathbb{E}[\mathcal{X}\mathbf{1}_{\mathcal{A}}]$ for all $\mathcal{A} \in \mathcal{F}$. It is linear in $\mathcal{X}$. A sequence of random vectors $\{m_i\}$ in a (separable) Hilbert space converges almost surely to a limit x if $P(\omega : \lim_{i\to\infty} \|m_i(\omega) - x(\omega)\| = 0) = 1$. Equivalently, with probability one the sample-paths satisfy $\|m_i - x\| \to 0$ as $i \to \infty$. A sequence $\{m_i\}$ is bounded almost surely if there exists a finite constant M such that $P(\omega : \sup_i \|m_i(\omega)\| \leq M) = 1$. In other words, with probability 1 the norms $\|m_i\|$ remain uniformly bounded.

Lemma 2.1 ([38], Theorem 1) *Consider an increasing sequence of sub-$\widehat{\sigma}$-algebras $\{\mathcal{F}_i\}$ of $\mathcal{F}$, and the $[0, +\infty)$-valued random sequences $\{a_i\}$, $\{b_i\}$, $\{n_i\}$ and $\{d_i\}$ so that, for each $i \geq 1$, a_i, b_i, n_i and d_i are $\mathcal{F}_i$-measurable. Assume moreover that $\sum_{i=1}^{\infty} d_i < +\infty$, $\sum_{i=1}^{\infty} n_i < +\infty$ a.s. and for all $i \geq 1$,*

$$\mathbb{E}[a_{i+1} \mid \mathcal{F}_i] \leq (1 + d_i)a_i + n_i - b_i \ \ a.s..$$

Then $\lim_{i\to\infty} a_i$ exists a.s. and $\sum_{i=1}^{\infty} b_i < +\infty$ a.s..

Lemma 2.2 ([34], Corollary 2.6) *Consider an increasing sequence of sub-$\widehat{\sigma}$-algebras $\{\mathcal{F}_i\}$ of $\mathcal{F}$, and the $[0, +\infty)$-valued random sequence $\{m_i\}$ such that for each $i \geq 1$, m_{i-1} is $\mathcal{F}_i$-measurable and*

$$\sum_{i=1}^{\infty} \mathbb{E}[m_i \mid \mathcal{F}_i] < +\infty \ \ a.s..$$

Then

$$\sum_{i=1}^{\infty} m_i < +\infty \ \ a.s..$$

Lemma 2.3 ([34]) *Consider a nonempty closed subset $\mathcal{Q}$ of $\mathcal{V}$ and a $\mathcal{V}$-valued random process $\{m_i\}$. Suppose that $\{\|m_i - e\|\}$ converges a.s. for all $e \in \mathcal{Q}$ and the set of weak sequential cluster points of $\{m_i\}$ is a subset of $\mathcal{Q}$ a.s.. Then $\{m_i\}$ converges weakly a.s. to a $\mathcal{Q}$-valued random vector.*

The following identities holds:

$$2\langle q, j\rangle = \|q\|^2 + \|j\|^2 - \|q - j\|^2 = \|q + j\|^2 - \|q\|^2 - \|j\|^2, \forall q, j \in \mathcal{V} \quad (6)$$

and

$$\|\bar{\xi}\, q + (1 - \bar{\xi})j\|^2 = \bar{\xi}\|q\|^2 + (1 - \bar{\xi})\|j\|^2 - \bar{\xi}(1 - \bar{\xi})\|q - j\|^2, \ \forall q, j \in \mathcal{V}, \ \bar{\xi} \in \mathbb{R}. \quad (7)$$

3 Proposed Algorithm

We begin this section by stating the assumptions for our convergence analysis. Then we provide and discuss our newly introduced algorithm.

Assumption 3.1

(a) The operator $\mathcal{J}$ is maximal monotone.
(b) The operator $\mathcal{K}$ is both monotone and $\mathcal{L}$-Lipschitz continuous.
(c) The solution set $(\mathcal{J}+\mathcal{K})^{-1}(0) \neq \emptyset$.

Algorithm 3.2 Inertial Stochastic Reflected-Forward-Backward Splitting Algorithm.

Let $\sigma > 0$ and $\varrho \in [0, 1)$. For arbitrary $m_0, m_1 \in \mathcal{V}$, let the sequence $\{m_i\}$ be generated by

$$\begin{cases} w_i = m_i + \varrho(m_i - m_{i-1}) \\ n_i = 2m_i - m_{i-1} \\ \text{Find } q_i \ : \mathbb{E}[\, q_i \mid \mathcal{F}_i] = \mathcal{K}n_i \text{ a.s.} \\ m_{i+1} = \text{Res}_{\sigma}^{\mathcal{J}}(w_i - \sigma q_i), \ i \geq 1, \end{cases}$$

where $\mathcal{F}_i = \widehat{\sigma}(m_0, m_1, m_2, \ldots, m_i)$.

Remark 3.1 Algorithm 3.2 is an inertial stochastic variant of the reflected-forward-backward splitting algorithm (5) studied in [7]. Also, it extends the inertial-type algorithms in [4, 8, 13, 17, 19, 23, 30, 41] from the deterministic settings to stochastic settings. Unlike in these papers, where the operator $\mathcal{K}$ is computed in deterministic sense, Algorithm 3.2 requires a single unbiased estimation of the operator $\mathcal{K}$ in each iteration.
When $\varrho = 0$, the structure of Algorithm 3.2 aligns with the structure of the standard stochastic forward-backward splitting algorithms in [9, 10, 39] but unlike in these papers, Algorithm 3.2 is capable of addressing a broader range of challenges associated with non-cocoercive operators (see [17] for examples where $\mathcal{K}$ is monotone Lipschitz continuous but not cocoercive). Thus, our algorithm has unique features among inertial splitting algorithms.

4 Convergence Analysis

We now provide the weak and rate of convergence of our newly introduced algorithm.

4.1 Weak Convergence

Lemma 4.1 *Let Algorithm 3.2 generate $\{m_i\}$ and let the conditions outlined in Assumption 3.1 be met. For all $e \in (\mathcal{J}+\mathcal{K})^{-1}(0)$, define*

$$\begin{aligned} a_i := &\|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + (1 + 2\varrho + \sigma\mathcal{L}(1+\sqrt{2}))\|m_i - m_{i-1}\|^2 \\ &+(\mathcal{L}\sigma + \varrho)\|m_i - n_{i-1}\|^2 \\ &+2\sigma\langle q_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle + \sigma\frac{1}{s}\|q_{i-1} - \mathcal{K}n_{i-1}\|^2, \end{aligned} \tag{8}$$

where $s \in (0, \frac{1+\sigma\mathcal{L}(\sqrt{2}-1)}{2})$.
Then

$$\begin{aligned} a_{i+1} \le a_i + 2\sigma\beta_i - \left[1 - (\sigma\mathcal{L} + \varrho) - 2s\sigma - \sqrt{2}\mathcal{L}\sigma - 2\varrho\right]\|m_{i+1} - n_i\|^2 \\ -(1 - 6\varrho - \sigma\mathcal{L}(1+\sqrt{2}))\|m_{i+1} - m_i\|^2 \\ +2\sigma\frac{1}{s}\|q_i - \mathcal{K}n_i\|^2, \end{aligned}$$

where $\beta_i = \langle q_i - \mathcal{K}n_i, e - n_i\rangle$.

Proof Let $e \in (\mathcal{J} + \mathcal{K})^{-1}(0)$ and

$$p_i := \frac{1}{\sigma}(w_{i-1} - m_i) - q_{i-1}. \tag{9}$$

From $m_i = \mathrm{Res}_\sigma^{\mathcal{J}}(w_{i-1} - \sigma q_{i-1})$, the definition of $\mathrm{Res}_\sigma^{\mathcal{J}}$ and (9), we get $p_{i+1} \in \mathcal{J}m_{i+1}$. Since $\mathcal{J}$ is monotone and $-\mathcal{K}e \in \mathcal{J}e$, we get

$$\langle p_{i+1} + \mathcal{K}e, m_{i+1} - e\rangle \ge 0.$$

That is,

$$\langle\frac{1}{\sigma}(w_i - m_{i+1}) - q_i + \mathcal{K}e, m_{i+1} - e\rangle \ge 0. \tag{10}$$

This implies

$$\langle w_i - m_{i+1}, m_{i+1} - e\rangle \ge \sigma\langle q_i - \mathcal{K}e, m_{i+1} - e\rangle. \tag{11}$$

Now, using $n_i = 2m_i - m_{i-1}$, we have

$$\begin{aligned} \langle q_i - \mathcal{K}e, m_{i+1} - e\rangle &= \langle q_i - \mathcal{K}e, m_{i+1} - n_i\rangle + \langle q_i - \mathcal{K}n_i, n_i - e\rangle \\ &\quad +\langle\mathcal{K}n_i - \mathcal{K}e, n_i - e\rangle \\ &= \langle q_i - \mathcal{K}e, m_{i+1} - m_i\rangle - \langle q_i - \mathcal{K}e, m_i - m_{i-1}\rangle \\ &\quad +\langle q_i - \mathcal{K}n_i, n_i - e\rangle \\ &\quad +\langle\mathcal{K}n_i - \mathcal{K}e, n_i - e\rangle \\ &= \langle q_i - \mathcal{K}e, m_{i+1} - m_i\rangle - \langle q_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle \end{aligned}$$

$$
\begin{aligned}
&+\langle q_{i-1}-q_i, m_i-m_{i-1}\rangle \\
&+\langle q_i-\mathcal{K}n_i, n_i-e\rangle+\langle \mathcal{K}n_i-\mathcal{K}e, n_i-e\rangle \\
=\ &\langle q_i-\mathcal{K}e, m_{i+1}-m_i\rangle-\langle q_{i-1}-\mathcal{K}e, m_i-m_{i-1}\rangle \\
&+\langle q_i-q_{i-1}, m_{i+1}-n_i\rangle \\
&+\langle q_{i-1}-q_i, m_{i+1}-m_i\rangle+\langle q_i-\mathcal{K}n_i, n_i-e\rangle \\
&+\langle \mathcal{K}n_i-\mathcal{K}e, n_i-e\rangle. \qquad (12)
\end{aligned}
$$

From the monotonicity of $\mathcal{J}$ and (9), we obtain

$$
\langle \frac{w_i-m_{i+1}}{\sigma}-q_i-\frac{w_{i-1}-m_i}{\sigma}+q_{i-1}, m_{i+1}-m_i\rangle \geq 0, \qquad (13)
$$

which implies that

$$
\begin{aligned}
\langle q_{i-1}-q_i, m_{i+1}-m_i\rangle \geq\ &\frac{1}{\sigma}\langle m_{i+1}-w_i, m_{i+1}-m_i\rangle \\
&+\frac{1}{\sigma}\langle w_{i-1}-m_i, m_{i+1}-m_i\rangle. \qquad (14)
\end{aligned}
$$

Also, from $n_i=2m_i-m_{i-1}$ and $w_i=m_i+\varrho(m_i-m_{i-1})$, we obtain

$$
\begin{aligned}
m_{i+1}-w_i &= m_{i+1}-m_i-\varrho(m_i-m_{i-1}) \\
&= m_{i+1}-m_i-\varrho(n_i-m_i) \\
&= (1-\varrho)(m_{i+1}-m_i)+\varrho(m_{i+1}-n_i) \\
&= (1-\varrho)(n_{i+1}-m_{i+1})+\varrho(m_{i+1}-n_i). \qquad (15)
\end{aligned}
$$

Using (15) in (14), we obtain

$$
\begin{aligned}
\langle q_{i-1}-q_i, m_{i+1}-m_i\rangle \geq\ &\frac{1}{\sigma}\big[(1-\varrho)\langle m_{i+1}-m_i, m_{i+1}-m_i\rangle \\
&+\varrho\langle m_{i+1}-n_i, m_{i+1}-m_i\rangle\big] \\
&-\frac{1}{\sigma}\big[(1-\varrho)\langle n_i-m_i, m_{i+1}-m_i\rangle \\
&+\varrho\langle m_i-n_{i-1}, m_{i+1}-m_i\rangle\big]. \qquad (16)
\end{aligned}
$$

By (6), we have

$$
2\langle m_i-n_i, m_{i+1}-m_i\rangle=\|m_{i+1}-n_i\|^2-\|m_i-n_i\|^2-\|m_{i+1}-m_i\|^2 \qquad (17)
$$

and

$$
2\langle w_i-m_{i+1}, m_{i+1}-e\rangle=\|w_i-e\|^2-\|w_i-m_{i+1}\|^2-\|m_{i+1}-e\|^2. \qquad (18)
$$

Combining (11), (12), (16), (17) and (18) with the monotonicity of $\mathcal{K}$, we get

$$\begin{aligned}
&\|w_i - e\|^2 - \|w_i - m_{i+1}\|^2 - \|m_{i+1} - e\|^2 \\
&\geq 2\sigma\left[\langle q_i - \mathcal{K}e, m_{i+1} - m_i\rangle - \langle q_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle\right] \\
&\quad +2\sigma\langle q_i - q_{i-1}, m_{i+1} - n_i\rangle \\
&\quad +2(1-\varrho)\|m_{i+1} - m_i\|^2 + 2\varrho\langle m_{i+1} - n_i, m_{i+1} - m_i\rangle \\
&\quad +(1-\varrho)\left[\|m_{i+1} - n_i\|^2 - \|m_i - n_i\|^2 - ||m_{i+1} - m_i\|^2\right] \\
&\quad -2\varrho\langle m_i - n_{i-1}, m_{i+1} - m_i\rangle + 2\sigma\langle q_i - \mathcal{K}n_i, n_i - e\rangle. \qquad (19)
\end{aligned}$$

Hence,

$$\begin{aligned}
&\|m_{i+1} - e\|^2 + (1-\varrho)\|m_{i+1} - m_i\|^2 + (1-\varrho)\|m_{i+1} - n_i\|^2 \\
&\quad +2\sigma\langle q_i - \mathcal{K}e, m_{i+1} - m_i\rangle \\
&\leq \|w_i - e\|^2 + 2\sigma\langle q_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle + 2\sigma\langle q_{i-1} - q_i, m_{i+1} - n_i\rangle \\
&\quad +\|m_i - n_i\|^2 + 2\sigma\langle q_i - \mathcal{K}n_i, e - n_i\rangle - \|w_i - m_{i+1}\|^2 \\
&\quad -2\varrho\langle m_{i+1} - n_i, m_{i+1} - m_i\rangle + 2\varrho\langle m_i - n_{i-1}, m_{i+1} - m_i\rangle. \qquad (20)
\end{aligned}$$

Using (6) again, we see that

$$\begin{aligned}
\|m_{i+1} - w_i\|^2 &= \|(m_{i+1} - m_i) - \varrho(m_i - m_{i-1})\|^2 \\
&= \|m_{i+1} - m_i\|^2 + \varrho^2\|m_i - m_{i-1}\|^2 - 2\varrho\langle m_{i+1} - m_i, m_i - m_{i-1}\rangle \\
&\geq (1-\varrho)\|m_{i+1} - m_i\|^2 + (\varrho^2 - \varrho)\|m_i - m_{i-1}\|^2. \qquad (21)
\end{aligned}$$

Also, by (7), we get

$$\begin{aligned}
\|w_i - e\|^2 &= \|(1+\varrho)(m_i - e) - \varrho(m_{i-1} - e)\|^2 \\
&= (1+\varrho)\|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + \varrho(1+\varrho)\|m_i - m_{i-1}\|^2. \qquad (22)
\end{aligned}$$

Putting (21) and (22) in (20), we obtain

$$\begin{aligned}
&\|m_{i+1} - e\|^2 + (1-\varrho)\|m_{i+1} - m_i\|^2 + (1-\varrho)\|m_{i+1} - n_i\|^2 \\
&\quad +2\sigma\langle q_i - \mathcal{K}e, m_{i+1} - m_i\rangle \\
&\leq (1+\varrho)\|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + \varrho(1+\varrho)\|m_i - m_{i-1}\|^2 \\
&\quad +2\sigma\langle q_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle \\
&\quad +2\sigma\langle q_{i-1} - q_i, m_{i+1} - n_i\rangle + \|m_i - n_i\|^2 \\
&\quad +2\sigma\langle q_i - \mathcal{K}n_i, e - n_i\rangle - (1-\varrho)\|m_i - m_{i+1}\|^2 - (\varrho^2 - \varrho)\|m_i - m_{i-1}\|^2 \\
&\quad -2\varrho\langle m_{i+1} - n_i, m_{i+1} - m_i\rangle + 2\varrho\langle m_i - n_{i-1}, m_{i+1} - m_i\rangle.
\end{aligned}$$

This implies

$$\begin{aligned}
&\|m_{i+1}-e\|^2-\varrho\|m_i-e\|^2+2(1-\varrho)\|m_{i+1}-m_i\|^2+(1-\varrho)\|m_{i+1}-n_i\|^2\\
&\quad+2\sigma\langle q_i-\mathcal{K}e,m_{i+1}-m_i\rangle\\
&\leq\|m_i-e\|^2-\varrho\|m_{i-1}-e\|^2+2\varrho\|m_i-m_{i-1}\|^2+2\sigma\langle q_{i-1}-\mathcal{K}e,m_i-m_{i-1}\rangle\\
&\quad+2\sigma\langle q_{i-1}-q_i,m_{i+1}-n_i\rangle+\|m_i-n_i\|^2+2\sigma\beta_i\\
&\quad-2\varrho\langle m_{i+1}-n_i,m_{i+1}-m_i\rangle+2\varrho\langle m_i-n_{i-1},m_{i+1}-m_i\rangle, \qquad (23)
\end{aligned}$$

where $\beta_i=\langle q_i-\mathcal{K}n_i,e-n_i\rangle$.
Note that

$$\begin{aligned}
&-2\varrho\langle m_{i+1}-n_i,m_{i+1}-m_i\rangle+2\varrho\langle m_i-n_{i-1},m_{i+1}-m_i\rangle\\
&\leq\varrho\left[\|m_{i+1}-n_i\|^2+\|m_{i+1}-m_i\|^2\right]+\varrho\left[\|m_i-n_{i-1}\|^2+\|m_{i+1}-m_i\|^2\right]\\
&=\varrho\|m_{i+1}-n_i\|^2+2\varrho\|m_{i+1}-m_i\|^2+\varrho\|m_i-n_{i-1}\|^2. \qquad (24)
\end{aligned}$$

Putting (24) in (23), we obtain

$$\begin{aligned}
&\|m_{i+1}-e\|^2-\varrho\|m_i-e\|^2+(1+2\varrho)\|m_{i+1}-m_i\|^2+\varrho\|m_{i+1}-n_i\|^2\\
&\quad+2\sigma\langle q_i-\mathcal{K}e,m_{i+1}-m_i\rangle\\
&\leq\|m_i-e\|^2-\varrho\|m_{i-1}-e\|^2+(1+2\varrho)\|m_i-m_{i-1}\|^2+\varrho\|m_i-n_{i-1}\|^2\\
&\quad+2\sigma\langle q_{i-1}-\mathcal{K}e,m_i-m_{i-1}\rangle\\
&\quad+2\sigma\langle q_{i-1}-q_i,m_{i+1}-n_i\rangle+2\sigma\beta_i-(1-3\varrho)\|m_{i+1}-n_i\|^2\\
&\quad-(1-6\varrho)\|m_{i+1}-m_i\|^2. \qquad (25)
\end{aligned}$$

Next, utilizing the fact that $\mathcal{K}$ is $\mathcal{L}$-Lipschitz continuous gives

$$\begin{aligned}
2\sigma\langle\mathcal{K}n_{i-1}-\mathcal{K}n_i,m_{i+1}-n_i\rangle&\leq 2\sigma\mathcal{L}\|n_i-n_{i-1}\|\|m_{i+1}-n_i\|\\
&\leq\sigma\mathcal{L}\left(\frac{1}{\sqrt{2}}\|n_i-n_{i-1}\|^2+\sqrt{2}\|m_{i+1}-n_i\|^2\right)\\
&\leq\sigma\mathcal{L}\frac{1}{\sqrt{2}}((2+\sqrt{2})\|n_i-m_i\|^2+\sqrt{2}\|m_i-n_{i-1}\|^2)\\
&\quad+\sqrt{2}\sigma\mathcal{L}\|m_{i+1}-n_i\|^2\\
&=\sigma\mathcal{L}(1+\sqrt{2})\|n_i-m_i\|^2+\sigma\mathcal{L}\|m_i-n_{i-1}\|^2\\
&\quad+\sqrt{2}\sigma\mathcal{L}\|m_{i+1}-n_i\|^2. \qquad (26)
\end{aligned}$$

Let $\tau=\mathcal{L}(\sqrt{2}-1)$. Then $\tau>0$. Hence, for any $s\in(0,\frac{\tau}{2})$, we get from Young's inequality and (26) that

$$\begin{aligned}
2\langle q_{i-1}-q_i,m_{i+1}-n_i\rangle&=2\langle q_{i-1}-\mathcal{K}n_{i-1}+\mathcal{K}n_{i-1}-\mathcal{K}n_i+\mathcal{K}n_i-q_i,m_{i+1}-n_i\rangle\\
&\leq\frac{1}{s}\|q_{i-1}-\mathcal{K}n_{i-1}\|^2+s\|m_{i+1}-n_i\|^2+\frac{1}{s}\|q_i-\mathcal{K}n_i\|^2
\end{aligned}$$

$$
\begin{aligned}
&+s\|m_{i+1}-n_i\|^2\\
&+\mathcal{L}(1+\sqrt{2})\|n_i-m_i\|^2+\mathcal{L}\|m_i-n_{i-1}\|^2\\
&+\sqrt{2}\mathcal{L}\|m_{i+1}-n_i\|^2. \qquad (27)
\end{aligned}
$$

Using (27) in (25), we obtain

$$
\begin{aligned}
&\|m_{i+1}-e\|^2-\varrho\|m_i-e\|^2+(1+2\varrho+\sigma\mathcal{L}(1+\sqrt{2}))\|m_{i+1}-m_i\|^2\\
&\quad+(\mathcal{L}\sigma+\varrho)\|m_{i+1}-n_i\|^2\\
&\quad+2\sigma\langle q_i-\mathcal{K}e,m_{i+1}-m_i\rangle\\
&\le\|m_i-e\|^2-\varrho\|m_{i-1}-e\|^2+(1+2\varrho+\sigma\mathcal{L}(1+\sqrt{2}))\|m_i-m_{i-1}\|^2\\
&\quad+(\mathcal{L}\sigma+\varrho)\|m_i-n_{i-1}\|^2\\
&\quad+2\sigma\langle q_{i-1}-\mathcal{K}e,m_i-m_{i-1}\rangle\\
&\quad+2\sigma\beta_i-\left[1-(\sigma\mathcal{L}+\varrho)-2s\sigma-\sqrt{2}\mathcal{L}\sigma-2\varrho\right]\|m_{i+1}-n_i\|^2\\
&\quad-(1-6\varrho-\sigma\mathcal{L}(1+\sqrt{2}))\|m_{i+1}-m_i\|^2\\
&\quad+\sigma\frac{1}{s}\|q_{i-1}-\mathcal{K}n_{i-1}\|^2+\sigma\frac{1}{s}\|q_i-\mathcal{K}n_i\|^2,
\end{aligned}
$$

which is the required conclusion. □

Assumption 4.1

(a) $0\le\varrho<\min\{\frac{1-\sigma\mathcal{L}}{2},\ \frac{1}{6}(1-\sigma\mathcal{L}(1+\sqrt{2})),\ \frac{1}{3}(1-\sigma(1+\sigma\mathcal{L}(\sqrt{2}-1))-\sigma\mathcal{L}(1+\sqrt{2}))\}$
and $0<\sigma<\frac{\sqrt{(\mathcal{L}(1+\sqrt{2})+1)^2+4\mathcal{L}(\sqrt{2}-1)}-(\mathcal{L}(1+\sqrt{2})+1)}{2\mathcal{L}(\sqrt{2}-1)}$,

(b) $\sum_{i=1}^{\infty}\mathbb{E}[\|q_i-\mathcal{K}n_i\|^2\mid\mathcal{F}_i]<+\infty$ a.s..

Theorem 4.2 *Let Algorithm 3.2 generate $\{m_i\}$ and let the conditions outlined in Assumptions 3.1 and 4.1 be met. Then $\{m_i\}$ is weakly convergent to a random variable $p^*:\Omega\to(\mathcal{J}+\mathcal{K})^{-1}(0)$ a.s..*

Proof Let $e\in(\mathcal{J}+\mathcal{K})^{-1}(0)$. Then by Lemma 4.1, we have

$$
\begin{aligned}
a_{i+1}\le a_i+2\sigma\beta_i-&\left[1-(\sigma\mathcal{L}+\varrho)-2s\sigma-\sqrt{2}\mathcal{L}\sigma-2\varrho\right]\|m_{i+1}-n_i\|^2\\
&-(1-6\varrho-\sigma\mathcal{L}(1+\sqrt{2}))\|m_{i+1}-m_i\|^2\\
&+2\sigma\frac{1}{s}\|q_i-\mathcal{K}n_i\|^2, \qquad (28)
\end{aligned}
$$

where

$$
a_i:=\|m_i-e\|^2-\varrho\|m_{i-1}-e\|^2+(1+2\varrho+\sigma\mathcal{L}(1+\sqrt{2}))\|m_i-m_{i-1}\|^2
$$

$$+(\mathcal{L}\sigma+\varrho)\|m_i-n_{i-1}\|^2$$
$$+2\sigma\langle q_{i-1}-\mathcal{K}e, m_i-m_{i-1}\rangle+\sigma\frac{1}{s}\|q_{i-1}-\mathcal{K}n_{i-1}\|^2.$$

Next, we show that $a_i \geq 0$ for all $i \geq 1$.
First, observe that

$$\begin{aligned}
&2\langle q_{i-1}-\mathcal{K}e, m_i-m_{i-1}\rangle\\
&\quad=2\langle q_{i-1}-\mathcal{K}n_{i-1}, m_i-m_{i-1}\rangle+2\langle \mathcal{K}n_{i-1}-\mathcal{K}e, m_i-m_{i-1}\rangle\\
&\quad\geq-\frac{1}{s}\|q_{i-1}-\mathcal{K}n_{i-1}\|^2-s\|m_i-m_{i-1}\|^2-2\mathcal{L}\|n_{i-1}-e\|\|m_i-m_{i-1}\|\\
&\quad\geq-\frac{1}{s}\|q_{i-1}-\mathcal{K}n_{i-1}\|^2-s\|m_i-m_{i-1}\|^2-\mathcal{L}\Big(\|m_i-n_{i-1}\|^2\\
&\qquad+\|m_i-e\|^2+2\|m_i-m_{i-1}\|^2\Big).
\end{aligned}\tag{29}$$

Second, we have

$$\|m_{i-1}-e\|^2\leq 2\|m_{i-1}-m_i\|^2+2\|m_i-e\|^2.\tag{30}$$

Now, using (29) and (30) in the definition of a_i, we obtain

$$\begin{aligned}
a_i\geq\;&\|m_i-e\|^2-\varrho\Big(2\|m_{i-1}-m_i\|^2+2\|m_i-e\|^2\Big)\\
&+(1+2\varrho+\sigma\mathcal{L}(1+\sqrt{2}))\|m_i-m_{i-1}\|^2\\
&+(\mathcal{L}\sigma+\varrho)\|m_i-n_{i-1}\|^2-\sigma\frac{1}{s}\|q_{i-1}-\mathcal{K}n_{i-1}\|^2-\sigma s\|m_i-m_{i-1}\|^2\\
&-\sigma\mathcal{L}\Big(\|m_i-n_{i-1}\|^2+\|m_i-e\|^2+2\|m_i-m_{i-1}\|^2\Big)+\sigma\frac{1}{s}\|q_{i-1}-\mathcal{K}n_{i-1}\|^2\\
=\;&(1-2\varrho-\sigma\mathcal{L})\|m_i-e\|^2\\
&+(1+(\sqrt{2}-1)\sigma\mathcal{L}-\sigma s)\|m_{i-1}-m_i\|^2+\varrho\|m_i-n_{i-1}\|^2.
\end{aligned}\tag{31}$$

Since $\tau=\mathcal{L}(\sqrt{2}-1)$ and $s\in(0,\frac{\tau}{2})$, we obtain that $1+(\sqrt{2}-1)\sigma\mathcal{L}-s>0$. Also, by the condition on ϱ, we see that $1-2\varrho-\sigma\mathcal{L}\geq 0$. Hence, by Assumption 4.1 (a), $a_i\geq 0$ for all $i\geq 1$. Now, applying the conditional expectation w.r.t $\mathcal{F}_i$ on both sides of (28), we arrive at the following expression:

$$\begin{aligned}
\mathbb{E}[a_{i+1}\mid\mathcal{F}_i]\leq\;&a_i+2\sigma\mathbb{E}[\beta_i\mid\mathcal{F}_i]-\Big[1-(\sigma\mathcal{L}+\varrho)-2s\sigma\\
&-\sqrt{2}\mathcal{L}\sigma-2\varrho\Big]\mathbb{E}[\|m_{i+1}-n_i\|^2\mid\mathcal{F}_i]\\
&-(1-6\varrho-\sigma\mathcal{L}(1+\sqrt{2}))\mathbb{E}[\|m_{i+1}-m_i\|^2\mid\mathcal{F}_i]+2\sigma\frac{1}{s}\mathbb{E}[\|q_i\\
&-\mathcal{K}n_i\|^2\mid\mathcal{F}_i]\ \text{a.s.}\\
=\;&a_i-\Big[1-(\sigma\mathcal{L}+\varrho)-2s\sigma-\sqrt{2}\mathcal{L}\sigma-2\varrho\Big]\mathbb{E}[\|m_{i+1}-n_i\|^2\mid\mathcal{F}_i]
\end{aligned}$$

$$-(1-6\varrho-\sigma\mathcal{L}(1+\sqrt{2}))\mathbb{E}[\|m_{i+1}-m_i\|^2 \mid \mathcal{F}_i]$$
$$+2\sigma\frac{1}{s}\mathbb{E}[\|q_i-\mathcal{K}n_i\|^2 \mid \mathcal{F}_i] \text{ a.s.}, \tag{32}$$

since by Algorithm 3.2, $\mathbb{E}[\beta_i \mid \mathcal{F}_i]=0$ a.s.. By the condition on ϱ and σ, and noting that $\tau=\mathcal{L}(\sqrt{2}-1)$ and $s\in(0,\frac{\tau}{2})$, we obtain $1-6\varrho-\sigma\mathcal{L}(1+\sqrt{2})>0$ and $1-(\sigma\mathcal{L}+\varrho)-2s\sigma-\sqrt{2}\mathcal{L}\sigma-2\varrho>0$. Thus, Lemma 2.1 yields

$$\lim_{i\to\infty} a_i \text{ exists, and } \lim_{i\to\infty}\|m_{i+1}-m_i\|=0 \text{ a.s..} \tag{33}$$

From Assumption 4.1(b) and Lemma 2.2, we obtain

$$\sum_{i=1}^{\infty}\|q_i-\mathcal{K}n_i\|^2<+\infty \quad \text{a.s.},$$

which implies that

$$\lim_{i\to\infty}\|q_i-\mathcal{K}n_i\|=0 \quad \text{a.s..} \tag{34}$$

Also, note that

$$\begin{aligned}
&2\langle q_{i-1}-\mathcal{K}e, m_i-m_{i-1}\rangle\\
&\le \frac{1}{s}\|q_{i-1}-\mathcal{K}n_{i-1}\|^2+s\|m_i-m_{i-1}\|^2+2\mathcal{L}\|n_{i-1}-e\|\|m_i-m_{i-1}\|\\
&\le \frac{1}{s}\|q_{i-1}-\mathcal{K}n_{i-1}\|^2+s\|m_i-m_{i-1}\|^2\\
&\quad +2\mathcal{L}\left(\|m_i-n_{i-1}\|+\|m_i-e\|\right)\|m_i-m_{i-1}\|.
\end{aligned} \tag{35}$$

By (33), $\{a_i\}$ is bounded. Thus, by (31), we obtain that $\{m_i\}$ is also bounded. Hence, $\{n_i\}$ and $\{w_i\}$ are bounded too. Using these facts, (33) and (34), we obtain that

$$\lim_{i\to\infty}\langle q_{i-1}-\mathcal{K}e, m_i-m_{i-1}\rangle=0 \quad \text{a.s..} \tag{36}$$

Now, since $n_i=2m_i-m_{i-1}$ and $w_i=m_i+\varrho(m_i-m_{i-1})$, we get from (33) that

$$\lim_{i\to\infty}\|n_i-m_i\|=\lim_{i\to\infty}\|m_i-m_{i-1}\|=0=\lim_{i\to\infty}\|w_i-m_i\|. \tag{37}$$

Also,

$$\|m_{i+1}-n_i\|\le\|m_{i+1}-m_i\|+\|n_i-m_i\|\to 0,\ i\to\infty \tag{38}$$

and

$$\|m_{i+1}-w_i\|\le\|m_{i+1}-m_i\|+\|m_i-w_i\|\to 0,\ i\to\infty. \tag{39}$$

Now, since $\lim_{i\to\infty} a_i$ exists, we obtain from (8), (33), (38), (34) and (36) that

$$\lim_{i\to\infty}\left[\|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2\right] \text{ exists a.s..} \tag{40}$$

Set

$$\begin{aligned} b_i^* &= 2\langle m_{i-1} - m_i, m_i - e\rangle + \|m_{i-1} - m_i\|^2, \\ d_i^* &= -\Big((1 + 2\varrho + \sigma\mathcal{L}(1+\sqrt{2}))\|m_i - m_{i-1}\|^2 + (\mathcal{L}\sigma + \varrho)\|m_i - n_{i-1}\|^2 \\ &\quad + 2\sigma\langle q_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle + \sigma\frac{1}{s}\|q_{i-1} - \mathcal{K}n_{i-1}\|^2\Big), \end{aligned}$$

and note that

$$\|m_{i-1} - e\|^2 = \|m_{i-1} - m_i\|^2 + 2\langle m_{i-1} - m_i, m_i - e\rangle + \|m_i - e\|^2.$$

Then

$$b_i^* = \|m_{i-1} - e\|^2 - \|m_i - e\|^2.$$

Thus,

$$a_i + d_i^* + \varrho b_i^* = (1 - \varrho)\|m_i - e\|^2.$$

By (33), (34), (36) and (38), we have that $d_i^* \to 0$, and $b_i^* \to 0$, and $\lim_{i\to\infty} a_i$ exists, a.s.. Thus,

$$\lim_{i\to\infty} \|m_i - e\| \quad \text{exists for each } e \in (\mathcal{J} + \mathcal{K})^{-1}(0) \text{ a.s..} \tag{41}$$

Let $z_i = \text{Res}_\sigma^{\mathcal{J}}(w_i - \sigma\mathcal{K}n_i)$. Then by the nonexpansivity of $\text{Res}_\sigma^{\mathcal{J}}$, we obtain

$$\|m_{i+1} - z_i\| \le \sigma\|\mathcal{K}n_i - q_i\| \to 0 \quad \text{a.s..} \tag{42}$$

Moreover, we obtain from (42), (38) and (39) that

$$\lim_{i\to\infty} \|w_i - z_i\| = \lim_{i\to\infty} \|n_i - z_i\| = 0. \tag{43}$$

Since $\{m_i\}$ is bounded, there exists $\{m_{i_k}\} \subset \{m_i\}$ such that $m_{i_k} \rightharpoonup \widehat{p} \in \mathcal{V}$. Thus, we obtain from (33) and (42) that $z_{i_k} \rightharpoonup \widehat{p}$. Also, we obtain from (43) that

$$\frac{1}{\sigma}(w_{i_k} - z_{i_k}) - \mathcal{K}n_{i_k} + \mathcal{K}z_{i_k} \to 0,\ i \to \infty, \quad \text{a.s..} \tag{44}$$

Note that

$$\frac{1}{\sigma}(w_{i_k} - z_{i_k}) - \mathcal{K}n_{i_k} + \mathcal{K}z_{i_k} \in (\mathcal{J} + \mathcal{K})z_{i_k}. \tag{45}$$

The graph of $\mathcal{J}+\mathcal{K}$ is closed in the space $\mathcal{V}^{\text{strong}} \times \mathcal{V}^{\text{weak}}$ since it is maximal monotone. Thus, by (44) and (45), we obtain that $\widehat{p} \in (\mathcal{J}+\mathcal{K})^{-1}(0)$ a.s.. Therefore, by (41) and Lemma 2.3, we can infer that the sequence $\{m_i\}$ is weakly convergent to $p^* \in (\mathcal{J}+\mathcal{K})^{-1}(0)$ as asserted. □

Remark 4.1 The conclusion of Theorem 4.2 still hold if we replace the constant stepsize σ in Algorithm 3.2 with variable stepsizes $\{\sigma_i\}$ satisfying the condition $\sigma_i \leq \sigma_{i-1}\ \forall i \geq 1$. In what follows, we shall replace the constant stepsize σ in Algorithm 3.2 with a nonincreasing sequence.

4.2 Rate of Convergence

Assumption 4.3

(a) The operator $\mathcal{J}$ is maximal monotone and ϵ-strongly monotone.
(b) The operator $\mathcal{K}$ is monotone and $\mathcal{L}$-Lipschitz continuous.
(c) The solution set $(\mathcal{J}+\mathcal{K})^{-1}(0)$ is nonempty.
(d) $0 \leq \sigma_i \leq \sigma_{i-1} < \widehat{\sigma}\ \forall i \geq 1$.
(e) $0 \leq \varrho < \min\{\frac{1}{4}(1-2\sigma_i\mathcal{L}),\ \frac{1}{12}(1+4\sigma_i\epsilon-2\sigma_i\mathcal{L}(1+\sqrt{2})),\ \frac{\sigma_i}{3\sigma_{i-1}}(1-\sigma_{i-1}\mathcal{L}(1+\sqrt{2})\}$ and $\sigma_i = \frac{1}{2\mathcal{L}(1+\sqrt{2})(i+1)}\ \forall i \geq 1$ (i.e., we replace σ with σ_i in Algorithm 3.2).
(f) For all $i \geq 1$, there is a constant $d > 0$ such that $\mathbb{E}[\|q_i - \mathcal{K}n_i\|^2 \mid \mathcal{F}_i] \leq d$ a.s..

Lemma 4.2 *Let the conditions outlined in Assumption 4.3(a)-(d) be met. Let* $k, p, t \in [0,1)$ *and* $r \in \mathbb{R}$ *such that*

$$\sigma_i\mathcal{L} \leq t \leq \sigma_i\mathcal{L}+\varrho,\ 2\sigma_i\mathcal{L}+4\varrho \leq k,\ \ 2\sigma_i\mathcal{L}+2\varrho \leq p,\ \forall i \geq 1, \tag{46}$$

and

$$1 \leq r \leq \min\ \Big\{1+\frac{1}{2(\varrho+\sigma_{i+1}\mathcal{L}-t)},\ 1+\frac{2\epsilon\sigma_i}{1-k},\ 1 + \frac{1}{2(1+2\varrho+\sigma_{i+1}\mathcal{L}(1+\sqrt{2})-p)}\Big\}. \tag{47}$$

For all $e \in (\mathcal{J}+\mathcal{K})^{-1}(0)$, *define*

$$\begin{aligned}\Gamma_i :=\ & (1-k)\|m_i-e\|^2 + \Big[1+2\varrho+\sigma_i\mathcal{L}(1+\sqrt{2})-p\Big]\|m_i-m_{i-1}\|^2 \\ & +(\varrho+\sigma_i\mathcal{L}-t)\|m_i-n_{i-1}\|^2\end{aligned}$$

and

$$\begin{aligned}\Upsilon_i :=\ & k\|m_i-e\|^2 - \varrho\|m_{i-1}-e\|^2 + p\|m_i-m_{i-1}\|^2 + t\|m_i-n_{i-1}\|^2 \\ & +2\sigma_i\langle\mathcal{K}n_{i-1}-\mathcal{K}e, m_i-m_{i-1}\rangle.\end{aligned}$$

Then we have

(i). $\Gamma_i \geq 0$ *and* $\Upsilon_i \geq 0$ *for all* $i \geq 1$.
(ii). *There exist* $\rho > 0$ *and* $i_0 \geq 1$ *such that* $\rho\Upsilon_i \leq \Gamma_i$ *for all* $i \geq i_0$.
(iii). $r\Gamma_{i+1} + \Upsilon_{i+1} \leq \Gamma_i + \Upsilon_i + 2\sigma_i k_i - T_i$ *for all* $i \geq 1$.
(iv). *For* $\alpha \geq 1$ *and* $\beta = \min\{\alpha,\ 1 + \rho(r-\alpha)\} \geq 1$, *we have that* $\beta(\Gamma_{i+1} + \Upsilon_{i+1}) \leq \Gamma_i + \Upsilon_i + 2\sigma_i k_i - T_i$ *for all* $i \geq i_0$,

where $k_i = \langle q_i - \mathcal{K}n_i, e - n_i\rangle + \langle q_{i-1} - \mathcal{K}n_{i-1}, m_i - m_{i-1}\rangle - \langle q_i - \mathcal{K}n_i, m_{i+1} - m_i\rangle + \langle q_{i-1} - \mathcal{K}n_{i-1} - q_i + \mathcal{K}n_i, m_{i+1} - n_i\rangle$ *and* $T_i = (\frac{\sigma_i}{\sigma_{i-1}}(1-\varrho) - 2\varrho - \sigma_{i+1}\mathcal{L} - \sqrt{2}\mathcal{L}\sigma_i - \frac{1}{2})\|m_{i+1} - n_i\|^2 + (\frac{1}{2} + 2\sigma_i\epsilon - \sigma_{i+1}\mathcal{L}(1+\sqrt{2}) - 5\varrho - \frac{\sigma_i}{\sigma_{i-1}}\varrho)\|m_{i+1} - m_i\|^2 + (1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - m_{i-1}\|^2 + \varrho(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - n_i\|^2$.

Proof Let $e \in (\mathcal{J} + \mathcal{K})^{-1}(0)$.
(i). Since $k, p \in [0, 1)$ and $t < \sigma_i L + \varrho$, we have that $\Gamma_i \geq 0$ for all $i \geq 1$. Also, we have

$$\begin{aligned}
\Upsilon_i &= k\|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + p\|m_i - m_{i-1}\|^2 + t\|m_i - n_{i-1}\|^2 \\
&\quad +2\sigma_i\langle \mathcal{K}n_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle \\
&\geq k\|m_i - e\|^2 - 2\varrho\left[\|m_{i-1} - m_i\|^2 + \|m_i - e\|^2\right] + p\|m_i - m_{i-1}\|^2 \\
&\quad +t\|m_i - n_{i-1}\|^2 - \sigma_i\mathcal{L}\|n_{i-1} - m_i\|^2 - \sigma_i\mathcal{L}\|m_i - e\|^2 - 2\sigma_i\mathcal{L}\|m_i - m_{i-1}\|^2 \\
&\geq \frac{k}{2}\|m_i - e\|^2 + (\frac{k}{2} - \sigma_i\mathcal{L} - 2\varrho)\|m_i - e\|^2 + (p - 2\sigma_i\mathcal{L} - 2\varrho)\|m_i - m_{i-1}\|^2 \\
&\quad +(t - \sigma_i\mathcal{L})\|m_i - n_{i-1}\|^2.
\end{aligned}$$

By (46), we obtain that $\Upsilon_i \geq 0$ for all $i \geq 1$.
(ii).

$$\begin{aligned}
\Upsilon_i &\leq k\|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + p\|m_i - m_{i-1}\|^2 + t\|m_i - n_{i-1}\|^2 \\
&\quad +\sigma_i\mathcal{L}t\|m_i - n_{i-1}\|^2 + \frac{\sigma_i\mathcal{L}}{t}\|m_i - e\|^2 + 2\sigma_i\mathcal{L}\|m_i - m_{i-1}\|^2 \\
&= (k + \frac{\sigma_i\mathcal{L}}{t})\|m_i - e\|^2 + (p + 2\sigma_i\mathcal{L})\|m_i - m_{i-1}\|^2 + (t + \sigma_i\mathcal{L}t)\|m_i - n_{i-1}\|^2 \\
&\quad -\varrho\|m_{i-1} - e\|^2 \\
&\leq (k + \frac{\widehat{\sigma}\mathcal{L}}{t})\|m_i - e\|^2 + (p + 2\widehat{\sigma}\mathcal{L})\|m_i - m_{i-1}\|^2 + (t + \widehat{\sigma}\mathcal{L}t)\|m_i - n_{i-1}\|^2 \\
&\quad -\varrho\|m_{i-1} - e\|^2. \qquad (48)
\end{aligned}$$

Define,

$$\rho_i := \min\left\{\frac{1-k}{k + \frac{\widehat{\sigma}\mathcal{L}}{t}}, \frac{1 + 2\varrho + \sigma_i\mathcal{L}(1+\sqrt{2}) - p}{p + 2\widehat{\sigma}\mathcal{L}}, \frac{\varrho + \sigma_i\mathcal{L} - t}{t + \widehat{\sigma}\mathcal{L}t}\right\}.$$

Thus, we obtain from (48) that

$$\rho_i \Upsilon_i \leq \Gamma_i \ \forall i \geq 1. \tag{49}$$

Since $0 < \sigma_i \leq \sigma_{i-1}$, limit of σ_i exists. Hence, limit of ρ_i exists. Then there exist $\rho > 0$ and $i_0 \geq 1$ such that $\rho_i > \rho, \ \forall i \geq i_0$. Thus, we obtain from (49) that

$$\rho \Upsilon_i \leq \Gamma_i \ \forall i \geq i_0.$$

(iii). Applying the ϵ-strongly monotonicity of $\mathcal{J}$ in (10) and (13), we obtain

$$\langle \frac{1}{\sigma_i}(w_i - m_{i+1}) - q_i + \mathcal{K}e, m_{i+1} - e\rangle \geq \epsilon \|m_{i+1} - e\|^2.$$

and

$$\langle \frac{w_i - m_{i+1}}{\sigma_i} - q_i - \frac{w_{i-1} - m_i}{\sigma_{i-1}} + q_{i-1}, m_{i+1} - m_i\rangle \geq \epsilon \|m_{i+1} - m_i\|^2,$$

respectively. Now, using these two inequalities in place of (10) and (13) and following the proof of Lemma 4.1 up to (25), we arrive at

$$\begin{aligned}
&\|m_{i+1} - e\|^2 - \varrho\|m_i - e\|^2 + (1 + 2\varrho)\|m_{i+1} - m_i\|^2 + \varrho\|m_{i+1} - n_i\|^2 \\
&\quad + 2\sigma_i\langle q_i - \mathcal{K}e, m_{i+1} - m_i\rangle \\
&\quad + 2\epsilon\sigma_i\|m_{i+1} - e\|^2 + 2\epsilon\sigma_i\|m_{i+1} - m_i\|^2 \\
&\leq \|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + (1 + 2\varrho)\|m_i - m_{i-1}\|^2 + \varrho\|m_i - n_{i-1}\|^2 \\
&\quad + 2\sigma_i\langle q_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle \\
&\quad + 2\sigma_i\langle q_{i-1} - q_i, m_{i+1} - n_i\rangle + 2\sigma_i\beta_i - (\frac{\sigma_i}{\sigma_{i-1}}(1 - \theta) - 2\theta)\|m_{i+1} - n_i\|^2 \\
&\quad - (1 - 5\varrho - \frac{\sigma_i}{\sigma_{i-1}}\varrho)\|m_{i+1} - m_i\|^2 \\
&\quad - (1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - m_{i-1}\|^2 - \varrho(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - n_i\|^2.
\end{aligned}$$

This implies

$$\begin{aligned}
&\|m_{i+1} - e\|^2 - \varrho\|m_i - e\|^2 + (1 + 2\varrho)\|m_{i+1} - m_i\|^2 + \varrho\|m_{i+1} - n_i\|^2 \\
&\quad + 2\sigma_i\langle \mathcal{K}n_i - \mathcal{K}e, m_{i+1} - m_i\rangle \\
&\quad + 2\epsilon\sigma_i\|m_{i+1} - e\|^2 + 2\epsilon\sigma_i\|m_{i+1} - m_i\|^2 \\
&\leq \|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + (1 + 2\varrho)\|m_i - m_{i-1}\|^2 + \varrho\|m_i - n_{i-1}\|^2 \\
&\quad + 2\sigma_i\langle \mathcal{K}n_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle \\
&\quad + 2\sigma_i\langle q_{i-1} - \mathcal{K}n_{i-1}, m_i - m_{i-1}\rangle - 2\sigma_i\langle q_i - \mathcal{K}n_i, m_{i+1} - m_i\rangle \\
&\quad + 2\sigma_i\langle q_{i-1} - q_i, m_{i+1} - n_i\rangle
\end{aligned}$$

$$
\begin{aligned}
&+2\sigma_i\langle q_i - \mathcal{K}n_i, e - n_i\rangle - (\frac{\sigma_i}{\sigma_{i-1}}(1-\theta) - 2\theta)\|m_{i+1} - n_i\|^2\\
&-(1 - 5\varrho - \frac{\sigma_i}{\sigma_{i-1}}\varrho)\|m_{i+1} - m_i\|^2\\
&-(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - m_{i-1}\|^2 - \varrho(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - n_i\|^2\\
=\ &\|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + (1+2\varrho)\|m_i - m_{i-1}\|^2 + \varrho\|m_i - n_{i-1}\|^2\\
&+2\sigma_i\langle\mathcal{K}n_{i-1} - \mathcal{K}e, m_i - m_{i-1}\rangle\\
&+2\sigma_i\left(\langle q_i - \mathcal{K}n_i, e - n_i\rangle + 2\sigma_i\langle q_{i-1} - \mathcal{K}n_{i-1}, m_i - m_{i-1}\rangle\right.\\
&\left.-2\sigma_i\langle q_i - \mathcal{K}n_i, m_{i+1} - m_i\rangle\right)\\
&+2\sigma_i\langle q_{i-1} - \mathcal{K}n_{i-1} - q_i + \mathcal{K}n_i, m_{i+1} - n_i\rangle + 2\sigma_i\langle\mathcal{K}n_{i-1} - \mathcal{K}n_i, m_{i+1} - n_i\rangle\\
&-(\frac{\sigma_i}{\sigma_{i-1}}(1-\theta) - 2\theta)\|m_{i+1} - n_i\|^2 - (1 - 5\varrho - \frac{\sigma_i}{\sigma_{i-1}}\varrho)\|m_{i+1} - m_i\|^2\\
&-(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - m_{i-1}\|^2 - \varrho(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - n_i\|^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&(1+2\epsilon\sigma_i)\|m_{i+1} - e\|^2 - \varrho\|m_i - e\|^2 + (1 + 2\varrho + 2\epsilon\sigma_i)\|m_{i+1} - m_i\|^2\\
&\quad+\varrho\|m_{i+1} - n_i\|^2 + 2\sigma_i\alpha_i\\
&\le \|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + (1+2\varrho)\|m_i - m_{i-1}\|^2 + \varrho\|m_i - n_{i-1}\|^2\\
&\quad+2\sigma_i\alpha_{i-1} + 2\sigma_i k_i + 2\sigma_i\langle\mathcal{K}n_{i-1} - \mathcal{K}n_i, m_{i+1} - n_i\rangle\\
&\quad-(\frac{\sigma_i}{\sigma_{i-1}}(1-\varrho) - 2\varrho)\|m_{i+1} - n_i\|^2 - (1 - 5\varrho - \frac{\sigma_i}{\sigma_{i-1}}\varrho)\|m_{i+1} - m_i\|^2\\
&\quad-(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - m_{i-1}\|^2 - \varrho(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - n_i\|^2,
\end{aligned}
\tag{50}
$$

where $\alpha_i = \langle\mathcal{K}n_i - \mathcal{K}e, m_{i+1} - m_i\rangle$.
Now, using (26) in (50), we obtain

$$
\begin{aligned}
&(1+2\epsilon\sigma_i)\|m_{i+1} - e\|^2 - \varrho\|m_i - e\|^2 + (1 + 2\varrho + \sigma_{i+1}\mathcal{L}(1+\sqrt{2}))\|m_{i+1} - m_i\|^2\\
&\quad+(\varrho + \sigma_{i+1}\mathcal{L})\|m_{i+1} - n_i\|^2 + 2\sigma_i\alpha_i\\
&\le \|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + (1 + 2\varrho + \sigma_i\mathcal{L}(1+\sqrt{2}))\|m_i - m_{i-1}\|^2\\
&\quad+(\varrho + \sigma_i\mathcal{L})\|m_i - n_{i-1}\|^2 + 2\sigma_i\alpha_{i-1} + 2\sigma_i k_i\\
&\quad-(\frac{\sigma_i}{\sigma_{i-1}}(1-\varrho) - 2\varrho - \sigma_{i+1}\mathcal{L} - \sqrt{2}\mathcal{L}\sigma_i)\|m_{i+1} - n_i\|^2\\
&\quad-(1 + 2\sigma_i\epsilon - \sigma_{i+1}\mathcal{L}(1+\sqrt{2}) - 5\varrho - \frac{\sigma_i}{\sigma_{i-1}}\varrho)\|m_{i+1} - m_i\|^2\\
&\quad-(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - m_{i-1}\|^2 - \varrho(1 - \frac{\sigma_i}{\sigma_{i-1}})\|m_i - n_i\|^2.
\end{aligned}
$$

Observe that

$$
\Gamma_i + \Upsilon_i = \|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + (1 + 2\varrho + \sigma_i\mathcal{L}(1+\sqrt{2}))\|m_i - m_{i-1}\|^2
$$

$$+(\varrho+\sigma_i\mathcal{L})\|m_i-n_{i-1}\|^2+2\sigma_i\alpha_{i-1}. \tag{51}$$

Thus, the above inequality becomes

$$\begin{aligned}
&(1+2\epsilon\sigma_i)\|m_{i+1}-e\|^2-\varrho\|m_i-e\|^2+(1+2\varrho+\sigma_{i+1}\mathcal{L}(1+\sqrt{2}))\|m_{i+1}-m_i\|^2\\
&\quad+(\varrho+\sigma_{i+1}\mathcal{L})\|m_{i+1}-n_i\|^2+2\sigma_i\alpha_i\\
&\leq \Gamma_i+\Upsilon_i+2\sigma_i k_i-(\frac{\sigma_i}{\sigma_{i-1}}(1-\varrho)-2\varrho-\sigma_{i+1}\mathcal{L}-\sqrt{2}\mathcal{L}\sigma_i)\|m_{i+1}-n_i\|^2\\
&\quad-(1+2\sigma_i\epsilon-\sigma_{i+1}\mathcal{L}(1+\sqrt{2})-5\varrho-\frac{\sigma_i}{\sigma_{i-1}}\varrho)\|m_{i+1}-m_i\|^2\\
&\quad-(1-\frac{\sigma_i}{\sigma_{i-1}})\|m_i-m_{i-1}\|^2-\varrho(1-\frac{\sigma_i}{\sigma_{i-1}})\|m_i-n_i\|^2.
\end{aligned} \tag{52}$$

Now, by the definition of Γ_i and Υ_i, we get

$$\begin{aligned}
r\Gamma_{i+1}+\Upsilon_{i+1}&=(r(1-k)+k)\|m_{i+1}-e\|^2-\varrho\|m_i-e\|^2\\
&\quad+\left[r(1+2\varrho+\sigma_{i+1}\mathcal{L}(1+\sqrt{2})-p)+p\right]\|m_{i+1}-m_i\|^2\\
&\quad+\left[r(\varrho+\sigma_{i+1}\mathcal{L}-t)+t\right]\|m_{i+1}-n_i\|^2+2\sigma_{i+1}\alpha_{i+1}.
\end{aligned} \tag{53}$$

By (47), we see that $r(1-k)+k<1+2\epsilon\sigma_i$, $r(1+2\varrho+\sigma_{i+1}\mathcal{L}(1+\sqrt{2})-p)+p<1+2\varrho+\sigma_{i+1}\mathcal{L}(1+\sqrt{2})+\frac{1}{2}$ and $r(\varrho+\sigma_{i+1}\mathcal{L}-t)+t<\varrho+\sigma_{i+1}\mathcal{L}+\frac{1}{2}$. Using these in (53) and noting (52), we obtain

$$\begin{aligned}
r\Gamma_{i+1}+\Upsilon_{i+1}&\leq(1+2\epsilon\sigma_i)\|m_{i+1}-e\|^2-\varrho\|m_i-e\|^2\\
&\quad+\left[1+2\varrho+\sigma_{i+1}\mathcal{L}(1+\sqrt{2})+\frac{1}{2}\right]\|m_{i+1}-m_i\|^2\\
&\quad+\left[\varrho+\sigma_{i+1}\mathcal{L}+\frac{1}{2}\right]\|m_{i+1}-n_i\|^2+2\sigma_i\alpha_i\\
&\leq\Gamma_i+\Upsilon_i+2\sigma_i k_i\\
&\quad-(\frac{\sigma_i}{\sigma_{i-1}}(1-\varrho)-2\varrho-\sigma_{i+1}\mathcal{L}-\sqrt{2}\mathcal{L}\sigma_i-\frac{1}{2})\|m_{i+1}-n_i\|^2\\
&\quad-(\frac{1}{2}+2\sigma_i\epsilon-\sigma_{i+1}\mathcal{L}(1+\sqrt{2})-5\varrho-\frac{\sigma_i}{\sigma_{i-1}}\varrho)\|m_{i+1}-m_i\|^2\\
&\quad-(1-\frac{\sigma_i}{\sigma_{i-1}})\|m_i-m_{i-1}\|^2-\varrho(1-\frac{\sigma_i}{\sigma_{i-1}})\|m_i-n_i\|^2.
\end{aligned} \tag{54}$$

Next, we prove part (iv).
(iv). According to part (ii) of Lemma 4.2, we have $\rho\Upsilon_{i+1}\leq\Gamma_{i+1}\forall i\geq i_0$. Therefore,

$$\begin{aligned}
r\Gamma_{i+1}&=\alpha\Gamma_{i+1}+(r-\alpha)\Gamma_{i+1}\\
&\geq\alpha\Gamma_{i+1}+\rho(r-\alpha)\Upsilon_{i+1}\\
&=(\alpha\Gamma_{i+1}+(1+\rho(r-\alpha))\Upsilon_{i+1})-\Upsilon_{i+1}\\
&\geq\beta(\Gamma_{i+1}+\Upsilon_{i+1})-\Upsilon_{i+1}.
\end{aligned} \tag{55}$$

Combining (55) with (iii), we obtain

$$\begin{aligned}\beta(\Gamma_{i+1}+\Upsilon_{i+1}) &\leq r\Gamma_{i+1}+\Upsilon_{i+1}\\ &\leq \Gamma_i+\Upsilon_i+2\sigma_i k_i-T_i. \end{aligned} \tag{56}$$

□

Theorem 4.4 *Let Algorithm 3.2 generate* $\{m_i\}$ *and let the conditions outlined in Assumption 4.3 be met. Then for all* $i \geq i_0$,

$$\mathbb{E}[\|m_i-\widehat{p}\|^2] = \mathcal{O}(\log(i+1)/(i+1)),$$

where $\widehat{p}$ *is a unique solution.*

Proof Let $e \in (\mathcal{J}+\mathcal{K})^{-1}(0)$ and $G_i := \Gamma_i+\Upsilon_i$. Then, by Lemma 4.2 (i), $G_i \geq 0$ for all $i \geq 1$. It follows from Lemma 4.2 (iv) (noting that $\beta \geq 1$) that

$$G_{i+1} \leq G_i + 2\sigma_i k_i - T_i. \tag{57}$$

By computing the conditional expectation w.r.t $\mathcal{F}_i$, we derive

$$\mathbb{E}[G_{i+1} \mid \mathcal{F}_i] \leq \mathbb{E}[G_i \mid \mathcal{F}_i] + 2\sigma_i\mathbb{E}[k_i \mid \mathcal{F}_i] - \mathbb{E}[T_i \mid \mathcal{F}_i]. \tag{58}$$

But,

$$\begin{aligned} 2\sigma_i\mathbb{E}[k_i \mid \mathcal{F}_i] &= 2\sigma_i\mathbb{E}[\langle q_{i-1}-\mathcal{K}n_{i-1}, m_{i+1}-m_i\rangle \mid \mathcal{F}_i]\\ &\quad -4\sigma_i\mathbb{E}[\langle q_i-\mathcal{K}n_i, m_{i+1}-m_i\rangle \mid \mathcal{F}_i]\\ &\leq \frac{\sigma_i}{\epsilon}\mathbb{E}[\|q_{i-1}-\mathcal{K}n_{i-1}\|^2 \mid \mathcal{F}_i] + \sigma_i\epsilon\mathbb{E}[\|m_{i+1}-m_i\|^2 \mid \mathcal{F}_i]\\ &\quad +2\sigma_i\frac{2}{\epsilon}\mathbb{E}[\|q_i-\mathcal{K}n_i\|^2 \mid \mathcal{F}_i] + 2\sigma_i\frac{\epsilon}{2}\mathbb{E}[\|m_{i+1}-m_i\|^2 \mid \mathcal{F}_i]. \end{aligned} \tag{59}$$

Using (59) in (58), we obtain

$$\begin{aligned}\mathbb{E}[G_{i+1} \mid \mathcal{F}_i] &\leq \mathbb{E}[G_i \mid \mathcal{F}_i] + \frac{\sigma_i}{\epsilon}\mathbb{E}[\|q_{i-1}-\mathcal{K}n_{i-1}\|^2 \mid \mathcal{F}_i]\\ &\quad +2\sigma_i\epsilon\mathbb{E}[\|m_{i+1}-m_i\|^2 \mid \mathcal{F}_i]\\ &\quad +\frac{4\sigma_i}{\epsilon}\mathbb{E}[\|q_i-\mathcal{K}n_i\|^2 \mid \mathcal{F}_i] - \mathbb{E}[T_i \mid \mathcal{F}_i]\\ &\leq \mathbb{E}[G_i \mid \mathcal{F}_i] + 5\frac{\sigma_i}{\epsilon}d - \mathbb{E}[T_i \mid \mathcal{F}_i]. \end{aligned} \tag{60}$$

Applying Assumption 4.3(e) and computing expectation on (60), yields

$$\mathbb{E}[G_{i+1}] \leq \mathbb{E}[G_1] + \sum_{k=i_0}^{i} 5\frac{\sigma_k}{\epsilon}d. \tag{61}$$

Now, by the definition of G_i and Assumption 4.3, we obtain

$$\begin{aligned}
G_i &= \|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + (1 + 2\varrho + \sigma_i\mathcal{L}(1+\sqrt{2}))\|m_i - m_{i-1}\|^2 \\
&\quad +(\varrho + \sigma_i\mathcal{L})\|m_i - n_{i-1}\|^2 + 2\sigma_i\alpha_{i-1} \\
&\geq \|m_i - e\|^2 - \varrho\|m_{i-1} - e\|^2 + (\varrho + \sigma_i\mathcal{L})\|m_i - n_{i-1}\|^2 \\
&\quad +(1 + 2\varrho + \sigma_i\mathcal{L}(1+\sqrt{2}))\|m_i - m_{i-1}\|^2 - 2\sigma_i\mathcal{L}\|n_{i-1} - e\|\|m_i - m_{i-1}\| \\
&\geq \|m_i - e\|^2 - 2\varrho\|m_i - e\|^2 - 2\varrho\|m_i - m_{i-1}\|^2 + (\varrho + \sigma_i\mathcal{L})\|m_i - n_{i-1}\|^2 \\
&\quad +(1 + 2\varrho + \sigma_i\mathcal{L}(1+\sqrt{2}))\|m_i - m_{i-1}\|^2 - \sigma_i\mathcal{L}\|n_{i-1} - m_i\|^2 \\
&\quad -\sigma_i\mathcal{L}\|m_i - e\|^2 - 2\sigma_i\mathcal{L}\|m_i - m_{i-1}\|^2 \\
&= \frac{1}{2}\|m_i - e\|^2 + (\frac{1}{2} - 2\varrho - \sigma_i\mathcal{L})\|m_i - e\|^2 \\
&\quad + \left(1 + \sigma_i\mathcal{L}(\sqrt{2}+1) - 2\sigma_i\mathcal{L}\right)\|m_i - m_{i-1}\|^2 + \varrho\|n_{i-1} - m_i\|^2 \\
&\geq \frac{1}{2}\|m_i - e\|^2. \qquad (62)
\end{aligned}$$

Using (61) and (62), and noting that $\sigma_k = \frac{1}{2\mathcal{L}(1+\sqrt{2})(k+1)}$ ($\sum_{k=i_0}^{i} \frac{1}{k} = \log(i)$), we arrive at the desired conclusion. □

5 Numerical Applications to Traffic Network Problems

A large portion of urban activities revolves around the movement of people and goods across the transportation network, making an efficient and well-functioning transportation system crucial for both the economic vitality and the quality of life in urban areas (see [46]). In the context of planning for future investments and operational policies, a thorough examination of the transportation system is a critical element of the planning process. Traffic flow models are commonly developed to manage and optimize these movements (see, for example, [15, 29, 31, 46]). This section provides applications to Traffic Equilibrium Problems (TEPs) both in the deterministic and stochastic settings. Drivers compete to find the most cost-effective routes based on a given travel demand in traffic-congested networks. The TEP model is widely used in transportation planning to identify flow patterns that satisfy the equilibrium condition, where no driver can reduce their travel time/cost by switching routes unilaterally. At equilibrium, the interactions among drivers stabilize into a consistent traffic flow pattern, which planners can then utilize to predict future traffic behavior.

We implement these applications via numerical computations and compare our proposed algorithm (Algorithm 3.2) with the stochastic forward-backward-forward splitting algorithm (SFBF) in [33], the stochastic forward-reflected-backward splitting algorithm (SFRB), called variance reduced forward-reflected-backward in [1] and the reflected-forward-backward splitting algorithm (SRFB) in [34].

Our code is implemented in Matlab 2023 (b) and executed on a personal computer equipped with an Intel(R) Core(TM) i5-2600 CPU operating at 2.30GHz, with 8.00 GB of RAM.
During the computations, we utilize the following:

- Algorithm 3.2 (Our Algorithm): $\sigma = 0.2$ and $\varrho = 0.1$.
- SFBF in [33]: $\epsilon_i = \frac{1}{i^2}$, $\lambda_i = \frac{0.9i}{i+1}$ and $\varrho = 0.1$.
- SFRB in [1]: $\tau = 0.2$.
- SRFB in [34]: $\sigma_i = \frac{10i}{25i+1}$.

We apply the stopping criterion $\text{TOL}_i := 0.5\|m_i - \text{Res}^{\mathcal{J}}(m_i - \mathcal{K}m_i)\|^2 < \epsilon$ across all algorithms, where ϵ represents the predefined error tolerance.

Example 5.1 **Deterministic Setting**

In this example, we consider a deterministic TEP. In the deterministic setting, it is assumed that all drivers have perfect knowledge of travel times and always choose the route with the shortest travel time. Travel times are deterministic and solely depend on traffic flow. A solution to a deterministic TEP signifies a stable state in a transportation network where all drivers have chosen routes such that no one can reduce their travel time by switching to a different route. At this point, all used routes between an origin and destination have equal and minimal travel times, and unused routes have greater travel times.

We focus our attention to a traffic flow network composed of $\mathcal{N}$ nodes interconnected by directed edges [20, Chapter 6]. We label the collection of edges within this network as D, while W represents the set of directed node pairs. For a given pair $w = (a, b) \in W$, where node a serves as the origin node and node b as the destination node, P_w represents the collection of all possible routes connecting node a to node b. The entire network's set of paths is denoted by $Q = \cup_{w \in W} P_w$. For each path $p \in Q$, the flow along that path is represented by e_p. Additionally, each pair $w \in W$ is linked to a positive value d_w, which specifies the flow demand from node a to node b. The set of all feasible flow configurations is denoted by K and is defined as follows:

$$K = \prod_{w \in W} K_w = \left\{ e \in \mathbb{R}^{|Q|} \ : \ \sum_{p \in P_w} e_p = d_w, \ \forall w \in W; \ e_p \geq 0, \ \forall p \in P_w \right\},$$

where

$$K_w = \left\{ e \in \mathbb{R}^{|P_w|} \ : \ \sum_{p \in P_w} e_p = d_w, \ e_p \geq 0, \ \forall p \in P_w \right\}.$$

The flow f_d associated with the edge $d \in D$ can be defined, provided that the flow vector e is known, as follows

$$f_d = \sum_{p \in Q} \xi_{pd} e_p = \sum_{p \in P_w} \sum_{w \in W} \xi_{pd} e_p,$$

where

$$\xi_{pd} = \begin{cases} 1 & \text{if } d \text{ is in the path } p, \\ 0 & \text{otherwise.} \end{cases}$$

Additionally, the cost associated with $d \in D$ can be determined by $t_d = C_d(f_d)$. Also, the cost corresponding to each path p can be computed as follows (see [16]):

$$F_p(e) = \sum_{d \in D} \xi_{pd} t_d.$$

A flow vector $e^* \in K$ is termed an equilibrium vector if it meets the condition:

$$\forall q \in P_w,\ e_q^* > 0 \implies F_q(e^*) = \min_{p \in P_w} F_p(e^*),\ \forall w \in W. \tag{63}$$

This implies that, at equilibrium, the traffic network ensures that among all paths in P_w, the path carrying traffic incurs the smallest cost. Problem (63) can be re-expressed as problem (1) (see, for instance, [20, Theorem 6.1]), where $\mathcal{J}$ represents the normal cone to K and $\mathcal{K}(e)$ is a vector in $\mathbb{R}^{|Q|}$ with components $F_p(e)$.
Note: In the context of the traffic equilibrium problem:

- The Hilbert space $\mathcal{V}$ serves as the underlying function space in which the problem is formulated.
- The set $K \subset \mathcal{V}$ denotes the feasible flows, representing all flow configurations that satisfy network constraints.
- The operator $\mathcal{K}$ is the cost operator, mapping flows to their corresponding travel costs.
- The operator $\mathcal{J}$ represents the normal cone to the set K, capturing directions in which movement would lead outside the feasible region. Mathematically, this is expressed as follows:

$$\mathcal{J}(e) = N_K(e) = \{f \in \mathcal{V} : \langle y - e, f \rangle \leq 0 \quad \forall y \in K\},$$

  where N_K is the normal cone of K at a point e.

For the numerical implementation, we examine a traffic network consisting of 5 nodes and 8 edges, denoted as di $(i = 1, 2, \ldots, 8)$, as illustrated in Figure 1 (refer to [16] for more details).
The cost function C_d is given by

$$C_d(f) = \begin{cases} \alpha_d f + \beta_d & \text{if } 0 \leq f \leq \kappa_d, \\ \gamma_d f + \alpha_d \kappa_d + \beta_d - \gamma_d \kappa_d & \text{if } f > \kappa_d, \end{cases} \tag{64}$$

where the values for α_d, β_d, γ_d, and κ_d are provided in Table 1.
The expression (64) is explained as follows (see also [16]):

1. As the traffic flow f rises yet stays under the edge's permissible threshold (i.e., $f \leq \kappa_d$), the cost rises at a modest rate α_d. In this case, β_d is the smallest cost for the edge d.

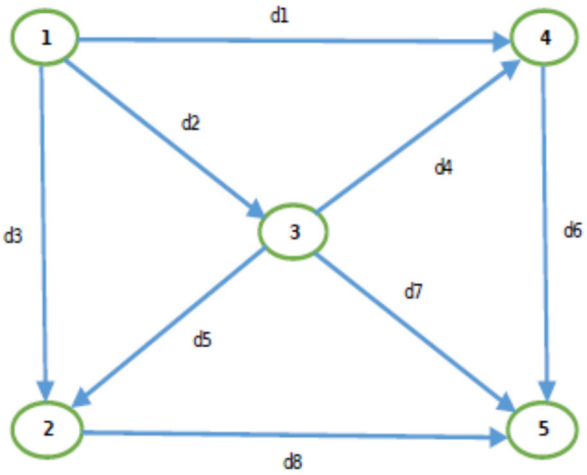


Fig. 1 A five-node traffic network

Table 1 Cost function parameters in the context of the transport equilibrium framework

d	α_d	β_d	γ_d	κ_d
$d1$	1	100	10	100
$d2$	1.1	120	11	120
$d3$	0.9	80	9	80
$d4$	0.1	150	8	150
$d5$	0.1	70	11	70
$d6$	0.7	140	12	210
$d7$	1.2	150	13	150
$d8$	0.6	160	14	250

2. Conversely, as the traffic flow f surpasses permissible threshold κ_d, the cost escalates at a heightened rate γ_d, signaling a traffic jam.

Now, consider the case where $W = \{(1,5)\}$ and $d_{(1,5)} = 1000$. The set $P_{(1,5)}$ consists of five paths: $p_1 = d1 \rightarrow d6$, $p_2 = d3 \rightarrow d8$, $p_3 = d2 \rightarrow d7$, $p_4 = d2 \rightarrow d5 \rightarrow d8$, and $p_5 = d2 \rightarrow d4 \rightarrow d6$ (where di $(i = 1, 2, \ldots, 8)$ represent the edges). Given that C_d is increasing, $\mathcal{K}$ exhibits monotonicity (see, [16]). For this example, we use the following initial points: Case 1: $m_1 = (100, 100, 50, 500, 250)$, $m_0 = (100, 100, 50, 500, 250)$; Case 2: $m_1 = (10, 190, 150, 400, 250)$, $m_0 = (100, 100, 50, 500, 250)$; Case 3: $m_1 = (5, 20, 175, 600, 200)$, $m_0 = (300, 200, 100, 100, 300)$ and Case 4: $m_1 = (500, 200, 35, 165, 100)$, $m_0 = (30, 2, 1, 17, 950)$.
The numerical outcomes in this case are presented in Figure 2 and Table 2. In the tables and figures, "Iter" refers to iteration numbers, while "CPU" represents the time from CPU recording in seconds.

Example 5.2 **Stochastic Setting**

In this example, we consider the TEP in a stochastic setting. Stochastic TEP accounts for uncertainty and variability in driver perception and behavior. Here, drivers

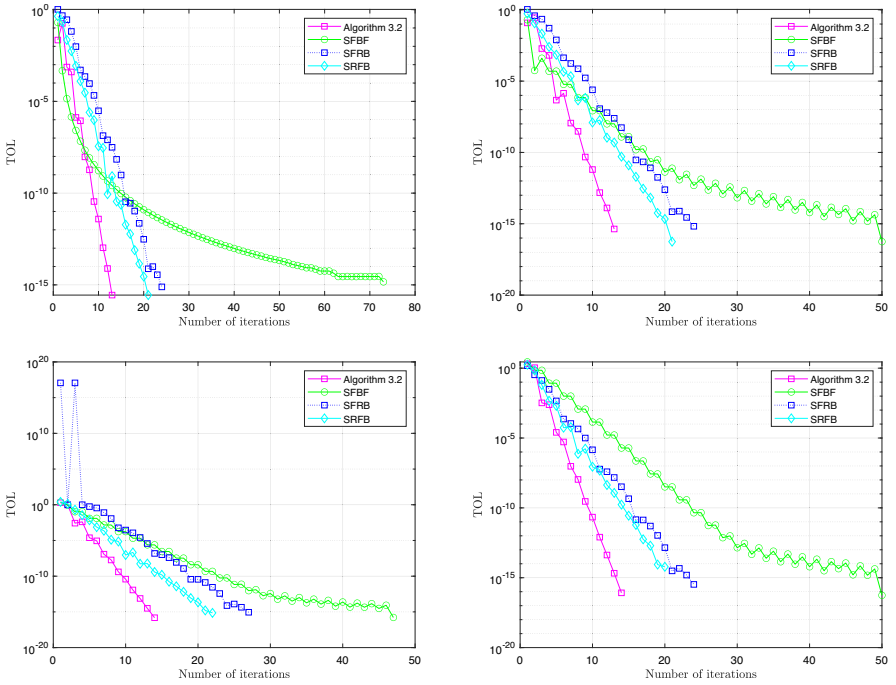


Fig. 2 Performance of TOL_i with $\epsilon = 10^{-15}$: Upper Left: "Case 1"; Upper Right: "Case 2"; Lower Left: "Case 3"; Lower Right: "Case 4"

Table 2 Our algorithm and comparison with others. $\epsilon = 10^{-15}$

Algorithms	**Case 1**		**Case 2**		**Case 3**		**Case 4**	
	CPU	Iter	CPU	Iter	CPU	Iter	CPU	Iter
Algorithm 3.2	0.0055	13	0.0021	13	0.0019	14	0.0081	14
SFBF in [33]	0.0195	74	0.0174	50	0.0197	47	0.0242	50
SFRB in [1]	0.0113	24	0.0105	24	0.0154	27	0.0209	24
SRFB in [34]	0.0101	21	0.0106	21	0.0110	22	0.0100	21

are assumed to have imperfect or incomplete information, and their route choices are influenced by perceived travel times, which are treated as random variables. As a result, route choice becomes probabilistic. Here, equilibrium is achieved when the expected perceived time of all used routes is equal, and no driver can improve their expected utility by changing routes. While deterministic TEP provides a simplified and idealized view of traffic flow, dynamic TEP offers a more realistic representation by incorporating human behavior and uncertainty, making it better suited for modeling real-world traffic conditions.

Consider a traffic network represented by $[\mathcal{N}, \mathcal{A}]$, where $\mathcal{N}$ is the collection of nodes and $\mathcal{A}$ is the collection of edges. Let Ω be the sample space representing variables that contribute to traffic network fluctuations, including accidents, weather conditions, and

so on. We let $\xi(\omega)$ be the realization for each $\omega \in \Omega$. For simplicity, we still denote the realization by ξ without any confusion. We present a summary of other notations used:

$\mathcal{I}$	set representing all origin-destination (OD) pairs,
$\mathcal{R}_i$	collection of routes available for the OD pair i,
Γ	incidence matrix linking OD pairs to their corresponding routes in the network, specific to each realization ξ,
$h_r(\xi)$	flow amount on route r,
$\hat{d}_i(\xi)$	demand function associated with the OD pair i,
$C_r(h(\xi), \xi)$	cost function corresponding to route r.

We define $\mathcal{R} = \bigcup_{i\in\mathcal{I}} \mathcal{R}_i$ as the union of all routes across the OD pairs. Additionally, the vectors $\hat{d}(\xi)$, $h(\xi)$, and $C(h(\xi), \xi)$ represent the collections of $\hat{d}_i(\xi)$, $h_r(\xi)$, and $C_r(h(\xi), \xi)$, respectively, for all $i \in \mathcal{I}$ and $r \in \mathcal{R}$.
The uncertain TEP model focuses on minimizing route costs in accordance with the Wardrop equilibrium principle, which is represented as the following stochastic variational inequality problem [46]: search $h^* \in K$ such that

$$\langle h - h^*, \mathbb{E}[C(h^*, \xi)]\rangle \geq 0, \quad \forall h \in K = \left\{h \geq 0 | \Gamma^T h = \hat{d}(\xi)\right\} \subset \mathbb{R}^{|\mathcal{R}|}, \tag{65}$$

where $\mathbb{R}^{|\mathcal{R}|}$ is the set of all flow assignments which are not necessarily feasible. The cost function for a route, C_r, is given as follows:

$$C_r(h, \xi) = \sum_{a\in\mathcal{A}} \kappa_{ar} t_a\Big((\hat{K}h)_a, \xi\Big), \tag{66}$$

where $\hat{K} = (\kappa_{ar})_{(a,r)\in\mathcal{A}\times\mathcal{R}}$ represents the incidence matrix of the link-route whose elements are given as follows:

$$\kappa_{ar} = \begin{cases} 1 & \text{if route } r \text{ passes through link } a \\ 0 & \text{otherwise,} \end{cases} \tag{67}$$

and the function t_a represents travel time, derived from the Bureau of Public Roads' 1964 travel time function, and is expressed as follows:

$$t_a((\hat{K}h)_a, \xi) = t_a^0(1 + 0.15((\hat{K}h)_a/c_a(\xi))^4), \tag{68}$$

where t_a^0 represents the free-flow travel time for link a, and $c_a(\xi)$ indicates the perturbed link capacity, where the probability $P\{c_a(\xi) = 0\} = 0$ for every $a \in \mathcal{A}$. Given the definition of $C(h, \xi)$, it is straightforward to observe that $\mathcal{K}(h) = \mathbb{E}[C(h, \xi)]$ is a monotone function.

For numerical implementation, we focus on a traffic network composed of 13 nodes, 19 edges, and 4 origin-destination (OD) pairs: $(1 \to 2)$, $(1 \to 3)$, $(4 \to 2)$, and $(4 \to 3)$. These details are illustrated in Table 5. The characteristics of the network and the corresponding OD demand are presented in Tables 3 and 4, while Table 5 provides an overview of the routes and edges within the network. In this scenario, we assume that the random variable ξ, which reflects the weather factor influencing link capacity, is distributed uniformly in $[0.49, 0.51]$. Furthermore, we define the sets as follows: $\mathcal{A} = \{1, \ldots, 19\}$ for the edges, $\mathcal{I} = \{(1 \to 2), (1 \to 3), (4 \to 2), (4 \to 3)\}$ for the OD pairs, and $\mathcal{R} = \{1, \ldots, 8, 9, \ldots, 14, 15, \ldots, 19, 20, \ldots, 23\}$ for the routes. The demand vector is given by $\hat{d}(\xi) = (400, 800, 600, 200)$,

$$\Gamma^T = \begin{bmatrix} 1&1&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&1&1&1&1&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1 \end{bmatrix},$$

and

$$\hat{K} = \begin{bmatrix} 0&0&0&1&1&1&1&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0 \\ 1&1&1&0&0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&1&1&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1 \\ 0&0&0&1&0&1&1&0&0&0&1&1&0&0&1&1&1&0&0&1&1&0&0 \\ 0&0&0&0&1&0&0&0&0&0&0&0&1&1&0&0&0&1&0&0&0&1&0 \\ 0&1&0&1&0&0&1&1&1&0&1&0&0&0&1&1&0&0&0&1&0&0&0 \\ 0&0&1&0&0&1&0&0&0&1&0&1&0&0&0&0&1&0&0&0&1&0&0 \\ 0&1&0&1&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&1&0&1&0&0&0&0&1&0&0&0&1&0&0&0 \\ 1&1&0&1&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0&0&0&1&0&0&0&0&1&1&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1 \\ 0&0&1&0&1&1&0&0&0&1&0&1&1&0&0&0&1&1&1&0&1&1&0 \\ 0&0&1&0&1&1&1&1&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&1&1&1&1&0&0&0&0&0&0&1&1&1&0 \\ 0&1&1&0&0&0&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&1 \end{bmatrix}.$$

For this example, we choose m_1 and m_0 randomly in $\mathbb{R}^{23}$ with different predetermined errors, $\epsilon = 10^{-15}, 10^{-20}, 10^{-25}, 10^{-30}$. The numerical outcomes in this case are presented in Figure 3 and Table 6.

6 Conclusion

We have studied an inertial stochastic reflected-forward-backward splitting algorithm designed to address monotone inclusion problems in real separable Hilbert spaces.

Table 3 Network characteristics

Edge a	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
t_a^0	7	9	9	12	3	9	5	13	5	9	9	10	9	6	9	8	7	14	11
$n_a(\xi)$	800	400	200	800	350	400	800	500 ξ	250	300	550	550	600	700	500	300	400 ξ	400	600

Table 4 Demand for travel between origin-destination (OD) pairs

O \ D	2	3
1	400	800
4	600	200

Table 5 Origin-destination pairs, routes and edges

OD pairs	r	Edges	OD pairs	r	Edges
$1 \to 2$	1	2, 18, 11	$4 \to 2$	15	3, 5, 7, 9, 11
	2	2, 17, 7, 9, 11		16	3, 5, 7, 10, 15
	3	2, 17, 8, 14, 15		17	3, 5, 8, 14, 15
	4	1, 5, 7, 9, 11		18	3, 6, 12, 14, 15
	5	1, 6, 12, 14, 15		19	4, 12, 14, 15
	6	1, 5, 8, 14, 15	$4 \to 3$	20	3, 5, 7, 10, 16
	7	1, 5, 7, 10, 15		21	3, 5, 8, 14, 16
	8	2, 17, 7, 10, 15		22	3, 6, 12, 14, 16
$1 \to 3$	9	2, 17, 7, 10, 16		23	4, 13, 19
	10	2, 17, 8, 14, 16			
	11	1, 5, 7, 10, 16			
	12	1, 5, 8, 14, 16			
	13	1, 6, 12, 14, 16			
	14	1, 6, 13, 19			

Table 6 Comparison of algorithms with different ϵ

Algorithms	$\epsilon = 10^{-15}$		$\epsilon = 10^{-20}$		$\epsilon = 10^{-25}$		$\epsilon = 10^{-30}$	
	CPU	Iter	CPU	Iter	CPU	Iter	CPU	Iter
Algorithm 3.2	0.0529	10	0.0719	11	0.2062	12	0.2061	12
SFBF in [33]	0.2048	30	0.6138	68	1.5723	68	1.5723	68
SFRB in [1]	0.1579	19	0.1774	20	1.1377	20	1.1315	20
SRFB in [34]	0.1021	18	0.1051	21	1.0768	21	1.0768	21

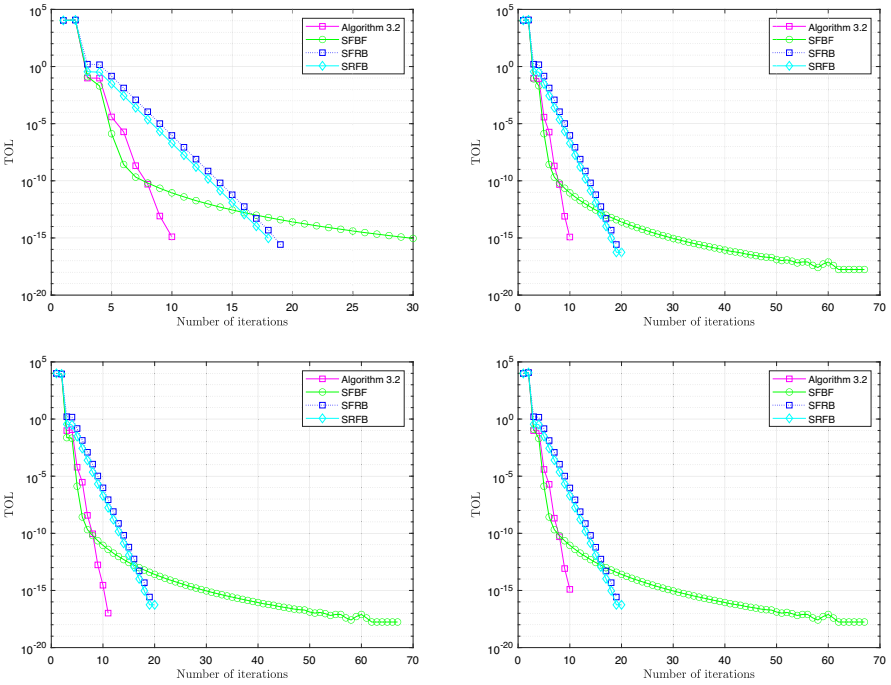


Fig. 3 The behavior of TOL_i: Upper Left: $\epsilon = 10^{-15}$; Upper Right: $\epsilon = 10^{-20}$; Lower Left: $\epsilon = 10^{-25}$; Lower Right: $\epsilon = 10^{-30}$

Our approach stands out by effectively combining the strengths of inertial techniques with stochastic methods, yielding a convergence rate of $\mathcal{O}(\log(i)/(i))$ in expectation for strongly monotone cases, and ensuring almost sure convergence in the general monotone scenario.

Unlike previous inertial methods, our algorithm requires only a single unbiased estimate of the single-valued operator and a backward evaluation of the set-valued operator at each step, making it computationally efficient and suitable for large-scale optimization problems where operator evaluations are costly. This approach also broadens the applicability of inertial stochastic algorithms to non-cocoercive operators, addressing a significant gap in the existing literature.

We have demonstrated the potential of our algorithm in practical applications by focusing on traffic flow network problems, a critical area in urban planning and transportation engineering. The ability of our method to provide solutions to traffic equilibrium problems in deterministic and stochastic settings highlights its versatility and effectiveness in real-world scenarios. Moreover, when compared with related algorithms, our algorithm performed better.

Acknowledgements This work was done when CI visited Queen's University Belfast, United Kingdom. CI was supported by the London Mathematical Society Grant Scheme 5 (Reference No 52303) and the National Graduate Academy for Mathematical and Statistical Sciences, South Africa (Reference No OMT-ECA-053) for this visit. It is also supported by the Natural Science Foundation of Guangxi Grant Nos. 2021GXNSFA196004 and 2024GXNSFBA010337, the NNSF of China Grant No. 12371312, the Natural

Science Foundation of Chongqing Grant No. CSTB2024NSCQ-JQX0033. The authors thank the reviewers for providing valuable comments and useful suggestions that helped improve the earlier version.

Author Contributions The authors contributed equally to this paper.

Data Availability Data sharing not applicable to this article as no data sets were generated or analysed during the current study.

Declarations

Ethical Approval Not applicable.

Competing interests There is no conflict of interests.

References

1. Alacaoglu, A., Malitsky, Y., Cevher, V.: Forward-reflected-backward method with variance reduction. Comput. Optim. Appl. **80**, 321–346 (2021)
2. Attouch, H., Cabot, A.: Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. Math. Program. **184**, 243–287 (2020)
3. Attouch, H., Cabot, A.: Convergence of a relaxed inertial forward–backward algorithm for structured monotone inclusions. Appl. Math. Optim. **80**, 547–598 (2019)
4. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
5. Briceño-Arias, L.M., Combettes, P.L.: A monotone skew splitting model for composite monotone inclusions in duality. SIAM J. Optim. **21**, 1230–1250 (2011)
6. Bui, H.T., Burachik, R.S., Nurminski, E.A., Tam, M.K.: Single-projection procedure for infinite dimensional convex optimization problems. SIAM J. Optim. **34**, 1646–1678 (2024)
7. Cevher, V., Vũ, B.C.: A reflected forward-backward splitting method for monotone inclusions involving Lipschitzian operators. Set-Valued Var. Anal. **29**, 163–174 (2021)
8. Cholamjiak, W., Cholamjiak, P., Suantai, S.: An inertial forward–backward splitting method for solving inclusion problems in Hilbert spaces. J. fixed point theory appl. **20**, 42 (2018)
9. Combettes, P.L., Pesquet, J.-C.: Stochastic approximations and perturbations in forward–backward splitting for monotone operators. Pure Appl. Funct. Anal. **1**, 13–37 (2016)
10. Combettes, P.L., Pesquet, J.-C.: Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping. SIAM J. Optim. **25**, 1221–1248 (2015)
11. Cui, S., Shanbhag, U.V.: Variance-reduced proximal and splitting schemes for monotone stochastic generalized equations. arXiv:2008.11348
12. Cui, S., Shanbhag, U.V., Staudigl, M., Vuong, P.: Stochastic relaxed inertial forward-backward-forward splitting for monotone inclusions in Hilbert spaces. Comput. Optim. Appl. **83**, 465–524 (2022)
13. Dong, Q.L., Lu, Y.Y., Yang, J.: The extragradient algorithm with inertial effects for solving the variational inequality. Optimization **65**, 2217–2226 (2016)
14. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, New York (2016)
15. Gottlich, S., Herty, M., Moutari, S., Weissen, J.: Second-order traffic flow models on networks. SIAM J. Appl. Math. **81**, 58–281 (2021)
16. Hieu, D.V., Muu, L.D., Quy, P.K.: Strong convergence of multi-parameter projection methods for variational inequality problems. Math. Model. Anal. **27**, 242–262 (2022)
17. Izuchukwu, C., Reich, S., Shehu, Y.: Relaxed inertial methods for solving the split monotone variational inclusion problem beyond co-coerciveness. Optimization **72**, 607–646 (2023)
18. Izuchukwu, C., Reich, S., Shehu, Y.: Convergence of two simple methods for solving monotone inclusion problems in reflexive Banach spaces. Results Math. **77**, 143 (2022)
19. Izuchukwu, C., Shehu, Y., Yao, J.C.: New inertial forward-backward type for variational inequalities with quasi-monotonicity. J. Global Optim. **84**, 441–464 (2022)
20. Konnov, I.V.: Equilibrium Models and Variational Inequalities. Amsterdam (2007)

21. Ledoux, M., Talagrand, M.: Probability in Banach Spaces: Isoperimetry and Processes. Springer, New York (1991)
22. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
23. Lorenz, D.A., Pock, T.: An inertial forward-backward algorithm for monotone inclusions. J. Math. Imaging Vis. **51**, 311–325 (2015)
24. Malitsky, Y.: Golden ratio algorithms for variational inequalities. Math. Program. **184**, 383–410 (2020)
25. Malitsky, Y.: Proximal extrapolated gradient methods for variational inequalities. Optim. Methods Softw. **33**, 140–164 (2018)
26. Malitsky, Y.V.: Projected reflected gradient methods for monotone variational inequalities. SIAM J. Optim. **25**, 502–520 (2015)
27. Malitsky, Y., Tam, M.K.: Resolvent splitting for sums of monotone operators with minimal lifting. Math. Program. **201**, 231–262 (2023)
28. Malitsky, Y., Tam, M.K.: A forward-backward splitting method for monotone inclusions without cocoercivity. SIAM J. Optim. **30**, 1451–1472 (2020)
29. Moutari, S., Herty, M., Klein, A., Oeser, M., Steinauer, B., Schleper, V.: Modelling road traffic accidents using macroscopic second-order models of traffic flow. IMA J. Appl. Math. **78**, 1087–1108 (2013)
30. Moudafi, A., Oliny, M.: Convergence of a splitting inertial proximal method for monotone operators. J. Comput. Appl. Math. **155**, 447–454 (2003)
31. Moutari, S., Rascle, M.: A hybrid Lagrangian model based on the Aw–Rascle traffic flow model. SIAM J. Appl. Math. **68**, 413–436 (2007)
32. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. Soviet Math. Doklady **27**, 372–376 (1983)
33. Nguyen, V.D., Vinh, N.T.: A Tseng type stochastic forward-backward algorithm for monotone inclusions. https://arxiv.org/abs/2202.09759
34. Nguyen, V.D., Vũ, B.C.: Convergence analysis of the stochastic reflected-forward–backward splitting algorithm. Optim. Lett. **16**, 2649–2679 (2022)
35. Ortúzar, J.D., Willumsen, L.: Modelling Transport. Wiley, Chichester, New York (2024)
36. Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert spaces. J. Math. Anal. Appl. **72**, 383–390 (1979)
37. Polyak, B.T.: Some methods of speeding up the convergence of iterates methods. U.S.S.R Comput. Math. Phys. **4**(5), 1–17 (1964)
38. Robbins, H., Siegmund, D.: A convergence theorem for non negative almost supermartingales and some applications. In: Rustagi, J.S. (ed.): Optimizing Methods in Statistics, pp. 233–257. Academic Press, New York (1971)
39. Rosasco, L., Villa, S., Vũ, B.C.: Stochastic forward-backward splitting for monotone inclusions. J. Optim. Theory Appl. **169**, 388–406 (2016)
40. Su, W., Boyd, S., Candes, E.J.: A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. J. Mach. Learn. Res. **17**, 1–43 (2016)
41. Tan, B., Qin, X., Yao, J.C.: Strong convergence of inertial projection and contraction methods for pseudomonotone variational inequalities with applications to optimal control problems. J. Glob. Optim. **82**, 523–557 (2022)
42. Tran-Dinh, Q., Zhu, Y.: Non-stationary first-order primal-dual algorithms with faster convergence rates. SIAM J. Optim. **30**, 866–2896 (2020)
43. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control. Optim. **38**, 431–446 (2000)
44. Vũ, B.C.: Almost sure convergence of the forward-backward-forward splitting algorithm. Optim. Lett. **10**, 781–803 (2016)
45. Vũ, B.C.: A variable metric extension of the forward–backward–forward algorithm for monotone operators. Numer. Funct. Anal. Optim. **34**, 1050–1065 (2013)
46. Wang, S., Tao, H., Lin, R., Cho, Y.J.: A self-adaptive stochastic subgradient extragradient algorithm for the stochastic pseudomonotone variational inequality problem with application. Z. Angew. Math. Phys. **73**, 164 (2022)
47. Zeng, S., Papageorgiou, N.S., Winkert, P.: Inverse problems for double-phase obstacle problems with variable exponents. J. Optim. Theory Appl. **196**, 666–699 (2023)
48. Zeng, S., Rădulescu, V.D., Winker, P.: Nonlocal double phase implicit obstacle problems with multivalued boundary conditions. SIAM J. Math. Anal. **56**, 877–91 (2024)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.