Journal of Computational Mathematics
Vol.xx, No.x, 2025, 1–25.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2503-m2024-0175

AN INERTIAL-RELAXED THREE-TERM CONJUGATE GRADIENT PROJECTION METHOD FOR LARGE-SCALE UNCONSTRAINED NONLINEAR PSEUDO-MONOTONE EQUATIONS WITH APPLICATIONS*

Jianghua Yin, Qiongxuan Huang, Jinbao Jian[1)] and Daolan Han
School of Mathematical Sciences, Center for Applied Mathematics of Guangxi,
Guangxi Minzu University, Nanning 530006, China
Emails: jianghuayin1017@126.com, hqx0710@163.com, jianjb@gxu.edu.cn, 546474207@qq.com

Abstract

In this paper, a three-term conjugate gradient projection method that employs both inertial and relaxed techniques is proposed to find approximate solutions for unconstrained nonlinear pseudo-monotone equations. The search direction generated at each iteration possesses the sufficient descent and trust region properties independent of the line search technique used. The global convergence of the proposed method is shown without the Lipschitz continuity of the underlying mapping. Moreover, the asymptotic and non-asymptotic convergence rates in terms of iteration complexity are established under the local Lipschitz continuity assumption. To our knowledge, this is the first time in the literature that an iteration-complexity analysis has been conducted for inertial-relaxed gradient-type projection methods. Numerical experiments on large-scale benchmark test problems are conducted to demonstrate the effectiveness and efficiency of the proposed algorithm. Furthermore, the applicability and practicality of the proposed method are also verified by applying it to solve sparse signal restoration problems.

Mathematics subject classification: 65K10, 90C56.
Key words: Inertial-relaxed technique, Gradient-type projection method, Global convergence, Iteration complexity, Sparse signal restoration.

1. Introduction

In this paper, we consider solving the system of unconstrained nonlinear equations of the form

$$F(x) = 0, \tag{1.1}$$

where F is a continuous pseudo-monotone mapping from $\mathbb{R}^n$ to itself. The pseudo-monotonicity of the mapping F implies that

$$F(y)^\top (x-y) \geq 0 \;\; \Rightarrow \;\; F(x)^\top (x-y) \geq 0, \quad \forall\, x, y \in \mathbb{R}^n. \tag{1.2}$$

As is known to all, the monotonicity implies the pseudo-monotonicity, and the converse of this statement, however, is not necessarily true. We refer the interested readers to the comprehensive monograph [11] for more details.

* Received August 15, 2024 / Revised version received January 4, 2025 / Accepted March 7, 2025 / Published online April 21, 2025 /
1) Corresponding author

In the optimization community, solving the system of nonlinear equations is a fundamental and challenging problem. Nonlinear pseudo-monotone equations (1.1), viewed as an important part of the system of nonlinear equations, frequently appear in various disciplines and engineering applications, such as economic and chemical equilibrium problems [32], compressive sensing [22, 46], image restoration [7, 15, 16, 52], to mention just a few.

One of the most popular methods for solving unconstrained nonlinear monotone equations is the so-called gradient-type projection method (since the mapping F can be viewed as the gradient of a certain function), including spectral gradient projection methods [26,55], conjugate gradient projection methods (CGPMs) [9,47], spectral CGPMs [28,41], and three-term CGPMs [2, 35, 53, 54]. The authors in [42] directly extended the projected Newton method in [38] to solve nonlinear monotone equations with convex constraints. Since then, it has attracted much attention, and many gradient-type projection methods for constrained nonlinear monotone equations have sprung up such as [5, 13, 14, 27, 48, 49, 51].

Nowadays, with the continuous development of inertial techniques, the idea of inertial extrapolation has been widely applied in optimization algorithms to numerically accelerate convergence. For example, [18] proposed three gradient-type projection methods with an inertial term to solve unconstrained nonlinear monotone equations, and established the global convergence of these methods under the Lipschitz continuity of F. Additionally, a large number of numerical results in [18] illustrated the behavior of the accelerated variants compared to the corresponding unaccelerated counterparts. Subsequently, a closely related work [56] presented a fast inertial self-adaptive projection method with a fixed relaxation factor for solving unconstrained nonlinear monotone equations. The numerical results reported in [56] showed the efficiency of the algorithm. Quite recently, combining the inertial-relaxed technique in [3] with the Armijo line search technique, [50] introduced a family of inertial-relaxed derivative-free projection methods to deal with unconstrained nonlinear monotone equations, and showed the global convergence of the family without the Lipschitz continuity. Numerical experiments in [50] confirmed the accelerated convergence speed achieved by the inertial-relaxed technique. Moreover, we refer to [6, 17, 19, 31, 44] for the related studies on inertial-based projection methods for solving nonlinear monotone equations with convex constraints.

Moreover, to our knowledge, there is a small amount of related research on solving unconstrained or constrained nonlinear pseudo-monotone equations. We refer the readers to [21, 29] and the references therein for some numerical algorithms in solving nonlinear pseudo-monotone equations with convex constraints. However, existing methods for solving such equations have certain limitations; see, e.g. [21, 29]. For instance, most of the previous methods did not analyze the iteration complexity. This motivates us to seek a new and more effective algorithm to address these theoretical challenges.

Recently, Alves and Marcavillaca [4] proposed an inertial under-relaxed hybrid proximal extragradient method for solving the monotone inclusion problem, and analyzed the asymptotic convergence and non-asymptotic convergence rates in terms of iteration complexity. It is natural to extend the just-mentioned method to solve the problem under discussion. However, solving the proximal subproblem inexactly, at each iteration, may be as difficult as the original problem. Hence, a question arises: Based on the benchmark line of analysis in [4], can we devise an inertial-relaxed gradient-type projection method with iteration-complexity guarantee? We answer this question affirmatively in this paper.

Motivated by the aforementioned discussion, in this paper, we propose an inertial-relaxed three-term CGPM for solving (1.1). Importantly, following the outstanding work [4], we analyze the iteration complexity of the proposed method. This paper makes several significant contributions, which can be summarized as follows:

- We propose a three-term CGPM that incorporates the inertial-relaxed technique, which is expected to numerically accelerate convergence. The search direction generated satisfies the sufficient descent and trust region properties, which are independent of line searches.
- Under a summability condition, the global convergence is established without the Lipschitz continuity. Moreover, we provide a sufficient condition to ensure the summability condition.
- Under the local Lipschitz continuity condition, we derive asymptotic and non-asymptotic convergence rates in terms of iteration complexity. It is worth emphasizing that, to the best of our knowledge, this is the first time in the literature that an iteration-complexity analysis has been conducted for inertial-relaxed gradient-type projection methods for solving (1.1).
- Numerical experiments on large-scale benchmark test problems are conducted to verify the performance of the proposed method. Furthermore, we apply the proposed algorithm to solve the problem of sparse signal restoration, showcasing its effectiveness and potential.

To further explain the advantages of the proposed method, Table 1.1 provides the difference among existing methods. From Table 1.1, the main advantages of the proposed method are that it has a lower computational cost, the global convergence is independent of the Lipschitz continuity of F, and that both asymptotic and non-asymptotic convergence rates are analyzed under the local Lipschitz continuity of F.

Table 1.1: A comparison of existing methods.

Existing methods	**The problem**	**The used technique**	**Convergence analysis**	**Computational cost**	**Iteration-complexity analysis**
Spectral gradient projection methods [26, 55]	Unconstrained nonlinear monotone equations	-	Global convergence under the Lipschitz continuity	Cheap	No
CGPMs [9, 47]	Unconstrained nonlinear monotone equations	-	Global convergence under the Lipschitz continuity	Cheap	No
Spectral CGPMs [28, 41]	Unconstrained nonlinear monotone equations	-	Global convergence under the Lipschitz continuity	Cheap	No
Three-Term CGPMs [2, 35, 53, 54]	Unconstrained nonlinear monotone equations	-	Global convergence under the Lipschitz continuity	Cheap	No
Projected Newton-type method [42]	Nonlinear monotone equations with convex constraints	-	Global convergence and local convergence rate under certain conditions	High (solving the linear equations at each iteration)	No
Gradient-type projection methods [5, 13, 14, 27, 29, 48, 49, 51]	Nonlinear (pseudo-) monotone equations with convex constraints	-	Convergence under certain conditions	Cheap	No

Table 1.1: A comparison of existing methods. (cont'd)

Existing methods	**The problem**	**The used technique**	**Convergence analysis**	**Computational cost**	**Iteration-complexity analysis**
Inertial gradient-type projection methods [18, 56]	Unconstrained nonlinear monotone equations	Inertial technique	Global convergence under the Lipschitz continuity	Cheap	No
Inertial-relaxed DFPMs [50]	Unconstrained nonlinear monotone equations	Inertial-relaxed technique	Global convergence without the Lipschitz continuity	Cheap	No
Inertial-based projection methods [6, 19, 21, 31, 44]	Nonlinear (pseudo-) monotone equations with convex constraints	Inertial technique	Convergence under certain conditions	Cheap	No
Inertial under-relaxed hybrid proximal extragradient method [4]	Monotone inclusion problem	Inertial under-relaxed technique	Asymptotic and non-asymptotic convergence rates in terms of iteration complexity	High (solving the proximal subproblem exactly at each iteration)	Yes
The proposed method	**Unconstrained nonlinear pseudo-monotone equations**	**Inertial-relaxed technique**	**Global convergence without the Lipschitz continuity; Asymptotic and non-asymptotic convergence rates under the local Lipschitz continuity condition**	**Cheap**	**Yes**

This paper is organized as follows. In the next section, we describe the motivation and introduce the algorithm. Section 3 analyzes the global convergence, and we provide the analysis of asymptotic and non-asymptotic convergence rates in terms of iteration complexity in Section 4. Section 5 gives numerical experiments to illustrate the performance of the proposed method. We give concluding remarks in the last section.

2. Motivation and Algorithm

In this section, in order to clarify the motivation of this paper, we first review some related methods and then introduce the algorithm.

As is known to all, the so-called three-term conjugate gradient method (TTCGM) can effectively solve unconstrained optimization problems

$$\min_{x\in\mathbb{R}^n} f(x),$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is a continuously differentiable function. By a slight abuse of notation, we denote the gradient of f at $x\in\mathbb{R}^n$ by $F(x)$, i.e. $F(x):=\nabla f(x)$. Given the current iteration x_k, the iterative scheme for the classical TTCGM can be read as

$$x_{k+1}=x_k+\alpha_k d_k,\quad d_k=-F(x_k)+\beta_k d_{k-1}+\theta_k\hat{y}_{k-1},\quad k\ge 1,\quad d_0=-F(x_0). \tag{2.1}$$

Here, $\alpha_k > 0$ represents the steplength derived from the line search, d_k denotes the search direction, β_k and θ_k are scalar parameters, and $\hat{y}_{k-1} := F(x_k) - F(x_{k-1})$. Clearly, the classical conjugate gradient method is a special case of the TTCGM when $\theta_k \equiv 0$ in (2.1).

Recently, based on the ideas of TTCGMs in [23–25] and the hybrid CGPM in [40], the authors in [49] constructed a hybrid three-term CGPM to solve nonlinear monotone equations with convex constraints, where the search direction is defined as

$$d_k = -F(x_k) + \beta_k^{\mathrm{HTP}} d_{k-1} + \tilde{\theta}_k \hat{y}_{k-1}, \quad k \geq 1, \quad d_0 = -F(x_0), \tag{2.2}$$

where

$$\begin{aligned}
\beta_k^{\mathrm{HTP}} &= \frac{F(x_k)^\top \hat{y}_{k-1}}{\varpi_k} - \frac{\|\hat{y}_{k-1}\|^2 F(x_k)^\top d_{k-1}}{\varpi_k^2}, \\
\tilde{\theta}_k &= t_k \frac{F(x_k)^\top d_{k-1}}{\varpi_k}, \\
\varpi_k &= \max\left\{\mu \|d_{k-1}\| \|\hat{y}_{k-1}\|, d_{k-1}^\top \hat{y}_{k-1}, \|F(x_{k-1})\|^2\right\}
\end{aligned}$$

with $0 \leq t_k \leq \bar{t} < 1$ and $\mu > 0$. Throughout this paper, the symbol $\|\cdot\|$ stands for the Euclidean norm on $\mathbb{R}^n$ unless otherwise is mentioned. As elaborated in [49], the search direction defined in (2.2) is close to that of the memoryless Broyden-Fletcher-Goldfarb-Shanno (BFGS) method in [33,37]. The previous works have laid a solid foundation for our research. Inspired by [49,55], the search direction in this paper can be read as

$$d_k = \begin{cases} -F_k, & k = 0, \\ -F_k + \beta_k d_{k-1} + \theta_k \bar{y}_{k-1}, & k \geq 1, \end{cases} \tag{2.3}$$

where

$$\begin{aligned}
& s_{k-1} = w_k - w_{k-1}, \quad y_{k-1} = F_k - F_{k-1}, \\
& \lambda_{k-1} = 1 + \max\left\{0, -\frac{y_{k-1}^\top s_{k-1}}{\|F_k\| \cdot \|s_{k-1}\|^2}\right\}, \\
& t_k = \min\left\{\bar{t}, \max\left\{0, 1 - \frac{y_{k-1}^\top s_{k-1}}{\|y_{k-1}\|^2}\right\}\right\}, \quad \bar{y}_{k-1} = y_{k-1} + \lambda_{k-1}\|F_k\| s_{k-1}, \\
& v_k = \max\left\{\|F_{k-1}\|^2, \ \mu\|d_{k-1}\| \|\bar{y}_{k-1}\|, \ d_{k-1}^\top \bar{y}_{k-1}\right\}, \\
& \beta_k = \frac{F_k^\top \bar{y}_{k-1}}{v_k} - \frac{\|\bar{y}_{k-1}\|^2 F_k^\top d_{k-1}}{v_k^2}, \quad \theta_k = t_k \frac{F_k^\top d_{k-1}}{v_k}.
\end{aligned}$$

Here, for convenience in description, $F(w_k)$ has been abbreviated to F_k, i.e. $F_k := F(w_k)$, where w_k is defined in (2.4). Clearly, the main difference between (2.2) and (2.3) is the definition of $\bar{y}_{k-1}$ even if $w_k \equiv x_k$ for all k. It is worthwhile to mention that the main purpose of this paper is to show asymptotic and non-asymptotic convergence rates of the inertial-relaxed gradient-type projection method under some mild conditions mentioned in Section 1.

Combining the inertial-relaxed technique in [3,50] with the search direction designed in (2.3), we propose a new inertial-relaxed three-term CGPM (for simplicity, IRTTCGPMN), which is summarized as Algorithm 2.1.

Remark 2.1. (i) Compared with the algorithm in [50], the proposed IRTTCGPMN adopts a new search direction. On the other hand, IRTTCGPMN also employs an alternative Armijo line search, which is originally introduced in [30].

Algorithm 2.1: IRTTCGPMN.

Step 0. Input the initial point $x_0 \in \mathbb{R}^n$, the parameters $\nu > 0, \hat{\eta} > 0, \mu > 0, \gamma > 0$, $0 \leq \alpha < 1, 0 < \tau < 1, \sigma > 0, \epsilon > 0, 0 < \underline{\rho} < \overline{\rho} < 2$ and $0 \leq \bar{t} < 1$. Set $x_{-1} := x_0$ and $k := 0$.

Step 1. Compute $F(x_k)$. If $\|F(x_k)\| \leq \epsilon$, then stop. Choose an inertial extrapolation steplength $\alpha_k \in [0, \alpha]$ and compute the inertial extrapolation step

$$w_k = x_k + \alpha_k(x_k - x_{k-1}). \tag{2.4}$$

If $\|F_k\| \leq \epsilon$, then stop.

Step 2. Compute the search direction by the relation in (2.3). If $\|d_k\| \leq \epsilon$, then stop.

Step 3. Set $z_k = w_k + \eta_k d_k$, where the steplength $\eta_k = \max\{\gamma\tau^i \mid i = 0, 1, 2, \ldots\}$ satisfies

$$-F(z_k)^\top d_k \geq \sigma\eta_k\Phi_k\|d_k\|^2 \tag{2.5}$$

with $\Phi_k := \min\{\nu,\ \hat{\eta}\|F(z_k)\|\}$. If $\|F(z_k)\| \leq \epsilon$, then stop.

Step 4. Pick a relaxation factor $\rho_k \in [\underline{\rho}, \overline{\rho}]$ and then compute the relaxation step

$$x_{k+1} = w_k - \rho_k\xi_k F(z_k),$$

where

$$\xi_k = \frac{F(z_k)^\top(w_k - z_k)}{\|F(z_k)\|^2} \geq 0. \tag{2.6}$$

Step 5. Set $k := k + 1$ and go to Step 1.

(ii) The main difference between the algorithm presented in [4] and our proposed algorithm is that the former requires solving the proximal subproblem inexactly at each iteration, while the latter performs an adaptive Armijo line search. Clearly, the computational cost of the proposed IRTTCGPMN is much cheaper.

(iii) The relaxation step (2.6) can be rewritten equivalently as

$$\begin{aligned} x_{k+1} &= (1 - \rho_k)w_k + \rho_k\left[w_k - \frac{F(z_k)^\top(w_k - z_k)}{\|F(z_k)\|^2}F(z_k)\right] \\ &=: (1 - \rho_k)w_k + \rho_k\tilde{x}_{k+1}. \end{aligned} \tag{2.7}$$

3. Convergence

In this section, we aim to examine the global convergence of the proposed method. To this end, we will rely on the following standard assumptions.

Assumption 3.1. The solution set, denoted by $\mathcal{S}$, of problem (1.1) is nonempty. The mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and pseudo-monotone.

Assumption 3.2. The sequences $\{\alpha_k\}$ and $\{x_k\}$ generated by IRTTCGPMN satisfy the following summability condition:

$$\sum_{k=1}^{\infty} \alpha_k \|x_k - x_{k-1}\|^2 < \infty. \tag{3.1}$$

For ease of analysis, we need the following two equalities:

(i) For any $a, b, c, d \in \mathbb{R}^n$, it holds that

$$2(a-b)^\top (c-d) = (\|a-d\|^2 - \|a-c\|^2) + (\|b-c\|^2 - \|b-d\|^2). \tag{3.2}$$

(ii) For any $a, b \in \mathbb{R}^n$ and $\rho \in \mathbb{R}$, we have

$$\|(1-\rho)a + \rho b\|^2 = (1-\rho)\|a\|^2 + \rho\|b\|^2 - \rho(1-\rho)\|a-b\|^2. \tag{3.3}$$

The following lemma shows that the search direction defined in (2.3) has some nice properties independent of the line search technique. These properties are the reason why the global convergence of the proposed method can be established without the Lipschitz continuity of F, while asymptotic and non-asymptotic convergence rates can be showed under the local Lipschitz continuity of F. We also refer the reader to [21,29,30,50] and the references therein for the global convergence analysis of gradient-type projection methods without the Lipschitz continuity. The proof is similar to that of [49, Lemma 2], and thus we omit it for simplicity.

Lemma 3.1. *The search direction sequence $\{d_k\}$ generated by IRTTCGPMN always satisfies, for all $k \geq 0$, the sufficient descent condition*

$$F_k^\top d_k \leq -\left(1 - \frac{(1+\bar{t})^2}{4}\right) \|F_k\|^2, \tag{3.4}$$

and the trust region property

$$\left(1 - \frac{(1+\bar{t})^2}{4}\right) \|F_k\| \leq \|d_k\| \leq \left(1 + \frac{1+\bar{t}}{\mu} + \frac{1}{\mu^2}\right) \|F_k\|. \tag{3.5}$$

The next lemma shows that IRTTCGPMN is well-defined, i.e. the adaptive Armijo line search in (2.5) can be terminated in a finite number of iterations.

Lemma 3.2. *Let $\{w_k\}$ and $\{d_k\}$ be the sequences generated by IRTTCGPMN. Suppose that the mapping F is continuous on $\mathbb{R}^n$. Then IRTTCGPMN is well-defined.*

Proof. We prove this statement by contradiction and assume that there exists an integer $k_0 \geq 0$ such that the Armijo line search (2.5) does not hold for any nonnegative integer i. Hence, for all $i \geq 0$, we have

$$\begin{aligned}
& -F\big(w_{k_0} + \gamma\tau^i d_{k_0}\big)^\top d_{k_0} \\
& < \sigma\gamma\tau^i \min\big\{\nu, \hat{\eta}\big\|F\big(w_{k_0} + \gamma\tau^i d_{k_0}\big)\big\|\big\} \cdot \|d_{k_0}\|^2 \\
& \leq \sigma\gamma\tau^i \nu \cdot \|d_{k_0}\|^2.
\end{aligned}$$

Since F is continuous and $\tau \in (0,1)$, taking the limit in the above inequality tells us that

$$-F_{k_0}^\top d_{k_0} \leq 0.$$

This together with (3.4) implies $\|F_{k_0}\| = 0$, which contradicts the fact $\|F_{k_0}\| > \epsilon > 0$. This completes the proof. □

The following lemma establishes an important relationship between $\|x_{k+1} - x^*\|^2$ and $\|w_k - x^*\|^2$, where $x^* \in \mathcal{S}$.

Lemma 3.3. *Suppose that Assumption* 3.1 *holds. Let* $\{x_k\}, \{w_k\}$ *and* $\{z_k\}$ *be the sequences generated by IRTTCGPMN. For any* $x^* \in \mathcal{S}$*, we have the following inequality:*

$$\|x_{k+1} - x^*\|^2 + h_{k+1} \leq \|w_k - x^*\|^2, \quad \forall\, k \geq 0, \tag{3.6}$$

where h_{k+1} *is defined as*

$$h_{k+1} := \max\left\{\underline{\rho}(2-\overline{\rho})\sigma^2 \frac{\Phi_k^2\|w_k - z_k\|^4}{\|F(z_k)\|^2},\ \overline{\rho}^{\,-1}(2-\overline{\rho})\|x_{k+1} - w_k\|^2\right\}, \quad k \geq 0. \tag{3.7}$$

Proof. We begin our proof by observing from the definition of z_k and (2.5) that

$$F(z_k)^\top(w_k - z_k) = -\eta_k F(z_k)^\top d_k \geq \sigma\Phi_k\|\eta_k d_k\|^2 = \sigma\Phi_k\|w_k - z_k\|^2 > 0. \tag{3.8}$$

On the other hand, using the fact that $x^* \in \mathcal{S}$, we have

$$F(x^*)^\top(z_k - x^*) = 0.$$

Combining this with (1.2), it can be inferred that

$$F(z_k)^\top(z_k - x^*) \geq 0,$$

which is equivalent to

$$F(z_k)^\top(x^* - z_k) \leq 0. \tag{3.9}$$

Denote the hyperplane H_k and its corresponding closed half-space H_k^- respectively by

$$H_k := \{x \in \mathbb{R}^n \mid F(z_k)^\top(x - z_k) = 0\},$$
$$H_k^- := \{x \in \mathbb{R}^n \mid F(z_k)^\top(x - z_k) \leq 0\}.$$

It is not hard to see from (3.9) that $\mathcal{S} \subseteq H_k^-$. Furthermore, we declare from (3.8) and (3.9) that H_k strictly separates the point $w_k \notin H_k^-$ from the set $\mathcal{S} \subseteq H_k^-$. Therefore, by the definition of $\tilde{x}_{k+1}$ in (2.7), it follows that $\tilde{x}_{k+1}$ is the projection of w_k onto H_k^-. Making use of the properties of projection and $x^* \in \mathcal{S} \subseteq H_k^-$, we obtain

$$(\tilde{x}_{k+1} - w_k)^\top(x^* - \tilde{x}_{k+1}) \geq 0.$$

Combining the above inequality with (3.2) leads to the inequality

$$-\frac{1}{2}\|\tilde{x}_{k+1} - x^*\|^2 + \frac{1}{2}\left(\|w_k - x^*\|^2 - \|\tilde{x}_{k+1} - w_k\|^2\right) \geq 0,$$

or, equivalently,

$$\|w_k - x^*\|^2 - \|\tilde{x}_{k+1} - x^*\|^2 \geq \|\tilde{x}_{k+1} - w_k\|^2. \tag{3.10}$$

From (2.7), it follows that

$$x_{k+1} - x^* = (1-\rho_k)(w_k - x^*) + \rho_k(\tilde{x}_{k+1} - x^*).$$

This along with (3.3) implies

$$\begin{aligned}\|x_{k+1}-x^*\|^2 &= \|(1-\rho_k)(w_k-x^*)+\rho_k(\tilde{x}_{k+1}-x^*)\|^2\\ &= (1-\rho_k)\|w_k-x^*\|^2+\rho_k\|\tilde{x}_{k+1}-x^*\|^2-\rho_k(1-\rho_k)\|\tilde{x}_{k+1}-w_k\|^2.\end{aligned}$$

Combining this with the definition of $\tilde{x}_{k+1}$ in (2.7), (3.8), (3.10) and $\rho_k \in [\underline{\rho}, \overline{\rho}] \subseteq (0,2)$, we conclude that

$$\begin{aligned}&\|w_k-x^*\|^2-\|x_{k+1}-x^*\|^2\\ &= \|w_k-x^*\|^2-(1-\rho_k)\|(w_k-x^*)\|^2\\ &\quad -\rho_k\|\tilde{x}_{k+1}-x^*\|^2+\rho_k(1-\rho_k)\|\tilde{x}_{k+1}-w_k\|^2\\ &= \rho_k\big(\|w_k-x^*\|^2-\|\tilde{x}_{k+1}-x^*\|^2\big)+\rho_k(1-\rho_k)\|\tilde{x}_{k+1}-w_k\|^2\\ &\geq \rho_k\|\tilde{x}_{k+1}-w_k\|^2+\rho_k(1-\rho_k)\|\tilde{x}_{k+1}-w_k\|^2\\ &= \rho_k(2-\rho_k)\|\tilde{x}_{k+1}-w_k\|^2\\ &= \rho_k(2-\rho_k)\left\|\frac{F(z_k)^\top(w_k-z_k)}{\|F(z_k)\|^2}F(z_k)\right\|^2 \qquad (3.11)\\ &= \rho_k(2-\rho_k)\left(\frac{F(z_k)^\top(w_k-z_k)}{\|F(z_k)\|}\right)^2\\ &\geq \rho_k(2-\rho_k)\frac{\sigma^2\Phi_k^2\|w_k-z_k\|^4}{\|F(z_k)\|^2}\\ &\geq \underline{\rho}(2-\overline{\rho})\frac{\sigma^2\Phi_k^2\|w_k-z_k\|^4}{\|F(z_k)\|^2}. \qquad (3.12)\end{aligned}$$

On the other hand, from (2.6), (3.11) and $\rho_k \in [\underline{\rho}, \overline{\rho}] \subseteq (0,2)$, we have

$$\begin{aligned}&\|w_k-x^*\|^2-\|x_{k+1}-x^*\|^2\\ &\geq \rho_k^{-1}(2-\rho_k)\|x_{k+1}-w_k\|^2\\ &\geq \overline{\rho}^{\,-1}(2-\overline{\rho})\|x_{k+1}-w_k\|^2. \qquad (3.13)\end{aligned}$$

Therefore, the desired result in (3.6) follows from (3.12), (3.13) and the definition of h_{k+1} in (3.7). The proof is complete. □

With Lemma 3.3 at hand, we give the following results, which are useful for convergence analysis. Since the proof follows essentially from [50, Lemma 3.2 and Theorem 3.1], we omit the details.

Lemma 3.4. *Let $\{x_k\}, \{w_k\}$ and $\{z_k\}$ be the sequences generated by IRTTCGPMN, and for convenience in analysis, define*

$$\begin{aligned}u_k &:= \|x_k-x^*\|^2, & k\geq -1,\\ \delta_k &:= \alpha_k(1+\alpha_k)\|x_k-x_{k-1}\|^2, & k\geq 0,\end{aligned} \qquad (3.14)$$

where $x^* \in \mathcal{S}$. If Assumption 3.1 holds, then we have $u_{-1}=u_0$ and

$$u_{k+1}-u_k+h_{k+1}\leq \alpha_k(u_k-u_{k-1})+\delta_k, \quad \forall\, k\geq 0. \qquad (3.15)$$

Theorem 3.1. *Let $\{x_k\}$, $\{w_k\}$ and $\{z_k\}$ be the sequences generated by IRTTCGPMN. If Assumptions 3.1 and 3.2 hold, then for any $x^* \in \mathcal{S}$, the following hold:*

(i) The limit $\lim_{k\to\infty} \|x_k - x^*\|$ *exists.*

(ii) $\sum_{k=1}^{\infty} h_k < \infty$.

Corollary 3.1. *Suppose that all the assumptions used in Theorem* 3.1 *hold. Then, the following statements are true:*

(i) The sequences $\{x_k\}$ *and* $\{w_k\}$ *are both bounded.*

(ii) $\lim_{k\to\infty} \frac{\Phi_k \|w_k - z_k\|^2}{\|F(z_k)\|} = \lim_{k\to\infty} \|x_{k+1} - w_k\| = 0.$

Proof. (i) Let $x^* \in \mathcal{S}$. It follows from Theorem 3.1(i) that the sequence $\{x_k\}$ is bounded. Combining the definition of w_k in (2.4), we claim that $\{w_k\}$ is also bounded.

(ii) By Theorem 3.1(ii) and the definition of h_{k+1} in (3.7), the desired relation is naturally true. This completes the proof. □

Now, we are able to establish the main convergence result of the proposed method under some mild conditions.

Theorem 3.2. *Suppose that Assumptions* 3.1 *and* 3.2 *hold. Let* $\{x_k\}, \{w_k\}$ *and* $\{z_k\}$ *be the sequences generated by IRTTCGPMN. Then the sequences* $\{x_k\}, \{w_k\}$ *and* $\{z_k\}$ *converge to a solution of problem* (1.1).

Proof. This theorem can be proved by the following three parts:

(1) We want to prove that

$$\lim_{k\to\infty} \|z_k - w_k\| = \lim_{k\to\infty} \eta_k \|d_k\| = 0.$$

It follows from Corollary 3.1(i) that the sequences $\{x_k\}$ and $\{w_k\}$ are bounded. Therefore, we know from the continuity of F that the sequence $\{F_k\}$ is bounded. Then, according to the relation between $\|F_k\|$ and $\|d_k\|$ in (3.5), we get that the sequence $\{d_k\}$ is also bounded. Hence, by the definition of z_k, we obtain the boundedness of $\{z_k\}$. Again, it follows from the continuity of F that the sequence $\{F(z_k)\}$ is bounded. Combining this with the definition of Φ_k, we deduce that $\{\|F(z_k)\|/\Phi_k\}$ is also bounded. Hence, for all $k \geq 0$, there exists a positive constant M_1 such that

$$\|w_k\| \leq M_1, \quad \|d_k\| \leq M_1, \quad \|F_k\| \leq M_1, \quad \|F(z_k)\| \leq M_1, \quad \frac{\|F(z_k)\|}{\Phi_k} \leq M_1. \tag{3.16}$$

By Corollary 3.1(ii) and the last relation in (3.16), we deduce that

$$\frac{\|w_k - z_k\|^2}{M_1} \leq \frac{\Phi_k \|w_k - z_k\|^2}{\|F(z_k)\|} \;\to\; 0 \quad \text{as} \quad k \;\to\; \infty.$$

Hence, we conclude from (3.7) and the definition of z_k in Step 3 of Algorithm 2.1 that

$$\lim_{k\to\infty} \|w_k - z_k\| = \lim_{k\to\infty} \eta_k \|d_k\| = 0. \tag{3.17}$$

(2) We aim to show that $\liminf_{k\to\infty}\|F_k\| = 0$. Assume by contradiction that

$$\liminf_{k\to\infty}\|F_k\| \neq 0.$$

This implies that there exists a constant $\varepsilon > 0$ such that

$$\|F_k\| \geq \varepsilon, \quad \forall\, k \geq 0. \tag{3.18}$$

This, along with the known relations (3.5) and (3.16), yields that

$$0 < \left(1 - \frac{(1+\bar{t})^2}{4}\right)\varepsilon \leq \left(1 - \frac{(1+\bar{t})^2}{4}\right)\|F_k\| \leq \|d_k\| \leq M_1, \quad \forall\, k \geq 0.$$

It then follows from the limit $\lim_{k\to\infty}\eta_k\|d_k\| = 0$ that $\lim_{k\to\infty}\eta_k = 0$. By the boundedness of $\{w_k\}$ and $\{d_k\}$, without loss of generality, we assume that

$$\lim_{i\to\infty} w_{k_i} = \overline{w}, \quad \lim_{i\to\infty} d_{k_i} = \overline{d} \neq 0. \tag{3.19}$$

Letting $k := k_i$ in (3.4), and making use of (3.18), (3.19) and the continuity of F, we deduce that

$$F(\overline{w})^\top\overline{d} \leq -\left(1 - \frac{(1+\bar{t})^2}{4}\right)\|F(\overline{w})\|^2 \leq -\left(1 - \frac{(1+\bar{t})^2}{4}\right)\varepsilon^2 < 0. \tag{3.20}$$

From the Armijo line search procedure, we declare that

$$\begin{aligned}
&- F\big(w_k + \tau^{-1}\eta_k d_k\big)^\top d_k \\
&< \sigma\tau^{-1}\eta_k \min\big\{\nu, \hat{\eta}\big\|F\big(w_k + \tau^{-1}\eta_k d_k\big)\big\|\big\} \cdot \|d_k\|^2 \\
&\leq \sigma\tau^{-1}\eta_k\nu\|d_k\|^2.
\end{aligned}$$

Similarly, combining the above inequality with (3.19), and using the relation $\lim_{k\to\infty}\eta_k = 0$ and the continuity of F, one can obtain $-F(\overline{w})^\top\overline{d} \leq 0$, which contradicts the relation in (3.20). Thus, we have proved that

$$\liminf_{k\to\infty}\|F_k\| = 0. \tag{3.21}$$

(3) We show that the sequences $\{x_k\}, \{w_k\}$ and $\{z_k\}$ generated by the proposed IRTTCGPMN converge to a solution of problem (1.1). From the boundedness of $\{w_k\}$, the continuity of F and the known relation in (3.21), we assume, without loss of generality, that

$$\lim_{i\to\infty} w_{k_i} = \overline{w}, \quad \lim_{i\to\infty}\|F_{k_i}\| = \|F(\overline{w})\| = 0.$$

These two relations imply $\overline{w} \in \mathcal{S}$. Combining this with Corollary 3.1(ii), we deduce that $\overline{w}$ is also an accumulation point of the sequence $\{x_k\}$. Hence there exists an infinite index set $\mathcal{K}$, such that $\lim_{k\in\mathcal{K},k\to\infty} x_k = \overline{w} \in \mathcal{S}$. Letting $x^* := \overline{w}$ in Theorem 3.1, we have

$$\lim_{k\to\infty}\|x_k - \overline{w}\| = \lim_{k\in\mathcal{K},\, k\to\infty}\|x_k - \overline{w}\| = 0,$$

which further implies that the whole sequence $\{x_k\}$ converges to $\overline{w} \in S$. Moreover, by the relation in (3.17) and Corollary 3.1(ii), we deduce that both $\{w_k\}$ and $\{z_k\}$ also converge to $\overline{w}$. Therefore, all the sequences $\{x_k\}$, $\{w_k\}$ and $\{z_k\}$ converge to a solution of problem (1.1). As a result, the proof is complete. □

As elaborated in [50], we know that the convergence of IRTTCGPMN depends on the summability condition (3.1). In the following theorem, we provide a sufficient condition for (3.1) by assuming that the sequence $\{\alpha_k\}$ satisfies certain easily implementable conditions.

Theorem 3.3. *Let sequences $\{x_k\}$ and $\{\alpha_k\}$ be generated by IRTTCGPMN. Suppose that Assumption 3.1 is valid, and that there exist $0 \le \alpha < \upsilon < 1$ and $\overline{\rho} \in (0,2)$ such that the sequence $\{\alpha_k\}$ satisfies*

$$0 \le \alpha_k \le \alpha_{k+1} \le \alpha < \upsilon < 1, \quad k \ge 0, \tag{3.22}$$

and

$$\overline{\rho} = \overline{\rho}(\upsilon) := \frac{2(\upsilon-1)^2}{2(\upsilon-1)^2 + 3\upsilon - 1}. \tag{3.23}$$

Then, we have

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

Furthermore, the summability condition (3.1) holds.

Proof. By the definition of w_k in (2.4) and the important inequality

$$2a^\top b \le \|a\|^2 + \|b\|^2, \quad a, b \in \mathbb{R}^n,$$

one can derive

$$\begin{aligned}
\|x_{k+1} - w_k\|^2 &= \|x_{k+1} - x_k\|^2 + \alpha_k^2 \|x_k - x_{k-1}\|^2 - 2\alpha_k (x_{k+1} - x_k)^\top (x_k - x_{k-1}) \\
&\ge \|x_{k+1} - x_k\|^2 + \alpha_k^2 \|x_k - x_{k-1}\|^2 - \alpha_k \big(\|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2\big) \\
&= (1-\alpha_k)\|x_{k+1} - x_k\|^2 - \alpha_k(1-\alpha_k)\|x_k - x_{k-1}\|^2.
\end{aligned}$$

Combining this with (3.7) and Lemma 3.4, we obtain

$$\begin{aligned}
& u_{k+1} - u_k - \alpha_k(u_k - u_{k-1}) \\
&\le \delta_k - h_{k+1} \le \alpha_k(1+\alpha_k)\|x_k - x_{k-1}\|^2 - (2-\overline{\rho})\overline{\rho}^{-1}\|x_{k+1} - w_k\|^2 \\
&\le \alpha_k(1+\alpha_k)\|x_k - x_{k-1}\|^2 \\
&\qquad - (2-\overline{\rho})\overline{\rho}^{-1}\big[(1-\alpha_k)\|x_{k+1} - x_k\|^2 - \alpha_k(1-\alpha_k)\|x_k - x_{k-1}\|^2\big] \\
&= -(2-\overline{\rho})\overline{\rho}^{-1}(1-\alpha_k)\|x_{k+1} - x_k\|^2 + s_k\|x_k - x_{k-1}\|^2,
\end{aligned} \tag{3.24}$$

where

$$s_k := -2(\overline{\rho}^{-1} - 1)\alpha_k^2 + 2\overline{\rho}^{-1}\alpha_k = 2\alpha_k^2 + 2\overline{\rho}^{-1}\alpha_k(1-\alpha_k) \ge 0.$$

Let

$$\zeta_k := u_k - \alpha_k u_{k-1} + s_k\|x_k - x_{k-1}\|^2, \quad k \ge 0.$$

It is easy to see that

$$\zeta_0 = (1-\alpha_0)u_0 \le u_0.$$

From (3.22), (3.24) and the definition of s_k, we deduce that

$$\begin{aligned}
\zeta_{k+1} - \zeta_k &= u_{k+1} - \alpha_{k+1}u_k + s_{k+1}\|x_{k+1} - x_k\|^2 - \big(u_k - \alpha_k u_{k-1} + s_k\|x_k - x_{k-1}\|^2\big) \\
&\le u_{k+1} - u_k - \alpha_k(u_k - u_{k-1}) + s_{k+1}\|x_{k+1} - x_k\|^2 - s_k\|x_k - x_{k-1}\|^2 \\
&\le \big[-2(\overline{\rho}^{-1} - 1)\alpha_{k+1}^2 + 2\overline{\rho}^{-1}\alpha_{k+1} - (2-\overline{\rho})\overline{\rho}^{-1}(1-\alpha_k)\big]\|x_{k+1} - x_k\|^2 \\
&\le \big[-2(\overline{\rho}^{-1} - 1)\alpha_{k+1}^2 + 2\overline{\rho}^{-1}\alpha_{k+1} - (2-\overline{\rho})\overline{\rho}^{-1}(1-\alpha_{k+1})\big]\|x_{k+1} - x_k\|^2 \\
&= -q(\alpha_{k+1})\|x_{k+1} - x_k\|^2,
\end{aligned} \tag{3.25}$$

where

$$q(t) := 2(\overline{\rho}^{-1} - 1)t^2 - (4\overline{\rho}^{-1} - 1)t + 2\overline{\rho}^{-1} - 1. \tag{3.26}$$

From [3, Lemma 4], we know that the inverse function of $\overline{\rho}(\cdot)$ defined in (3.23) can be written as

$$\upsilon = \frac{2(2-\overline{\rho})}{4-\overline{\rho}+\sqrt{\overline{\rho}(16-7\overline{\rho})}} \in (0,1), \quad \overline{\rho} \in (0,2).$$

It is easy to verify that υ is a root of the equation $q(t) = 0$. By [3, Lemma 5], we get that the function $q(t)$ is strictly decreasing on $[0,\, \upsilon]$. Hence, we have

$$q(t) \geq q(\alpha) > q(\upsilon) = 0, \quad \forall\, t \in [0, \alpha] \subseteq [0, \upsilon).$$

Combining (3.22) and (3.25) yields

$$\zeta_{k+1} - \zeta_k \leq -q(\alpha)\|x_{k+1} - x_k\|^2 \leq 0, \quad \forall\, k \geq 0, \tag{3.27}$$

which further implies that the sequence $\{\zeta_k\}$ is monotonically nonincreasing. Using $u_k \geq 0$ (see (3.14)), (3.22) and the definition of ζ_k, we obtain

$$-\alpha u_{k-1} \leq u_k - \alpha u_{k-1} \leq u_k - \alpha_k u_{k-1} \leq \zeta_k \leq \zeta_0 \leq u_0.$$

It then follows that $\zeta_k \geq -\alpha u_{k-1}$ and

$$u_k \leq \alpha u_{k-1} + u_0 \leq \alpha^k u_0 + u_0 \sum_{j=0}^{k-1} \alpha^j \leq \alpha^k u_0 + \frac{u_0}{1-\alpha}. \tag{3.28}$$

On the other hand, it follows from (3.27) that

$$q(\alpha)\|x_k - x_{k-1}\|^2 \leq \zeta_{k-1} - \zeta_k, \quad \forall\, k \geq 1.$$

Further, using $\zeta_k \geq -\alpha u_{k-1}$, (3.28) and the fact that $0 \leq \alpha < 1$, we derive that

$$q(\alpha) \sum_{j=1}^{k} \|x_j - x_{j-1}\|^2 \leq \zeta_0 - \zeta_k \leq u_0 + \alpha u_{k-1} \leq \alpha^k u_0 + \frac{u_0}{1-\alpha} \leq \frac{(2-\alpha)u_0}{1-\alpha}.$$

Therefore, for all $k \geq 1$, it follows from (3.22) that

$$\sum_{j=1}^{k} \alpha_j \|x_j - x_{j-1}\|^2 \leq \sum_{j=1}^{k} \|x_j - x_{j-1}\|^2 \leq \frac{(2-\alpha)u_0}{(1-\alpha)q(\alpha)}. \tag{3.29}$$

Thus, the proof is complete. □

4. Iteration-complexity Analysis

In this section, we establish the asymptotic and non-asymptotic convergence rates for the proposed method. To this end, the following assumption is necessary.

Assumption 4.1. The mapping F is locally Lipschitz continuous on $\mathbb{R}^n$.

To facilitate the subsequent analysis, let us define

$$\widehat{\eta}_k = \tau^{-1}\eta_k, \quad \widehat{z}_k = w_k + \widehat{\eta}_k d_k. \tag{4.1}$$

It then follows from (3.17) that

$$\|\widehat{z}_k - w_k\| = \tau^{-1}\eta_k\|d_k\| \ \to \ 0 \quad \text{as} \quad k \ \to \ \infty.$$

By Theorem 3.2, the convergence of the sequence $\{w_k\}$ implies that the sequence $\{\widehat{z}_k\}$ is also convergent. Consequently, for the convergent sequences $\{w_k\}, \{z_k\}$, and $\{\widehat{z}_k\}$, there exists a compact set $\Omega \subseteq \mathbb{R}^n$ such that $\{w_k, z_k, \widehat{z}_k\} \subseteq \Omega$. Combining this with Assumption 4.1 and [39, Theorem 2.1.6], there exists a constant $L > 0$ such that

$$\|F(p) - F(q)\| \leq L\|p - q\|, \quad \forall\, p, q \in \Omega. \tag{4.2}$$

The following lemma states that the steplength sequence $\{\eta_k\}$ has a positive lower bound.

Lemma 4.1. *Suppose that Assumptions* 3.1, 3.2 *and* 4.1 *hold. Then the steplength sequence* $\{\eta_k\}$ *generated by* (2.5) *satisfies that*

$$\eta_k \geq \tilde{\eta} := \min\left\{\gamma, \frac{\tau\mu^4[4 - (1 + \bar{t})^2]}{4(L + \sigma\nu)[\mu^2 + \mu(1 + \bar{t}) + 1]^2}\right\} > 0, \quad \forall\, k \geq 0. \tag{4.3}$$

Proof. Clearly, if $\eta_k = \gamma$, then (4.3) holds. Otherwise, $\widehat{\eta}_k = \tau^{-1}\eta_k$ does not satisfy the adaptive Armijo line search (2.5), namely,

$$-F(\widehat{z}_k)^\top d_k < \sigma\widehat{\eta}_k \min\{\nu, \hat{\eta}\|F(\widehat{z}_k)\|\}\|d_k\|^2 \leq \sigma\nu\widehat{\eta}_k\|d_k\|^2,$$

where $\widehat{z}_k$ is defined in (4.1). This together with (3.4), (4.2) and $\{w_k, \widehat{z}_k\} \subseteq \Omega$ yields

$$\begin{aligned} &\left(1 - \frac{(1 + \bar{t})^2}{4}\right)\|F_k\|^2 \\ \leq\ & -F_k^\top d_k = \big(F(\widehat{z}_k) - F_k\big)^\top d_k - F(\widehat{z}_k)^\top d_k \\ <\ & (L + \sigma\nu)\widehat{\eta}_k\|d_k\|^2 = (L + \sigma\nu)\tau^{-1}\eta_k\|d_k\|^2. \end{aligned}$$

Therefore, it holds that

$$\begin{aligned} \eta_k &> \frac{\tau[4 - (1 + \bar{t})^2]}{4(L + \sigma\nu)} \frac{\|F_k\|^2}{\|d_k\|^2} \\ &\geq \frac{\tau[4 - (1 + \bar{t})^2]}{4(L + \sigma\nu)} \frac{\|F_k\|^2}{\big(1 + (1 + \bar{t})/\mu + 1/\mu^2\big)^2\|F_k\|^2} \\ &= \frac{\tau\mu^4[4 - (1 + \bar{t})^2]}{4(L + \sigma\nu)[\mu^2 + \mu(1 + \bar{t}) + 1]^2}, \end{aligned}$$

where the second inequality follows from the second inequality in (3.5). This yields the desired result and the proof is complete. □

The asymptotic convergence rate characterizes the behavior of the algorithm as the number of iterations approaches infinity. With Theorem 3.1 and Lemma 4.1 at hand, one can establish the asymptotic convergence rate for the proposed method.

Theorem 4.1. *Let sequences* $\{w_k\}$ *and* $\{z_k\}$ *be generated by IRTTCGPMN. Under Assumptions* 3.1*,* 3.2 *and* 4.1*, we have*

$$\sum_{k=1}^{\infty} \|F_k\|^4 < \infty. \tag{4.4}$$

Further, as $k \to \infty$*, it follows that*

$$\min_{1\le i\le k} \|F_i\| = o\left(\frac{1}{\sqrt[4]{k}}\right). \tag{4.5}$$

Proof. Based on Theorem 3.1(ii), it follows from the definition of h_{k+1} in (3.7) that

$$\sum_{k=1}^{\infty} \underline{\rho}(2-\overline{\rho})\sigma^2 \frac{\Phi_k^2 \|w_k - z_k\|^4}{\|F(z_k)\|^2} \le \sum_{k=1}^{\infty} h_k < \infty.$$

Hence, one can deduce from (3.16) and the definition of z_k that

$$\sum_{k=1}^{\infty} \|w_k - z_k\|^4 = \sum_{k=1}^{\infty} \eta_k^4 \|d_k\|^4 < \infty.$$

Furthermore, by (3.5), we get

$$\sum_{k=1}^{\infty} \eta_k^4 \|F_k\|^4 < \infty.$$

Combining this with Lemma 4.1, we obtain the desired result in (4.4). Then, by (4.4) and the Cauchy principle, one can deduce that (4.5) holds. □

In practical applications, it is impossible to perform infinite iterations, making it crucial to understand the performance of the algorithm within a finite number of iterations. This is the significance of the non-asymptotic convergence rate. To establish the non-asymptotic convergence rate of the proposed algorithm, we need the following lemma, which can be derived from Lemma 3.4.

Lemma 4.2. *Suppose that Assumption* 3.1 *holds. Let the sequences* $\{u_k\}, \{h_k\}, \{\alpha_k\}$ *and* $\{\delta_k\}$ *be generated by IRTTCGPMN. For all* $k \ge 1$*, the following holds:*

$$u_k + \sum_{j=1}^{k} h_j \le u_0 + \frac{1}{1-\alpha} \sum_{j=0}^{k-1} \delta_j. \tag{4.6}$$

Proof. To begin with, we obtain from (3.7) and (3.15) that

$$u_{k+1} - u_k \le u_{k+1} - u_k + h_{k+1} \le \alpha_k(u_k - u_{k-1}) + \delta_k.$$

It then follows from $\alpha_k \in [0, \alpha] \subseteq [0, 1)$ that

$$u_{k+1} - u_k \le \alpha_k(u_k - u_{k-1}) + \delta_k \le \alpha[u_k - u_{k-1}]_+ + \delta_k,$$

where $[\cdot]_+ := \max\{\cdot, 0\}$. This, together with the known relation $u_0 = u_{-1}$ from Lemma 3.4, further tells us that

$$\begin{aligned}
[u_{k+1} - u_k]_+ &\le \alpha[u_k - u_{k-1}]_+ + \delta_k \\
&\le \alpha^{k+1}[u_0 - u_{-1}]_+ + \sum_{j=0}^{k} \alpha^j \delta_{k-j} = \sum_{j=0}^{k} \alpha^j \delta_{k-j}.
\end{aligned}$$

Therefore, we claim that for all $k \geq 1$,

$$\sum_{i=0}^{k-1}[u_{i+1}-u_i]_+ \leq \sum_{i=0}^{k-1}\left(\sum_{j=0}^{i}\alpha^j\delta_{i-j}\right) \leq \frac{1}{1-\alpha}\sum_{j=0}^{k-1}\delta_j.$$

Hence, using $u_0 = u_{-1}$ again, $\alpha_k \in [0,\alpha] \subseteq [0,1)$ and (3.15), after some algebraic manipulations we find that for all $k \geq 1$,

$$\begin{aligned} u_k + \sum_{j=1}^{k} h_j &\leq u_0 + \alpha\sum_{j=0}^{k-2}[u_{j+1}-u_j]_+ + \sum_{j=0}^{k-1}\delta_j \\ &\leq u_0 + \alpha\left(\frac{1}{1-\alpha}\sum_{j=0}^{k-1}\delta_j\right) + (1-\alpha)\left(\frac{1}{1-\alpha}\sum_{j=0}^{k-1}\delta_j\right) \\ &= u_0 + \frac{1}{1-\alpha}\sum_{j=0}^{k-1}\delta_j. \end{aligned}$$

This completes the proof. □

Combining Lemma 4.2 with Theorem 3.3, we can deduce the following result.

Corollary 4.1. *Under the assumptions of Theorem* 3.3*, let $q(\cdot)$ be defined as in* (3.26) *and $x^* \in \mathcal{S}$. Then, for all $k \geq 1$, we obtain*

$$\begin{aligned} &u_k + \sum_{j=1}^{k}(2-\overline{\rho})\left(\max\left\{\underline{\rho}\sigma^2\frac{\Phi_{j-1}^2\|w_{j-1}-z_{j-1}\|^4}{\|F(z_{j-1})\|^2},\ \overline{\rho}^{-1}\|x_j-w_{j-1}\|^2\right\}\right) \\ &\leq \left(1+\frac{\alpha(2-\alpha)(1+\alpha)}{(1-\alpha)^2q(\alpha)}\right)u_0. \end{aligned}$$

Proof. It follows from (3.29) and the definition of δ_k in Lemma 3.4 that

$$\begin{aligned} \frac{1}{1-\alpha}\sum_{j=0}^{k}\delta_j &= \frac{1}{1-\alpha}\sum_{j=1}^{k}\alpha_j(1+\alpha_j)\|x_j-x_{j-1}\|^2 \\ &\leq \frac{\alpha(2-\alpha)(1+\alpha)}{(1-\alpha)^2q(\alpha)}u_0. \end{aligned} \tag{4.7}$$

Thus, the desired result follows from the definition of h_k in (3.7), (4.6) and (4.7). □

We are now in the position to establish the non-asymptotic convergence rate (iteration complexity) of the proposed IRTTCGPMN.

Theorem 4.2 ($\mathcal{O}(1/\sqrt[4]{k})$ Convergence Rate) *Under the assumptions of Theorem* 3.3 *and Lemma* 4.1*, let $q(\cdot)$ be given as in* (3.26)*. Then for every $k \geq 1$, there exists $i \in \{0,1,\ldots,k-1\}$ such that for some constant $c_1 > 0$,*

$$\|F_i\| \leq c_1\frac{1}{\sqrt[4]{k}}.$$

Proof. It follows from Corollary 4.1 that, for every $k \geq 1$, there exists $i \in \{0,1,\ldots,k-1\}$ such that

$$(2-\overline{\rho})k\max\left\{\underline{\rho}\sigma^2\frac{\Phi_i^2\|w_i-z_i\|^4}{\|F(z_i)\|^2},\ \overline{\rho}^{\,-1}\|x_{i+1}-w_i\|^2\right\} \leq \left(1+\frac{\alpha(2-\alpha)(1+\alpha)}{(1-\alpha)^2q(\alpha)}\right)u_0.$$

Combining this with the definition of z_k and the last relation in (3.16), we deduce that

$$\begin{aligned}
&(2-\overline{\rho})k\max\left\{\underline{\rho}\sigma^2\frac{\|\eta_i d_i\|^4}{M_1^2},\ \overline{\rho}^{\,-1}\|x_{i+1}-w_i\|^2\right\}\\
&\le (2-\overline{\rho})k\max\left\{\underline{\rho}\sigma^2\frac{\Phi_i^2\|\eta_i d_i\|^4}{\|F(z_i)\|^2},\ \overline{\rho}^{\,-1}\|x_{i+1}-w_i\|^2\right\}\\
&\le \left(1+\frac{\alpha(2-\alpha)(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)u_0.
\end{aligned}$$

From the relation in (4.3), it then follows that

$$\|d_i\|^4 \le \frac{\big(1+\alpha(2-\alpha)(1+\alpha)/\big((1-\alpha)^2 q(\alpha)\big)\big)M_1^2 u_0}{(2-\overline{\rho})\underline{\rho}\sigma^2\tilde{\eta}^4 k}.$$

This leads to the inequality

$$\begin{aligned}
\|d_i\| &\le \sqrt[4]{\frac{\big(1+(2-\alpha)\alpha(1+\alpha)/\big((1-\alpha)^2 q(\alpha)\big)\big)u_0}{(2-\overline{\rho})\underline{\rho}}}\sqrt{\frac{M_1}{\sigma}}\tilde{\eta}^{-1}\frac{1}{\sqrt[4]{k}}\\
&= \sqrt[4]{\frac{\big[(1-\alpha)^2\big(q(\alpha)-(1+\alpha)\big)+(1+\alpha)\big]u_0}{(1-\alpha)^2 q(\alpha)(2-\overline{\rho})\underline{\rho}}}\sqrt{\frac{M_1}{\sigma}}\tilde{\eta}^{-1}\frac{1}{\sqrt[4]{k}}.
\end{aligned}$$

Hence, by Lemma 3.1, we obtain

$$\|F_i\| \le \sqrt[4]{\frac{\big[(1-\alpha)^2\big(q(\alpha)-(1+\alpha)\big)+(1+\alpha)\big]u_0}{(1-\alpha)^2 q(\alpha)(2-\overline{\rho})\underline{\rho}}}\sqrt{\frac{M_1}{\sigma}}\left(\tilde{\eta}\left(1-\frac{(1+\bar{t})^2}{4}\right)\sqrt[4]{k}\right)^{-1}.$$

Letting

$$c_1 := \sqrt[4]{\frac{\big[(1-\alpha)^2\big(q(\alpha)-(1+\alpha)\big)+(1+\alpha)\big]u_0}{(1-\alpha)^2 q(\alpha)(2-\overline{\rho})\underline{\rho}}}\sqrt{\frac{M_1}{\sigma}}\left(\tilde{\eta}\left(1-\frac{(1+\bar{t})^2}{4}\right)\right)^{-1},$$

one can conclude

$$\|F_i\| \le c_1\frac{1}{\sqrt[4]{k}},$$

which completes the proof. □

5. Numerical Experiments

In this section, we validate the effectiveness of the proposed IRTTCGPMN, and apply it to solve large-scale benchmark test instances and sparse signal restoration problems. Meanwhile, we compare the proposed algorithm with four closely related algorithms: TTCGPM [26], FISAP [56], IRTTCGPM [50], and EIRCGPM [20]. All codes are written in MATLAB R2016a and run on a 64-bit Lenovo PC with an Intel(R) Xeon(R) E-2224G, CPU 3.50 GHz, 64 GB RAM, and Windows 10.0 OS. Throughout this section, we stop the program when one of the following five termination conditions is met:

(i) $\|F(x_k)\| \le 10^{-6}$,

(ii) $\|F_k\| \le 10^{-6}$,

(iii) $\|F(z_k)\| \leq 10^{-6}$,

(iv) $\|d_k\| \leq 10^{-7}$,

(v) Itr ≥ 2000,

where the symbol "Itr" represents the number of iterations. If the last condition is satisfied, we declare that the algorithm is unsuccessful for the corresponding test problem, and the numerical results obtained are denoted by "NaN".

The parameters for IRTTCGPMN are set as follows: $\mu = 0.2, \bar{t} = 0.2, \nu = 1$, and $\hat{\eta} = 1$, $\gamma = 1, \sigma = 0.01, \tau = 0.4362, \alpha_k \equiv \alpha = 0.18966, \upsilon = 0.18976$, and $\rho_k \equiv 1.4882 \leq \bar{\rho}(\upsilon)$. Clearly, the inertial extrapolation steplength α_k and the relaxation factor ρ_k satisfy (3.22) and (3.23), respectively. The parameter settings for other four algorithms are consistent with those in the respective original papers. Specifically, we set $\gamma = 1, \sigma = 0.01, \tau = 0.5$, and $\rho_k \equiv 1$ for TTCGPM; $\mu = 1.4, \gamma = 1, \sigma = 0.01, \tau = 0.5, \alpha_k \equiv \alpha = 0.18966, \upsilon = 0.18976$, and $\rho_k \equiv 1.4882 \leq \bar{\rho}(\upsilon)$ for IRTTCGPM; $\gamma = 1, \sigma = 0.01, \tau = 0.5, \alpha_k \equiv \alpha = 0.1$, and $\rho_k \equiv 1.7$ for FISAP; and $\gamma = 0.392, \sigma = 0.01, \tau = 0.4028, \alpha_k \equiv \alpha = 0.1$, and $\rho_k \equiv 1.7$ for EIRCGPM.

5.1. Benchmark test instances

We first list some benchmark test instances, where the form of the mapping F defined in (1.1) is as follows:

$$F(x) = \big(f_1(x), f_2(x), \ldots, f_n(x)\big)^\top.$$

Problem 5.1 ([36]). Set

$$f_i(x) = e^{x_i} - 1, \quad i = 1, 2, \ldots, n.$$

Problem 5.2 ([57]). Set

$$f_i(x) = 2x_i - \sin(|x_i|), \quad i = 1, 2, \ldots, n.$$

Clearly, this problem is nonsmooth at $x = (0, 0, \ldots, 0)^\top$.

Problem 5.3 ([58]). Set

$$\begin{aligned} f_1(x) &= 2x_1 - x_2 + e^{x_1} - 1, \\ f_i(x) &= -x_{i-1} + 2x_i - x_{i+1} + e^{x_i} - 1, \quad i = 2, 3, \ldots, n-1, \\ f_n(x) &= 2x_n - x_{n-1} + e^{x_n} - 1. \end{aligned}$$

Problem 5.4 ([1]). Set

$$\begin{aligned} f_1(x) &= 2.5x_1 + x_2 - 1, \\ f_i(x) &= x_{i-1} + 2.5x_i + x_{i+1} - 1, \quad i = 2, 3, \ldots, n-1, \\ f_n(x) &= x_{n-1} + 2.5x_n - 1. \end{aligned}$$

Problem 5.5 ([47]). Set

$$f_i(x) = x_i - \frac{1}{n}x_i^2 + \frac{1}{n}\sum_{i=1}^{n} x_i + i, \quad i = 1, 2, \ldots, n.$$

Problem 5.6 ([8]). Set

$$\begin{aligned} f_1(x) &= x_1 - e^{\left(\frac{x_1+x_2}{n+1}\right)}, \\ f_i(x) &= x_i - e^{\left(\frac{x_{i-1}+x_i+x_{i+1}}{n+1}\right)}, \quad i = 2, 3, \ldots, n-1, \\ f_n(x) &= x_n - e^{\left(\frac{x_{n-1}+x_n}{n+1}\right)}. \end{aligned}$$

Problem 5.7 ([43]). Set

$$\begin{aligned} f_1(x) &= x_1\left(x_1^2 + x_2^2\right) - 1, \\ f_i(x) &= x_i\left(x_{i-1}^2 + 2x_i^2 + x_{i+1}^2\right) - 1, \quad i = 2, 3, \ldots, n-1, \\ f_n(x) &= x_n\left(x_{n-1}^2 + x_n^2\right) - 1. \end{aligned}$$

Problem 5.8 ([54]). Set

$$f_1(x) = e^{x_1} - 1, \quad f_i(x) = e^{x_i} + x_i - 1, \quad i = 2, 3, \ldots, n.$$

To minimize the impact of the initial point x_0 on the algorithm, we generate x_0 for all test problems using the MATLAB script `(rand(n,1))*4-1`. In our experiments, we set $n = 1000, 3000, 5000, 8000, 10000, 30000, 50000, 80000,$ and 100000 for all problems except Problem 5.5. For Problem 5.5 we only consider $n = 1000, 3000$ and 5000 as the solution process is time-consuming for all methods. The corresponding numerical results are presented in `https://github.com/QiongxuanH/Numerical-Result_tables/blob/master/Numerical_Result_tables.pdf`, where we report Itr, the number of function evaluations (NF), the CPU time in seconds (Tcpu) and the final value ($\|F^*\|$), averaged over 5 random instances. It is worth to mention that the first column of each table stands for the problem dimension multiplied by 10^4. Furthermore, we utilize the performance profile introduced by [10] to effectively characterize the performance of these algorithms.

Based on the results listed in `https://github.com/QiongxuanH/Numerical-Result_tables/blob/master/Numerical_Result_tables.pdf` and Figs. 5.1-5.3, it is evident that the proposed IRTTCGPMN outperforms the other four methods in terms of Tcpu, NF and Itr, while it also achieves better solution quality than TTCGPM and IRTTCGPM. There may be two underlying factors which lead to the superior performance for the proposed method. Firstly, the inertial-relaxed technique does significantly accelerate the convergence of the algorithm numerically. Secondly, IRTTCGPMN inherits the effectiveness and stability of the hybrid three-term CGPM.

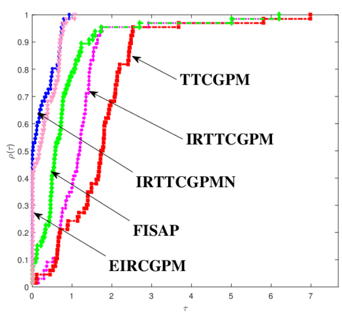

Fig. 5.1. Performance profiles on Tcpu.

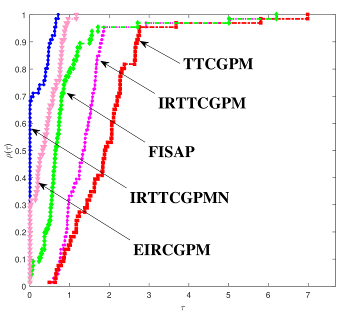

Fig. 5.2. Performance profiles on NF.

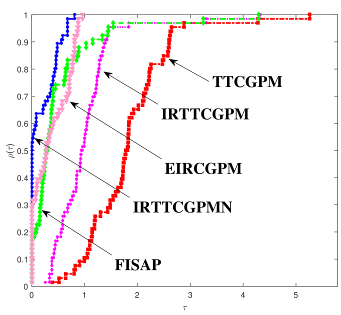

Fig. 5.3. Performance profiles on Itr.

5.2. Sparse signal restoration

In this subsection, we focus on evaluating and discussing the performance of the proposed method for solving the sparse signal recovery problem, which can be modeled as a penalized least squares problem

$$\min_{x\in\mathbb{R}^n} f(x) := \frac{1}{2}\|Ax-b\|^2 + \varrho\|x\|_1, \tag{5.1}$$

where $A \in \mathbb{R}^{m\times n}(m \ll n)$, $b \in \mathbb{R}^m$ is an observation and $\varrho > 0$ is a regularization parameter that balances data-fitting and sparsity. Letting $u := \max(x, 0)$ and $v := \max(-x, 0)$, where the operator "max" denotes componentwise maximum, one can deduce that $x = u - v$ and $\|x\|_1 = e_n^\top u + e_n^\top v$ with $e_n := (1, 1, \ldots, 1)^\top \in \mathbb{R}^n$. Following the studies in [12, 45], the penalized least squares problem (5.1) can be equivalently reformulated as a system of nonlinear equations

$$F(z) = \min\{z,\ Hz + c\} = 0, \tag{5.2}$$

where the operator "min" is interpreted as componentwise minimum,

$$z = \begin{bmatrix} u \\ v \end{bmatrix}, \quad c = \begin{bmatrix} \varrho e_n - A^\top b \\ \varrho e_n + A^\top b \end{bmatrix}, \quad H = \begin{bmatrix} A^\top A & -A^\top A \\ -A^\top A & A^\top A \end{bmatrix}.$$

Furthermore, based on [45, Lemma 2.2] and [34, Lemma 3], one can conclude that the mapping $F : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is Lipschitz continuous and pseudo-monotone. As a result, we can apply the proposed IRTTCGPMN to solve (5.2).

By convention, the quality of signal recovery is evaluated by adopting the so-called mean squared error (MSE)

$$\text{MSE} = \frac{1}{n}\|x^* - x\|^2,$$

where x^* represents the restored signal and x is the original sparse signal. Clearly, the smaller the MSE value, the better the corresponding method. In our experiments, we use N to represent the number of non-zero entries in x. For a given triplet (m, n, N), we use the following MATLAB codes to generate the test data:

```
x = zeros(n, 1);                    % initialize the original sparse signal
q = randperm(n);
x(q(1 : N)) = randn(N, 1);
A = randn(m, n);
A = orth(A')';                      % orthonormalize the rows of A
b = A * x + 0.001 * randn(m, 1);
varrho = 0.005 * max(abs(A' * b));  % the regularization parameter
```

In this part, the parameters for the proposed algorithm are the same as those of IRTTCGPM. All methods start their iterations with $x_0 = A^\top b$, which further implies that

$$z_0 = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} = \begin{bmatrix} \max(A^\top b,\ 0) \\ \max(-A^\top b,\ 0) \end{bmatrix}.$$

Note that we only compare our proposed method with IRTTCGPM, FISAP and EIRCGPM because TTCGPM is failed to solve (5.2) within the maximum number of iterations. From Figs. 5.4-5.6, we again illustrate that the proposed IRTTCGPMN is superior to the other three methods in terms of Tcpu and Itr.

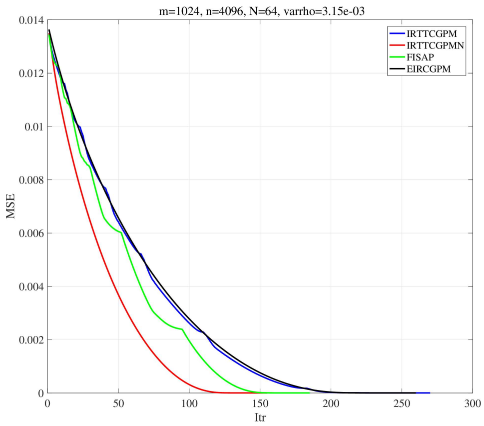

Fig. 5.4. MSE in terms of Itr.

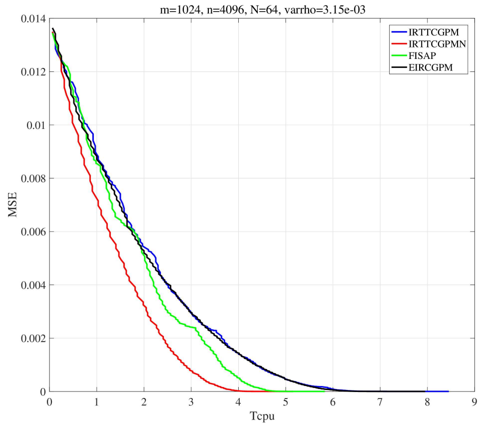

Fig. 5.5. MSE in terms of Tcpu.

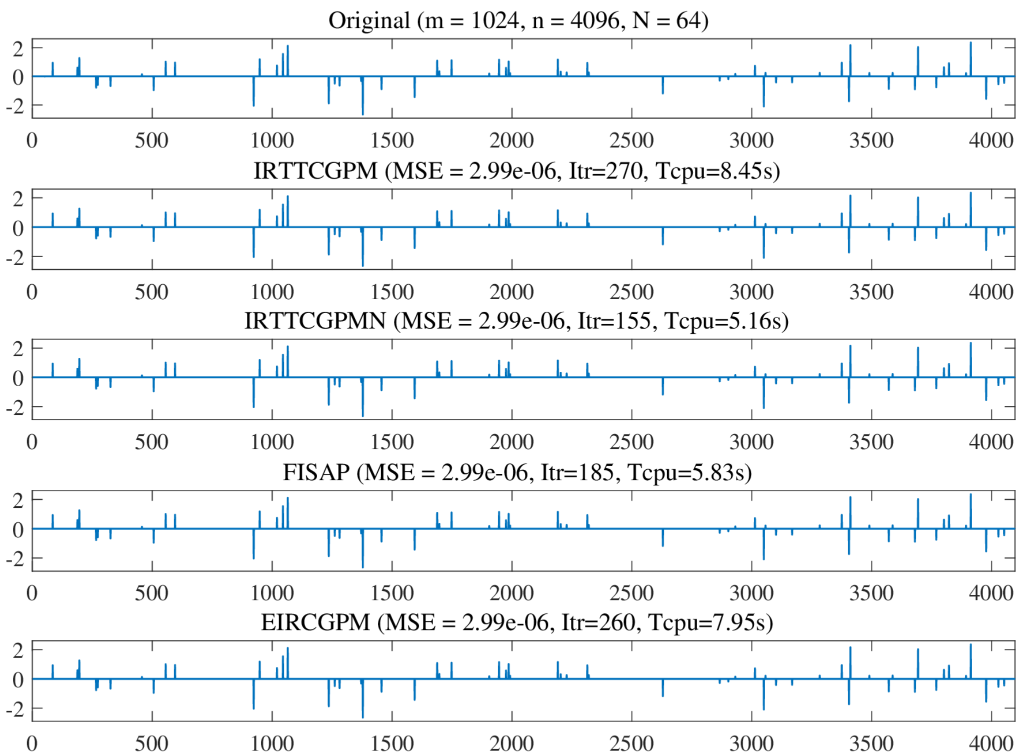

Fig. 5.6. The original sparse signal and the restored signals.

6. Conclusions

In this paper, we propose an inertial-relaxed three-term conjugate gradient projection method for solving large-scale unconstrained nonlinear pseudo-monotone equations. The search direction generated by the proposed method owns the sufficient descent and trust region properties independent of the line search technique. Under some mild conditions, we establish the global convergence of the proposed method, and analyze the asymptotic and non-asymptotic convergence rates in terms of iteration complexity. The numerical results demonstrate that our method not only absorbs the strength of the inertial-relaxed technique and inherits the good properties of the three-term conjugate gradient projection method, but also benefits from the adaptive Armijo line search, resulting in significantly improving the computational efficiency of the proposed method compared to four state-of-the-art algorithms.

Acknowledgments. The authors would like to thank the Associate Editor and two anonymous referees for their constructive and pertinent suggestions and comments that improved the quality of the paper significantly.

This work was supported by the National Natural Science Foundation of China (Grant No. 12401403), by the Guangxi Science and Technology Program (Grant No. AD23023001), by the Natural Science Foundation of Guangxi Province (Grant No. 2023GXNSFBA026029), by the Research Project of Guangxi Minzu University (Grant No. 2022KJQD03) and by the Middle-aged and Young Teachers' Basic Ability Promotion Project of Guangxi Province (Grant No. 2023KY0168).

References

[1] A.B. Abubakar and P. Kumam, A descent Dai-Liao conjugate gradient method for nonlinear equations, *Numer. Algorithms*, **81** (2019), 197–210.

[2] M. Ahookhosh, K. Amini, and S. Bahrami, Two derivative-free projection approaches for systems of large-scale nonlinear monotone equations, *Numer. Algorithms*, **64** (2013), 21–42.

[3] M.M. Alves, J. Eckstein, M. Geremia, and J.G. Melo, Relative-error inertial-relaxed inexact versions of Douglas-Rachford and ADMM splitting algorithms, *Comput. Optim. Appl.*, **75** (2020), 389–422.

[4] M.M. Alves and R.T. Marcavillaca, On inexact relative-error hybrid proximal extragradient, forward-backward and Tseng's modified forward-backward methods with inertial effects, *Set-Valued Var. Anal.*, **28** (2020), 301–325.

[5] A.M. Awwal, P. Kumam, and A.B. Abubakar, A modified conjugate gradient method for monotone nonlinear equations with convex constraints, *Appl. Numer. Math.*, **145** (2019), 507–520.

[6] A.M. Awwal, P. Kumam, L. Wang, S. Huang, and W. Kumam, Inertial-based derivative-free method for system of monotone nonlinear equations and application, *IEEE Access*, **8** (2020), 226921–226930.

[7] A.M. Awwal, L. Wang, P. Kumam, and H. Mohammad, A two-step spectral gradient projection method for system of nonlinear monotone equations and image deblurring problems, *Symmetry*, **12**:6 (2020), 874.

[8] Y. Bing and G. Lin, An efficient implementation of Merrills method for sparse or partially separable systems of nonlinear equations, *SIAM J. Optim.*, **1**:2 (1991), 206–221.

[9] W. Cheng, A PRP type method for systems of monotone equations, *Math. Comput. Modell.*, **50**:1-2 (2009), 15–20.

[10] E.D. Dolan and J.J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.*, **91** (2002), 201–213.

[11] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, 2003.

[12] M.A. Figueiredo, R.D. Nowak, and S.J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, *IEEE J. Sel. Top. Signal Process.*, **1**:4 (2007), 586–597.

[13] P. Gao and C. He, An efficient three-term conjugate gradient method for nonlinear monotone equations with convex constraints, *Calcolo*, **55** (2018), 53.

[14] P. Gao, C. He, and Y. Liu, An adaptive family of projection methods for constrained monotone nonlinear equations with applications, *Appl. Math. Comput.*, **359** (2019), 1–16.

[15] A.S. Halilu, A. Majumder, M.Y. Waziri, A.M. Awwal, and K. Ahmed, On solving double direction methods for convex constrained monotone nonlinear equations with image restoration, *Comput. Appl. Math.*, **40** (2021), 239.

[16] W. Hu, J. Wu, and G. Yuan, Some modified Hestenes-Stiefel conjugate gradient algorithms with application in image restoration, *Appl. Numer. Math.*, **158** (2020), 360–376.

[17] A.H. Ibrahim and S. Al-Homidan, Two-step inertial derivative-free projection method for solving nonlinear equations with application, *J. Comput. Appl. Math.*, (2024), 116071.

[18] A.H. Ibrahim, P. Kumam, S. Rapajić, Z. Papp, and A.B. Abubakar, Approximation methods with inertial term for large-scale nonlinear monotone equations, *Appl. Numer. Math.*, **181** (2022), 417–435.

[19] A.H. Ibrahim, P. Kumam, M. Sun, P. Chaipunya, and A.B. Abubakar, Projection method with inertial step for nonlinear equations: Application to signal recovery, *J. Ind. Manag. Optim.*, **19**:1 (2022), 30–55.

[20] J. Jian, Z. Ren, J. Yin, D. Han, and X. Wu, An effective inertial-relaxed CGPM for nonlinear monotone equations, *J. Appl. Math. Comput.*, **70**:1 (2024), 689–710.

[21] J. Jian, J. Yin, C. Tang, and D. Han, A family of inertial derivative-free projection methods for constrained nonlinear pseudo-monotone equations with applications, *Comput. Appl. Math.*, **41**:7 (2022), 309.

[22] M. Koorapetse, P. Kaelo, S. Lekoko, and T. Diphofu, A derivative-free RMIL conjugate gradient projection method for convex constrained nonlinear monotone equations with applications in compressive sensing, *Appl. Numer. Math.*, **165** (2021), 431–441.

[23] M. Li, A family of three-term nonlinear conjugate gradient methods close to the memoryless BFGS method, *Optim. Lett.*, **12** (2018), 1911–1927.

[24] M. Li, A modified Hestense-Stiefel conjugate gradient method close to the memoryless BFGS quasi-Newton method, *Optim. Methods Software*, **33**:2 (2018), 336–353.

[25] M. Li, A three term Polak-Ribière-Polyak conjugate gradient method close to the memoryless BFGS quasi-Newton method, *J. Ind. Manag. Optim.*, **16**:1 (2018), 245–260.

[26] Q. Li and D. Li, A class of derivative-free methods for large-scale nonlinear monotone equations, *IMA J. Numer. Anal.*, **31**:4 (2011), 1625–1635.

[27] J. Liu and Y. Feng, A derivative-free iterative method for nonlinear monotone equations with convex constraints, *Numer. Algorithms*, **82** (2019), 245–262.

[28] J. Liu and S. Li, Spectral DY-type projection method for nonlinear monotone system of equations, *J. Comput. Math.*, **33**:4 (2015), 341–355.

[29] J. Liu, Z. Lu, J. Xu, S. Wu, and Z. Tu, An efficient projection-based algorithm without Lipschitz continuity for large-scale nonlinear pseudo-monotone equations, *J. Comput. Appl. Math.*, **403** (2022), 113822.

[30] P. Liu, H. Shao, Y. Wang, and X. Wu, A three-term CGPM-based algorithm without Lipschitz continuity for constrained nonlinear monotone equations with applications, *Appl. Numer. Math.*, **175** (2022), 98–107.

[31] P. Liu, H. Shao, Z. Yuan, X. Wu, and T. Zheng, A family of three-term conjugate gradient projection methods with a restart procedure and their relaxed-inertial extensions for the constrained nonlinear pseudo-monotone equations with applications, *Numer. Algorithms*, **94**:3 (2023), 1055–1083.

[32] K. Meintjes and A.P. Morgan, A methodology for solving chemical equilibrium systems, *Appl. Math. Comput.*, **22**:4 (1987), 333–361.

[33] J. Nocedal, Updating quasi-Newton matrices with limited storage, *Math. Comp.*, **35**:151 (1980), 773–782.

[34] J.S. Pang, Inexact Newton methods for the nonlinear complementarity problem, *Math. Program.*,

36:1 (1986), 54–71.

[35] Z. Papp and S. Rapajić, FR type methods for systems of large-scale nonlinear monotone equations, *Appl. Math. Comput.*, **269** (2015), 816–823.

[36] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM J. Optim.*, **7**:1 (1997), 26–33.

[37] D.F. Shanno, Conjugate gradient methods with inexact searches, *Math. Oper. Res.*, **3**:3 (1978), 244–256.

[38] M.V. Solodov and B.F. Svaiter, A globally convergent inexact Newton method for systems of monotone equations, in: *Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Applied Optimization*, Vol. 22, Springer, 355–369, 1998.

[39] C. Ştefan, M. Radu, and N. Adriana, *Lipschitz Functions*, Springer, 2019.

[40] M. Sun and J. Liu, New hybrid conjugate gradient projection method for the convex constrained equations, *Calcolo*, **53** (2016), 399–411.

[41] Z. Wan, J. Guo, J. Liu, and W. Liu, A modified spectral conjugate gradient projection method for signal recovery, *Signal Image Video Process.*, **12** (2018), 1455–1462.

[42] C. Wang, Y. Wang, and C. Xu, A projection method for a system of nonlinear monotone equations with convex constraints, *Math. Methods Oper. Res.*, **66** (2007), 33–46.

[43] M.Y. Waziri, K.A. Hungu, and J. Sabiu, Descent Perry conjugate gradient methods for systems of monotone nonlinear equations, *Numer. Algorithms*, **85** (2020), 763–785.

[44] X. Wu, H. Shao, P. Liu, and Y. Zhuo, An inertial spectral CG projection method based on the memoryless BFGS update, *J. Optim. Theory Appl.*, **198**:3 (2023), 1130–1155.

[45] Y. Xiao, Q. Wang, and Q. Hu, Non-smooth equations based method for ℓ_1-norm problems with applications to compressed sensing, *Nonlinear Anal. Theory Methods Appl.*, **74**:11 (2011), 3570–3577.

[46] Y. Xiao and H. Zhu, A conjugate gradient method to solve convex constrained monotone equations with applications in compressive sensing, *J. Math. Anal. Appl.*, **405**:1 (2013), 310–319.

[47] Q. Yan, X. Peng, and D. Li, A globally convergent derivative-free method for solving large-scale nonlinear monotone equations, *J. Comput. Appl. Math.*, **234**:3 (2010), 649–657.

[48] J. Yin, J. Jian, and X. Jiang, A generalized hybrid CGPM-based algorithm for solving large-scale convex constrained equations with applications to image restoration, *J. Comput. Appl. Math.*, **391** (2021), 113423.

[49] J. Yin, J. Jian, X. Jiang, M. Liu, and L. Wang, A hybrid three-term conjugate gradient projection method for constrained nonlinear monotone equations with applications, *Numer. Algorithms*, **88** (2021), 389–418.

[50] J. Yin, J. Jian, X. Jiang, and X. Wu, A family of inertial-relaxed DFPM-based algorithms for solving large-scale monotone nonlinear equations with application to sparse signal restoration, *J. Comput. Appl. Math.*, **419** (2023), 114674.

[51] Z. Yu, J. Lin, J. Sun, Y. Xiao, L. Liu, and Z. Li, Spectral gradient projection method for monotone nonlinear equations with convex constraints, *Appl. Numer. Math.*, **59**:10 (2009), 2416–2423.

[52] G. Yuan, Y. Li, and W. Hu, A conjugate gradient algorithm for large-scale nonlinear equations and image restoration problems, *Appl. Numer. Math.*, **147** (2020), 129–141.

[53] G. Yuan, Z. Meng, and Y. Li, A modified Hestenes and Stiefel conjugate gradient algorithm for large-scale nonsmooth minimizations and nonlinear equations, *J. Optim. Theory Appl.*, **168** (2016), 129–152.

[54] G. Yuan and M. Zhang, A three-terms Polak-Ribière-Polyak conjugate gradient algorithm for large-scale nonlinear equations, *J. Comput. Appl. Math.*, **286** (2015), 186–195.

[55] L. Zhang and W. Zhou, Spectral gradient projection method for solving nonlinear monotone equations, *J. Comput. Appl. Math.*, **196**:2 (2006), 478–484.

[56] N. Zhang, J. Liu, L. Zhang, and Z. Lu, A fast inertial self-adaptive projection based algorithm for solving large-scale nonlinear monotone equations, *J. Comput. Appl. Math.*, **426** (2023), 115087.

[57] W. Zhou and D. Li, Limited memory BFGS method for nonlinear monotone equations, *J. Comput. Math.*, **25**:1 (2007), 89–96.

[58] W. Zhou and D. Li, A globally convergent BFGS method for nonlinear monotone equations without any merit functions, *Math. Comp.*, **77**:264 (2008), 2231–2240.