FIXED POINT METHODS AND OPTIMIZATION
Volume 2 (2025), No. 2, 158–166
https://doi.org/10.69829/fpmo-025-0202-ta05
Tulipa Opera Scholarum

ON THE LINEAR CONVERGENCE RATE OF THE GENERALIZED PROXIMAL POINT ALGORITHM

NA XIAN[1] AND KE GUO[1,*]

[1] *Key Laboratory of Optimization Theory and Applications at China West Normal University of Sichuan Province, School of Mathematics and Information, China West Normal University, Nanchong 637009, China*

Dedicated to Professor Hari Mohan Srivastava on the Occasion of His 85th Birthday

Abstract. The proximal point algorithm (PPA) has been extensively studied in the literature, with its linear convergence rate well established. Recent studies have demonstrated that PPA retains linear convergence under specific regularity conditions, including subdifferential error bound, the Polyak–Łojasiewicz inequality, and quadratic growth. The generalized proximal point algorithm (GPPA), a relaxed variant of PPA, offers numerical acceleration compared to the classical scheme. In this paper, we focus on examining the linear convergence of GPPA under the same regularity conditions that apply to PPA.

Keywords. Generalized proximal point algorithm, Error bound, Polyak–Łojasiewicz inequality, Quadratic growth.

© Fixed Point Methods and Optimization

1. Introduction

In this paper, we consider the following optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where f is a proper closed and convex function. Let $S = \arg\min_{x\in\mathbb{R}^n} f(x)$ be the optimal solution set of problem (1.1), with f^* denoting the optimal function value. We assume that $S \neq \emptyset$ throughout the paper. The proximal point algorithm (PPA), originally introduced by Martinet [14] and later generalized by Rockafellar [18, 19], provides a fundamental framework for solving problem (1.1). Moreover, the PPA serves as a unifying foundation for several classical optimization algorithms. In particular, well-known iterative schemes such as the Douglas–Rachford splitting method [3, 12], the Peaceman–Rachford splitting method [12, 16], and the augmented Lagrangian method [7] can all be interpreted as special cases of the PPA. This unifying property highlights the significance of the PPA, as it offers a comprehensive framework for various important splitting algorithms. Starting with an arbitrary initial point x_0, the PPA generates a sequence $\{x_k\}$ via the iterative scheme

$$x_{k+1} = \mathrm{prox}_{\lambda,f}(x_k), \tag{1.2}$$

where $\mathrm{prox}_{\lambda,f}$ is the proximal mapping [15] of f, defined as

$$\mathrm{prox}_{\lambda,f}(x_k) := \arg\min_{x\in\mathbb{R}^n} \left\{ f(x) + \frac{1}{2\lambda}\|x - x_k\|^2 \right\},$$

with the proximal parameter $\lambda > 0$. The convergence properties of the PPA for convex optimization have been extensively studied since the 1970s, and various sufficient conditions have been proposed

*Corresponding author.
E-mail address: Xianna0105@163.com (N. Xian), keguo1989@cwnu.edu.cn (K. Guo)
2020 Mathematics Subject Classification: 90C25; 65K10; 90C30.
Accepted: June 22, 2025.

to guarantee its linear convergence. For example, Rockafellar [19] first established the linear convergence of the PPA under the assumption that $(\partial f)^{-1}$ is Lipschitz continuous at 0. Leventhal [10] later relaxed this requirement to metric subregularity, while still ensuring the linear convergence of the PPA. More recently, Liao et al. [11] proved the linear convergence of the PPA under several regularity conditions, including subdifferential error bound, the Polyak–Łojasiewicz inequality, and quadratic growth. These conditions are strictly weaker than those assumed in [19]. It is worth noting that for convex functions, the quadratic growth condition is equivalent to the metric subregularity of ∂f as shown in [10]. However, from an analytical perspective, Liao et al. [11] observed that the quadratic growth condition is more geometrically intuitive for establishing the linear convergence of the PPA, leading to a simpler and more transparent proof. Furthermore, these regularity conditions also ensure the linear convergence of a broad class of first-order methods [2, 5, 8, 17, 24].

A common approach to accelerate convergence is through relaxation techniques, which originate from classical methods such as successive over-relaxation for solving linear systems [21, 23] and the Krasnosel'skii–Mann iteration for fixed-point algorithms [9, 13]. The so-called generalized proximal point algorithm (GPPA) is a relaxed version of the PPA [4], whose iterative scheme is given by

$$x_{k+1} = (1-\gamma_k)x_k + \gamma_k \operatorname{prox}_{\lambda_k, f}(x_k), \tag{1.3}$$

where the relaxation parameter $\gamma_k \in (0, 1]$. Note that for $\gamma_k = 1$, this scheme reduces to classical PPA (1.2). The linear convergence of the generalized PPA (GPPA) has been studied under various conditions. Corman and Yuan [1] established the linear convergence of the GPPA under the strong convexity assumption. Tao and Yuan [22] showed that the condition that $(\partial f)^{-1}$ is Lipschitz continuous at 0 also guarantees the linear convergence for the GPPA. Shen and Pan [20] further relaxed this requirement to metric subregularity, while still ensuring the linear convergence of the algorithm.

The primary objective of this work is to extend the linear convergence analysis presented in [11] to the GPPA. We establish the linear convergence rate of the GPPA for convex optimization problems under the same regularity conditions. The remainder of this paper is organized as follows. In Section 2, we summarize some preliminaries that are useful for the subsequent analysis. In Section 3, we establish the linear convergence rate of the GPPA for convex optimization problems by employing these regularity conditions. Finally, Section 4 concludes the paper.

2. Preliminaries

In this section, we recall several essential concepts and lemmas that will be used in the subsequent analysis. Let $\mathbb{R}^n$ denote the n-dimensional Euclidean space, and let $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ denote the extended real line. The symbols $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ represent the standard inner product and the Euclidean norm on $\mathbb{R}^n$, respectively. For a closed set $S \subseteq \mathbb{R}^n$, the distance from a point $x \in \mathbb{R}^n$ to S is defined as $\operatorname{dist}(x, S) := \min_{y \in S} \|x - y\|$, and the projection of x onto S is denoted by $\Pi_S(x) := \arg\min_{y \in S} \|x - y\|$. The notation $[f \leq \nu] := \{x \in \mathbb{R}^n \mid f(x) \leq \nu\}$ denotes the ν-sublevel set of a function f.

Definition 2.1. Let $f(x) : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper closed convex function. Then for any given $x \in \mathbb{R}^n$, the subdifferential of f at x is defined by

$$\partial f(x) = \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + \langle s, y - x \rangle, \forall y \in \mathbb{R}^n\}.$$

For problem (1.1), we summarize the following regularity conditions.

Definition 2.2. Let $f(x) : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper closed convex function, and let S be the optimal solution set of problem (1.1), with f^* denoting the optimal function value. Let $\nu > 0$,

(i) Restricted Secant Inequality (RSI): we say the restricted secant inequality holds, if there exists a positive constant $\mu_{\mathrm{r}} > 0$ such that

$$\mu_{\mathrm{r}} \cdot \operatorname{dist}^2(x, S) \leq \langle g, x - \Pi_S(x) \rangle, \quad \forall x \in [f \leq f^* + \nu] \text{ and } g \in \partial f(x). \tag{RSI}$$

(ii) Error Bound (EB): we say the error bound condition holds, if there exists a positive constant $\mu_{\mathrm{e}} > 0$ such that

$$\operatorname{dist}(x, S) \leq \mu_{\mathrm{e}} \cdot \operatorname{dist}(0, \partial f(x)), \quad \forall x \in [f \leq f^* + \nu]. \tag{EB}$$

(iii) Polyak–Łojasiewicz inequality (PL): we say the Polyak–Łojasiewicz inequality holds, if there exists a positive constant $\mu_{\mathrm{p}} > 0$ such that

$$\mu_{\mathrm{p}} \cdot (f(x) - f^*) \leq \operatorname{dist}^2(0, \partial f(x)), \quad \forall x \in [f \leq f^* + \nu]. \tag{PL}$$

(iv) Quadratic Growth (QG): we say the quadratic growth condition holds, if there exists a positive constant $\mu_{\mathrm{q}} > 0$ such that

$$\mu_{\mathrm{q}} \cdot \operatorname{dist}^2(x, S) \leq f(x) - f^*, \quad \forall x \in [f \leq f^* + \nu]. \tag{QG}$$

Note that the above four regularity conditions are defined on the sublevel set $[f \leq f^* + \nu]$. If $\nu = +\infty$, then they are global. The relationships between these regularity conditions have been extensively studied for different function classes in the literature. In particular, Liao et al. [11] demonstrated the equivalence between (EB), (QG), (PL), and (RSI) for nonsmooth convex functions, and extended these results to weakly convex functions. We present the following conclusion in [11], which is useful in subsequent analysis.

Lemma 2.3. [11, Theorem 3.1] *Let f be a proper closed ρ-weakly convex function. The following relationship holds*

$$\text{(RSI)} \to \text{(EB)} \equiv \text{(PL)} \to \text{(QG)}.$$

Furthermore, if any of the following two conditions is satisfied

- *$f(x)$ is convex*
- *the (QG) coefficient satisfies $\mu_q > \frac{\rho}{2}$,*

then the following equivalence holds

$$\text{(RSI)} = \text{(EB)} = \text{(PL)} = \text{(QG)}.$$

3. Linear Convergence Rate of the Generalized Proximal Point Algorithm

When applying the regularity conditions introduced in Section 2 to prove the linear convergence of the GPPA, it is important to first establish the sublinear convergence of the GPPA in terms of the sequence of function values. This step ensures that the iterates eventually enter a sublevel set where the regularity conditions are valid and can be effectively applied. While the sublinear convergence of the PPA (1.2) in terms of function value gaps was well established by Güler [6] over three decades ago, the sublinear convergence of the GPPA (1.3) has, to the best of our knowledge, not yet been formally analyzed in the existing literature. Therefore, in this work, we aim to establish the sublinear convergence of the GPPA.

Theorem 3.1 (Sublinear convergence rate)**.** *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper, closed, convex function. Assume that the sequence $\{x_k\}$ is generated by the GPPA (1.3) with a positive stepsize sequence $\{\lambda_k\}_{k\geq 0}$ that is uniformly bounded above by a constant $\tilde{\lambda} > 0$. Then, we have*

$$f(x_k) - f(x^*) \leq \frac{1}{2(\tilde{\lambda} + \sum_{t=0}^{k-1} \lambda_t \gamma_t)} \|x_0 - x^*\|^2 + \frac{\tilde{\lambda}}{\tilde{\lambda} + \sum_{t=0}^{k-1} \lambda_t \gamma_t} (f(x_0) - f(x^*)). \tag{3.1}$$

Proof. Let $p(x_k) := \operatorname{prox}_{\lambda_k, f}(x_k)$. According to the iteration of GPPA (1.3), we have

$$\frac{1}{\lambda_k}(x_k - p(x_k)) \in \partial f(p(x_k)).$$

By the definition of the subdifferential for convex functions, we know

$$f(x^*) \geq f(p(x_k)) + \left\langle \frac{1}{\lambda_k}(x_k - p(x_k)), x^* - p(x_k) \right\rangle. \tag{3.2}$$

On the other hand, we have

$$\begin{aligned} \|p(x_k) - x^*\|^2 &= \|x_k - x^*\|^2 - \|x_k - p(x_k)\|^2 + 2\langle p(x_k) - x_k, p(x_k) - x^* \rangle \\ &\leq \|x_k - x^*\|^2 + 2\langle p(x_k) - x_k, p(x_k) - x^* \rangle, \end{aligned} \tag{3.3}$$

where the first identity follows from simple algebraic manipulations

$$\begin{aligned} \|x_k - x^*\|^2 &= \|x_k - p(x_k) + p(x_k) - x^*\|^2 \\ &= \|x_k - p(x_k)\|^2 - 2\langle p(x_k) - x_k, p(x_k) - x^* \rangle + \|p(x_k) - x^*\|^2, \end{aligned}$$

and the second inequality drops a nonnegative term $\|x_k - p(x_k)\|^2$. Combing (3.2) with (3.3) leads to

$$f(p(x_k)) - f(x^*) \leq \frac{1}{2\lambda_k}\|x_k - x^*\|^2 - \frac{1}{2\lambda_k}\|p(x_k) - x^*\|^2. \tag{3.4}$$

According to the convexity of the function $\|\cdot\|^2$ and the iteration rule of the GPPA (1.3), when $0 < \gamma_k \leq 1$, we have

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &= \|(1-\gamma_k)x_k + \gamma_k p(x_k) - x^*\|^2 \\ &= \|(1-\gamma_k)(x_k - x^*) + \gamma_k(p(x_k) - x^*)\|^2 \\ &\leq (1-\gamma_k)\|x_k - x^*\|^2 + \gamma_k\|p(x_k) - x^*\|^2. \end{aligned} \tag{3.5}$$

Combining (3.4) with (3.5), we obtain

$$f(p(x_k)) - f(x^*) \leq \frac{1}{2\lambda_k\gamma_k}\left(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2\right). \tag{3.6}$$

Moreover, by the convexity of f and the GPPA iteration (1.3), when $0 < \gamma_k \leq 1$, we have

$$\begin{aligned} f(x_{k+1}) - f(x^*) &= f((1-\gamma_k)x_k + \gamma_k p(x_k)) - f(x^*) \\ &\leq (1-\gamma_k)f(x_k) + \gamma_k f(p(x_k)) - f(x^*) \\ &= (1-\gamma_k)(f(x_k) - f(x^*)) + \gamma_k(f(p(x_k)) - f(x^*)). \end{aligned}$$

Rearranging the terms yields

$$\frac{1}{\gamma_k}(f(x_{k+1}) - f(x_k)) + f(x_k) - f(x^*) \leq f(p(x_k)) - f(x^*). \tag{3.7}$$

Combining (3.6) and (3.7), we obtain

$$\lambda_k(f(x_{k+1}) - f(x_k)) + \lambda_k\gamma_k(f(x_k) - f(x^*)) \leq \frac{1}{2}\left(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2\right). \tag{3.8}$$

Summing inequality (3.8) from $t = 0$ to $k-1$, we obtain a telescoping sum

$$\begin{aligned} \sum_{t=0}^{k-1}\lambda_t(f(x_{t+1}) - f(x_t)) + \sum_{t=0}^{k-1}\lambda_t\gamma_t(f(x_t) - f(x^*)) &\leq \sum_{t=0}^{k-1}\frac{1}{2}\left(\|x_t - x^*\|^2 - \|x_{t+1} - x^*\|^2\right) \\ &\leq \frac{1}{2}\|x_0 - x^*\|^2. \end{aligned} \tag{3.9}$$

According to the definition of the proximal mapping

$$p(x_k) = \arg\min_{x\in\mathbb{R}^n}\left\{f(x) + \frac{1}{2\lambda_k}\|x - x_k\|^2\right\},$$

we have

$$f(p(x_k)) \le f(p(x_k)) + \frac{1}{2\lambda_k}\|p(x_k) - x_k\|^2 \le f(x_k).$$

By the convexity of f, it follows that

$$\begin{aligned} f(x_{k+1}) &= f\left((1-\gamma_k)x_k + \gamma_k p(x_k)\right) \\ &\le (1-\gamma_k) f(x_k) + \gamma_k f(p(x_k)) \\ &\le f(x_k). \end{aligned}$$

Hence, the sequence $\{f(x_k)\}_{k\ge 0}$ is nonincreasing. Therefore, we can deduce that

$$\begin{aligned} \tilde{\lambda}(f(x_k) - f(x_0)) + (f(x_k) - f(x^*))\sum_{t=0}^{k-1}\lambda_t\gamma_t &\le \tilde{\lambda}(f(x_k) - f(x_0)) + (f(x_{k-1}) - f(x^*))\sum_{t=0}^{k-1}\lambda_t\gamma_t \\ &\le \sum_{t=0}^{k-1}\lambda_t(f(x_{t+1}) - f(x_t)) + \sum_{t=0}^{k-1}\lambda_t\gamma_t(f(x_t) - f(x^*)) \\ &\le \frac{1}{2}\|x_0 - x^*\|^2, \end{aligned}$$

where the last inequality follows from (3.9). By applying simple algebraic manipulations to the inequality above, we finally obtain

$$f(x_k) \le f(x^*) + \frac{1}{2\left(\tilde{\lambda} + \sum_{t=0}^{k-1}\lambda_t\gamma_t\right)}\|x_0 - x^*\|^2 + \frac{\tilde{\lambda}}{\tilde{\lambda} + \sum_{t=0}^{k-1}\lambda_t\gamma_t}(f(x_0) - f(x^*)).$$

This completes the proof. □

Remark 3.2. Note that by choosing a constant proximal parameter $\lambda_k = \lambda > 0$ and a constant relaxation parameter $\gamma_k = \gamma > 0$ in (3.1), the GPPA (1.3) directly achieves the standard sublinear convergence rate of $O(1/k)$.

Next, we extend the results in [11] by proving that the GPPA (1.3) also achieves linear convergence under the same regularity conditions. In particular, we establish the linear convergence of the GPPA through three distinct approaches:

(i) leveraging the Polyak–Łojasiewicz (PL) inequality to show the linear convergence of the function value sequence;
(ii) utilizing the error bound (EB) condition to prove the linear convergence of the iterate sequence;
(iii) applying the quadratic growth (QG) condition to establish the linear convergence of the iterate sequence.

Theorem 3.3 (Linear convergence rate)**.** *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper, closed, convex function. Assume that f satisfies one of the regularity conditions* (PL), (EB), (QG), *or* (RSI) *over the sublevel set $[f \le f^* + \nu]$. Assume that the sequence $\{x_k\}$ is generated by the GPPA* (1.3) *with a positive stepsize sequence $\{\lambda_k\}_{k\ge 0}$ that is uniformly bounded above by a constant $\tilde{\lambda} > 0$. Then, for all $k \ge k_0$, the iterates $\{x_k\}$ of the GPPA* (1.3) *exhibit a linear convergence rate, that is,*

$$\begin{aligned} f(x_{k+1}) - f^* &\le \omega_k(f(x_k) - f^*), \\ \mathrm{dist}(x_{k+1}, S) &\le \theta_k \mathrm{dist}(x_k, S), \end{aligned}$$

where the constants are

$$\omega_k = \frac{2 + \mu_p\lambda_k - \mu_p\lambda_k\gamma_k}{2 + \mu_p\lambda_k} < 1, \;\; \theta_k = \min\left\{1 - (1 - \frac{1}{\sqrt{1 + \lambda_k^2/\mu_e^2}})\gamma_k, 1 - (1 - \frac{1}{\sqrt{2\lambda_k\mu_q + 1}})\gamma_k\right\} < 1,$$

$$k_0 = \frac{\|x_0 - x^*\|^2 + 2\tilde{\lambda}(f(x_0) - f^* - \nu)}{2\nu \min_{k\geq 0}\{\lambda_k \gamma_k\}}.$$

Proof. The result in (3.1) ensures that the iterate x_k reaches the sublevel set $[f \leq f^* + \nu]$ after at most k_0 iterations. (If $f(x_0) \leq f^* + \nu$, then $f(x_k) \leq f(x_0) \leq f^* + \nu$ for all $k \geq 0$; If $f(x_0) > f^* + \nu$, then x_k reaches the set $[f \leq f^* + \nu]$ after at most k_0 iterations.) According to Lemma 2.3, once x_k is within $[f \leq f^* + \nu]$, the regularity conditions (EB), (PL), (QG) and (RSI) are equivalent. For the analysis below, we assume $x_k \in [f \leq f^* + \nu]$.

(i) Firstly, we use (PL) to show the linear convergence of the function value sequence of GPPA (1.3). According to the iteration of GPPA (1.3) and the convexity of f, we have

$$\frac{1}{\lambda_k}(x_k - p(x_k)) \in \partial f(p(x_k)), \tag{3.10}$$

and when $0 < \gamma_k \leq 1$, we have

$$\begin{aligned} f(x_{k+1}) - f^* &= f\left((1-\gamma_k)x_k + \gamma_k p(x_k)\right) - f^* \\ &\leq (1-\gamma_k) f(x_k) + \gamma_k f(p(x_k)) - f^* \\ &= (1-\gamma_k)(f(x_k) - f^*) + \gamma_k (f(p(x_k)) - f^*). \end{aligned} \tag{3.11}$$

Then, the following inequalities hold

$$\begin{aligned} f(x_k) - f(p(x_k)) \geq \frac{1}{2\lambda_k}\|p(x_k) - x_k\|^2 &\geq \frac{\lambda_k}{2}\operatorname{dist}^2\left(0, \partial f(p(x_k))\right) \\ &\geq \frac{\lambda_k \mu_p}{2}\left(f(p(x_k)) - f^*\right), \end{aligned} \tag{3.12}$$

where the first inequality follows from the definition of proximal mapping, the second comes from the optimality in (3.10), and the last inequality applies the (PL) inequality. Rearranging and subtracting f^* from both sides of (3.12) leads to the result

$$f(p(x_k)) - f^* \leq \frac{2}{2 + \mu_p \lambda_k} \cdot \left(f(x_k) - f^*\right). \tag{3.13}$$

Combining (3.11) and (3.13), we have

$$f(x_{k+1}) - f^* \leq \frac{2 + \mu_p\lambda_k - \mu_p\lambda_k\gamma_k}{2 + \mu_p\lambda_k} \cdot (f(x_k) - f^*),$$

where $\frac{2+\mu_p\lambda_k - \mu_p\lambda_k\gamma_k}{2+\mu_p\lambda_k} < 1$.

(ii) Now, we employ (EB) to demonstrate the linear convergence of the iterate sequence of GPPA (1.3). According to the firmly nonexpensive of the proximal point operator for a convex function f, we have

$$\|x - y\|^2 \geq \|p(x) - p(y)\|^2 + \|(x - p(x)) - (y - p(y))\|^2.$$

Setting $x = x_k$ and $y = x^*$, we have

$$\|x_k - x^*\|^2 \geq \|p(x_k) - x^*\|^2 + \|x_k - p(x_k)\|^2. \tag{3.14}$$

Then, the following inequalities hold

$$\|x_k - p(x_k)\|^2 \geq \lambda_k^2 \operatorname{dist}^2(0, \partial f(p(x_k))) \geq \frac{\lambda_k^2}{\mu_e^2}\operatorname{dist}^2(p(x_k), S), \tag{3.15}$$

where the first inequality comes from the optimality (3.10), the second applies the (EB) condition. Combining (3.14) with (3.15), we have

$$\|p(x_k) - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{\lambda_k^2}{\mu_e^2}\operatorname{dist}^2(p(x_k), S).$$

Using the fact that

$$\text{dist}^2(p(x_k), S) \leq \|p(x_k) - x^*\|^2 .$$

So we have

$$\text{dist}^2(p(x_k), S) \leq \frac{\mu_e^2}{\lambda_k^2 + \mu_e^2} \|x_k - x^*\|^2 .$$

Setting $x^* = \Pi_S(x_k) = \arg\min_{y \in S} \|x_k - y\|$, we have

$$\text{dist}(p(x_k), S) \leq \frac{\mu_e}{\sqrt{\lambda_k^2 + \mu_e^2}} \text{dist}(x_k, S). \tag{3.16}$$

According to the convexity of the distance function and the iteration of GPPA (1.3), we have

$$\begin{aligned}\text{dist}(x_{k+1}, S) &= \text{dist}((1-\gamma_k)x_k + \gamma_k p(x_k), S) \\ &\leq (1-\gamma_k)\text{dist}(x_k, S) + \gamma_k \text{dist}(p(x_k), S).\end{aligned} \tag{3.17}$$

Combining (3.16) with (3.17), we have

$$\text{dist}(x_{k+1}, S) \leq \left\{ 1 - (1 - \frac{1}{\sqrt{1 + \lambda_k^2/\mu_e^2}})\gamma_k \right\} \cdot \text{dist}(x_k, S),$$

where

$$1 - \left(1 - \frac{1}{\sqrt{1 + \lambda_k^2/\mu_e^2}} \right) \gamma_k < 1.$$

(iii) Finally, we apply (QG) to establish the linear convergence of the iterate sequence of GPPA (1.3). By definition, we have

$$f(\Pi_S(x_k)) = f^* \text{ and } \|\Pi_S(x_k) - x_k\|^2 = \text{dist}^2(x_k, S).$$

Since $g(x) := f(x) + \frac{1}{2\lambda_k}\|x - x_k\|^2$ is $\frac{1}{\lambda_k}$ strongly convex, its first-order lower bound at $p(x_k)$ is

$$\begin{aligned}f^* + \frac{1}{2\lambda_k}\text{dist}^2(x_k, S) &= f(\Pi_S(x_k)) + \frac{1}{2\lambda_k}\|\Pi_S(x_k) - x_k\|^2 \\ &\geq f(p(x_k)) + \frac{1}{2\lambda_k}\|p(x_k) - x_k\|^2 + \langle 0, \Pi_S(x_k) - p(x_k)\rangle \\ &\quad + \frac{1}{2\lambda_k}\|\Pi_S(x_k) - p(x_k)\|^2 .\end{aligned} \tag{3.18}$$

From (3.18), we drop the positive term $\frac{1}{2\lambda_k}\|p(x_k) - x_k\|^2$ and use the fact that

$$\|\Pi_S(x_k) - p(x_k)\| \geq \text{dist}(p(x_k), S),$$

leading to

$$f^* - f(p(x_k)) + \frac{1}{2\lambda_k}\text{dist}^2(x_k, S) \geq \frac{1}{2\lambda_k}\text{dist}^2(p(x_k), S).$$

Combining this inequality with (QG) and simple rearrangement, we obtain

$$\text{dist}(p(x_k), S) \leq \frac{1}{\sqrt{2\lambda_k\mu_q + 1}} \text{dist}(x_k, S). \tag{3.19}$$

By the convexity of the distance function and the iteration of the GPPA (1.3), combining (3.17) with (3.19), we get

$$\operatorname{dist}(x_{k+1}, S) \leq \left\{1 - \left(1 - \frac{1}{\sqrt{2\lambda_k \mu_q + 1}}\right)\gamma_k\right\} \cdot \operatorname{dist}(x_k, S),$$

where

$$1 - \left(1 - \frac{1}{\sqrt{2\lambda_k \mu_q + 1}}\right)\gamma_k < 1.$$

□

Remark 3.4. In Theorem 3.3, if x_0 is in the sublevel set $[f \leq f^* + \nu]$, combining this with the fact that the function values $\{f(x_k)\}_{k\geq 0}$ are nonincreasing, we immediately have $f(x_k) \leq f^* + \nu$ for all $k \geq 0$. Otherwise, if $f(x_0) > f^* + \nu$, then the result in (3.1) ensures that the iterate x_k reaches the set $[f \leq f^* + \nu]$ after at most

$$k_0 = \frac{\|x_0 - x^*\|^2 + 2\tilde{\lambda}(f(x_0) - f^* - \nu)}{2\nu \min_{k\geq 0}\{\lambda_k \gamma_k\}} > 0$$

iterations.

4. Conclusion

In this work, we have studied the linear convergence of the generalized proximal point algorithm (GPPA) for convex optimization problems. Our main theoretical contribution demonstrates that the GPPA attains linear convergence under certain regularity conditions, including subdifferential error bound, the Polyak–Łojasiewicz inequality, and quadratic growth.

Statements and Declarations

The authors declare no competing interests. All authors contributed to the study conception and design. Material preparation and data analysis were performed jointly by all authors. All authors read and approved the final manuscript.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 11801455), the Sichuan Science and Technology Program (Grant No. 2023NSFSC1922), the Innovation Team Funds of China West Normal University (Grant No. KCXTD2023-3), and the Fundamental Research Funds of China West Normal University (Grant No. 23kc010).

References

[1] E. Corman and X. M. Yuan. A generalized proximal point algorithm and its convergence rate. *SIAM Journal on Optimization*, 24(4):1614–1638, 2014.

[2] M. Díaz and B. Grimmer. Optimal convergence rates for the proximal bundle method. *SIAM Journal on Optimization*, 33(2):424–454, 2023.

[3] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82(2):421–439, 1956.

[4] E. G. Gol'shtein and N. V. Tret'yakov. Modified lagrangians in convex programming and their generalizations. In P. Huard, editor, *Point-to-Set Maps and Mathematical Programming*, pages 86–97. Springer, Berlin, Heidelberg, 1979.

[5] C. Guille-Escuret, A. Ibrahim, B. Goujaud, and I. Mitliagkas. Gradient descent is optimal under lower restricted secant inequality and upper error bound. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc.

[6] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991.

[7] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4(5):303–320, 1969.

[8] H. Karimi, J. Nutini, and M. Schmidt. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. In P. Frasconi, N. Landwehr, G. Manco, and J. Vreeken, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 9851, pages 795–811, ECML PKDD 2016, Italy, 2016. Springer, Cham.

[9] M. A. Krasnosel'skii. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10(1):123–127, 1955.

[10] D. Leventhal. Metric subregularity and the proximal point method. *Journal of Mathematical Analysis and Applications*, 360(2):681–688, 2009.

[11] F. Y. Liao, L. J. Ding, and Y. Zheng. Error bounds, pl condition, and quadratic growth for weakly convex functions, and linear convergences of proximal point methods. *Proceedings of Machine Learning Research*, 242:993–1005, 2024.

[12] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[13] W. R. Mann and W. R. Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4:506–510, 1953.

[14] B. Martinet. Brève communication. Régularisation d'inéquations variationnelles par approximations successives. *Revue Française D'informatique et de Recherche Opérationnelle. Série Rouge*, 4(R3):154–158, 1970.

[15] J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.

[16] D. W. Peaceman and H. H. Rachford, Jr. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics*, 3(1):28–41, 1955.

[17] B. Polyak. Gradient methods for the minimisation of functionals. *USSR Computational Mathematics and Mathematical Physics*, 3(4):864–878, 1963.

[18] R. T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1(2):97–116, 1976.

[19] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.

[20] L. Shen and S. H. Pan. Linear convergence of the generalized ppa and several splitting methods for the composite inclusion problem. *ArXiv preprint, arXiv:1508.05156*, 2015.

[21] G. Shortley. Relaxation methods in theoretical physics. *Science*, 105(2730):455–456, 1947.

[22] M. Tao and X. M. Yuan. On the optimal linear convergence rate of a generalized proximal point algorithm. *Journal of Scientific Computing*, 74:826–850, 2018.

[23] R. S. Varga. *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs, 1962.

[24] Y. C. Zhang and L. Xiao. Stochastic primal-dual coordinate method for regularized empirical risk minimization. *Journal of Machine Learning Research*, 18:1–42, 2017.