

Relaxed Proximal Point Algorithm: Tight Complexity Bounds and Acceleration without Momentum

Bofan Wang*, Shiqian Ma^{†‡}, Junfeng Yang^{*‡}, Danqing Zhou*

Abstract

In this paper, we focus on the *relaxed proximal point algorithm* (RPPA) for solving convex (possibly nonsmooth) optimization problems. We conduct a comprehensive study on three types of relaxation schedules: (i) constant schedule with relaxation parameter $\alpha_k \equiv \alpha \in (0, \sqrt{2}]$, (ii) the dynamic schedule put forward by Teboulle and Vaisbourd [TV23], and (iii) the silver stepsize schedule proposed by Altschuler and Parrilo [AP23b]. The latter two schedules were initially investigated for the *gradient descent* (GD) method and are extended to the RPPA in this paper. For type (i), we establish tight non-ergodic $O(1/N)$ convergence rate results measured by function value residual and subgradient norm, where N denotes the iteration counter. For type (ii), we establish a convergence rate that is tight and approximately $\sqrt{2}$ times better than the constant schedule of type (i). For type (iii), aside from the original silver stepsize schedule put forward by Altschuler and Parrilo, we propose two new modified silver stepsize schedules, and for all the three silver stepsize schedules, $O(1/N^{1.2716})$ accelerated convergence rate results with respect to three different performance metrics are established. Furthermore, our research affirms the conjecture in [LG24, Conjecture 3.2] on GD method with the original silver stepsize schedule.

Mathematical Subject Classifications: 90C25

1 Introduction

In this paper, we consider the following general convex optimization problem

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad f(x), \quad (\mathcal{Q})$$

where $f: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ is an extended-real-valued proper, closed, and convex (possibly nonsmooth) function. Denote the optimal solution set of (\mathcal{Q}) by \mathcal{X}^* , which is assumed to be nonempty. Throughout this paper, we let $x^* \in \mathcal{X}^*$ be arbitrarily fixed. The *proximal mapping* [Mor65] of f is defined by

$$\text{Prox}_{\lambda f}(x) = \arg \min_{y \in \mathbb{R}^d} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}, \quad \forall x \in \mathbb{R}^d, \forall \lambda > 0. \quad (1)$$

A classic approach for solving (\mathcal{Q}) is the *proximal point algorithm* (PPA), which, starting at $x^0 \in \mathbb{R}^d$, iterates for $k \geq 0$ as $x^{k+1} = \text{Prox}_{\lambda f}(x^k)$. PPA was first introduced in [Mar70] and was further studied by Rockafellar [Roc76b, Roc76a].

Convergence rate results of PPA with varying proximal parameters $\{\lambda_k\}_{k=0}^{N-1}$ for solving (\mathcal{Q}) was first established by Güler [Gül91], where $O(1/\sum_{k=0}^{N-1} \lambda_k)$ complexity bounds for both function value residual and subgradient norm were derived. Recently, the results derived by [Gül91] were tightened by improving a constant factor in [THG17a].

A common technique for accelerating convergence is relaxation, which originates from methods like successive over-relaxation for solving linear systems [Sou46, Var62] and Krasnosel'skiĭ-Mann (KM) iteration for fixed-point algorithms [Kra55, Man53]. Given a positive integer $N \in \mathbb{N}_+$, an initialization point $x^0 \in \mathbb{R}^d$ and N relaxation parameters $\{\alpha_0, \alpha_1, \dots, \alpha_{N-1}\} \subseteq \mathbb{R}_{++}$, the N -step *relaxed proximal point algorithm* (RPPA) for solving (\mathcal{Q}) is summarized in Algorithm 1. The sequence of relaxation

*School of Mathematics, Nanjing University, Nanjing, P. R. China. Research supported by National Natural Science Foundation of China (12371301, 12431011). Email: wangbf@smail.nju.edu.cn, zhoudanqing@smail.nju.edu.cn.

[†]Department of Computational Applied Mathematics and Operations Research, Rice University, Houston, TX, USA.

[‡]Corresponding authors: Shiqian Ma (sqma@rice.edu) and Junfeng Yang (jfyang@nju.edu.cn).

parameters, denoted by vector $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{N-1}) \in \mathbb{R}_{++}^N$, is also referred to as the relaxation schedule. It has been widely recognized empirically that relaxation, especially over-relaxation, i.e., $\alpha_i > 1$, can enhance the numerical performance of fixed-point algorithms and optimization methods, see, e.g., [EB92, Boy⁺11, CGH14].

Algorithm 1: Relaxed Proximal Point Algorithm

input : $x^0 \in \mathbb{R}^d$, $\lambda > 0$, $N \in \mathbb{N}_+$, $\alpha \in \mathbb{R}_{++}^N$
for $k = 0, 1, 2, \dots, N - 1$ **do**
 $z^k = \text{Prox}_{\lambda f}(x^k)$ (Proximal step)
 $x^{k+1} = x^k + \alpha_k(z^k - x^k)$ (Relaxation step)
end
output: $z^N = \text{Prox}_{\lambda f}(x^N)$

Various studies have analyzed the convergence of RPPA in different settings. For instance, it has been applied to solving variational inequality problems in [LY13, GY24], where ergodic sublinear convergence rates were established. However, the performance measures used in these works cannot be applied to our setting. In [CY14, GHY14], RPPA was analyzed for solving monotone inclusion and mixed variational inequality problems. While the performance measures used in these works can be adapted to our setting, only ergodic convergence rates were provided. Recently, with the fixed point residual being used as the performance measure, tight non-ergodic convergence rate results of the KM iteration were established in [Lie18], and these results also apply to RPPA. Detailed convergence bound results of RPPA under different measures for different stepsize schedules, along with the corresponding references, are summarized in Table 1.

Despite the various studies on RPPA, current research has several limitations. As far as we know, there has been no analysis regarding the tight convergence rate of RPPA with a constant relaxation schedule for solving (Q). Moreover, existing studies fail to demonstrate a clear advantage of RPPA over PPA in terms of the order of convergence rate. In this paper, we conduct a theoretical examination of how different relaxation schedules impact the worst-case guarantees of PPA. We will study existing relaxation schedules and also propose new ones. In particular, we establish tight convergence rate results for RPPA and show that RPPA can achieve improved complexity bounds compared to PPA. As byproducts, we also solve some conjectures in the literature by providing their formal proofs. A majority of the main results derived in this paper are summarized in Table 1.

1.1 Connection with gradient descent methods

Our study is closely related to the recent advances in the analysis of *gradient descent* (GD) methods [DT14, TV23, AP23b, GSW24]. This is because RPPA can be viewed as applying GD method to the Moreau envelope of f , which is defined as

$$f^\lambda(x) := \min_{u \in \mathbb{R}^d} \left\{ f(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\}, \quad x \in \mathbb{R}^d, \lambda > 0.$$

It is well known that f^λ is convex, $(1/\lambda)$ -smooth, i.e., it is differentiable and its gradient is $(1/\lambda)$ -Lipschitz continuous, and furthermore, it shares the same set of minimizers as f , see, e.g., [Bec17, Sec. 6.7]. In particular, for $k \in \{\star, 0, 1, 2, \dots, N - 1\}$, the sequence generated by Algorithm 1 together with $z^\star := \text{Prox}_{\lambda f}(x^\star) = x^\star$ satisfy

$$\frac{x^k - z^k}{\lambda} = \frac{x^k - \text{Prox}_{\lambda f}(x^k)}{\lambda} = \nabla f^\lambda(x^k) \in \partial f(z^k), \tag{2}$$

$$f^\lambda(x^k) = f(z^k) + \frac{1}{2\lambda} \|x^k - z^k\|^2 = f(z^k) + \frac{\lambda}{2} \|\nabla f^\lambda(x^k)\|^2. \tag{3}$$

As a result, the iterative scheme of RPPA can be equivalently expressed as

$$x^{k+1} = x^k + \alpha_k(\text{Prox}_{\lambda f}(x^k) - x^k) \equiv x^k - \alpha_k \lambda \nabla f^\lambda(x^k), \quad k = 0, 1, \dots, N - 1, \tag{4}$$

which is precisely the N -step GD method applied to the $(1/\lambda)$ -smooth function f^λ with stepsize schedule $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{N-1})$ normalized by the Lipschitz constant $(1/\lambda)$. For this reason, we will also refer to the relaxation schedule α of RPPA as the stepsize schedule.

Measure	Stepsize Schedule α	N arbitrary?	Upper Bound	Lower Bound
	$\alpha_k \equiv 1$ (reduce to PPA)	Yes	$\frac{1}{1+N} \frac{1}{4\lambda}$ [THG17a, Thm. 4.1]	
	$\alpha_k \equiv \alpha \in (0, \sqrt{2}]$	Yes	$\frac{1}{1+N} \frac{1}{\alpha 4\lambda}$ [Ours, Thm. 3.1]	$\frac{1}{1+\sum_{k=0}^{N-1} \alpha_k} \frac{1}{4\lambda}$
$\frac{f(x^N) - f(x^*)}{\ x^0 - x^*\ ^2}$	Teboulle and Vaisbourd's schedule (9) [TV23]	Yes	$\frac{1}{1+\sum_{k=0}^{N-1} \alpha_k} \frac{1}{4\lambda} \sim \frac{1}{1+2N} \frac{1}{4\lambda}$ [Ours, Thm. 4.1]	[Ours, Lem. 2.3 (ii)]
	Right silver schedule $\bar{\pi}^{(m)}$ (38) [Ours]	$N = 2^{m-1}$	$\frac{1}{1+\sum_{k=0}^{N-1} \alpha_k} \frac{1}{4\lambda} \sim \frac{1}{N^{\log_2 \rho}} \frac{1}{4\lambda}$ [Ours, Thm. 5.4]	
$\frac{\ \nabla f^\lambda(x^N)\ ^2}{f(x^0) - f(x^*)}$	Left silver schedule $\underline{\pi}^{(m)}$ (40) [Ours]	$N = 2^m$	$\frac{1}{1+\sum_{k=0}^{N-1} \alpha_k} \frac{1}{\lambda} \sim \frac{1}{N^{\log_2 \rho}} \frac{1}{\lambda}$ [Ours, Thm. 5.5]	$\frac{1}{1+\sum_{k=0}^{N-1} \alpha_k} \frac{1}{\lambda}$
	$\alpha_k \equiv \alpha \in [1, 2)$	Yes	$\frac{1}{\sqrt{N+1}} \left(\frac{N}{N+1} \right)^{N/2} \frac{1}{\lambda \sqrt{\alpha(2-\alpha)}}$, if $1 \leq \alpha \leq 1 + \sqrt{\frac{N}{N+1}}$ $(\alpha - 1)^{N \frac{1}{\lambda}}$, if $1 + \sqrt{\frac{N}{N+1}} < \alpha < 2$	
$\frac{\ \nabla f^\lambda(x^N)\ }{\ x^0 - x^*\ }$	$\alpha_k \equiv 1$ (reduce to PPA)	Yes	[Lie18, Thm. 4.9]	$\frac{1}{1+\sum_{k=0}^{N-1} \alpha_k} \frac{1}{\lambda}$
	$\alpha_k \equiv \alpha \in (0, \sqrt{2}]$	Yes	$\frac{1}{1+N} \frac{1}{\lambda}$ [Gül91, Thm. 2.2]	[Ours, Lem. 2.3 (i)]
	Silver schedule $\pi^{(m)}$ (31) [AP23b]	$N = 2^m - 1$	$\frac{1}{1+\alpha N} \frac{1}{\lambda}$ [Ours, Rmk. 3.1]	
			$\frac{1}{1+\sum_{k=0}^{N-1} \alpha_k} \frac{1}{\lambda} \sim \frac{1}{N^{\log_2 \rho}} \frac{1}{\lambda}$ [Ours, Thm. 5.3]	

Table 1: Convergence bound results of RPPA with different stepsize schedules under different measures for solving (Q). $\rho = 1 + \sqrt{2} \approx 2.414$ is the silver ratio.

Given the connection expressed in (4) between RPPA and GD methods, it appears that the convergence results of GD methods can be extended to RPPA in a straightforward manner. However, if extended in this way, the resulting results are characterized by the Moreau envelope f^λ , whereas our primary interest lies in f itself. Moreover, the results obtained in this manner are usually worse than the tight complexity bound that we desire. Next, we will review the key results from the GD literature. This will help place our contributions in context.

1.2 Prior arts on stepsize design for gradient descent methods

Let $x^0 \in \mathbb{R}^d$ be a given starting point. Consider minimizing an L -smooth convex function h via the N -step GD method with stepsize schedule $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{N-1}) \in \mathbb{R}_{++}^N$ normalized by L , that is

$$x^{k+1} = x^k - \frac{\alpha_k}{L} \nabla h(x^k), \quad k = 0, 1, \dots, N-1. \quad (5)$$

Traditional analysis of GD is confined to the scenario of “short stepsizes”. In this case, the (normalized) stepsizes α_k are chosen in $(0, 2)$, as seen [DT14, Kim+23, THG17c, TV23]. This strategy ensures a decrease in the function value at each iteration. For constant stepsizes $\alpha_k \equiv \alpha \in (0, 1]$, tight convergence rate results have been established under three optimality measures, which are $\|\nabla h(x^N)\| / \|x^0 - x^*\|$ in [TV23], $(h(x^N) - h(x^*)) / \|x^0 - x^*\|^2$ in [DT14, TV23] and $\|\nabla h(x^N)\|^2 / (h(x^0) - h(x^*))$ in [Kim+23]. Very recently, an exact full-range analysis of GD method for any constant stepsizes $\alpha_k \equiv \alpha \in (0, 2)$ was conducted in [RGP24]. In [TV23], a dynamic stepsize schedule with $\alpha_k \rightarrow 2$ was proposed. This schedule improves the complexity bound of the GD method with $\alpha_k \equiv 1$ by a factor of 2 asymptotically for the measure $(h(x^N) - h(x^*)) / \|x^0 - x^*\|^2$. Similarly, another dynamic schedule was introduced in [RGP24], which achieved an accelerated rate under the measure $\|\nabla h(x^N)\|^2 / (h(x^0) - h(x^*))$. In this “short stepsize” case, the best possible convergence rate without momentum is $O(1/N)$, see [DVR24, Sec. 6.1.1].

Recent research has delved into “long stepsizes”, where the (normalized) stepsizes α_k can surpass the bound of 2. Although this approach may result in an increase in function value at some iterations, a meticulous design of the stepsizes can potentially enhance the final complexity bound. Altschuler’s master’s thesis [Alt18] first identified optimal stepsize schedules for strongly convex problems with $N = 2, 3$. Das Gupta et al. proposed in [DVR24] the use of a branch-and-bound technique to solve the nonconvex performance estimation problem. This approach, referred to as BnB-PEP in the rest of this paper, can numerically find the optimal stepsize schedules for any $N \geq 1$. However, due to the nonconvexity of the stepsize optimization problem, only optimal schedules for $N \leq 50$ were provided as only small-scale problems can be solved by the branch-and-bound approach due to their nonconvex nature. Grimmer [Gri24] achieved asymptotic convergence rates by cycling through finite “straightforward” schedules, thereby improving the classical convergence rates by a constant factor.

All the works reviewed above maintain the convergence rate of $O(1/N)$ for the GD method without altering the order. However, recent studies [AP23a, AP23b, GSW23, GSW24] have developed stepsize sequences that boost the convergence rate of GD method to a higher order. Motivated by the numerically optimal step sizes in [DVR24], Grimmer et al. in [GSW23] proposed an infinite “straightforward” stepsize schedule, which achieves an $O(1/N^{1.0564})$ rate in the non-strongly convex setting. In [AP23a], Altschuler and Parrilo introduced the *silver stepsize schedule* for the GD method in the μ -strongly convex setting, which is named after its use of the silver ratio $\rho := 1 + \sqrt{2}$ in the stepsizes. In the limit as μ approaches 0, an accelerated $O(1/N^{1.2716})$ rate for the function value residual in the non-strongly convex case was derived in [AP23b]. Later, Grimmer et al. [GSW24] proposed a “right-side heavy” stepsize schedule based on the silver stepsize schedule, which improves the original result by a constant factor of approximately 2.3244. Furthermore, motivated by the H-duality of gradient-based algorithms as described in [Kim+23], Grimmer et al. in [GSW24] also introduced a “left-side heavy” stepsize schedule. This schedule mirrors the “right-side heavy” stepsize schedule and achieves an $O(1/N^{1.2716})$ rate for the gradient norm.

The above-reviewed stepsize schedules, along with their corresponding convergence rate results and references, are summarized in Table 2.

1.3 Contributions

The contributions of this paper are summarized below.

Measure	Stepsize Schedule α	N arbitrary?	Upper Bound	Lower Bound
$\frac{h(x^N) - h(x^*)}{\ x^0 - x^*\ ^2}$	$\alpha_k \equiv \alpha \in (0, 1]$	Yes	$\frac{L}{2+4N\alpha}$ [DT14, TV23]	
	Teboulle and Vaisbourd's schedule (9) [TV23]	Yes	$\frac{L}{2+4\sum_{k=0}^{N-1} \alpha_k} \sim \frac{L}{2+8N}$ [TV23]	$\frac{L}{2+4\sum_{k=0}^{N-1} \alpha_k}$ [LG24, Lem. 3.1]
	Silver schedule $\pi^{(m)}$ (31) [AP23b]	$N = 2^m - 1$	$\frac{L}{1+\sqrt{4\rho^{2m}-3}} \sim \frac{L}{2N^{\log_2 \rho}}$ [AP23b]	
			$\frac{L}{2+4\sum_{k=0}^{N-1} \alpha_k} = \frac{L}{4\rho^m - 2} \sim \frac{L}{4N^{\log_2 \rho}}$ [Ours, Thm. 5.2]	
	Right-side heavy schedule $\mathfrak{h}_{\text{right}}^{(m)}$ [GSW24]	$N = 2^m - 1$	$\frac{L}{2+4\sum_{k=0}^{N-1} \alpha_k} \sim \frac{1}{2(\rho-1)(1+\rho^{-1/2})} \frac{L}{N^{\log_2 \rho}}$ [GSW24]	
	$\alpha_k \equiv \alpha \in (0, 1]$	Yes	$\frac{2L}{1+2N\alpha}$ [Kim+23]	
$\frac{\ \nabla h(x^N)\ ^2}{h(x^0) - h(x^*)}$	$\alpha_k \equiv \alpha \in (0, 2)$	Yes	$2L \max \left\{ \frac{1}{1+2N\alpha}, (1-\alpha)^{2N} \right\}$ [RGP24]	$\frac{2L}{1+2\sum_{k=0}^{N-1} \alpha_k}$ [GSW24, Rmk. 2]
	Teodor, François and Panagiotis's schedule [RGP24, Sec. 2.6.2]	Yes	$\frac{2L}{1+2\sum_{k=0}^{N-1} \alpha_k} \sim \frac{2L}{1+4N}$ [RGP24]	
	Left-side heavy schedule $\mathfrak{h}_{\text{left}}^{(m)}$ [GSW24]	$N = 2^m - 1$	$\frac{2L}{1+2\sum_{k=0}^{N-1} \alpha_k} \sim \frac{2}{(\rho-1)(1+\rho^{-1/2})} \frac{L}{N^{\log_2 \rho}}$ [GSW24]	
	$\alpha_k \equiv \alpha \in (0, 1]$	Yes	$\frac{L}{1+N\alpha}$ [TV23]	$\frac{L}{1+\sum_{k=0}^{N-1} \alpha_k}$ [LG24, Lem. 3.1]
	Silver schedule $\pi^{(m)}$ (31) [AP23b]	$N = 2^m - 1$	$\frac{L}{1+\sum_{k=0}^{N-1} \alpha_k} = \frac{L}{\rho^m} \sim \frac{L}{N^{\log_2 \rho}}$ [Ours, Thm. 5.2]	
	Mixed schedule $[\mathfrak{h}_{\text{right}}^{(m)}, \mathfrak{h}_{\text{left}}^{(m)}]$ [GSW24]	$N = 2^{m+1} - 2$	$\frac{L}{1+\sum_{k=0}^{N-1} \alpha_k} \sim \frac{\rho}{(\rho-1)(1+\rho^{-1/2})} \frac{L}{N^{\log_2 \rho}}$ [GSW24]	

Table 2: Convergence bound results of GD method with different stepsize schedules under different measures. Here, the objective function h is L -smooth and convex, and $\rho = 1 + \sqrt{2} \approx 2.4142$ denotes the silver ratio. Our Theorem 5.2 establishes the tight convergence rates of GD method with silver stepsize schedule that have been conjectured in [LG24]. The tight rate of silver stepsize schedule under the measure $(h(x^N) - h(x^*)) / \|x^0 - x^*\|^2$ is very close to that of the right-side heavy schedule $\mathfrak{h}_{\text{right}}^{(m)}$ as $(\rho - 1)(1 + \rho^{-1/2}) \approx 2.3244$. For the measure $\|\nabla h(x^N)\| / \|x^0 - x^*\|$, the silver stepsize schedule achieves even better rate than the mixed schedule $[\mathfrak{h}_{\text{right}}^{(m)}, \mathfrak{h}_{\text{left}}^{(m)}]$.

1. We establish tight complexity bounds for RPPA with the constant relaxation schedule $\alpha_k \equiv \alpha \in (0, \sqrt{2}]$ in terms of function value residual and subgradient norm, see Theorem 3.1 and Remark 3.1, respectively.
2. We establish a tight complexity bound for RPPA with the dynamic relaxation schedule proposed by Teboulle and Vaisbourd in [TV23]. Once again, we use the function value residual as the optimality measure. The main result is presented in Theorem 4.1, which improves the result in Theorem 3.1 by a constant factor of approximately $\sqrt{2}$ as $N \rightarrow \infty$.
3. The original silver stepsize schedule, denoted by $\pi^{(m)}$, was proposed in [AP23b] for GD method. In this paper, we propose two extensions of it, i.e., the left and the right silver stepsize schedules, denoted respectively by $\underline{\pi}^{(m)}$ and $\overline{\pi}^{(m)}$. For $\pi^{(m)}$, $\underline{\pi}^{(m)}$ and $\overline{\pi}^{(m)}$, we establish tight complexity bounds based on three different optimality measures $\|\nabla f^\lambda(x^N)\| / \|x^0 - x^*\|$ (Theorem 5.3), $\|\nabla f^\lambda(x^N)\|^2 / (f(x^0) - f(x^*))$ (Theorem 5.5) and $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$ (Theorem 5.4), respectively. The established bounds show that RPPA with all these silver stepsize schedules converges at the same rate $O(1/N^{1.2716})$, which is faster than the classical rate $O(1/N)$ for standard GD method with “short stepsizes” $\alpha_k \in (0, 2)$. Furthermore, the established upper bounds are tight since they match the lower bounds derived in Lemma 2.3. See Section 5 for details.
4. As a byproduct of our research, we establish in Theorem 5.2 tight convergence rate results for the GD method with the silver stepsize schedule $\pi^{(m)}$ under two different optimality measures for solving smooth convex optimization problems. These results provide an affirmative answer to the conjecture in [LG24, Conjecture 3.2].

1.4 Notation and organization

Our notation is standard. Here, we clarify some of it. We denote \mathbb{R}_{++} as the set of all positive real numbers; \mathbb{R}^d represents the standard Euclidean space with dimension d ; \mathbb{R}_{++}^d is the subset of \mathbb{R}^d consisting of all vectors in \mathbb{R}^d with positive components; $\mathbb{R}^{m \times n}$ stands for the set of $m \times n$ matrices; and \mathbb{N}_+ is the set of positive integers. The standard inner product and the induced norm on \mathbb{R}^d are denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ respectively. The *subgradient* of a convex function f at $x \in \mathbb{R}^d$, denoted by $\partial f(x)$, is given by $\partial f(x) = \{g \in \mathbb{R}^d \mid f(y) \geq f(x) + \langle g, y - x \rangle \text{ for all } y \in \mathbb{R}^d\}$. The *effective domain* of f is given by $\text{dom}(f) = \{x \in \mathbb{R}^d \mid f(x) < \infty\}$. For any positive sequences $\{a_k\}$ and $\{b_k\}$, we employ the notation $a_k \sim b_k$ to signify that $\lim_{k \rightarrow \infty} a_k/b_k = 1$. Other notations will be explained as the paper proceeds.

The remainder of this paper is structured as follows. In Section 2, we gather preliminary results that will be useful in the subsequent analysis. Specifically, we derive lower complexity bounds of RPPA with any relaxation schedule under different optimality measures. Section 3 is dedicated to a thorough analysis of RPPA with a constant relaxation schedule $\alpha_k \equiv \alpha \in (0, \sqrt{2}]$. Section 4 deals with Teboulle and Vaisbourd’s relaxation schedule as presented in [TV23]. Section 5 examines the silver stepsize schedule of Altschuler and Parrilo [AP23b] and two extensions thereof. Finally, concluding remarks are presented in Section 6.

2 Preliminaries

The following identities and inequalities are elementary and easy to verify. They will be employed in our analysis. For any $x, y, z, w \in \mathbb{R}^d$, $\theta \in \mathbb{R}$ and $\kappa > 0$, there hold

$$2\langle x - y, z - w \rangle = \|x - w\|^2 + \|y - z\|^2 - \|x - z\|^2 - \|y - w\|^2, \quad (6)$$

$$\|\theta x + (1 - \theta)y\|^2 = \theta\|x\|^2 + (1 - \theta)\|y\|^2 - \theta(1 - \theta)\|x - y\|^2, \quad (7)$$

$$(1 + 1/\kappa)\|x\|^2 + (1 + \kappa)\|y\|^2 \geq \|x + y\|^2 \geq (1 - 1/\kappa)\|x\|^2 + (1 - \kappa)\|y\|^2. \quad (8)$$

Next, we present two lemmas. Lemma 2.1 is an extension of [TV23, Lemma 4], and Lemma 2.2 recalls some properties of the stepsize schedule proposed in [TV23]. It should be noted that in [TV23], it is specified that $\gamma \in \{0, 1\}$. However, for the conclusion of Lemma 2.1 to be valid, it is sufficient to ensure that $\gamma \geq -s$.

Lemma 2.1 (Extension of [TV23, Lemma 4]). *For any $x, y, z \in \mathbb{R}^d$, $s > 0$, $\gamma \geq -s$ and $v \in \mathbb{R}$. If $\|x\|^2 \geq \|y\|^2 + s(s + \gamma)\|z\|^2 + sv$ and $v \leq 2\langle y, z \rangle - \gamma\|z\|^2$, then there holds $v \leq \|x\|^2 / (2s + \gamma)$.*

Lemma 2.2 (Collected from [TV23]). For $k \geq 1$, define α_k recursively as

$$\alpha_k = \frac{1}{2} \left(-A_{k-1} + \sqrt{A_{k-1}^2 + 8(A_{k-1} + 1)} \right) \quad \text{with } \alpha_0 = \sqrt{2} \quad \text{and } A_{k-1} := \sum_{i=0}^{k-1} \alpha_i. \quad (9)$$

Then, it holds for any $k \geq 0$ that

$$\sqrt{2} \leq \alpha_k = \frac{2(1 + A_k)}{2 + A_k} < 2 \quad \text{and} \quad \alpha_k^2 - 2 = (2 - \alpha_k)A_{k-1}. \quad (10)$$

Furthermore, we also have $A_{k-1} \sim 2k$.

In fact, the first equality in (10) is established by [TV23, Lemma 13], while the second equality follows directly from the definition of α_k . Moreover, it can be verified that α_k is monotonically increasing and approaches 2 as $k \rightarrow \infty$. Thus, $\sqrt{2} \leq \alpha_k < 2$ and $A_{k-1}/(2k)$ converges to 1. For simplicity, below we refer to the relaxation schedule defined in (9) as TV's schedule.

Next, we establish some lower bounds that are applicable to Algorithm 1 with any relaxation schedule α . These bounds serve as limitations and set the stage for determining the best possible upper bounds of RPPA when carefully designed relaxation schedules are employed. The proofs of these lower bounds are based on constructing examples that match these lower bounds. Particularly, the examples have been recognized in [THG17a, TV23] as worst-case instances for PPA.

Lemma 2.3 (Lower bounds for RPPA). Let $N \in \mathbb{N}_+$ be any positive integer, $\{(x^k, z^k)\}_{k=0}^N$ be the sequence generated by Algorithm 1 with any positive relaxation schedule $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{N-1}) \in \mathbb{R}_{++}^N$ for solving (Q). Then, there hold

$$(i) \quad \|\nabla f^\lambda(x^N)\| = \left\| \frac{1}{\lambda}(x^N - z^N) \right\| \geq \frac{1}{1 + \sum_{i=0}^{N-1} \alpha_i} \frac{1}{\lambda} \|x^0 - x^*\|,$$

$$(ii) \quad f(z^N) - f(x^*) \geq \frac{1}{2(1 + \sum_{i=0}^{N-1} \alpha_i)} \frac{1}{2\lambda} \|x^0 - x^*\|^2, \quad \text{and}$$

$$(iii) \quad \|\nabla f^\lambda(x^N)\|^2 = \left\| \frac{1}{\lambda}(x^N - z^N) \right\|^2 \geq \frac{1}{1 + \sum_{i=0}^{N-1} \alpha_i} \frac{1}{\lambda} (f(x^0) - f(x^*)).$$

Proof. The proof is accomplished by constructing specific functions that match the lower bounds. In all three cases, the constructed functions have the form $f(\cdot) = \eta \|\cdot\|$ for some $\eta > 0$. Note that f is convex, $\mathcal{X}^* = \{0\}$, $f(x^*) = 0$ and $\partial f(x) = \{\eta x / \|x\|\}$ for any $x \neq 0$. Let x^0 satisfy $\|x^0\| = 1$.

To prove (i) and (iii), we let $\eta = (\lambda(1 + \sum_{i=0}^{N-1} \alpha_i))^{-1}$. Using (4), (2) and $\|x^0\| = 1$, by induction we can show that $x^k = x^0 - \sum_{i=0}^{k-1} \alpha_i \lambda \eta x^i / \|x^i\| = (1 - \sum_{i=0}^{k-1} \alpha_i \lambda \eta) x^0 \neq 0$ and $z^k = x^k - \lambda \eta x^k / \|x^k\| = (1 - \lambda \eta - \sum_{i=0}^{k-1} \alpha_i \lambda \eta) x^0$ for any $0 \leq k \leq N$. Therefore, we have shown that $\nabla f^\lambda(x^N) = (\lambda(1 + \sum_{i=0}^{N-1} \alpha_i))^{-1} x^0$, which proves (i). Moreover, by noting $\|x^0\| = 1$, $f(x^*) = 0$ and $f(x^0) = \eta$, it is easy to verify that $\|\nabla f^\lambda(x^N)\|^2 = (\lambda(1 + \sum_{i=0}^{N-1} \alpha_i))^{-1} (f(x^0) - f(x^*))$, which proves (iii). To prove (ii), we set $\eta = (2\lambda(1 + \sum_{i=0}^{N-1} \alpha_i))^{-1}$. In this case, we have $z^N = x^0/2$, and thus $f(z^N) = (4\lambda(1 + \sum_{i=0}^{N-1} \alpha_i))^{-1}$, which proves (ii) since $\|x^0 - x^*\| = 1$. \square

If a complexity upper bound established for RPPA with any relaxation schedule coincides with any of the lower bounds provided in Lemma 2.3, it is considered to be tight.

3 Tight analysis of RPPA with $\alpha_k \equiv \alpha \in (0, \sqrt{2}]$

In this section, we establish a tight worst-case complexity bound for RPPA as presented in Algorithm 1 with constant relaxation schedule $\alpha_k \equiv \alpha \in (0, \sqrt{2}]$. Drawing inspiration from [TV23], we identify the key monotonically decreasing terms within the analysis that are crucial for our analysis and introduce a double sufficient decrease inequality to simplify the proof. We now present a useful lemma, which is applicable beyond the case of constant relaxation schedule.

Lemma 3.1. Let $x^* \in \mathcal{X}^*$ be arbitrarily fixed and $\{(z^k, x^k)\}_{k \geq 0}$ be the sequence generated by Algorithm 1 for solving (Q). Then, for $k \geq 0$, we have

$$f(x^*) \geq f(z^k) + \frac{1}{2\lambda\alpha_k} \|x^{k+1} - x^*\|^2 - \frac{1}{2\lambda\alpha_k} \|x^k - x^*\|^2 + \frac{2 - \alpha_k}{2\lambda} \|z^k - x^k\|^2, \quad (11)$$

$$f(z^k) \geq f(z^{k+1}) + \frac{1}{2\lambda} (\|z^{k+1} - x^{k+1}\|^2 + \|z^{k+1} - z^k\|^2 - (1 - \alpha_k)^2 \|z^k - x^k\|^2), \quad (12)$$

$$\begin{aligned} & \frac{\alpha_k - 2}{2(1 - \alpha_k)} \|z^{k+1} - z^k\|^2 + \frac{\alpha_k}{2(1 - \alpha_k)} \|z^{k+1} - x^{k+1}\|^2 + \frac{\alpha_k(\alpha_k - 1)}{2} \|z^k - x^k\|^2 \\ & = \langle (x^{k+1} - z^{k+1}) - (x^k - z^k), z^{k+1} - z^k \rangle \geq 0. \quad (\alpha_k \neq 1) \end{aligned} \quad (13)$$

Proof. For any $k \geq 0$, from $z^k = \text{Prox}_{\lambda f}(x^k)$, we obtain $0 \in \partial f(z^k) + \frac{z^k - x^k}{\lambda}$, and thus

$$f(x) \geq f(z^k) + \frac{1}{\lambda} \langle x^k - z^k, x - z^k \rangle, \quad \forall x \in \mathbb{R}^d. \quad (14)$$

Substituting $x = x^*$ in (14), we get

$$f(x^*) \geq f(z^k) + \frac{1}{\lambda} \langle x^k - z^k, x^* - z^k \rangle \stackrel{(6)}{=} f(z^k) + \frac{1}{2\lambda} (\|z^k - x^k\|^2 + \|z^k - x^*\|^2 - \|x^k - x^*\|^2). \quad (15)$$

Since $x^{k+1} = \alpha_k z^k + (1 - \alpha_k)x^k$, using (7) and the fact that $x^{k+1} - x^k = \alpha_k(z^k - x^k)$, we deduce

$$\|z^k - x^*\|^2 = \frac{1}{\alpha_k} \|x^{k+1} - x^*\|^2 + \frac{\alpha_k - 1}{\alpha_k} \|x^k - x^*\|^2 + (1 - \alpha_k) \|z^k - x^k\|^2.$$

Substituting this into (15) directly gives (11). Setting $x = z^{k-1}$ in (14) and utilizing $z^{k-1} - x^k = (1 - \alpha_{k-1})(z^{k-1} - x^{k-1})$, we obtain

$$\begin{aligned} f(z^{k-1}) & \geq f(z^k) + \frac{1}{\lambda} \langle x^k - z^k, z^{k-1} - z^k \rangle \stackrel{(6)}{=} f(z^k) + \frac{1}{2\lambda} (\|z^k - x^k\|^2 + \|z^k - z^{k-1}\|^2 - \|x^k - z^{k-1}\|^2) \\ & = f(z^k) + \frac{1}{2\lambda} (\|z^k - x^k\|^2 + \|z^k - z^{k-1}\|^2 - (1 - \alpha_{k-1})^2 \|z^{k-1} - x^{k-1}\|^2), \end{aligned}$$

which verifies (12). The equality in (13) follows by

$$\begin{aligned} & \langle (x^{k+1} - z^{k+1}) - (x^k - z^k), z^{k+1} - z^k \rangle = \langle x^{k+1} - z^{k+1}, z^{k+1} - z^k \rangle - \langle x^k - z^k, z^{k+1} - z^k \rangle \\ & \stackrel{(\text{Alg1})}{=} \langle x^{k+1} - x^{k+1}, z^{k+1} - z^k \rangle + \frac{1}{\alpha_k - 1} \langle x^{k+1} - z^k, z^{k+1} - z^k \rangle \\ & \stackrel{(6)}{=} \frac{\alpha_k - 2}{2(1 - \alpha_k)} \|z^{k+1} - z^k\|^2 + \frac{\alpha_k}{2(1 - \alpha_k)} \|z^{k+1} - x^{k+1}\|^2 + \frac{\alpha_k}{2(\alpha_k - 1)} \|z^k - x^{k+1}\|^2 \\ & \stackrel{(\text{Alg1})}{=} \frac{\alpha_k - 2}{2(1 - \alpha_k)} \|z^{k+1} - z^k\|^2 + \frac{\alpha_k}{2(1 - \alpha_k)} \|z^{k+1} - x^{k+1}\|^2 + \frac{\alpha_k(\alpha_k - 1)}{2} \|z^k - x^k\|^2. \end{aligned}$$

Recall that $\frac{x^k - z^k}{\lambda} \in \partial f(z^k)$. Leveraging the monotonicity property of the proximal operator, we can deduce that $\langle (x^{k+1} - z^{k+1}) - (x^k - z^k), z^{k+1} - z^k \rangle \geq 0$, thereby confirming the assertion of (13). \square

Next, we prove the monotonicity of $\|z^k - x^k\|$, which shows that the norm of a specifically selected subgradient element decreases monotonically as k increases.

Lemma 3.2 (Monotonicity of $\|z^k - x^k\|$). *Let $\{(z^k, x^k)\}_{k \geq 0}$ be the sequence generated by Algorithm 1 for solving (Q) with $\alpha_k \equiv \alpha \in (0, 2]$. Then, for any $k \geq 0$, it holds that*

$$\|z^{k+1} - x^{k+1}\| \leq \|z^k - x^k\|. \quad (16)$$

Proof. We consider three cases: (i) $\alpha \in (0, 1)$, (ii) $\alpha \in (1, 2]$, and (iii) $\alpha = 1$.

(i) For $\alpha \in (0, 1)$, applying (8) with $x = z^{k+1} - x^{k+1}$, $y = x^{k+1} - z^k$ and $\kappa = \frac{1}{1 - \alpha} > 0$, we obtain

$$\begin{aligned} \|z^{k+1} - z^k\|^2 & \geq \alpha \|z^{k+1} - x^{k+1}\|^2 - \frac{\alpha}{1 - \alpha} \|x^{k+1} - z^k\|^2 \\ & \stackrel{(\text{Alg1})}{=} \alpha \|z^{k+1} - x^{k+1}\|^2 - \alpha(1 - \alpha) \|z^k - x^k\|^2. \end{aligned} \quad (17)$$

(ii) For $\alpha \in (1, 2]$, similarly, applying (8) with $x = z^{k+1} - x^{k+1}$, $y = x^{k+1} - z^k$ and $\kappa = \frac{1}{\alpha - 1} > 0$ gives

$$\begin{aligned} \|z^{k+1} - z^k\|^2 & \leq \alpha \|z^{k+1} - x^{k+1}\|^2 + \frac{\alpha}{\alpha - 1} \|x^{k+1} - z^k\|^2 \\ & \stackrel{(\text{Alg1})}{=} \alpha \|z^{k+1} - x^{k+1}\|^2 + \alpha(\alpha - 1) \|z^k - x^k\|^2. \end{aligned} \quad (18)$$

For both cases (i) and (ii), we derive

$$\begin{aligned}
& \alpha \|z^k - x^k\|^2 - \alpha \|z^{k+1} - x^{k+1}\|^2 \\
&= \frac{\alpha - 2}{1 - \alpha} \left(\alpha \|z^{k+1} - x^{k+1}\|^2 - \alpha(1 - \alpha) \|z^k - x^k\|^2 \right) + \frac{\alpha}{1 - \alpha} \|z^{k+1} - x^{k+1}\|^2 + \alpha(\alpha - 1) \|z^k - x^k\|^2 \\
&\stackrel{(17) \text{ or } (18)}{\geq} \frac{\alpha - 2}{1 - \alpha} \|x^{k+1} - x^k\|^2 + \frac{\alpha}{1 - \alpha} \|z^{k+1} - x^{k+1}\|^2 + \alpha(\alpha - 1) \|z^k - x^k\|^2 \stackrel{(13)}{\geq} 0,
\end{aligned}$$

where the “=” can be verified straightforwardly. This confirms (16) for $\alpha \in (0, 1) \cup (1, 2]$.

(iii) For $\alpha = 1$, we have $x^{k+1} = z^k$. Using the monotonicity of the proximal operator, we derive

$$\begin{aligned}
0 &\leq \langle (z^k - z^{k+1}) - (z^{k-1} - z^k), z^{k+1} - z^k \rangle = \langle (2z^k - z^{k-1}) - z^{k+1}, z^{k+1} - z^k \rangle \\
&\leq \|z^k - z^{k-1}\|^2 - \|z^{k+1} - z^k\|^2 = \|z^k - x^k\|^2 - \|z^{k+1} - x^{k+1}\|^2.
\end{aligned}$$

In summary, we have shown that (16) holds for all $k \geq 0$. \square

Lemma 3.3 (Double sufficient decrease). *Let $\{(z^k, x^k)\}_{k \geq 0}$ be the sequence generated by Algorithm 1 for solving (Q) with $\alpha_k \equiv \alpha \in (0, 2)$. Then, for any $k \geq 0$, there hold*

$$1. f(z^k) \geq f(z^{k+1}) + \frac{\alpha+1}{2\lambda} \|z^{k+1} - x^{k+1}\|^2 + \frac{\alpha-1}{2\lambda} \|z^k - x^k\|^2 \text{ if } \alpha \in (0, 1], \text{ and } \dots \dots \dots (2\text{SD-1})$$

$$2. f(z^k) \geq f(z^{k+1}) + \frac{1}{2\lambda} \frac{2}{2-\alpha} \|z^{k+1} - x^{k+1}\|^2 + \frac{1}{2\lambda} \frac{2(\alpha-1)^2}{\alpha-2} \|z^k - x^k\|^2 \text{ if } \alpha \in (1, 2). \dots \dots \dots (2\text{SD-2})$$

Proof. We first address the case $\alpha \in (0, 1]$, the proof of which is divided into two subcases: $\alpha \in (0, 1)$ and $\alpha = 1$. For $\alpha \in (0, 1)$, combining (12) with (17) directly yields (2SD-1). For $\alpha = 1$, we have $x^{k+1} = z^k$. Thus, (12) reduces to $f(z^k) \geq f(z^{k+1}) + \frac{1}{\lambda} \|z^{k+1} - x^{k+1}\|^2$, which is exactly (2SD-1).

Now, consider the case $\alpha \in (1, 2)$. It follows from (13) that

$$\|z^{k+1} - z^k\|^2 \geq \frac{\alpha}{2-\alpha} \|z^{k+1} - x^{k+1}\|^2 + \frac{\alpha(\alpha-1)^2}{\alpha-2} \|z^k - x^k\|^2.$$

Combining the above inequality with (12) directly demonstrates (2SD-2). \square

Now, we are prepared to establish the tight complexity bound for RPPA with $\alpha_k \equiv \alpha \in (0, \sqrt{2}]$ by using the function value residual as the optimality measure.

Theorem 3.1 (Tight complexity bound for RPPA with $\alpha_k \equiv \alpha \in (0, \sqrt{2}]$). *Let $x^* \in \mathcal{X}^*$ and $N \in \mathbb{N}_+$ be arbitrarily fixed. Let z^N be the output after N iterations of Algorithm 1 for solving (Q) with constant relaxation schedule $\alpha_k \equiv \alpha \in (0, \sqrt{2}]$. Then, it holds that*

$$f(z^N) - f(x^*) \leq \frac{1}{\alpha N + 1} \frac{1}{4\lambda} \|x^0 - x^*\|^2. \quad (19)$$

Furthermore, this bound is tight.

Proof. Reorganizing the terms in (11) gives

$$\frac{1}{2\lambda\alpha} \|x^k - x^*\|^2 - \frac{1}{2\lambda\alpha} \|x^{k+1} - x^*\|^2 \geq f(z^k) - f(x^*) + \frac{2-\alpha}{2\lambda} \|z^k - x^k\|^2, \quad \forall k \geq 0.$$

Summing up the above inequality over $k = 0, \dots, N$, we obtain

$$\begin{aligned}
\frac{1}{2\lambda\alpha} \|x^0 - x^*\|^2 - \frac{1}{2\lambda\alpha} \|x^{N+1} - x^*\|^2 &\geq \sum_{i=0}^N \left(f(z^i) - f(x^*) + \frac{2-\alpha}{2\lambda} \|z^i - x^i\|^2 \right) \\
&= \sum_{i=0}^N \left(f(z^i) - f(z^N) + \frac{2-\alpha}{2\lambda} \|z^i - x^i\|^2 \right) + (N+1)(f(z^N) - f(x^*)). \quad (20)
\end{aligned}$$

Next, we separate the proof into two cases: $\alpha \in (0, 1]$ and $\alpha \in (1, \sqrt{2}]$. For $\alpha \in (0, 1]$, utilizing (2SD-1) in Lemma 3.3, we obtain for all $0 \leq i < N$ that

$$f(z^i) - f(z^N) + \frac{2-\alpha}{2\lambda} \|z^i - x^i\|^2 \geq \frac{\alpha+1}{2\lambda} \|z^N - x^N\|^2 + \frac{\alpha}{\lambda} \sum_{k=i+1}^{N-1} \|z^k - x^k\|^2 + \frac{1}{2\lambda} \|z^i - x^i\|^2$$

$$\stackrel{(16)}{\geq} \left(\frac{(N-i)\alpha}{\lambda} + \frac{2-\alpha}{2\lambda} \right) \|z^N - x^N\|^2. \quad (21)$$

For $\alpha \in (1, \sqrt{2}]$, we can also establish (21). In fact, in this case, by utilizing (2SD-2) in Lemma 3.3, it yields for all $0 \leq i < N$ that

$$\begin{aligned} f(z^i) - f(z^N) + \frac{2-\alpha}{2\lambda} \|z^i - x^i\|^2 &\geq \frac{1}{2\lambda} \frac{2}{2-\alpha} \|z^N - x^N\|^2 + \frac{\alpha}{\lambda} \sum_{k=i+1}^{N-1} \|z^k - x^k\|^2 + \frac{1}{2\lambda} \frac{\alpha^2 - 2}{\alpha - 2} \|z^i - x^i\|^2 \\ &\stackrel{(16)}{\geq} \left(\frac{(N-i)\alpha}{\lambda} + \frac{2-\alpha}{2\lambda} \right) \|z^N - x^N\|^2, \end{aligned}$$

where the second “ \geq ” also uses $\alpha \leq \sqrt{2}$. Substituting (21) into (20), we obtain

$$\begin{aligned} \frac{1}{2\lambda\alpha} \|x^0 - x^*\|^2 - \frac{1}{2\lambda\alpha} \|x^{N+1} - x^*\|^2 \\ \geq (N+1) \left[\frac{\alpha(N+1) + 2(1-\alpha)}{2\lambda} \|z^N - x^N\|^2 + (f(z^N) - f(x^*)) \right]. \end{aligned} \quad (22)$$

Substituting $k = N$ into (15), we drive

$$\begin{aligned} f(x^*) &\geq f(z^N) + \frac{1}{\lambda} \langle x^N - z^N, x^* - z^N \rangle = f(z^N) + \frac{1}{\lambda} \langle x^N - z^N, x^* - x^{N+1} + x^{N+1} - z^N \rangle \\ &\stackrel{(\text{Alg1})}{=} f(z^N) + \frac{1}{\lambda} \langle x^N - z^N, x^* - x^{N+1} \rangle + \frac{1-\alpha}{\lambda} \|z^N - x^N\|^2. \end{aligned} \quad (23)$$

Combining (22) and (23), setting $v = 2\lambda(f(z^N) - f(x^*))$, $x = x^0 - x^*$, $y = x^{N+1} - x^*$, $z = z^N - x^N$, $s = \alpha(N+1)$ and $\gamma = 2(1-\alpha)$ in Lemma 2.1, we directly obtain the desired result (19) by noting that $s + \gamma > 0$ holds for any $\alpha \in (0, \sqrt{2}]$ and $N \geq 1$. Finally, the tightness of (19) can be easily verified by noticing that it matches the lower bound given in item (ii) of Lemma 2.3. \square

Remark 3.1 (Complexity bound for subgradient norm $\|\nabla f^\lambda(x^N)\|$). *Let $\alpha_k \equiv \alpha \in (0, \sqrt{2}]$. Since, by (2), we have $\nabla f^\lambda(x^N) \in \partial f(z^N)$, $\|\nabla f^\lambda(x^N)\|$ can be viewed as the subgradient norm residual of f at z^N . By slightly modifying the previous analysis, we can derive a tight complexity bound for $\|\nabla f^\lambda(x^N)\|$. First, by summing the inequality (11) over $k = 0, 1, \dots, N-1$, we can obtain a result analogous to (20). Then, by following the same reasoning as in Theorem 3.1, we can derive an inequality similar to (22):*

$$\frac{1}{2\lambda\alpha} \|x^0 - x^*\|^2 - \frac{1}{2\lambda\alpha} \|x^N - x^*\|^2 \geq N \left[\frac{\alpha N + 2}{2\lambda} \|z^N - x^N\|^2 + (f(z^N) - f(x^*)) \right]. \quad (24)$$

Next, by setting $k = N$ in (15), leveraging the fact that $f(z^N) - f(x^*) \geq 0$ and rearranging terms, we obtain

$$\|x^N - x^*\|^2 \geq \|z^N - x^N\|^2 + \|z^N - x^*\|^2. \quad (25)$$

By combining (25) with (24), followed by eliminating certain non-negative terms from the right-hand side of (24), we achieve the bound $\|\nabla f^\lambda(x^N)\|^2 \leq \frac{1}{\lambda^2(\alpha N + 1)^2} \|x^0 - x^*\|^2$ for any $\alpha \in (0, \sqrt{2}]$. The tightness of this bound can be verified by item (i) of Lemma 2.3 as well.

4 Tight analysis of RPPA with TV’s schedule (9)

In this section, we establish a tight worst-case complexity bound for RPPA with TV’s relaxation schedule as presented in (9). As in Theorem 3.1, function value residual is used as the optimality measure, and the established bound is approximately $\sqrt{2}$ times better as $N \rightarrow \infty$.

Theorem 4.1. *Let $x^* \in \mathcal{X}^*$ and $N \in \mathbb{N}_+$ be arbitrarily fixed. Let z^N be the output after N iterations of Algorithm 1 for solving (Q) with TV’s relaxation schedule defined in (9). Then, it holds that*

$$f(z^N) - f(x^*) \leq \frac{1}{A_{N-1} + 1} \frac{1}{4\lambda} \|x^0 - x^*\|^2, \quad (26)$$

where A_{N-1} is defined in (9). Furthermore, this bound is tight.

Proof. Similar to the proof of (2SD-2) in Lemma 3.3, by utilizing (12) and (13), we can show

$$f(z^k) \geq f(z^{k+1}) + \frac{1}{2\lambda} \frac{2}{2-\alpha_k} \|z^{k+1} - x^{k+1}\|^2 + \frac{1}{2\lambda} \frac{2(\alpha_k-1)^2}{\alpha_k-2} \|z^k - x^k\|^2, \quad \forall k \geq 0,$$

which further implies for all $0 \leq i < N$ that

$$\begin{aligned} f(z^i) - f(z^N) + \frac{2-\alpha_i}{2\lambda} \|z^i - x^i\|^2 &\geq \frac{1}{2\lambda} \frac{2}{2-\alpha_{N-1}} \|z^N - x^N\|^2 + \frac{1}{2\lambda} \frac{\alpha_i^2-2}{\alpha_i-2} \|z^i - x^i\|^2 \\ &\quad + \frac{1}{2\lambda} \sum_{k=i+1}^{N-1} \left(\frac{2}{2-\alpha_{k-1}} + \frac{2(\alpha_k-1)^2}{\alpha_k-2} \right) \|z^k - x^k\|^2. \end{aligned} \quad (27)$$

Rearranging (11) gives $\frac{1}{2\lambda} \|x^k - x^*\|^2 - \frac{1}{2\lambda} \|x^{k+1} - x^*\|^2 \geq \alpha_k (f(z^k) - f(x^*) + \frac{2-\alpha_k}{2\lambda} \|z^k - x^k\|^2)$ for all $k \geq 0$. Summing up this inequality over $k = 0, 1, \dots, N$, we obtain

$$\frac{1}{2\lambda} \|x^0 - x^*\|^2 - \frac{1}{2\lambda} \|x^{N+1} - x^*\|^2 \geq \sum_{i=0}^N \alpha_i \left(f(z^i) - f(z^N) + \frac{2-\alpha_i}{2\lambda} \|z^i - x^i\|^2 \right) + A_N (f(z^N) - f(x^*)).$$

Making use of (27), the above inequality further implies

$$\begin{aligned} \frac{1}{2\lambda} \|x^0 - x^*\|^2 - \frac{1}{2\lambda} \|x^{N+1} - x^*\|^2 &\geq \frac{\alpha_N(2-\alpha_N)}{2\lambda} \|z^N - x^N\|^2 + A_N (f(z^N) - f(x^*)) \\ &\quad + \frac{1}{2\lambda} \sum_{i=0}^{N-1} \alpha_i \left[\frac{2}{2-\alpha_{N-1}} \|z^N - x^N\|^2 + \frac{\alpha_i^2-2}{\alpha_i-2} \|z^i - x^i\|^2 + \sum_{k=i+1}^{N-1} \left(\frac{2}{2-\alpha_{k-1}} + \frac{2(\alpha_k-1)^2}{\alpha_k-2} \right) \|z^k - x^k\|^2 \right] \\ &= \frac{1}{2\lambda} \left[\left(\frac{2A_{N-1}}{2-\alpha_{N-1}} + \alpha_N(2-\alpha_N) \right) \|z^N - x^N\|^2 + \sum_{k=1}^{N-1} \left(\frac{2A_{k-1}}{2-\alpha_{k-1}} + \frac{2A_{k-1}(\alpha_k-1)^2}{\alpha_k-2} + \frac{\alpha_k(\alpha_k^2-2)}{\alpha_k-2} \right) \|z^k - x^k\|^2 \right. \\ &\quad \left. + \frac{\alpha_0(\alpha_0^2-2)}{\alpha_0-2} \|z^0 - x^0\|^2 \right] + A_N (f(z^N) - f(x^*)). \end{aligned} \quad (28)$$

Applying the results from (10), we can verify for $k = 1, \dots, N-1$ that

$$\begin{aligned} \frac{2A_{k-1}}{2-\alpha_{k-1}} + \frac{2A_{k-1}(\alpha_k-1)^2}{\alpha_k-2} + \frac{\alpha_k(\alpha_k^2-2)}{\alpha_k-2} &= A_{k-1}(2 + A_{k-1}) + \frac{1}{\alpha_k-2} [2A_{k-1}(\alpha_k-1)^2 + \alpha_k(2A_{k-1} - \alpha_k A_{k-1})] \\ &= A_{k-1}(2 + A_{k-1}) + \frac{1}{\alpha_k-2} [A_{k-1}(\alpha_k^2 - 2\alpha_k) + 2A_{k-1}] = A_{k-1}(2 + A_{k-1}) + A_{k-1}(\alpha_k + \frac{2}{\alpha_k-2}) \\ &= A_{k-1}(2 + A_{k-1}) + A_{k-1}(\alpha_k - 2 - A_k) = 0, \end{aligned}$$

as well as

$$\frac{2A_{N-1}}{2-\alpha_{N-1}} + \alpha_N(2-\alpha_N) = A_{N-1}(2 + A_{N-1}) + 2\alpha_N - \alpha_N^2 = A_{N-1}^2 - \alpha_N^2 + 2A_N = A_N(A_N + 2(1-\alpha_N)).$$

Hence, (28) can be simplified to

$$\frac{1}{2\lambda} \|x^0 - x^*\|^2 - \frac{1}{2\lambda} \|x^{N+1} - x^*\|^2 \geq \frac{A_N(A_N + 2(1-\alpha_N))}{2\lambda} \|z^N - x^N\|^2 + A_N (f(z^N) - f(x^*)). \quad (29)$$

Similar to (23), we can derive

$$f(x^*) \geq f(z^N) + \frac{1}{\lambda} \langle x^N - z^N, x^* - x^{N+1} \rangle + \frac{1-\alpha_N}{\lambda} \|z^N - x^N\|^2. \quad (30)$$

By combining (29) and (30), and setting $v = 2\lambda(f(z^N) - f(x^*))$, $x = x^0 - x^*$, $y = x^{N+1} - x^*$, $z = z^N - x^N$, $s = A_N$ and $\gamma = 2(1-\alpha_N)$ in Lemma 2.1, the desired upper bound result (26) can be derived straightforwardly. Finally, the tightness of (26) follows from item (ii) of Lemma 2.3 and the definition of A_{N-1} in Lemma 2.1. \square

5 Accelerate RPPA via silver stepsize schedules

In this section, we analyze the performance of RPPA using the silver stepsize schedule from [AP23b]. Additionally, we propose two modified schedules based on the silver stepsize schedule. These schedules permit the stepsizes to be considerably larger than 2 and enable higher-order convergence rates of

$O(1/N^{1.2716})$ for either the function value residual or the subgradient norm, surpassing the standard $O(1/N)$ rate.

Before deriving the results for RPPA, we first establish the tight complexity bound for GD method with the silver stepsize schedule for minimizing smooth convex functions. Part of these results can be directly utilized to establish the tight convergence rate of RPPA with the silver stepsize schedule.

5.1 Tight convergence rate of GD method with the silver stepsize schedule

Consider minimizing an L -smooth convex function h via the N -step GD method (5) with stepsize schedule $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{N-1}) \in \mathbb{R}_{++}^N$ normalized by L . For any $m \geq 1$, the silver stepsize schedule, denoted by $\pi^{(m)}$, was first proposed for the GD method in [AP23b]. The m -th silver stepsize schedule $\pi^{(m)}$ is a vector in $\mathbb{R}_{++}^{2^m-1}$, and it is recursively defined as follows:

$$\pi^{(1)} := [\sqrt{2}] \text{ and } \pi^{(m+1)} := [\pi^{(m)}, 1 + \rho^{m-1}, \pi^{(m)}], \quad \forall m \geq 1, \quad (31)$$

where $\rho = 1 + \sqrt{2}$ is the silver ratio satisfying $\rho^2 = 2\rho + 1$. For convenience, the components of $\pi^{(m)}$ are indexed by $i \in \{0, 1, \dots, 2^m - 2\}$. By this definition, we have $\sum_{i=0}^{2^m-2} \pi_i^{(m)} = \rho^m - 1$. In [AP23b], the following convergence rate for GD method with silver stepsize schedule was established.

Theorem 5.1 ([AP23b, Theorem 1.1]). *For any fixed $m \geq 1$, let $N = 2^m - 1$. Then, for any L -smooth convex function $h : \mathbb{R}^d \rightarrow \mathbb{R}$ and any $x^0 \in \mathbb{R}^d$, there holds*

$$h(x^N) - h(x^*) \leq \frac{L \|x^0 - x^*\|^2}{1 + \sqrt{4\rho^{2m-3} - 3}} \leq \frac{L \|x^0 - x^*\|^2}{2\rho^{\log_2 N}} \approx \frac{L \|x^0 - x^*\|^2}{2N^{1.2716}},$$

where x^* denotes an arbitrarily fixed minimizer of h , and x^N denotes the output of the N -step GD method (5) using the silver stepsize schedule $\alpha = \pi^{(m)}$.

However, as discussed in [LG24, Section 3.1], the above convergence rate is not tight. By [LG24, Conjecture 3.2], GD method with the silver stepsize schedule $\pi^{(m)}$ was conjectured to have the following tight rates:

$$h(x^N) - h(x^*) \leq \frac{L \|x^0 - x^*\|^2}{4\rho^m - 2} \text{ and } \|\nabla h(x^N)\| \leq \frac{L \|x^0 - x^*\|}{\rho^m}.$$

In the following, we will prove this conjecture is true. To do this, we first introduce the following definition of $Q_{x,y}^h$, whose nonnegativity plays a fundamental role in the proof.

Definition 5.1. *Given any L -smooth convex function h and $x, y \in \mathbb{R}^d$, we define*

$$Q_{x,y}^h := h(x) - h(y) - \langle \nabla h(y), x - y \rangle - \frac{1}{2L} \|\nabla h(x) - \nabla h(y)\|^2, \quad (32)$$

which is nonnegative due to L -smoothness and convexity of h [Nes18, Theorem 2.1.5].

Next, we recall [GSW24, Lemma 1], which is useful in our analysis.

Lemma 5.1 ([GSW24, Lemma 1]). *For any fixed $m \geq 1$, let $N = 2^m - 1$. Furthermore, define a matrix $A^{(m)} \in \mathbb{R}^{2^m \times 2^m}$, whose components are indexed by $(i, j) \in \{0, 1, \dots, 2^m - 1\}^2$, recursively by*

$$A^{(1)} := \begin{bmatrix} 0 & \rho \\ 1 & 0 \end{bmatrix}, \text{ and } A^{(i+1)} := \begin{bmatrix} A^{(i)} & \\ & \rho^2 A^{(i)} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{(2^i-1) \times (2^i-1)} & & \\ & 0 & \rho \pi^{(i)} & \rho \\ & & \mathbf{0}_{(2^i-1) \times (2^i-1)} & \\ & \rho^i & \rho \pi^{(i)} & 0 \end{bmatrix}, \quad (33)$$

for $i = 1, 2, \dots, m-1$. Let $h : \mathbb{R}^d \rightarrow \mathbb{R}$ be any L -smooth convex function, x^* be any fixed minimizer of h and $x^0 \in \mathbb{R}^d$ be any initial point. Let $\{x^k\}_{k=0}^N$ be the sequence generated by the N -step GD method (5) with stepsize schedule $\alpha = \pi^{(m)}$. Then, we have

$$\begin{aligned} 0 \leq \sum_{i,j \in \{0,1,\dots,2^m-1\}} A_{i,j}^{(m)} Q_{x^i,x^j}^h &= - \sum_{i=0}^{2^m-2} \alpha_i \left(h(x^{2^m-1}) - h(x^i) - \frac{1}{2L} \|\nabla h(x^i)\|^2 - \langle \nabla h(x^i), x^0 - x^i \rangle \right) \\ &\quad - \frac{L}{2} \|x^{2^m-1} - x^0\|^2 - \frac{\rho^m(\rho^m - 1)}{2L} \|\nabla h(x^{2^m-1})\|^2. \end{aligned} \quad (34)$$

Using Lemma 5.1, we can then prove the validity of Conjecture 3.2 in [LG24].

Theorem 5.2 ([LG24, Conjecture 3.2]). *For any fixed $m \geq 1$, let $N = 2^m - 1$. Let $h : \mathbb{R}^d \rightarrow \mathbb{R}$ be any L -smooth convex function, x^* be any fixed minimizer of h , $x^0 \in \mathbb{R}^d$ be any initial point. Let $\{x^k\}_{k=0}^N$ be the sequence generated by the N -step GD method (5) with stepsize schedule $\alpha = \pi^{(m)}$. Then, we have*

$$h(x^N) - h(x^*) \leq \frac{L \|x^0 - x^*\|^2}{2 \left(2 \sum_{i=0}^{2^m-2} \pi_i^{(m)} + 1 \right)} = \frac{L \|x^0 - x^*\|^2}{4\rho^m - 2}, \quad (35)$$

$$\|\nabla h(x^N)\| \leq \frac{L \|x^0 - x^*\|}{\sum_{i=0}^{2^m-2} \pi_i^{(m)} + 1} = \frac{L \|x^0 - x^*\|}{\rho^m}. \quad (36)$$

Furthermore, these bounds are tight.

Proof. To prove (35), we note that $0 \leq Q_{x^*, x^i}^h = h(x^*) - h(x^i) - \langle \nabla h(x^i), x^* - x^i \rangle - \frac{1}{2L} \|\nabla h(x^i)\|^2$. Thus, Lemma 5.1 implies

$$\begin{aligned} 0 &\leq \sum_{i,j \in \{0,1,\dots,2^m-1\}} A_{i,j}^{(m)} Q_{x^i, x^j}^h + \sum_{i=0}^{2^m-2} Q_{x^*, x^i}^h + \rho^m Q_{x^*, x^{2^m-1}}^h \\ &\stackrel{(34)}{=} \sum_{i=0}^{2^m-2} \pi_i^{(m)} \left(h(x^*) - h(x^{2^m-1}) - \langle \nabla h(x^i), x^* - x^0 \rangle \right) - \frac{L}{2} \|x^{2^m-1} - x^0\|^2 - \frac{\rho^{2m}}{2L} \|\nabla h(x^{2^m-1})\|^2 \\ &\quad + \rho^m \left(h(x^*) - h(x^{2^m-1}) - \langle \nabla h(x^{2^m-1}), x^* - x^{2^m-1} \rangle \right) \\ &\stackrel{(*)}{=} (2\rho^m - 1) \left(h(x^*) - h(x^{2^m-1}) \right) + L \langle x^{2^m-1} - x^0, x^* - x^0 \rangle + \frac{L}{2} \|x^0 - x^*\|^2 - \frac{L}{2} \|x^0 - x^*\|^2 \\ &\quad - \frac{L}{2} \|x^{2^m-1} - x^0\|^2 - \frac{\rho^{2m}}{2L} \|\nabla h(x^{2^m-1})\|^2 - \rho^m \langle \nabla h(x^{2^m-1}), x^* - x^{2^m-1} \rangle \\ &= (2\rho^m - 1) \left(h(x^*) - h(x^{2^m-1}) \right) + \frac{L}{2} \|x^0 - x^*\|^2 - \frac{L}{2} \left\| x^{2^m-1} - \frac{\rho^m}{L} \nabla h(x^{2^m-1}) - x^* \right\|^2, \end{aligned}$$

where we have used $\sum_{i=0}^{2^m-2} \pi_i^{(m)} = \rho^m - 1$ and $-\sum_{i=0}^{2^m-2} \pi_i^{(m)} \nabla h(x^i) = L(x^{2^m-1} - x^0)$ in (*). Reorganizing the terms yields the desired upper bound in (35).

To prove (36), we note that $0 \leq Q_{x^*, x^i}^h = h(x^*) - h(x^i) - \langle \nabla h(x^i), x^* - x^i \rangle - \frac{1}{2L} \|\nabla h(x^i)\|^2$ and $0 \leq Q_{x^{2^m-1}, x^*}^h = h(x^{2^m-1}) - h(x^*) - \frac{1}{2L} \|\nabla h(x^{2^m-1})\|^2$. Thus, Lemma 5.1 implies

$$\begin{aligned} 0 &\leq \sum_{i,j \in \{0,1,\dots,2^m-1\}} A_{i,j}^{(m)} Q_{x^i, x^j}^h + \sum_{i=0}^{2^m-2} Q_{x^*, x^i}^h + Q_{x^*, x^{2^m-1}}^h + \rho^m Q_{x^{2^m-1}, x^*}^h \\ &\stackrel{(34)}{=} \sum_{i=0}^{2^m-2} \pi_i^{(m)} \left(h(x^*) - h(x^{2^m-1}) - \langle \nabla h(x^i), x^* - x^0 \rangle \right) - \frac{L}{2} \|x^{2^m-1} - x^0\|^2 - \frac{\rho^{2m} + 1}{2L} \|\nabla h(x^{2^m-1})\|^2 \\ &\quad - \langle \nabla h(x^{2^m-1}), x^* - x^{2^m-1} \rangle + (\rho^m - 1) (h(x^{2^m-1}) - h(x^*)) \\ &\stackrel{(*)}{=} L \langle x^{2^m-1} - x^0, x^* - x^0 \rangle + \frac{L}{2} \|x^0 - x^*\|^2 - \frac{L}{2} \|x^0 - x^*\|^2 - \frac{L}{2} \|x^{2^m-1} - x^0\|^2 \\ &\quad - \frac{L}{2} \left\| \frac{\rho^m}{L} \nabla h(x^{2^m-1}) \right\|^2 - \frac{1}{2L} \|\nabla h(x^{2^m-1})\|^2 - \langle \nabla h(x^{2^m-1}), x^* - x^{2^m-1} \rangle \\ &= -\frac{L}{2} \left\| \frac{\rho^m}{L} \nabla h(x^{2^m-1}) \right\|^2 + \frac{L}{2} \|x^0 - x^*\|^2 - \frac{L}{2} \left\| x^{2^m-1} - \frac{1}{L} \nabla h(x^{2^m-1}) - x^* \right\|^2, \end{aligned}$$

where we have used $\sum_{i=0}^{2^m-2} \pi_i^{(m)} = \rho^m - 1$ and $-\sum_{i=0}^{2^m-2} \pi_i^{(m)} \nabla h(x^i) = L(x^{2^m-1} - x^0)$ in (*). Reorganizing the terms yields the desired upper bound (36).

Finally, the tightness of the bounds in (35) and (36) can be easily confirmed by considering the Huber function as discussed in [LG24, Section 3.1]. We omit the details for brevity. \square

5.2 Accelerated convergence rates for RPPA with silver stepsize schedules

We now return to the convex optimization problem (\mathcal{Q}) . Let $\{(x^k, z^k)\}_{k \geq 0}$ be the sequence generated by Algorithm 1. Consider the connection between RPPA and GD method as given in (4), we can apply Theorem 5.2 with $h = f^\lambda$ and $L = 1/\lambda$ and obtain the following convergence results.

Theorem 5.3 (Convergence rate of RPPA with silver stepsize schedule $\pi^{(m)}$). *Let $x^* \in \mathcal{X}^*$ be arbitrarily fixed and $z^* = \text{Prox}_{\lambda f}(x^*) = x^*$. For any fixed $m \geq 1$, let $N = 2^m - 1$ and $\{(x^k, z^k)\}_{k=0}^N$ be the sequence generated by Algorithm 1 for solving (\mathcal{Q}) with stepsize schedule $\alpha = \pi^{(m)}$ from any starting point $x^0 \in \mathbb{R}^d$. Then, we have*

$$f(z^N) + \frac{\lambda}{2} \|\nabla f^\lambda(x^N)\|^2 - f(x^*) \leq \frac{\|x^0 - x^*\|^2}{(4\rho^m - 2)\lambda} \quad \text{and} \quad \|\nabla f^\lambda(x^N)\| \leq \frac{\|x^0 - x^*\|}{\rho^m \lambda}. \quad (37)$$

Note that $\rho^m = (N + 1)^{\log_2 \rho} \approx (N + 1)^{1.2716}$. Therefore, Theorem 5.3 has already provided an accelerated convergence rate for RPPA using the silver stepsize schedule. Moreover, the upper bound on $\|\nabla f^\lambda(x^N)\|$ given in (37) is tight due to item (i) of Lemma 2.3 and the fact that $\sum_{i=0}^{2^m-2} \pi_i^{(m)} = \rho^m - 1$. On the other hand, the first inequality in (37) also yields an accelerated complexity bound for the function value residual $f(z^N) - f(x^*)$. However, this bound is no longer tight. Deriving a tight upper bound for $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$ from the first inequality in (37) remains challenging. This challenge motivates us to construct two modified silver stepsize schedules: the ‘‘right silver’’ stepsize schedule $\bar{\pi}^{(m)}$ and the ‘‘left silver’’ stepsize schedule $\underline{\pi}^{(m)}$. These two modified silver stepsize schedules allow us to achieve accelerated and tight convergence rates for RPPA under the measures $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$ and $\|\nabla f^\lambda(x^N)\|^2 / (f(x^0) - f(x^*))$, respectively.

Our ‘‘right silver’’ stepsize schedule $\bar{\pi}^{(m)}$ is defined by extending the silver stepsize schedule with an additional element on the right, i.e.,

$$\bar{\pi}^{(0)} := [\gamma_0] \quad \text{and} \quad \bar{\pi}^{(m)} := [\pi^{(m)}, \gamma_m], \quad \forall m \geq 1, \quad (38)$$

where $\gamma_m = \frac{1}{2} (1 + \sqrt{1 + 4\rho^m}) \geq 1$ for any $m \geq 0$ is a solution to

$$\gamma_m^2 = \gamma_m + \rho^m. \quad (39)$$

On the other hand, our ‘‘left silver’’ stepsize schedule $\underline{\pi}^{(m)}$ is defined as

$$\underline{\pi}^{(0)} := [\gamma_0] \quad \text{and} \quad \underline{\pi}^{(m)} := [\gamma_m, \pi^{(m)}], \quad \forall m \geq 1. \quad (40)$$

Define $T_m := 1 + \gamma_m + \sum_{i=0}^{2^m-2} \pi_i^{(m)}$ for $m \geq 1$. Then, it follows from (39) and $\sum_{i=0}^{2^m-2} \pi_i^{(m)} = \rho^m - 1$ that

$$T_m = 1 + \gamma_m + \sum_{i=0}^{2^m-2} \pi_i^{(m)} = \gamma_m + \rho^m \stackrel{(39)}{=} \gamma_m^2 = \frac{1}{2} + \frac{1}{2} \sqrt{1 + 4\rho^m} + \rho^m. \quad (41)$$

For $N = 2^m$, we have $T_m = \frac{1}{2} + \frac{1}{2} \sqrt{1 + 4N^{\log_2 \rho}} + N^{\log_2 \rho} \sim N^{\log_2 \rho}$.

To prove convergence results of RPPA with the silver stepsize schedules $\bar{\pi}^{(m)}$ and $\underline{\pi}^{(m)}$, we need to adapt Lemma 5.1 from GD method to RPPA (see Lemma 5.3). This adaptation is based on the relations (2) and (3). Before proceeding, we introduce the notation P_{y,z^j}^f , which is crucial to our proof.

Definition 5.2. *Given $N \in \mathbb{N}_+$, $\alpha \in \mathbb{R}_{++}^N$ and $x^0 \in \mathbb{R}^d$. Suppose that $\{(x^k, z^k)\}_{k=0}^N$ is the sequence generated by Algorithm 1 with stepsize schedule α for solving (\mathcal{Q}) . Let $x^* \in \mathcal{X}^*$ be any minimizer of f and $z^* = \text{Prox}_{\lambda f}(x^*) = x^*$. For any $y \in \mathbb{R}^d$ and $j \in \{\star, 0, 1, \dots, N\}$, define*

$$P_{y,z^j}^f := f(y) - f(z^j) - \langle \nabla f^\lambda(x^j), y - z^j \rangle. \quad (42)$$

From (2), we have $\nabla f^\lambda(x^j) \in \partial f(z^j)$ and thus $P_{y,z^j}^f \geq 0$.

Next, we show that $Q_{x^i, x^j}^{f^\lambda} \equiv P_{z^i, z^j}^f$ for any $i, j \in \{\star, 0, 1, \dots, N\}$.

Lemma 5.2. *Given any fixed $N \in \mathbb{N}_+$, $\alpha \in \mathbb{R}_{++}^N$ and $x^0 \in \mathbb{R}^d$. Suppose that $\{(x^k, z^k)\}_{k=0}^N$ is the sequence generated by Algorithm 1 with stepsize schedule α for solving (\mathcal{Q}) . Let $x^* \in \mathcal{X}^*$ be any minimizer of f and $z^* = \text{Prox}_{\lambda f}(x^*) = x^*$. Then, we have*

$$Q_{x^i, x^j}^{f^\lambda} \equiv P_{z^i, z^j}^f, \quad \forall i, j \in \{\star, 0, 1, \dots, N\}. \quad (43)$$

Proof. Recall that f^λ is convex and $(1/\lambda)$ -smooth. Further considering the notation defined in (32) and (42) and the relations (2) and (3), we obtain

$$\begin{aligned} Q_{x^i, x^j}^{f^\lambda} &\stackrel{(32)}{=} f^\lambda(x^i) - f^\lambda(x^j) - \langle \nabla f^\lambda(x^j), x^i - x^j \rangle - \frac{\lambda}{2} \|\nabla f^\lambda(z^i) - \nabla f^\lambda(z^j)\|^2 \\ &\stackrel{(2),(3)}{=} f(z^i) - f(z^j) - \langle \nabla f^\lambda(x^j), z^i - z^j + \lambda(\nabla f^\lambda(z^i) - \nabla f^\lambda(z^j)) \rangle \\ &\quad + \frac{\lambda}{2} \langle \nabla f^\lambda(x^i) - \nabla f^\lambda(x^j), \nabla f^\lambda(x^i) + \nabla f^\lambda(x^j) \rangle - \frac{\lambda}{2} \|\nabla f^\lambda(z^i) - \nabla f^\lambda(z^j)\|^2 \\ &= f(z^i) - f(z^j) - \langle \nabla f^\lambda(x^j), z^i - z^j \rangle = P_{z^i, z^j}^f, \end{aligned}$$

for any $i, j \in \{\star, 0, 1, \dots, N\}$. This completes the proof. \square

By invoking $h = f^\lambda$ in Lemma 5.1, we can derive the following lemma, which plays a key role in proving the convergence results for RPPA with the modified silver stepsize schedules $\bar{\pi}^{(m)}$ and $\underline{\pi}^{(m)}$.

Lemma 5.3 (Adapted from Lemma 5.1). *For any fixed $m \geq 1$, let $N = 2^m - 1$ and $\{(x^k, z^k)\}_{k=0}^N$ be the sequence generated by Algorithm 1 for solving (Q) with the silver stepsize schedule $\alpha = \pi^{(m)}$ from an arbitrarily starting point $x^0 \in \mathbb{R}^d$. Then, we have*

$$\begin{aligned} 0 \leq \sum_{i,j \in \{0,1,\dots,2^m-1\}} A_{i,j}^{(m)} P_{z^i, z^j}^f &= \sum_{i=0}^{2^m-2} \alpha_i \left(f(z^i) - f(z^{2^m-1}) + \langle \nabla f^\lambda(x^i), x^0 - z^i \rangle \right) \\ &\quad - \frac{(\rho^{2^m} - 1)\lambda}{2} \|\nabla f^\lambda(x^{2^m-1})\|^2 - \frac{\lambda}{2} \left\| \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i) \right\|^2, \end{aligned} \quad (44)$$

where $A^{(m)} \in \mathbb{R}^{2^m \times 2^m}$, whose components are indexed by $(i, j) \in \{0, 1, \dots, 2^m - 1\}^2$, is defined in (33).

Proof. Due to the connection (4), we can invoke $h = f^\lambda$ and $L = 1/\lambda$ in Lemma 5.1 and then deduce

$$\begin{aligned} 0 \leq \sum_{i,j \in \{0,1,\dots,2^m-1\}} A_{i,j}^{(m)} Q_{x^i, x^j}^{f^\lambda} &= \sum_{i=0}^{2^m-2} \alpha_i \left(f^\lambda(x^i) - f^\lambda(x^{2^m-1}) + \langle \nabla f^\lambda(x^i), x^0 - x^i \rangle + \frac{\lambda}{2} \|\nabla f^\lambda(x^i)\|^2 \right) \\ &\quad - \frac{1}{2\lambda} \|x^{2^m-1} - x^0\|^2 - \frac{\rho^m(\rho^m - 1)\lambda}{2} \|\nabla f^\lambda(x^{2^m-1})\|^2. \end{aligned}$$

Then, by leveraging Lemma 5.2 and the relations (2) and (3), we derive

$$\begin{aligned} 0 \leq \sum_{i,j \in \{0,1,\dots,2^m-1\}} A_{i,j}^{(m)} P_{z^i, z^j}^f &\stackrel{(43)}{=} \sum_{i,j \in \{0,1,\dots,2^m-1\}} A_{i,j}^{(m)} Q_{x^i, x^j}^{f^\lambda} \\ &\stackrel{(3)}{=} \sum_{i=0}^{2^m-2} \alpha_i \left(f(z^i) + \frac{\lambda}{2} \|\nabla f^\lambda(x^i)\|^2 - f(z^{2^m-1}) - \frac{\lambda}{2} \|\nabla f^\lambda(x^{2^m-1})\|^2 + \langle \nabla f^\lambda(x^i), x^0 - x^i \rangle \right) \\ &\quad + \frac{\lambda}{2} \|\nabla f^\lambda(x^i)\|^2 - \frac{1}{2\lambda} \|x^{2^m-1} - x^0\|^2 - \frac{(\rho^{2^m} - \rho^m)\lambda}{2} \|\nabla f^\lambda(x^{2^m-1})\|^2 \\ &\stackrel{(*)}{=} \sum_{i=0}^{2^m-2} \alpha_i \left(f(z^i) - f(z^{2^m-1}) + \langle \nabla f^\lambda(x^i), x^0 - x^i \rangle + \lambda \|\nabla f^\lambda(x^i)\|^2 \right) \\ &\quad - \frac{\lambda}{2} \left\| \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i) \right\|^2 - \frac{(\rho^{2^m} - 1)\lambda}{2} \|\nabla f^\lambda(x^{2^m-1})\|^2 \\ &\stackrel{(2)}{=} \sum_{i=0}^{2^m-2} \alpha_i \left(f(z^i) - f(z^{2^m-1}) + \langle \nabla f^\lambda(x^i), x^0 - z^i \rangle \right) \\ &\quad - \frac{\lambda}{2} \left\| \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i) \right\|^2 - \frac{(\rho^{2^m} - 1)\lambda}{2} \|\nabla f^\lambda(x^{2^m-1})\|^2, \end{aligned}$$

where we have used $\sum_{i=0}^{2^m-2} \alpha_i = \sum_{i=0}^{2^m-2} \pi_i^{(m)} = \rho^m - 1$ and $x^{2^m-1} - x^0 = -\sum_{i=0}^{2^m-2} \lambda \alpha_i \nabla f^\lambda(x^i)$ in (*). This completes the proof. \square

The following technical lemma will be used in subsequent analysis. Although it may appear complex initially, its verification is elementary as it comprises equations.

Lemma 5.4. *For any vectors $a, b, c, d \in \mathbb{R}^d$ and positive scalars $r, s \in \mathbb{R}_{++}$, we have*

$$\begin{aligned} & \langle b, a \rangle - \frac{1}{2}(r^2 - 1) \|c\|^2 - \frac{1}{2} \|b\|^2 + (r+1) \langle c, a - b - c \rangle + 2r \langle d, c - d - sc \rangle + 2s \langle d, a - b - sc - d \rangle \\ &= \frac{1}{2} (\|a\|^2 - \|a - b - (1+r)c - 2sd\|^2) + 2(s^2 - s - r) \langle d, d - c \rangle \end{aligned} \quad (45)$$

and

$$\begin{aligned} & \frac{s+r}{r^2} (\langle b, (1-s)c + d \rangle - (r^2 - 1) \|d\|^2 - \langle d, b + (s-1)c + d \rangle) - \|c\|^2 + \frac{s}{r} \langle c, b + (s-1)c + d \rangle \\ &= -(s+r) \|d\|^2 + \frac{s^2 - s - r}{r^2} \langle c, -b + rc - d \rangle. \end{aligned} \quad (46)$$

Then for RPPA with the right silver stepsize schedule $\bar{\pi}^{(m)}$, we establish an upper bound for the measure $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$ in the following theorem.

Theorem 5.4 (Tight convergence rate of RPPA with the right silver stepsize schedule $\bar{\pi}^{(m)}$). *For any fixed $m \geq 0$, let $N = 2^m$ and $\{(x^k, z^k)\}_{k=0}^N$ be the sequence generated by Algorithm 1 with stepsize schedule $\alpha = \bar{\pi}^{(m)}$ for solving (Q) from any starting point $x^0 \in \mathbb{R}^d$. Let $x^* \in \mathcal{X}^*$ be any minimizer of f and $z^* = \text{Prox}_{\lambda f}(x^*) = x^*$. Then, there holds*

$$f(z^N) - f(x^*) \leq \frac{1}{2T_m} \frac{1}{2\lambda} \|x^0 - x^*\|^2 \sim \frac{1}{2N^{\log_2 \rho}} \frac{1}{2\lambda} \|x^0 - x^*\|^2,$$

where T_m is defined in (41). Furthermore, this bound is tight.

Proof. For $m = 0$, we have $N = 1$ and $\alpha = \bar{\pi}^{(0)} = [\gamma_0]$. Note that $0 \leq P_{z^*, z^i}^f = f(z^*) - f(z^i) - \langle \nabla f^\lambda(x^i), z^* - z^i \rangle$, $0 \leq P_{z^0, z^1}^f = f(z^0) - f(z^1) - \langle \nabla f^\lambda(x^1), z^0 - z^1 \rangle$, and $T_0 = 1 + \gamma_0$. We thus have

$$\begin{aligned} 0 &\leq 2P_{z^*, z^0}^f + 2P_{z^0, z^1}^f + 2\gamma_0 P_{z^*, z^1}^f \\ &\stackrel{(2),(4)}{=} 2T_0 (f(z^*) - f(z^1)) + 2 \langle \nabla f^\lambda(x^0), x^0 - z^* - \lambda \nabla f^\lambda(x^0) \rangle \\ &\quad + 2 \langle \nabla f^\lambda(x^1), -\gamma_0 \lambda \nabla f^\lambda(x^0) - \lambda (\nabla f^\lambda(x^1) - \nabla f^\lambda(x^0)) \rangle \\ &\quad + 2\gamma_0 \langle \nabla f^\lambda(x^1), x^0 - z^* - \gamma_0 \lambda \nabla f^\lambda(x^0) - \lambda \nabla f^\lambda(x^1) \rangle, \\ &\stackrel{(45)}{=} 2T_0 (f(z^*) - f(z^1)) + \frac{1}{2\lambda} (\|x^0 - z^*\|^2 - \|x^0 - z^* - 2\lambda \nabla f^\lambda(x^0) - 2\gamma_0 \lambda \nabla f^\lambda(x^1)\|^2), \end{aligned} \quad (47)$$

where in the second “=” we have utilized (45) with $a = \lambda^{-1/2}(x^0 - z^*)$, $b = \mathbf{0}$, $c = \lambda^{1/2} \nabla f^\lambda(x^0)$, $d = \lambda^{1/2} \nabla f^\lambda(x^1)$, $r = 1$, $s = \gamma_0$, and the fact that $\gamma_0^2 - \gamma_0 - 1 = 0$. Hence, (47) implies that the desired bound holds for $m = 0$.

For $m \geq 1$, we have $N = 2^m$ and $\alpha = \bar{\pi}^{(m)} = [\pi^{(m)}, \gamma_m]$. Since the first $2^m - 1$ stepsizes are identical to the silver stepsize schedule $\pi^{(m)}$, the inequality (44) in Lemma 5.3 still holds. Then, by noting $0 \leq P_{z^*, z^i}^f = f(z^*) - f(z^i) - \langle \nabla f^\lambda(x^i), z^* - z^i \rangle$, $0 \leq P_{z^{2^m-1}, z^{2^m}}^f = f(z^{2^m-1}) - f(z^{2^m}) - \langle \nabla f^\lambda(x^{2^m}), z^{2^m-1} - z^{2^m} \rangle$ and $T_m = 1 + \gamma_m + \sum_{i=0}^{2^m-2} \alpha_i = \gamma_m + \rho^m$, we derive

$$\begin{aligned} 0 &\leq \sum_{i,j \in \{0,1,\dots,2^m-1\}} A_{i,j}^{(m)} P_{z^i, z^j}^f + \sum_{i=0}^{2^m-2} \alpha_i P_{z^*, z^i}^f + (\rho^m + 1) P_{z^*, z^{2^m-1}}^f + 2\rho^m P_{z^{2^m-1}, z^{2^m}}^f + 2\gamma_m P_{z^*, z^{2^m}}^f \\ &\stackrel{(44)}{=} \sum_{i=0}^{2^m-2} \alpha_i \left(f(z^*) - f(z^{2^m-1}) + \langle \nabla f^\lambda(x^i), x^0 - z^* \rangle \right) - \frac{(\rho^{2^m} - 1)\lambda}{2} \left\| \nabla f^\lambda(x^{2^m-1}) \right\|^2 \\ &\quad + (\rho^m + 1) \left(f(z^*) - f(z^{2^m-1}) - \langle \nabla f^\lambda(x^{2^m-1}), z^* - z^{2^m-1} \rangle \right) \\ &\quad + 2\rho^m \left(f(z^{2^m-1}) - f(z^{2^m}) - \langle \nabla f^\lambda(x^{2^m}), z^{2^m-1} - z^{2^m} \rangle \right) - \frac{\lambda}{2} \left\| \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i) \right\|^2 \end{aligned}$$

$$\begin{aligned}
& + 2\gamma_m \left(f(z^*) - f(z^{2^m}) - \langle \nabla f^\lambda(x^{2^m}), z^* - z^{2^m} \rangle \right) \\
\stackrel{(2),(4)}{=} & 2T_m \left(f(z^*) - f(z^{2^m}) \right) + \left\langle \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i), x^0 - z^* \right\rangle - \frac{(\rho^{2^m} - 1)\lambda}{2} \left\| \nabla f^\lambda(x^{2^m-1}) \right\|^2 \\
& + (\rho^m + 1) \left\langle \nabla f^\lambda(x^{2^m-1}), x^0 - z^* - \lambda \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i) - \lambda \nabla f^\lambda(x^{2^m-1}) \right\rangle \\
& + 2\rho^m \lambda \left\langle \nabla f^\lambda(x^{2^m}), \nabla f^\lambda(x^{2^m-1}) - \nabla f^\lambda(x^{2^m}) - \gamma_m \nabla f^\lambda(x^{2^m-1}) \right\rangle - \frac{\lambda}{2} \left\| \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i) \right\|^2 \\
& + 2\gamma_m \left\langle \nabla f^\lambda(x^{2^m}), x^0 - z^* - \lambda \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i) - \gamma_m \lambda \nabla f^\lambda(x^{2^m-1}) - \lambda \nabla f^\lambda(x^{2^m}) \right\rangle \\
\stackrel{(45)}{=} & - \frac{1}{2} \left\| \lambda^{-1/2}(x^0 - z^*) - \lambda^{1/2} \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i) - (1 + \rho^m) \lambda^{1/2} \nabla f^\lambda(x^{2^m-1}) - 2\gamma_m \lambda^{1/2} \nabla f^\lambda(x^{2^m}) \right\|^2 \\
& + 2T_m \left(f(z^*) - f(z^{2^m}) \right) + \frac{1}{2\lambda} \|x^0 - z^*\|^2, \tag{48}
\end{aligned}$$

where in the last “=” we have utilized (45) with $a = \lambda^{-1/2}(x^0 - z^*)$, $b = \lambda^{1/2} \sum_{i=0}^{2^m-2} \alpha_i \nabla f^\lambda(x^i)$, $c = \lambda^{1/2} \nabla f^\lambda(x^{2^m-1})$, $d = \lambda^{1/2} \nabla f^\lambda(x^{2^m})$, $r = \rho^m$, $s = \gamma_m$, and the fact that $\gamma_m^2 - \gamma_m - \rho^m = 0$. Hence, (48) implies that the desired bound also holds for $m \geq 1$. Finally, the tightness of the bound can be verified by using item (ii) of Lemma 2.3 and the fact that $T_m = 1 + \sum_{i=0}^{2^m-1} \alpha_i$. \square

Our final result concerns RPPA with the left silver stepsize schedule $\underline{\pi}^{(m)}$. Here, we establish a tight upper bound for the measure $\|\nabla f^\lambda(x^N)\|^2 / (f(x^0) - f(x^*))$. This measure is especially relevant in overparameterized models, which are frequently encountered in machine learning applications. In such models, $f(x^*)$ is often a known constant, such as 0. In these cases, the initial function value residual is directly computable, whereas the initial distance to a minimizer can be difficult to estimate.

Theorem 5.5 (Tight convergence rate of RPPA with the left silver stepsize schedule $\underline{\pi}^{(m)}$). *For any fixed $m \geq 0$, let $N = 2^m$ and $\{(x^k, z^k)\}_{k=0}^N$ be the sequence generated by Algorithm 1 with stepsize schedule $\alpha = \underline{\pi}^{(m)}$ for solving (Q) from any starting point $x^0 \in \text{dom}(f)$. Let $x^* \in \mathcal{X}^*$ be any minimizer of f and $z^* = \text{Prox}_{\lambda f}(x^*) = x^*$. Then, there holds*

$$\frac{\lambda}{2} \|\nabla f^\lambda(x^N)\|^2 \leq \frac{1}{2T_m} (f(x^0) - f(x^*)) \sim \frac{1}{2N^{\log_2 \rho}} (f(x^0) - f(x^*)),$$

where T_m is defined in (41). Furthermore, this bound is tight.

Proof. For $m = 0$, we have $N = 1$ and $\alpha = [\gamma_0]$. Note that $0 \leq P_{x^0, z^0}^f = f(x^0) - f(z^0) - \lambda \|\nabla f^\lambda(x^0)\|^2$, $0 \leq P_{z^1, z^*}^f = f(z^1) - f(z^*)$, $0 \leq P_{z^1, z^0}^f = f(z^1) - f(z^0) - \langle \nabla f^\lambda(x^0), z^1 - z^0 \rangle$, $0 \leq P_{z^0, z^1}^f = f(z^0) - f(z^1) - \langle \nabla f^\lambda(x^1), z^0 - z^1 \rangle$, and $T_0 = 1 + \gamma_0$. We thus have

$$\begin{aligned}
0 & \leq T_0 P_{z^0, z^1}^f + \gamma_0 P_{z^1, z^0}^f + P_{x^0, z^0}^f + P_{z^1, z^*}^f \\
& \stackrel{(2),(4)}{=} f(x^0) - f(z^*) - T_0 \lambda \langle \nabla f^\lambda(x^1), (\gamma_0 - 1) \nabla f^\lambda(x^0) + \nabla f^\lambda(x^1) \rangle \\
& \quad + \gamma_0 \lambda \langle \nabla f^\lambda(x^0), (\gamma_0 - 1) \nabla f^\lambda(x^0) + \nabla f^\lambda(x^1) \rangle - \lambda \|\nabla f^\lambda(x^0)\|^2 \\
& \stackrel{(46)}{=} f(x^0) - f(z^*) - T_0 \lambda \|\nabla f^\lambda(x^1)\|^2 + (\gamma_0^2 - \gamma_0 - 1) \lambda \langle \nabla f^\lambda(x^0), \nabla f^\lambda(x^0) - \nabla f^\lambda(x^1) \rangle, \tag{49}
\end{aligned}$$

where in the second “=” we have utilized (46) with $a = b = \mathbf{0}$, $c = \lambda^{1/2} \nabla f^\lambda(x^0)$, $d = \lambda^{1/2} \nabla f^\lambda(x^1)$, $r = 1$ and $s = \gamma_0$. Since $\gamma_0^2 - \gamma_0 - 1 = 0$, (49) implies that the desired bound holds for $m = 0$.

For $m \geq 1$, we have $N = 2^m$ and $\alpha = \underline{\pi}^{(m)} = [\gamma_m, \pi^{(m)}]$. We define the multiplier matrix $C^{(m)} := \frac{2T_m}{\rho^{2^m}} A^{(m)} \in \mathbb{R}^{2^m \times 2^m}$, whose components are indexed by $(i, j) \in \{1, 2, \dots, 2^m\}^2$. Since that the last $2^m - 1$ components of $\underline{\pi}^{(m)}$ are identical to the silver stepsize schedule $\pi^{(m)}$, the sequence $\{x^1\} \cup \{z^k\}_{k=1}^{2^m}$ is generated by Algorithm 1 with stepsize schedule $\pi^{(m)}$ from starting point x^1 . Thus, by Lemma 5.3, we

have

$$0 \leq \sum_{i,j \in \{1,2,\dots,2^m\}} C_{i,j}^{(m)} P_{z^i,z^j}^f = \frac{2T_m}{\rho^{2m}} \left[\sum_{i=1}^{2^m-1} \alpha_i \left(f(z^i) - f(z^{2^m}) + \langle \nabla f^\lambda(x^i), x^1 - z^i \rangle \right) - \frac{(\rho^{2^m} - 1)\lambda}{2} \left\| \nabla f^\lambda(x^{2^m}) \right\|^2 - \frac{\lambda}{2} \left\| \sum_{i=1}^{2^m-1} \alpha_i \nabla f^\lambda(x^i) \right\|^2 \right]. \quad (50)$$

Note that $0 \leq P_{x^0,z^0}^f = f(x^0) - f(z^0) - \lambda \left\| \nabla f^\lambda(x^0) \right\|^2$, $0 \leq P_{z^{2^m},z^*}^f = f(z^{2^m}) - f(z^*)$, $0 \leq P_{z^i,z^0}^f = f(z^i) - f(z^0) - \langle \nabla f^\lambda(x^0), z^i - z^0 \rangle$, $0 \leq P_{z^0,z^i}^f = f(z^0) - f(z^i) - \langle \nabla f^\lambda(x^i), z^0 - z^i \rangle$, and $T_m = 1 + \gamma_m + \sum_{i=1}^{2^m-1} \alpha_i = \gamma_m + \rho^m$. We thus have

$$\begin{aligned} 0 &\leq \sum_{i,j \in \{1,\dots,2^m\}} C_{i,j}^{(m)} P_{z^i,z^j}^f + \frac{T_m}{\rho^{2m}} \left[\sum_{i=1}^{2^m} \alpha_i \left(P_{z^0,z^i}^f + P_{z^{2^m},z^i}^f \right) + P_{z^0,z^{2^m}}^f \right] + \frac{\gamma_m}{\rho^m} P_{z^{2^m},z^0}^f + P_{x^0,z^0}^f + P_{z^{2^m},z^*}^f \\ &\stackrel{(50)}{=} \frac{T_m}{\rho^{2m}} \left(1 + \sum_{i=1}^{2^m-1} \alpha_i \right) \left(f(z^0) - f(z^{2^m}) \right) + \frac{\gamma_m}{\rho^m} \left(f(z^{2^m}) - f(z^0) \right) + f(x^0) - f(z^0) + f(z^{2^m}) - f(z^*) \\ &\quad + \frac{T_m}{\rho^{2m}} \left(\sum_{i=1}^{2^m} \alpha_i \langle \nabla f^\lambda(x^i), 2x^1 - z^0 - z^{2^m} \rangle - (\rho^{2^m} - 1)\lambda \left\| \nabla f^\lambda(x^{2^m}) \right\|^2 - \lambda \left\| \sum_{i=1}^{2^m-1} \alpha_i \nabla f^\lambda(x^i) \right\|^2 \right. \\ &\quad \left. + \langle \nabla f^\lambda(x^{2^m}), z^{2^m} - z^0 \rangle \right) + \frac{\gamma_m}{\rho^m} \langle \nabla f^\lambda(x^0), z^0 - z^{2^m} \rangle - \lambda \left\| \nabla f^\lambda(x^0) \right\|^2 \\ &\stackrel{(2),(4)}{=} f(x^0) - f(z^*) + \frac{T_m}{\rho^{2m}} \left(\lambda \left\langle \sum_{i=1}^{2^m} \alpha_i \nabla f^\lambda(x^i), (1 - \gamma_m) \nabla f^\lambda(x^0) + \nabla f^\lambda(x^{2^m}) \right\rangle \right. \\ &\quad \left. - \lambda \left\langle \nabla f^\lambda(x^{2^m}), \sum_{i=1}^{2^m} \alpha_i \nabla f^\lambda(x^i) + (\gamma_m - 1) \nabla f^\lambda(x^0) + \nabla f^\lambda(x^{2^m}) \right\rangle \right) - \lambda \left\| \nabla f^\lambda(x^0) \right\|^2 \\ &\quad + \frac{\gamma_m}{\rho^m} \left\langle \nabla f^\lambda(x^0), \sum_{i=1}^{2^m} \alpha_i \nabla f^\lambda(x^i) + (\gamma_m - 1) \nabla f^\lambda(x^0) + \nabla f^\lambda(x^{2^m}) \right\rangle - (\rho^{2^m} - 1)\lambda \left\| \nabla f^\lambda(x^{2^m}) \right\|^2 \\ &\stackrel{(46)}{=} f(x^0) - f(z^*) - T_m \lambda \left\| \nabla f^\lambda(x^{2^m}) \right\|^2, \end{aligned} \quad (51)$$

where in the second “=” we also used $\sum_{i=1}^{2^m-1} \alpha_i = \sum_{i=0}^{2^m-2} \pi_i^{(m)} = \rho^m - 1$ and in the last “=” we have utilized (46) with $a = \mathbf{0}$, $b = \lambda^{1/2} \sum_{i=1}^{2^m} \alpha_i \nabla f^\lambda(x^i)$, $c = \lambda^{1/2} \nabla f^\lambda(x^0)$, $d = \lambda^{1/2} \nabla f^\lambda(x^{2^m})$, $r = \rho^m$, $s = \gamma_m$, and the fact that $\gamma_m^2 - \gamma_m - \rho^m = 0$. Hence, (51) implies that the desired bound also holds for $m \geq 1$. Finally, the tightness of the bound can be verified by using item (iii) of Lemma 2.3 and the fact that $T_m = 1 + \sum_{i=0}^{2^m-1} \alpha_i$. \square

5.3 Discussion on the optimality of silver stepsize schedules

Given a fixed stepsize schedule, we obtain a specific instance of the RPPA and can establish its tight convergence rate under certain performance measures. This naturally raises the following questions: What is the optimal stepsize schedule that yields the best tight convergence rate for RPPA? Are the silver stepsize schedules $\pi^{(m)}$, $\bar{\pi}^{(m)}$, and $\underline{\pi}^{(m)}$ that we previously analyzed optimal in this regard? In this section, we aim to address these questions.

First, we consider the optimality of the order of the accelerated convergence rate $O(1/N^{\log_2 \rho}) \approx O(1/N^{1.2716})$, which we obtained for the RPPA using the (modified) silver stepsize schedules across different measures. This order of convergence has been conjectured optimal for GD method without momentum for solving smooth convex optimization problems [AP23b, Section 1.1] [GSW24, Remark 1]. Given the close relationship between RPPA and GD method, it is reasonable to conjecture that this order of rate is also optimal for RPPA without momentum, i.e. Algorithm 1.

Next, we examine the exact optimal stepsize schedules, that is, those that achieve the optimal rate even up to the constant. For any fixed $N \in \mathbb{N}_+$ and $\alpha \in \mathbb{R}_{++}^N$, the tight worst-case rate of Algorithm 1 with the stepsize schedule α under different measures can be computed numerically using the

performance estimation approach proposed in [DT14, THG17c, THG17b]. Additionally, for any fixed performance measure, we can conduct a numerical search for the optimal stepsize schedules that minimize the tight worst-case rate. The problem of finding the optimal stepsize schedules can be formulated as a nonconvex quadratically constrained quadratic programming problem and be solved numerically by a customized branch-and-bound method (abbreviated as BnB-PEP for later use), as described in [DVR24]. In this paper, we have considered the following three measures: $\|\nabla f^\lambda(x^N)\| / \|x^0 - x^*\|$, $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$ and $\|\nabla f^\lambda(x^N)\|^2 / (f(x^0) - f(x^*))$. In the following, we denote the numerically optimal stepsize schedules for these measures as $\alpha^{\text{gd,dist},*}$, $\alpha^{\text{fval,dist},*}$ and $\alpha^{\text{gd,fval},*}$, respectively. In Table 3, we compare $\alpha^{\text{gd,dist},*}$ with $\pi^{(m)}$, for which the measure $\|\nabla f^\lambda(x^N)\| / \|x^0 - x^*\|$ is used, and $\alpha^{\text{fval,dist},*}$ with $\bar{\pi}^{(m)}$, for which the measure $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$ is employed.

For the measure $\|\nabla f^\lambda(x^N)\| / \|x^0 - x^*\|$, we observe from Table 3 that the silver stepsize schedule $\pi^{(m)}$ approximates the optimal stepsize schedule $\alpha^{\text{gd,dist},*}$ found via BnB-PEP for $m = 1, 2, 3$ (corresponding to $N = 1, 3, 7$, respectively) quite well. Therefore, we conjecture that the silver stepsize schedule $\pi^{(m)}$ is already optimal for RPPA for the measure $\|\nabla f^\lambda(x^N)\| / \|x^0 - x^*\|$. For the measure $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$, the constructed right silver stepsize schedule $\bar{\pi}^{(m)}$ approximates the optimal stepsize schedule found by BnB-PEP quite well when $m = 0, 2$ (corresponding to $N = 1, 4$, respectively). In Table 4, we present the tight worst-case upper bound of $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$ corresponding to $\bar{\pi}^{(m)}$ and $\alpha^{\text{fval,dist},*}$ with $m = 0, 1, 2, 3$ (corresponding to $N = 1, 2, 4, 8$, respectively). It can be seen from Table 4 that, for $m = 1$ (and $N = 2$), our constructed schedule $\bar{\pi}^{(1)} \approx [1.414214, 2.132242]$ gives a tight worst-case bound of $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$ to be 0.054988. This value is larger than that corresponding to $\alpha^{\text{fval,dist},*}$, which is 0.54900. The discrepancy between the two is somehow nonnegligible. A similar discrepancy is observed for $m = 3$ (and $N = 8$). This indicates that $\bar{\pi}^{(m)}$ may be suboptimal, though it is quite close to the optimal stepsize schedule.

The right and left silver stepsize schedules $\bar{\pi}^{(m)}$ and $\underline{\pi}^{(m)}$ mirror each other and provide convergence rates that mirror the relationship between function value residual and subgradient norm (see Theorems 5.4 and 5.5). This symmetry has been observed in optimal gradient methods for solving smooth convex optimization problems [GSW24, Kim⁺23]. It appears closely related to the H-duality theory presented in [Kim⁺23]. Therein, it was shown that reversing the order of steps in methods using certain inductive proofs leads to a conversion between (sub)gradient norm and function value residual convergence. Due to this fundamental connection, the left silver stepsize schedule $\underline{\pi}^{(m)}$ may also be suboptimal.

6 Conclusion

In this paper, we have examined RPPA for solving convex optimization problems. Our analysis encompasses three types of relaxation schedules. For both the constant and the TV's relaxation schedules, we have established non-ergodic and tight $O(1/N)$ convergence rate results by using the function value residual or subgradient norm as the optimality measure. For the original silver relaxation schedule and two extensions, which permit α_k to exceed 2, we have attained accelerated convergence rates of $O(1/N^{1.2716})$ across three prevalent performance measures.

Some limitations of our analysis are as follows. First, our tight analysis for the constant case can only cover the range $\alpha \in (0, \sqrt{2}]$, rather than the traditional range $(0, 2)$. This gap might be addressed by leveraging recent advances in GD method in [RGP24], where tight bounds for the GD method with a constant stepsize across the full range $\alpha \in (0, 2)$ were established. Second, we have maintained the proximal parameter $\lambda > 0$ as constant throughout all the iterations. This may be overly restrictive in practice. Hence, conducting an in-depth analysis of RPPA with varying proximal parameters λ_k presents an interesting avenue for future research. It remains an open task for further studies. Moreover, there is also room for improvement in the constructed silver stepsize schedules $\bar{\pi}^{(m)}$ and $\underline{\pi}^{(m)}$, as they do not precisely match the numerical optimal stepsize schedules. Determining the exact optimal stepsize schedule also remains an open problem for future study.

N	Measure: $\frac{\ \nabla f^\lambda(x^N)\ }{\ x^0 - x^*\ }$		Measure: $\frac{f(z^N) - f(x^*)}{\ x^0 - x^*\ ^2}$	
	$\pi^{(m)}$	$\alpha^{\text{gd,dist},\star}$	$\bar{\pi}^{(m)}$	$\alpha^{\text{fval,dist},\star}$
1	[1.414214]	[1.414334]	[1.618034]	[1.618035]
2	–	[1.600519 1.414482]	[1.414214 2.132242]	[1.515086 2.038664]
3	[1.414214 2.0 1.414214]	[1.414745 2.000451 1.414115]	–	[1.414211 1.601232 2.565298]
4	–	[1.414214 1.601232 2.260578 1.414214]	[1.414214 2.0 1.414214 2.965447]	[1.414214 2.0 1.414213 2.965447]
7	[1.414214 2.0 1.414214 3.414214 1.414214 2.0 1.414214]	[1.414402 2.000943 1.413279 3.411508 1.413362 2.000698 1.414120]	–	[1.414214 1.601232 3.989651 1.414214 2.745299 1.515087 2.038664]
8	–	[1.414215 2.000003 1.414214 3.754383 1.414213 1.601232 2.260578 1.414214]	[1.414214 2.0 1.414214 3.414214 1.414214 2.0 1.414214 4.284319]	[1.414214 2.0 1.414214 5.004706 1.414214 2.0 1.414214 2.965447]

Table 3: Comparison between $\alpha^{\text{gd,dist},\star}$ and $\pi^{(m)}$, as well as $\alpha^{\text{fval,dist},\star}$ and $\bar{\pi}^{(m)}$, where $\alpha^{\text{gd,dist},\star}$ and $\alpha^{\text{fval,dist},\star}$ denote the optimal stepsize schedules computed by BnB-PEP under the measures $\|\nabla f^\lambda(x^N)\| / \|x^0 - x^*\|$ and $(f(z^N) - f(x^*)) / \|x^0 - x^*\|^2$, respectively. Recall that $\pi^{(m)}$ is available only for $N = 2^m - 1$ ($m \geq 1$), and $\bar{\pi}^{(m)}$ is available only for $N = 2^m$ ($m \geq 0$).

References

- [Alt18] Jason M. Altschuler. Greed, hedging, and acceleration in convex optimization. Master’s thesis, Massachusetts Institute of Technology, 2018.
- [AP23a] Jason M. Altschuler and Pablo A. Parrilo. Acceleration by stepsize hedging I: Multi-step descent and the silver stepsize schedule. *arXiv preprint arXiv:2309.07879*, September 2023.
- [AP23b] Jason M. Altschuler and Pablo A. Parrilo. Acceleration by stepsize hedging II: Silver stepsize schedule for smooth convex optimization. *arXiv preprint arXiv:2309.16530*, September 2023.
- [Bec17] Amir Beck. *First-order methods in optimization*, volume 25 of *MOS-SIAM Ser. Optim.* SIAM, Philadelphia, PA, 2017.

N	$\bar{\pi}^{(m)}$	$\alpha^{\text{fval,dist},\star}$
	Tight upper bound of $\frac{f(z^N) - f(x^*)}{\ x^0 - x^*\ ^2}$	
1	0.095492	0.095492
2	0.054988	0.054900
4	0.028429	0.028429
8	0.013620	0.013422

Table 4: Comparison between the tight upper bound of $\frac{f(z^N) - f(x^*)}{\|x^0 - x^*\|^2}$ corresponding to $\bar{\pi}^{(m)}$ and $\alpha^{\text{fval,dist},\star}$ with $m = 0, 1, 2, 3$ (corresponding to $N = 1, 2, 4, 8$, respectively). In the computation, we set $\lambda = 1$ for RPPA. Each table entry is obtained through solving the corresponding performance estimation problem. We observe that $\bar{\pi}^{(m)}$ attains larger tight worst-case bounds than $\alpha^{\text{fval,dist},\star}$ when $m = 1, 3$ (corresponding to $N = 2, 8$, respectively). This may indicate that $\bar{\pi}^{(m)}$ is suboptimal.

- [Boy⁺11] Stephen Boyd, Navdeep Parikh, Eric Chu, Jonathan Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.
- [CGH14] Xingju Cai, Guoyong Gu, and Bingsheng He. On the $O(1/t)$ convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators. *Comput. Optim. Appl.*, 57(2):339–363, March 2014.
- [CY14] Etienne Corman and Xiaoming Yuan. A generalized proximal point algorithm and its convergence rate. *SIAM J. Optim.*, 24(4):1614–1638, January 2014.
- [DT14] Yoel Drori and Marc Teboulle. Performance of first-order methods for smooth convex minimization: A novel approach. *Math. Program.*, 145(1):451–482, June 2014.
- [DVR24] Shuvomoy Das Gupta, Bart P. G. Van Parys, and Ernest K. Ryu. Branch-and-bound performance estimation programming: A unified methodology for constructing optimal optimization methods. *Math. Program.*, 204(1-2):567–639, 2024.
- [EB92] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55(3):293–318, 1992.
- [GHY14] Guoyong Gu, Bingsheng He, and Xiaoming Yuan. Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: A unified approach. *Comput. Optim. Appl.*, 59(1-2):135–161, October 2014.
- [Gri24] Benjamin Grimmer. Provably faster gradient descent via long steps. *SIAM J. Optim.*, 34(3):2588–2608, 2024.
- [GSW23] Benjamin Grimmer, Kevin Shu, and Alex L. Wang. Accelerated gradient descent via long steps. *arXiv preprint arXiv:2309.09961*, September 2023.
- [GSW24] Benjamin Grimmer, Kevin Shu, and Alex L. Wang. Accelerated objective gap and gradient norm convergence for gradient descent via long steps. *arXiv preprint arXiv:2403.14045*, April 2024.
- [Gül91] Osman Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.*, 29(2):403–419, March 1991.
- [GY24] Guoyong Gu and Junfeng Yang. Tight ergodic sublinear convergence rate of the relaxed proximal point algorithm for monotone variational inequalities. *J. Optim. Theory Appl.*, 202(1):373–387, 2024.
- [Kim⁺23] Jaeyeon Kim, Asuman Ozdaglar, Chanwoo Park, and Ernest Ryu. Time-reversed dissipation induces duality between minimizing gradient norm and function value. In *Adv. Neural Inform. Process. Syst.*, volume 36, pages 23389–23440, 2023.
- [Kra55] M. A. Krasnosel’skii. Two remarks on the method of successive approximations. *Uspehi Mat. Nauk*, 10(1(63)):123–127, 1955.
- [LG24] Alan Luner and Benjamin Grimmer. On averaging and extrapolation for gradient descent. *arXiv preprint arXiv:2402.12493*, February 2024.
- [Lie18] Felix Lieder. *Projection based methods for conic linear programming — optimal first order complexities and norm constrained quasi Newton methods*. PhD thesis, Heinrich-Heine-Universität Düsseldorf, September 2018.
- [LY13] Min Li and Yanfei You. Convergence analysis of the relaxed proximal point algorithm. *Abstr. Appl. Anal.*, 2013:Art. ID 912846, 6, June 2013.
- [Man53] W. Robert Mann. Mean value methods in iteration. *Proc. Amer. Math. Soc.*, 4:506–510, 1953.

- [Mar70] B. Martinet. Brève communication. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4(R3):154–158, 1970.
- [Mor65] Jean-Jacques Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.
- [Nes18] Yurii Nesterov. *Lectures on convex optimization*, volume 137 of *Springer Optimization and Its Applications*. Springer, Cham, second edition, 2018.
- [RGP24] Teodor Rotaru, François Glineur, and Panagiotis Patrinos. Exact worst-case convergence rates of gradient descent: A complete analysis for all constant stepsizes over nonconvex and convex functions. *arXiv preprint arXiv:2406.17506*, June 2024.
- [Roc76a] R. Tyrrell Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, May 1976.
- [Roc76b] R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, August 1976.
- [Sou46] R. V. Southwell. *Relaxation methods in theoretical physics*. Clarendon Press, Oxford, 1946.
- [THG17a] Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Exact worst-case performance of first-order methods for composite convex optimization. *SIAM J. Optim.*, 27(3):1283–1313, January 2017.
- [THG17b] Adrien B. Taylor, Julien M. Hendrickx, and Francois Glineur. Performance estimation toolbox (PESTO): Automated worst-case analysis of first-order optimization methods. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1278–1283, Melbourne, Australia, December 2017. IEEE.
- [THG17c] Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Math. Program.*, 161(1-2):307–345, 2017.
- [TV23] Marc Teboulle and Yakov Vaisbourd. An elementary approach to tight worst case complexity analysis of gradient based methods. *Math. Program.*, 201(1-2):63–96, 2023.
- [Var62] Richard S Varga. *Matrix iterative analysis*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1962.