AN IMPLEMENTABLE HALPERN-ACCELERATED PROXIMAL POINT METHOD WITH SHARP RESIDUAL RATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Halpern-accelerated proximal point methods (HPPMs) improve the fixed-point residual rate from $\mathcal{O}(k^{-1/2})$ to $\mathcal{O}(k^{-1})$ under exact resolvent evaluations. In practice, however, exact resolvent evaluations can be computationally expensive. To address this issue, we study an implementable HPPM with a relaxation parameter $\rho \in (0, 2]$ and resolvent errors satisfying $\|e_k\| = \mathcal{O}(k^{-p})$, aiming to determine the error decay required to preserve the accelerated $\mathcal{O}(k^{-1})$ residual rate. Specifically, for every fixed $\rho \in (0, 2)$, we establish the sharp worst-case residual rate $\Theta(k^{1/2-p})$ for $0 < p < 3/2$ and $\Theta(k^{-1})$ for $p \geq 3/2$. Thus, $p = 3/2$ is the exact threshold for preserving the accelerated $\mathcal{O}(k^{-1})$ residual rate, removing the extraneous $\sqrt{\log k}$ factor in existing upper bounds. By contrast, at the boundary $\rho = 2$, we obtain the sharp rates $\Theta(k^{1-p})$ for $0 < p < 2$, $\Theta((\log k)/k)$ for $p = 2$, and $\Theta(k^{-1})$ for $p > 2$. Matching lower-bound constructions establish the sharpness of all these rates. These results precisely characterize the resolvent accuracy required to retain Halpern acceleration under approximate resolvent evaluations.

1 INTRODUCTION

We consider the maximally monotone inclusion problem

$$0 \in A(w), \tag{1}$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximally monotone on a real Hilbert space $\mathcal{H}$ and $A^{-1}(0) \neq \varnothing$. A fundamental method for solving (1) is the proximal point method (PPM), whose general convergence theory was developed by Rockafellar (1976b). Writing $J_A := (I + A)^{-1}$ for the resolvent of A, the PPM with unit proximal parameter takes the form

$$w^{k+1} = J_A(w^k), \qquad k \geq 0.$$

For general maximally monotone operators with a nonempty zero set, the PPM satisfies the fixed-point residual bound $\|w^k - J_A(w^k)\| = \mathcal{O}(k^{-1/2})$, which is sharp in the worst-case sense (Brézis & Lions, 1978; Gu & Yang, 2020).

Kim (2021) developed an accelerated PPM that improves the $\mathcal{O}(k^{-1/2})$ residual rate of the classical PPM to $\mathcal{O}(1/k)$. This accelerated PPM is equivalent to the classical Halpern iteration (Halpern, 1967) applied to the nonexpansive reflected resolvent $2J_A - I$; see also (Contreras & Cominetti, 2023; Ryu & Yin, 2022). With unit proximal parameter, it can be written as

$$\begin{cases} \bar{w}^{k+1} = J_A(w^k), \\ w^{k+1} = \dfrac{k+1}{k+2}\big(2\bar{w}^{k+1} - w^k\big) + \dfrac{1}{k+2}w^0, \end{cases} \qquad k \geq 0.$$

From the fixed-point perspective, Lieder (2021) established a sharp $\mathcal{O}(1/k)$ residual bound for Halpern iteration applied to nonexpansive mappings in Hilbert spaces, and Park & Ryu (2022) subsequently established matching complexity lower bounds. Together, these results show that Halpern iteration, with appropriate anchoring weights, attains the optimal worst-case $\mathcal{O}(1/k)$ fixed-point residual rate in the nonexpansive setting.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

In many applications, however, evaluating the resolvent J_A exactly requires solving a nontrivial subproblem, rendering the exact Halpern-accelerated PPM computationally expensive or even impractical. This issue arises naturally in optimization methods and problems relevant to modern machine learning, including augmented-Lagrangian methods (Rockafellar, 1976a; Zhang et al., 2025a) for constrained learning (Ramirez & Lacoste-Julien, 2026), distributionally robust optimization (Liang et al., 2025) under distribution shift (Mehta et al., 2024; Zhang et al., 2025b), and finite-sum monotone inclusions for empirical equilibrium problems (Cai et al., 2024). In these settings, the resolvent is typically unavailable in closed form and is instead approximated using an iterative inner solver.

An implementable variant instead replaces the exact resolvent with an approximate one, allowing the underlying subproblem to be solved only to finite accuracy. Approximate resolvent evaluations can substantially reduce the computational cost, but may deteriorate the accelerated residual rate. This motivates a precise characterization of the resolvent accuracy required to preserve the $\mathcal{O}(1/k)$ residual rate of Halpern-accelerated PPM (HPPM).

To this end, we consider the following implementable relaxed HPPM with approximate resolvent evaluations. Given $w^0 \in \mathcal{H}$ and a fixed relaxation parameter $\rho \in (0, 2]$, each iteration $k \geq 0$ computes

$$\begin{cases} \widetilde{w}^{k+1} \approx J_A(w^k), \\ w^{k+1} = \dfrac{k+1}{k+2}\Big(\rho\widetilde{w}^{k+1} + (1-\rho)w^k\Big) + \dfrac{1}{k+2}w^0. \end{cases} \tag{2}$$

We define the resolvent approximation error by

$$e_{k+1} := \widetilde{w}^{k+1} - J_A(w^k) \tag{3}$$

and consider the polynomial error model

$$\|e_j\| \leq cj^{-p}, \qquad j \geq 1, \tag{4}$$

for some $c \geq 0$ and $p > 0$. Our goal is to characterize the sharp worst-case fixed-point residual rate as a function of p and ρ, and in particular, to determine the minimal error decay required to preserve the $\mathcal{O}(1/k)$ rate of the exact method.

Several existing analyses have studied approximate resolvent evaluations in Halpern-type methods. For any $\epsilon > 0$, Diakonikolas (2020) obtained a residual of at most ϵ within $\mathcal{O}(\epsilon^{-1})$ outer iterations under a target-dependent resolvent accuracy of order $\mathcal{O}(\epsilon k^{-2})$. Subsequent deterministic analyses established residual bounds under various decaying error conditions (Alacaoglu et al., 2024; Liang et al., 2025; Zhang et al., 2025a; Bravo & Contreras, 2026). Complementary stochastic analyses study Halpern-type methods under various moment-based and relative inexactness criteria; see, e.g., (Cai et al., 2022; 2024; Pischke & Powell, 2026). Table 1 summarizes the deterministic results most relevant to our setting and compares them with the sharp rates established in this paper.

Despite these advances, a sharp characterization under the deterministic polynomial error model (4) remains unavailable. In particular, existing results neither identify the minimal error decay required to preserve the $\mathcal{O}(1/k)$ residual rate nor provide matching upper and lower bounds across error regimes and relaxation parameters.

In this paper, we provide a complete sharp characterization of how approximation errors affect the residual convergence rate and reveal a phase transition between the interior regime $0 < \rho < 2$ and the boundary case $\rho = 2$. For each fixed $0 < \rho < 2$ and error constant $c > 0$, the $\mathcal{O}(1/k)$ residual rate is preserved in the worst case if and only if $p \geq 3/2$; in particular, the $\sqrt{\log k}$ loss in previous upper bounds at the critical exponent $p = 3/2$ is removed. In contrast, at $\rho = 2$, the sharp worst-case residual rate is $\Theta(k^{-1}\log k)$ when $p = 2$, and preserving the $\mathcal{O}(1/k)$ rate requires $p > 2$. The improvement for $0 < \rho < 2$ stems from the strict averagedness of the relaxed resolvent, which enables stronger control of accumulated approximation errors. At the boundary $\rho = 2$, only nonexpansiveness is guaranteed, leading to a higher accuracy threshold and the unavoidable logarithmic factor at $p = 2$.

Matching lower-bound constructions establish the sharpness of these rates for all $p > 0$ and each fixed $\rho \in (0, 2]$. Finally, we illustrate the practical implications of our theoretical results through applications to augmented Lagrangian methods, distributionally robust optimization, and finite-sum monotone inclusions.

Table 1: Comparison of deterministic residual bounds for implementable HPPM.

Reference	Relaxation	Sufficient error condition	Residual order	Sharpness
Alacaoglu et al. (2024)†	$\rho = 1$	$\mathcal{O}(k^{-3/2}/\log k)$	$\mathcal{O}(k^{-1})$	
Liang et al. (2025)	$0 < \rho \leq 1$	$\mathcal{O}(k^{-3/2}/\log k)$	$\mathcal{O}(k^{-1})$	
Zhang et al. (2025a)	$\rho = 1$	$1 < p < 3$	$\mathcal{O}(k^{1/2-p/2})$	
		$p = 3$	$\mathcal{O}(k^{-1}\sqrt{\log k})$	
		$p > 3$	$\mathcal{O}(k^{-1})$	
Bravo & Contreras (2026)	$0 < \rho \leq 2$	$1 \leq p < 2$	$\mathcal{O}(k^{1-p})$	
		$p \geq 2$	$\mathcal{O}(k^{-1}\log k)$	
This paper	$0 < \rho < 2$	$0 < p < 3/2$	$\Theta(k^{1/2-p})$	✓
		$p \geq 3/2$	$\Theta(k^{-1})$	✓
This paper	$\rho = 2$	$0 < p < 2$	$\Theta(k^{1-p})$	✓
		$p = 2$	$\Theta(k^{-1}\log k)$	✓
		$p > 2$	$\Theta(k^{-1})$	✓

A checkmark indicates a residual order established by matching upper and lower bounds. †The absolute-error condition listed for (Alacaoglu et al., 2024) is a sufficient condition derived from their Lemma 2.5; their implementation uses the residual-relative accuracy criterion in their equation (2.1).

The remainder of the paper is organized as follows. Section 2 introduces the preliminary estimates. Section 3 establishes the residual upper bounds, and Section 4 proves their sharpness through matching lower-bound constructions. Section 5 presents applications of the implementable HPPM, and Section 6 concludes the paper.

Notation. For two nonnegative sequences $\{a_k\}_{k\geq 0}$ and $\{b_k\}_{k\geq 0}$, we write $a_k = \mathcal{O}(b_k)$ if there exist constants $C > 0$ and $k_0 \in \mathbb{N}$ such that $a_k \leq Cb_k$ for all $k \geq k_0$, and $a_k = \Omega(b_k)$ if there exist constants $c > 0$ and $k_0 \in \mathbb{N}$ such that $a_k \geq cb_k$ for all $k \geq k_0$. We write $a_k = \Theta(b_k)$ if both $a_k = \mathcal{O}(b_k)$ and $a_k = \Omega(b_k)$ hold.

2 PRELIMINARIES

We impose the following standard assumption throughout the paper.

Assumption 1. *The solution set $A^{-1}(0)$ of (1) is nonempty.*

Unless otherwise stated, let $w^0 \in \mathcal{H}$ be the initial point and $w^* \in A^{-1}(0)$ be an arbitrary solution of (1). Define

$$R := I - J_A, \qquad T_\rho := (1-\rho)I + \rho J_A = I - \rho R. \tag{5}$$

Thus, $R(w) = w - J_A(w)$ is the fixed-point residual associated with J_A. Moreover, for $0 < \rho \leq 2$, the firm nonexpansiveness of J_A implies

$$\|T_\rho x - T_\rho y\|^2 \leq \|x-y\|^2 - \rho(2-\rho)\|R(x) - R(y)\|^2, \qquad x, y \in \mathcal{H}. \tag{6}$$

In particular, T_ρ is nonexpansive for every $\rho \in (0, 2]$ and is strictly averaged when $\rho \in (0, 2)$.

For comparison with the inexact iteration (2), let $u^0 = w^0$ and consider its exact counterpart

$$\begin{cases} \bar{u}^{k+1} = J_A(u^k), \\ u^{k+1} = \dfrac{k+1}{k+2}\Big(\rho\bar{u}^{k+1} + (1-\rho)u^k\Big) + \dfrac{1}{k+2}w^0, \qquad k \geq 0. \end{cases} \tag{7}$$

Equivalently,

$$u^{k+1} = \frac{k+1}{k+2}T_\rho u^k + \frac{1}{k+2}w^0,$$

so (7) is the classical Halpern iteration applied to T_ρ.

Define the discrepancy between the inexact and exact sequences by

$$\Delta_k := w^k - u^k, \qquad k \geq 0. \tag{8}$$

We first establish in Lemma 2.1 a basic estimate for this discrepancy using only the nonexpansiveness of T_ρ. The proofs of the results in this section are given in Appendix A.1.

Lemma 2.1. *Fix $w^0 \in \mathcal{H}$ and $\rho \in (0, 2]$. Let $\{w^k\}$ be generated by* (2)*, let $\{u^k\}$ be the exact Halpern iteration* (7) *with $u^0 = w^0$, and let Δ_k be defined by* (8)*. Then,*

$$\|\Delta_k\| \leq \frac{\rho}{k+1} \sum_{j=1}^{k} j\|e_j\|, \quad k \geq 0.$$

If, in addition, $\|e_j\| \leq cj^{-3/2}$ for some $c \geq 0$ and all $j \geq 1$, then

$$\|\Delta_k\| \leq \frac{2\rho c}{\sqrt{k+1}}, \quad k \geq 0. \tag{9}$$

We next establish in Proposition 2.2 a residual bound for the exact Halpern iteration.

Proposition 2.2. *Suppose that Assumption 1 holds. Fix $w^0 \in \mathcal{H}$, $\rho \in (0, 2]$, and $w^* \in A^{-1}(0)$. Let $\{u^k\}$ be generated by the exact Halpern iteration* (7) *with $u^0 = w^0$. Then*

$$\|R(u^k)\| \leq \frac{2\|w^0 - w^*\|}{\rho k + 2}, \qquad k \geq 0. \tag{10}$$

Remark 2.3 (Comparison with existing results)**.** *Since* (7) *is the classical Halpern iteration applied to T_ρ, (Lieder, 2021, Theorem 2.1) gives*

$$\|R(u^k)\| \leq \frac{2\|w^0 - w^*\|}{\rho(k+1)}, \qquad k \geq 0,$$

where we used $I - T_\rho = \rho R$. The bound in Proposition 2.2 is strictly sharper for $0 < \rho < 2$ and coincides with the above bound for $\rho = 2$.

3 Residual rate analysis

In this section, we establish residual bounds for the implementable Halpern PPM (2) under the error condition (4). For $0 < \rho < 2$, the operator T_ρ is strictly averaged, and the descent term in (6) provides additional control that is unavailable at the boundary $\rho = 2$. We therefore treat the two regimes separately in Subsections 3.1 and 3.2.

3.1 Residual rate for $0 < \rho < 2$

We first derive a general residual bound from a Lyapunov inequality. This bound gives the desired rate for $0 < p < 3/2$ but incurs a logarithmic loss at the critical exponent $p = 3/2$. We then establish a square-tail estimate to remove this loss and conclude with a unified residual bound for all $p > 0$. The proofs of the results in this subsection are given in Appendix A.2.

Inspired by the analyses in (Diakonikolas, 2020; Cai et al., 2022), we introduce the Lyapunov function

$$\Phi_k := (k+1)\langle R(w^k), w^k - w^0\rangle + \frac{\rho k(k+1)}{2}\|R(w^k)\|^2, \qquad k \geq 0. \tag{11}$$

The following lemma establishes a one-step estimate for Φ_k and a lower bound in terms of the residual.

Lemma 3.1. *Suppose that Assumption 1 holds. Let $w^0 \in \mathcal{H}$, $w^* \in A^{-1}(0)$, and $\rho \in (0, 2)$. Let $\{w^k\}$ be generated by* (2)*. Then, for every $k \geq 0$,*

$$\Phi_{k+1} - \Phi_k \leq \frac{\rho^2(k+1)(k+2)}{2(2-\rho)}\|e_{k+1}\|^2 + \rho(k+1)\|R(w^k)\|\,\|e_{k+1}\|. \tag{12}$$

Moreover,

$$\Phi_k \geq \frac{\rho(k+1)^2 + (2-\rho)(k+1)}{2}\|R(w^k)\|^2 - (k+1)\|w^0 - w^*\|\|R(w^k)\|. \tag{13}$$

Using Lemma 3.1, we obtain the following general residual bound.

Proposition 3.2. *Suppose that Assumption 1 holds. Let $w^0 \in \mathcal{H}$, $w^* \in A^{-1}(0)$, and $\rho \in (0,2)$. Let $\{w^k\}$ be generated by* (2)*. Then, for every $k \geq 0$,*

$$\|R(w^k)\| \leq \frac{1}{\rho k+2}\left\{\|w^0-w^*\|+\rho S_k+\sqrt{\left(\|w^0-w^*\|+\rho S_k\right)^2+\frac{\rho^2(\rho k+2)}{(2-\rho)(k+1)}Q_k}\right\}, \tag{14}$$

where

$$S_k := \sum_{j=1}^{k}\|e_j\|, \qquad Q_k := \sum_{j=1}^{k} j(j+1)\|e_j\|^2, \qquad S_0 = Q_0 = 0.$$

If, in addition, the errors satisfy (4) *for some $c \geq 0$ and $p \in (0, 3/2)$, then, for every $k \geq 0$,*

$$\|R(w^k)\| \leq \frac{2\|w^0-w^*\|}{\rho k+2} + c\sqrt{\frac{3}{2(3-2p)}}\left(2+\sqrt{\frac{\rho}{2-\rho}}\right)(k+1)^{\frac{1}{2}-p}.$$

For $0 < p < 3/2$, Proposition 3.2 gives the desired residual rate. At the critical exponent $p = 3/2$, however, (14) yields only $\mathcal{O}(\sqrt{\log k}/k)$, since $Q_k = \mathcal{O}(\log k)$. To eliminate this logarithmic loss, we exploit the strict averagedness of T_ρ, and in particular the descent term in (6), to establish the following square-tail estimate.

Lemma 3.3 (Square-tail estimate for the residual discrepancy)**.** *Suppose that Assumption 1 holds. Let $w^0 \in \mathcal{H}$ and $\rho \in (0,2)$. Let $\{w^k\}$ be generated by* (2)*, and let $\{u^k\}$ be the exact iterates generated by* (7) *with $u^0 = w^0$. Suppose that the errors satisfy* (4) *with $p = 3/2$ for some $c \geq 0$. Then, for every $m \geq 0$,*

$$\sum_{k=m}^{\infty}\|R(w^k)-R(u^k)\|^2 \leq \frac{9\rho c^2}{2(2-\rho)(m+1)}. \tag{15}$$

Using the square-tail estimate, we refine the Lyapunov analysis to obtain the sharp $\mathcal{O}(1/k)$ rate at $p = 3/2$ while retaining the exact-iteration term from Proposition 2.2.

Proposition 3.4. *Suppose that Assumption 1 holds. Let $w^0 \in \mathcal{H}$, $w^* \in A^{-1}(0)$, and $\rho \in (0,2)$. Let $\{w^k\}$ be generated by* (2)*, and suppose that the errors satisfy* (4) *with $p = 3/2$ for some $c \geq 0$. Then, for every $k \geq 0$,*

$$\|R(w^k)\| \leq \frac{2\|w^0-w^*\|}{\rho k+2} + \frac{6c}{k+1}\left(1+\sqrt{\frac{2\rho}{2-\rho}}\right).$$

Proposition 3.2 covers $0 < p < 3/2$, while Proposition 3.4 establishes the critical case $p = 3/2$. For $p > 3/2$, the error condition (4) implies the corresponding condition with exponent $3/2$, so Proposition 3.4 applies directly. The following theorem summarizes the resulting residual bounds for all $p > 0$.

Theorem 3.5. *Suppose that Assumption 1 holds. Let $w^0 \in \mathcal{H}$, $w^* \in A^{-1}(0)$, and $\rho \in (0,2)$. Let $\{w^k\}$ be generated by* (2)*, and suppose that the errors satisfy* (4) *for some $c \geq 0$ and $p > 0$. Then, for every $k \geq 0$,*

$$\|R(w^k)\| \leq \frac{2\|w^0-w^*\|}{\rho k+2} + 6c\left(1+\sqrt{\frac{2\rho}{2-\rho}}\right)(k+1)^{-1+(\frac{3}{2}-p)_+},$$

where $(a)_+ := \max\{a, 0\}$ for $a \in \mathbb{R}$. When $c = 0$, this bound reduces to the exact residual estimate in Proposition 2.2. Consequently, for each fixed $\rho \in (0,2)$,

$$\|R(w^k)\| = \begin{cases} \mathcal{O}(k^{1/2-p}), & 0 < p < \frac{3}{2}, \\ \mathcal{O}(k^{-1}), & p \geq \frac{3}{2}. \end{cases}$$

3.2 RESIDUAL RATE FOR $\rho = 2$

We now consider the boundary case $\rho = 2$. The descent term in (6) vanishes, so the square-tail estimate used for $0 < \rho < 2$ is no longer available. Instead, the errors accumulate with increasing weights, requiring $p > 2$ to preserve the $\mathcal{O}(1/k)$ residual rate. The proof of the following theorem is given in Appendix A.3.

Theorem 3.6. *Suppose that Assumption 1 holds. Let $w^0 \in \mathcal{H}$, $w^* \in A^{-1}(0)$, and $\rho = 2$. Let $\{w^k\}$ be generated by* (2)*, and suppose that the errors satisfy* (4) *for some $c \geq 0$ and $p > 0$. Then, for every $k \geq 0$,*

$$\|R(w^k)\| \leq \frac{\|w^0 - w^*\| + 2\sum_{j=1}^{k} j\|e_j\|}{k+1} \leq \frac{\|w^0 - w^*\| + 2c\sum_{j=1}^{k} j^{1-p}}{k+1}. \tag{16}$$

Consequently,

$$\|R(w^k)\| = \begin{cases} \mathcal{O}(k^{1-p}), & 0 < p < 2, \\ \mathcal{O}((\log k)/k), & p = 2, \\ \mathcal{O}(k^{-1}), & p > 2. \end{cases}$$

Theorems 3.5 and 3.6 establish the residual rates for the interior regime $0 < \rho < 2$ and the boundary case $\rho = 2$, respectively. The next section shows that these rates are sharp up to constant factors.

4 SHARPNESS OF THE RESIDUAL RATES

We now show that the residual rates established in Section 3 are sharp in the worst-case sense for both $0 < \rho < 2$ and $\rho = 2$. For $\rho = 2$, we further show that the bound in (16) is attained and identify the optimal leading constants in the corresponding residual rates.

4.1 SHARPNESS ANALYSIS FOR $0 < \rho < 2$

For $0 < p < \frac{3}{2}$, we construct a two-dimensional rotation example in which the operator and error directions are chosen so that the approximation errors accumulate coherently, yielding the matching $\Omega(k^{1/2-p})$ lower bound.

For $p \geq \frac{3}{2}$, the matching $\Omega(1/k)$ lower bound follows directly from the worst-case lower bound for the exact method in Park & Ryu (2022, Corollary 4.2 with $\mu = 0$). Thus we only need to prove that the rate is sharp for $0 < p < 3/2$, which is stated in the following proposition. The construction of the example and the proof are given in Appendix A.4.

Proposition 4.1. *Let $\rho \in (0, 2)$, $p \in (0, 3/2)$, and set $w^0 = (1, 0)$, $w^* = (0, 0)$. Then there exists a constant $\xi > 0$, depending only on ρ and p, such that, for every sufficiently large k, there exist a maximally monotone operator $A : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ satisfying $0 \in A(w^*)$ and an error sequence satisfying* (4) *with some $c > 0$ for which*

$$\|R(w^k)\| \geq \xi c\, k^{1/2-p},$$

where $\{w^k\}$ is generated by (2)*. Consequently, the residual rate $\mathcal{O}(k^{1/2-p})$ is sharp in the worst-case sense.*

Theorem 3.5, Proposition 4.1, and the $\Omega(1/k)$ lower bound in Park & Ryu (2022, Corollary 4.2 with $\mu = 0$) establish the sharp residual orders for $0 < \rho < 2$ in Table 1.

4.2 SHARPNESS ANALYSIS FOR $\rho = 2$

For $\rho = 2$, we construct a one-dimensional example that attains the bound in (16) at every even iteration and yields the optimal leading constants in the corresponding residual rates. The construction and proof are given in Appendix A.5.

Proposition 4.2 (Attainment of the bound for $\rho = 2$)**.** *Let $\rho = 2$, $p > 0$, and set $w^0 > 0$, $w^* = 0$. Then there exist a maximally monotone operator $A : \mathbb{R} \rightrightarrows \mathbb{R}$ satisfying $0 \in A(w^*)$ and an error*

sequence satisfying (4) *with some* $c > 0$*, such that, for every even* k*, it holds*

$$(k+1)\|R(w^k)\| = \|w^0 - w^*\| + 2c\sum_{j=1}^{k} j^{1-p}, \tag{17}$$

where $\{w^k\}$ *is generated by* (2)*. Hence, the bound in* (16) *is attained. Moreover, as* $k \to \infty$ *along the even integers, it follows*

$$\|R(w^k)\| = \begin{cases} \left(\dfrac{2c}{2-p} + o(1)\right) k^{1-p}, & 0 < p < 2, \\ (2c + o(1))\,\dfrac{\log k}{k}, & p = 2, \\ \left(\|w^0 - w^*\| + 2c\zeta(p-1) + o(1)\right)\dfrac{1}{k}, & p > 2, \end{cases} \tag{18}$$

where $\zeta(x) = \sum_{j=1}^{\infty} j^{-x}$ *is the Riemann zeta function. Consequently, the residual rates in Theorem 3.6 are sharp, and the leading constants in* (18) *are optimal in the worst-case sense.*

5 Applications

We now specialize our framework to three examples in which evaluating the resolvent exactly requires solving a nontrivial inner problem. The key question is how accurately these inner problems need to be solved to retain the accelerated outer rate. Throughout this section, we focus on (2) with a fixed $\rho \in (0, 2)$, for which the critical accuracy requirement for retaining the $\mathcal{O}(1/k)$ residual rate is weaker than that when $\rho = 2$.

5.1 Inexact Halpern augmented Lagrangian method

Consider the constrained convex optimization problem

$$\min_{x \in \mathcal{X}} \{f_0(x) : h(x) \le 0\}, \tag{19}$$

where $h = (f_1, \ldots, f_m)$, each $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 0, 1, ..., m$ is closed and convex, and $\mathcal{X}$ is nonempty, closed, and convex. Assume that a saddle point (x^*, y^*) exists.

For $\sigma > 0$, define the augmented Lagrangian

$$L_\sigma(x, y) := f_0(x) + \frac{1}{2\sigma}\|[y + \sigma h(x)]_+\|^2 - \frac{1}{2\sigma}\|y\|^2, \qquad x \in \mathcal{X}, \tag{20}$$

where $[\cdot]_+$ denotes the projection onto $\mathbb{R}^m_+$. Assume that the subproblems below attain their minima.

Starting from $y^0 \in \mathbb{R}^m_+$, the inexact Halpern augmented Lagrangian method in Zhang et al. (2025a, equation (4.12)), specialized to $\rho = 1$, takes the form, for any $k \ge 0$,

$$\begin{aligned} \widetilde{x}^{k+1} &\approx \arg\min_{x \in \mathcal{X}} L_\sigma(x, y^k), \\ \widetilde{y}^{k+1} &= [y^k + \sigma h(\widetilde{x}^{k+1})]_+, \\ y^{k+1} &= \frac{1}{k+2} y^0 + \frac{k+1}{k+2}\widetilde{y}^{k+1}, \end{aligned} \tag{21}$$

where the approximate minimization satisfies for some $\varepsilon_{k+1} \ge 0$ that

$$L_\sigma(\widetilde{x}^{k+1}, y^k) - \inf_{x \in \mathcal{X}} L_\sigma(x, y^k) \le \frac{\varepsilon_{k+1}^2}{2\sigma}, \quad k \ge 0. \tag{22}$$

As a consequence of Proposition 3.4, we obtain the following result.

Corollary 5.1. *Under the above assumptions, suppose that*

$$L_\sigma(\widetilde{x}^{k+1}, y^k) - \inf_{x \in \mathcal{X}} L_\sigma(x, y^k) \le \frac{c^2}{2\sigma(k+1)^3}, \quad k \ge 0 \tag{23}$$

for some $c \ge 0$*. Then*

$$\|[h(\widetilde{x}^{k+1})]_+\| + |f_0(\widetilde{x}^{k+1}) - f_0(x^*)| = \mathcal{O}((k+1)^{-1}). \tag{24}$$

Comparison with existing results. For a fixed penalty parameter and an inner objective-gap tolerance of order $\mathcal{O}((k+1)^{-3})$, Corollary 5.1 establishes $\mathcal{O}(1/k)$ bounds on the constraint violation and objective error at the current primal iterate $\widetilde{x}^{k+1}$. Under the same tolerance order, Zhang et al. (2025a, Proposition 4.1) gives $\mathcal{O}((\log k)/k)$ bounds for averaged primal iterates. Thus, our analysis removes the logarithmic factor and avoids primal averaging. The proof is given in Appendix A.6.

5.2 Distributionally robust optimization

Consider the finite-dimensional saddle-point problem

$$\min_{x\in X}\max_{q\in Q}\ell(x,q), \tag{25}$$

where $X\subseteq\mathbb{R}^n$ and $Q\subseteq\mathbb{R}^m$ are nonempty closed convex sets, $\ell:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ is continuously differentiable, convex in x, concave in q, and a saddle point exists. Such formulations arise, for example, in Wasserstein distributionally robust optimization (Liang et al., 2025, Section 5.2).

Let $w=(x,q)$, $C=X\times Q$, and define

$$F(w):=\big(\nabla_x\ell(x,q),-\nabla_q\ell(x,q)\big),\qquad G(w):=\frac{1}{\alpha}\big(w-\Pi_C(w-\alpha F(w))\big), \tag{26}$$

where $\alpha>0$ and Π_C denotes the Euclidean projection onto C. Following Liang et al. (2025), assume additionally that $G:\mathbb{R}^{n+m}\to\mathbb{R}^{n+m}$ is $1/L$-cocoercive for some $L>0$.

Starting from $w^0\in C$, the deterministic inexact Halpern iteration in Liang et al. (2025) takes the form, for $k\geq 0$,

$$\begin{aligned}&\|\widetilde{G}^k-G(w^k)\|\leq\gamma_k,\\ &w^{k+1}=\frac{1}{k+2}w^0+\frac{k+1}{k+2}\left(w^k-\frac{1}{L}\widetilde{G}^k\right).\end{aligned} \tag{27}$$

Since $I-G/L$ is firmly nonexpansive, (27) can be viewed as an instance of (2) with $\rho=1$. As a consequence of Proposition 3.4, we obtain the following result.

Corollary 5.2. *Under the above assumptions, suppose that*

$$\gamma_k\leq c(k+1)^{-3/2},\qquad k\geq 0,$$

for some $c\geq 0$. Then

$$\|G(w^k)\|=\mathcal{O}((k+1)^{-1}). \tag{28}$$

Comparison with existing results. Under the same accuracy $\gamma_k=\mathcal{O}((k+1)^{-3/2})$, Liang et al. (2025, Theorem 3.1) obtains a residual rate of $\mathcal{O}(\sqrt{\log k}/k)$, whereas Corollary 5.2 gives $\mathcal{O}(1/k)$. The proof is given in Appendix A.7.

5.3 Finite-sum monotone inclusions

Consider the finite-sum monotone inclusion

$$0\in F(w)+G(w),\qquad F=\frac{1}{n}\sum_{i=1}^{n}F_i, \tag{29}$$

where $F:\mathbb{R}^d\to\mathbb{R}^d$ is single valued, monotone and Lipschitz continuous, G is maximally monotone, and the solution set is nonempty. Cai et al. (2024, Section 4) study this problem using inexact Halpern iteration with variance-reduced resolvent evaluations. We consider the same outer iteration and weaken the relative mean-square accuracy requirement while preserving the accelerated fixed-point residual rate. Fix $\eta>0$ and set $A:=\eta(F+G)$.

Given w^0, the iteration with $\rho=1$ takes the form

$$\begin{aligned}&\widetilde{w}^{k+1}\approx J_A(w^k),\\ &w^{k+1}=\frac{1}{k+2}w^0+\frac{k+1}{k+2}\widetilde{w}^{k+1},\quad k\geq 0.\end{aligned} \tag{30}$$

Let $\mathcal{F}_k$ be the σ-algebra generated by the first k iterations, so that w^k is $\mathcal{F}_k$-measurable. Denote the conditional expectation by $\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot \mid \mathcal{F}_k]$. Define the residual $R(w) = w - J_A(w)$ and the resolvent error $e_{k+1} := \widetilde{w}^{k+1} - J_A(w^k)$, which is equivalent to that defined in (Cai et al., 2024).

Taking conditional expectations in the Lyapunov inequality of Lemma 3.1 yields the following result.

Corollary 5.3. *Under the above assumptions, suppose that the errors satisfy, almost surely,*

$$\mathbb{E}_k\|e_{k+1}\|^2 \le \frac{\|R(w^k)\|^2}{(k+2)^2}, \qquad k \ge 0. \tag{31}$$

Then the iterates generated by (30) *satisfy*

$$\mathbb{E}\|R(w^k)\| \le \left(\mathbb{E}\|R(w^k)\|^2\right)^{1/2} = \mathcal{O}((k+1)^{-1}). \tag{32}$$

Comparison with existing results. In our notation, Cai et al. (2024, Lemmas C.2–C.3) assume

$$\mathbb{E}_k\|e_{k+1}\|^2 \le \frac{\|R(w^k)\|^2}{(k+2)^8}, \qquad k \ge 1,$$

with a separate initial-error condition. Corollary 5.3 reduces the denominator exponent from 8 to 2 while preserving the $\mathcal{O}(1/k)$ expected fixed-point residual rate. This relaxes the sufficient inner accuracy requirement; total oracle complexity still requires further analysis of the inner solver. The proof is given in Appendix A.8.

6 Conclusion

We studied an implementable Halpern-accelerated proximal point method and provided a sharp characterization of how computational errors affect the residual convergence rate. In particular, for every fixed $0 < \rho < 2$, we showed that the $\mathcal{O}(1/k)$ residual rate is preserved at the critical exponent $p = 3/2$, removing the $\sqrt{\log k}$ loss in previous analyses. Together with matching lower bounds, our results establish sharp residual rates across all polynomial error-decay regimes for $\rho \in (0, 2]$. We further applied our theory to augmented Lagrangian methods, distributionally robust optimization, and finite-sum monotone inclusions, improving existing convergence guarantees in these settings.

AI Use Statement

Generative AI tools were used to assist with language editing, improving the clarity and readability of the manuscript, checking minor notational inconsistencies, and discussing proof strategies. In particular, an AI-assisted discussion provided an insight that contributed to the square-tail estimate in Lemma 3.3. All AI-assisted content and reasoning were reviewed and verified by the authors, who take full responsibility for the final content of the paper.

References

Ahmet Alacaoglu, Donghwan Kim, and Stephen J Wright. Revisiting inexact fixed-point iterations for min-max problems: Stochasticity and structured nonconvexity. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 840–878. PMLR, 2024.

Mario Bravo and Juan Pablo Contreras. Stochastic Halpern iteration in normed spaces and applications to reinforcement learning. *Mathematical Programming*, pp. 1–39, 2026. doi: 10.1007/s10107-026-02345-1.

H. Brézis and P.-L. Lions. Produits infinis de résolvantes. *Israel Journal of Mathematics*, 29(4): 329–345, 1978. doi: 10.1007/BF02761171.

Xufeng Cai, Chaobing Song, Cristóbal Guzmán, and Jelena Diakonikolas. Stochastic Halpern iteration with variance reduction for stochastic monotone inclusions. *Advances in Neural Information Processing Systems*, 35:24766–24779, 2022.

Xufeng Cai, Ahmet Alacaoglu, and Jelena Diakonikolas. Variance reduced Halpern iteration for finite-sum monotone inclusions. In *The Twelfth International Conference on Learning Representations*, 2024.

Juan Pablo Contreras and Roberto Cominetti. Optimal error bounds for non-expansive fixed-point iterations in normed spaces. *Mathematical Programming*, 199(1):343–374, 2023.

Jelena Diakonikolas. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. In *Conference on Learning Theory*, pp. 1428–1451. PMLR, 2020.

Guoyong Gu and Junfeng Yang. Tight sublinear convergence rate of the proximal point algorithm for maximal monotone inclusion problems. *SIAM Journal on Optimization*, 30(3):1905–1921, 2020. doi: 10.1137/19M1299049.

Benjamin Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73(6):957–961, 1967. doi: 10.1090/S0002-9904-1967-11864-0.

Donghwan Kim. Accelerated proximal point method for maximally monotone operators. *Mathematical Programming*, 190(1–2):57–87, 2021. doi: 10.1007/s10107-021-01643-0.

Ling Liang, Zusen Xu, Kim-Chuan Toh, and Jia-Jie Zhu. An inexact Halpern iteration with application to distributionally robust optimization. *Journal of Optimization Theory and Applications*, 206(3):58, 2025. doi: 10.1007/s10957-025-02744-y.

Felix Lieder. On the convergence rate of the Halpern-iteration. *Optimization Letters*, 15:405–418, 2021. doi: 10.1007/s11590-020-01617-9.

Ronak Mehta, Vincent Roulet, Krishna Pillutla, and Zaid Harchaoui. Distributionally robust optimization with bias and variance reduction. In *International Conference on Learning Representations*, volume 2024, pp. 43672–43737, 2024.

Jisun Park and Ernest K. Ryu. Exact optimal accelerated complexity for fixed-point iterations. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 17420–17457. PMLR, 2022.

Nicholas Pischke and Thomas Powell. Asymptotic regularity of a generalised stochastic Halpern scheme. *Journal of Optimization Theory and Applications*, 210:3, 2026. doi: 10.1007/s10957-026-03032-z.

Juan Ramirez and Simon Lacoste-Julien. Dual optimistic ascent (pi control) is the augmented lagrangian method in disguise. In *International Conference on Learning Representations*, volume 2026, pp. 122017–122051, 2026.

R. Tyrrell Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1(2):97–116, 1976a. doi: 10.1287/moor.1.2.97.

R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976b. doi: 10.1137/0314056.

Ernest K. Ryu and Wotao Yin. *Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators*. Cambridge University Press, 2022. doi: 10.1017/9781009160865. URL `https://large-scale-book.mathopt.com/`.

Liwei Zhang, Fanli Zhuang, and Ning Zhang. Halpern acceleration of the inexact proximal point method of Rockafellar. *arXiv preprint arXiv:2511.10372*, 2025a.

Qi Zhang, Yi Zhou, Simon Khan, Ashley Prater-Bennette, Lixin Shen, and Shaofeng Zou. Revisiting large-scale non-convex distributionally robust optimization. In *International Conference on Learning Representations*, volume 2025, pp. 76275–76301, 2025b.

A PROOFS

A.1 PROOFS FOR SECTION 2

Proof of Lemma 2.1. Subtracting the exact iteration (7) from the implementable one (2) gives

$$\Delta_{k+1} = \frac{k+1}{k+2}\left(T_\rho w^k - T_\rho u^k + \rho e_{k+1}\right). \tag{33}$$

Hence, nonexpansiveness of T_ρ gives

$$(k+2)\|\Delta_{k+1}\| \le (k+1)\|\Delta_k\| + \rho(k+1)\|e_{k+1}\|.$$

Since $\Delta_0 = 0$, telescoping the above inequality yields

$$(k+1)\|\Delta_k\| \le \rho\sum_{j=1}^{k} j\|e_j\|,$$

which proves the first claim. The second follows from

$$\sum_{j=1}^{k} j\|e_j\| \le c\sum_{j=1}^{k} j^{-1/2} \le 2c\sqrt{k}.$$

Since $\Delta_0 = 0$, the second estimate also holds for $k = 0$. □

Proof of Proposition 2.2. Define

$$\mathcal{L}_k := (k+1)\langle R(u^k), u^k - w^0\rangle + \frac{\rho k(k+1)}{2}\|R(u^k)\|^2. \tag{34}$$

From (7), we have

$$(k+2)(u^{k+1} - w^0) = (k+1)(u^k - w^0) - \rho(k+1)R(u^k),$$
$$u^k - w^0 = -(k+2)(u^{k+1} - u^k) - \rho(k+1)R(u^k).$$

Following a similar analysis of Halpern iteration in Diakonikolas (2020), by substituting the above two equations into $\mathcal{L}_{k+1} - \mathcal{L}_k$, and using

$$2\langle R(u^{k+1}) - R(u^k), R(u^k)\rangle = \|R(u^{k+1})\|^2 - \|R(u^k)\|^2 - \|R(u^{k+1}) - R(u^k)\|^2,$$

we obtain

$$\mathcal{L}_{k+1} - \mathcal{L}_k = (k+1)(k+2)\Big[-\langle R(u^{k+1}) - R(u^k), u^{k+1} - u^k\rangle + \frac{\rho}{2}\|R(u^{k+1}) - R(u^k)\|^2\Big].$$

Since R is firmly nonexpansive: $\langle R(u^{k+1}) - R(u^k), u^{k+1} - u^k\rangle \ge \|R(u^{k+1}) - R(u^k)\|^2$, it follows

$$\mathcal{L}_{k+1} - \mathcal{L}_k \le -\frac{(2-\rho)(k+1)(k+2)}{2}\|R(u^{k+1}) - R(u^k)\|^2 \le 0. \tag{35}$$

Together with $\mathcal{L}_0 = 0$, we have $\mathcal{L}_k \le 0$.

It remains to derive the lower bound of $\mathcal{L}_k$. By the resolvent identity $R(u^k) \in A(u^k - R(u^k))$, and $0 \in A(w^*)$, monotonicity of A gives

$$\langle R(u^k), u^k - R(u^k) - w^*\rangle \ge 0,$$

and hence

$$\langle R(u^k), u^k - w^0\rangle = \langle R(u^k), u^k - R(u^k) - w^*\rangle + \|R(u^k)\|^2 + \langle R(u^k), w^* - w^0\rangle$$
$$\ge \|R(u^k)\|^2 - \|w^0 - w^*\|\|R(u^k)\|.$$

Substituting this bound into the definition of $\mathcal{L}_k$ yields

$$0 \ge \mathcal{L}_k \ge \frac{(k+1)(\rho k+2)}{2}\|R(u^k)\|^2 - (k+1)\|w^0 - w^*\|\|R(u^k)\|.$$

If $R(u^k) = 0$, the conclusion is immediate. Otherwise, dividing by $(k+1)\|R(u^k)\|$ gives

$$\frac{\rho k+2}{2}\|R(u^k)\| \le \|w^0 - w^*\|,$$

which is exactly (10). □

A.2 Proofs for Subsection 3.1

Proof of Lemma 3.1. From (2), we have

$$(k+2)(w^{k+1}-w^0) = (k+1)(w^k-w^0) - \rho(k+1)R(w^k) + \rho(k+1)e_{k+1}. \tag{36}$$

Subtracting $(k+2)(w^k-w^0)$ from both sides yields

$$w^k - w^0 = -(k+2)(w^{k+1}-w^k) - \rho(k+1)R(w^k) + \rho(k+1)e_{k+1}. \tag{37}$$

By the construction of the Lyapunov function (11) and (36), we have

$$\begin{aligned}\Phi_{k+1}-\Phi_k = {}& (k+1)\left\langle R(w^{k+1})-R(w^k), w^k-w^0\right\rangle \\ & - \rho(k+1)\left\langle R(w^{k+1}), R(w^k)\right\rangle + \rho(k+1)\left\langle R(w^{k+1}), e_{k+1}\right\rangle \\ & + \frac{\rho(k+1)(k+2)}{2}\|R(w^{k+1})\|^2 - \frac{\rho k(k+1)}{2}\|R(w^k)\|^2.\end{aligned} \tag{38}$$

We begin by estimating the first inner-product term in (38). Using (37), we obtain

$$\begin{aligned}& (k+1)\left\langle R(w^{k+1})-R(w^k), w^k-w^0\right\rangle \\ = {}& -(k+1)(k+2)\left\langle R(w^{k+1})-R(w^k), w^{k+1}-w^k\right\rangle \\ & - \rho(k+1)^2\left\langle R(w^{k+1})-R(w^k), R(w^k)\right\rangle \\ & + \rho(k+1)^2\left\langle R(w^{k+1})-R(w^k), e_{k+1}\right\rangle \\ \le {}& -(k+1)(k+2)\|R(w^{k+1})-R(w^k)\|^2 \\ & - \rho(k+1)^2\left\langle R(w^{k+1})-R(w^k), R(w^k)\right\rangle \\ & + \rho(k+1)^2\left\langle R(w^{k+1})-R(w^k), e_{k+1}\right\rangle,\end{aligned} \tag{39}$$

where the last inequality follows from the firm nonexpansiveness of R: $\langle R(w^{k+1})-R(w^k), w^{k+1}-w^k\rangle \ge \|R(w^{k+1})-R(w^k)\|^2$.

Note that the following two identities hold:

$$\begin{aligned}& -\rho(k+1)^2\left\langle R(w^{k+1})-R(w^k), R(w^k)\right\rangle - \rho(k+1)\left\langle R(w^{k+1}), R(w^k)\right\rangle \\ & \quad + \frac{\rho(k+1)(k+2)}{2}\|R(w^{k+1})\|^2 - \frac{\rho k(k+1)}{2}\|R(w^k)\|^2 \\ & = \frac{\rho(k+1)(k+2)}{2}\|R(w^{k+1})-R(w^k)\|^2,\end{aligned}$$

and

$$\begin{aligned}& \rho(k+1)^2\left\langle R(w^{k+1})-R(w^k), e_{k+1}\right\rangle + \rho(k+1)\left\langle R(w^{k+1}), e_{k+1}\right\rangle \\ & \quad = \rho(k+1)\left\langle (k+2)R(w^{k+1}) - (k+1)R(w^k), e_{k+1}\right\rangle.\end{aligned}$$

Substituting (39) into (38), and using the above two identities gives

$$\begin{aligned}\Phi_{k+1}-\Phi_k \le {}& -\frac{(2-\rho)(k+1)(k+2)}{2}\|R(w^{k+1})-R(w^k)\|^2 \\ & + \rho(k+1)\left\langle (k+2)R(w^{k+1}) - (k+1)R(w^k), e_{k+1}\right\rangle \\ = {}& -\frac{(2-\rho)(k+1)(k+2)}{2}\|R(w^{k+1})-R(w^k)\|^2 \\ & + \rho(k+1)(k+2)\left\langle R(w^{k+1})-R(w^k), e_{k+1}\right\rangle + \rho(k+1)\left\langle R(w^k), e_{k+1}\right\rangle.\end{aligned} \tag{40}$$

Since $0<\rho<2$, Young's inequality then gives

$$\begin{aligned}& \rho(k+1)(k+2)\langle R(w^{k+1})-R(w^k), e_{k+1}\rangle \\ & \quad \le \frac{(2-\rho)(k+1)(k+2)}{2}\|R(w^{k+1})-R(w^k)\|^2 + \frac{\rho^2(k+1)(k+2)}{2(2-\rho)}\|e_{k+1}\|^2.\end{aligned}$$

Applying this to (40) proves (12).

For the lower bound, using the firm nonexpansiveness of R and $0 = R(w^*)$ gives $\langle R(w^k), w^k - w^*\rangle \geq \|R(w^k)\|^2$. Consequently,

$$\langle R(w^k), w^k - w^0\rangle \geq \|R(w^k)\|^2 - \|w^0 - w^*\|\|R(w^k)\|.$$

Substitution into (11) proves the lower bound (13). □

Proof of Proposition 3.2. Let

$$M_k := \max_{0\leq \ell \leq k} (\ell+1)\|R(w^\ell)\|,$$

and choose $r \in \{0, \ldots, k\}$ at which this maximum is attained. By (13), we have

$$\begin{aligned}\Phi_r &\geq \frac{\rho r + 2}{2(r+1)} M_k^2 - \|w^0 - w^*\| M_k \\ &\geq \frac{\rho k + 2}{2(k+1)} M_k^2 - \|w^0 - w^*\| M_k,\end{aligned} \tag{41}$$

where the second inequality is due to $r \leq k$ and $0 < \rho < 2$. Telescoping (12) from 0 to $r-1$ (with $r \geq 1$) and using $\Phi_0 = 0$ gives

$$\Phi_r \leq \frac{\rho^2}{2(2-\rho)} \sum_{j=1}^{r} j(j+1)\|e_j\|^2 + \rho \sum_{j=1}^{r} j\|R(w^{j-1})\|\,\|e_j\|.$$

When $r = 0$, we set $\sum_{j=1}^{0}(\cdot) = 0$, and the above inequality still holds due to the definition of Lyapunov function (11). Since $j\|R(w^{j-1})\| \leq M_k$ for $1 \leq j \leq r \leq k$,

$$\Phi_r \leq \rho M_k S_k + \frac{\rho^2}{2(2-\rho)} Q_k. \tag{42}$$

Combining (41) and (42) yields

$$\frac{\rho k + 2}{2(k+1)} M_k^2 \leq \big(\|w^0 - w^*\| + \rho S_k\big) M_k + \frac{\rho^2}{2(2-\rho)} Q_k.$$

Solving this quadratic inequality and using $(k+1)\|R(w^k)\| \leq M_k$ proves (14).

We next estimate the bound for S_k and Q_k. By Cauchy–Schwarz inequality, it follows

$$\begin{aligned}S_k &= \sum_{j=1}^{k} \frac{1}{\sqrt{j(j+1)}} \big(\sqrt{j(j+1)}\|e_j\|\big) \\ &\leq \left(\sum_{j=1}^{k} \frac{1}{j(j+1)}\right)^{1/2} Q_k^{1/2} \\ &= \sqrt{\frac{k}{k+1}}\, Q_k^{1/2} \leq Q_k^{1/2}.\end{aligned} \tag{43}$$

Suppose now that $\|e_j\| \leq cj^{-p}$ with $0 < p < 3/2$, and set $q := 3 - 2p \in (0, 3)$. For $k \geq 1$, the inequality $j + 1 \leq 3j/2$ for $j \geq 2$ gives

$$\begin{aligned}Q_k &\leq c^2 \sum_{j=1}^{k} \big(j^{q-1} + j^{q-2}\big) = c^2 \left(2 + \sum_{j=2}^{k} j^{q-2}(j+1)\right) \\ &\leq c^2 \left(2 + \frac{3}{2} \sum_{j=2}^{k} j^{q-1}\right).\end{aligned} \tag{44}$$

If $0 < q < 1$, t^{q-1} is a convex function, Jensen's inequality gives

$$j^{q-1} = \left(\int_{j-1/2}^{j+1/2} t\, dt\right)^{q-1} \leq \int_{j-1/2}^{j+1/2} t^{q-1}\, dt.$$

Then,

$$\sum_{j=2}^{k} j^{q-1} \le \sum_{j=2}^{k} \int_{j-1/2}^{j+1/2} t^{q-1}\,dt = \int_{3/2}^{k+1/2} t^{q-1}\,dt = \frac{(k+\frac{1}{2})^q - (\frac{3}{2})^q}{q}.$$

Consequently, we have

$$\begin{aligned} 2 + \frac{3}{2}\sum_{j=2}^{k} j^{q-1} &\le 2 + \frac{3}{2q}\left((k+\tfrac{1}{2})^q - (\tfrac{3}{2})^q\right) \\ &= \frac{3}{2q}\left(\frac{4q}{3} + (k+\tfrac{1}{2})^q - (\tfrac{3}{2})^q\right). \end{aligned}$$

A direct calculation shows that $\frac{4q}{3} \le \left(\frac{3}{2}\right)^q$ for $0 < q < 1$. It then follows that

$$2 + \frac{3}{2}\sum_{j=2}^{k} j^{q-1} \le \frac{3}{2q}\left(k+\tfrac{1}{2}\right)^q \le \frac{3}{2q}(k+1)^q.$$

If $1 \le q < 3$, since t^{q-1} is nondecreasing, we have

$$\sum_{j=2}^{k} j^{q-1} \le \int_{2}^{k+1} t^{q-1}\,dt = \frac{(k+1)^q - 2^q}{q}.$$

Hence,

$$2 + \frac{3}{2}\sum_{j=2}^{k} j^{q-1} \le 2 + \frac{3}{2q}\big((k+1)^q - 2^q\big) \le \frac{3}{2q}(k+1)^q,$$

where the last inequality follows from $2^q \ge 1 + q \ge 4q/3$ for $1 \le q < 3$.

Combining the above cases with (44) and substituting back $q = 3 - 2p$, we have

$$Q_k \le \frac{3c^2}{2(3-2p)}(k+1)^{3-2p}, \quad k \ge 0. \tag{45}$$

Applying $\sqrt{a^2+b^2} \le a + b$ to the square-root term in (14), we obtain

$$\|R(w^k)\| \le \frac{2\|w^0 - w^*\|}{\rho k + 2} + \frac{2\rho S_k}{\rho k + 2} + \frac{\rho}{\rho k + 2}\sqrt{\frac{\rho k + 2}{(2-\rho)(k+1)}Q_k}.$$

Since $0 < \rho < 2$, we have

$$\frac{\rho}{\rho k + 2} \le \frac{1}{k+1}.$$

Moreover,

$$\begin{aligned} &\frac{\rho}{\rho k + 2}\sqrt{\frac{\rho k + 2}{(2-\rho)(k+1)}Q_k} \\ &\quad= \sqrt{\frac{\rho}{2-\rho}}\sqrt{\frac{\rho(k+1)}{\rho k + 2}}\,\frac{\sqrt{Q_k}}{k+1} \\ &\quad\le \sqrt{\frac{\rho}{2-\rho}}\,\frac{\sqrt{Q_k}}{k+1}. \end{aligned}$$

Thus we further have

$$\begin{aligned} \|R(w^k)\| &\le \frac{2\|w^0 - w^*\|}{\rho k + 2} + \frac{2S_k}{k+1} + \sqrt{\frac{\rho}{2-\rho}}\,\frac{\sqrt{Q_k}}{k+1} \\ &\le \frac{2\|w^0 - w^*\|}{\rho k + 2} + \left(2 + \sqrt{\frac{\rho}{2-\rho}}\right)\frac{\sqrt{Q_k}}{k+1}, \end{aligned}$$

where the last inequality is due to $S_k \le \sqrt{Q_k}$ (43). Combining this estimate with (45) yields

$$\|R(w^k)\| \le \frac{2\|w^0 - w^*\|}{\rho k + 2} + c\sqrt{\frac{3}{2(3-2p)}}\left(2 + \sqrt{\frac{\rho}{2-\rho}}\right)(k+1)^{\frac{1}{2}-p},$$

which proves the second bound in Proposition 3.2.

□

Proof of Lemma 3.3. Let $\{u^k\}$ be the exact iterates from (7), and recall $\Delta_k = w^k - u^k$. From (33), we have

$$\begin{aligned}\left(\frac{k+2}{k+1}\right)^2 \|\Delta_{k+1}\|^2 &= \|T_\rho w^k - T_\rho u^k + \rho e_{k+1}\|^2 \\ &= \|T_\rho w^k - T_\rho u^k\|^2 + 2\rho\left\langle T_\rho w^k - T_\rho u^k, e_{k+1}\right\rangle + \rho^2\|e_{k+1}\|^2 \\ &\le \|\Delta_k\|^2 - \rho(2-\rho)\|R(w^k) - R(u^k)\|^2 \\ &\quad + 2\rho\|\Delta_k\|\,\|e_{k+1}\| + \rho^2\|e_{k+1}\|^2,\end{aligned}$$

where the last inequality follows from (6). Rearranging the terms yields

$$\begin{aligned}&\rho(2-\rho)\|R(w^k) - R(u^k)\|^2 \\ &\quad\le \|\Delta_k\|^2 - \left(\frac{k+2}{k+1}\right)^2 \|\Delta_{k+1}\|^2 + 2\rho\|\Delta_k\|\,\|e_{k+1}\| + \rho^2\|e_{k+1}\|^2.\end{aligned}$$

Fix integers $N \ge m \ge 0$. Summing from $k = m$ to N, using $\|e_{k+1}\| \le c(k+1)^{-3/2}$, and dropping the nonpositive terminal term yields

$$\begin{aligned}&\rho(2-\rho)\sum_{k=m}^{N}\|R(w^k) - R(u^k)\|^2 \\ &\quad\le \left(\|\Delta_m\| + \frac{\rho c}{(m+1)^{3/2}}\right)^2 \\ &\qquad + \sum_{k=m+1}^{N}\left[-\frac{2k+1}{k^2}\|\Delta_k\|^2 + \frac{2\rho c}{(k+1)^{3/2}}\|\Delta_k\| + \frac{\rho^2 c^2}{(k+1)^3}\right].\end{aligned}$$

Note that the summand is a concave quadratic function of $\|\Delta_k\|$; maximizing it yields

$$\begin{aligned}&-\frac{2k+1}{k^2}\|\Delta_k\|^2 + \frac{2\rho c}{(k+1)^{3/2}}\|\Delta_k\| + \frac{\rho^2 c^2}{(k+1)^3} \\ &\quad\le \frac{\rho^2 c^2 k^2}{(2k+1)(k+1)^3} + \frac{\rho^2 c^2}{(k+1)^3} = \frac{\rho^2 c^2}{(2k+1)(k+1)}.\end{aligned}$$

By Lemma 2.1 and $\sum_{j=1}^{m+1} j^{-1/2} \le \int_0^{m+1} t^{-1/2}\,dt = 2\sqrt{m+1}$, we have

$$\begin{aligned}\|\Delta_m\| + \frac{\rho c}{(m+1)^{3/2}} &\le \frac{\rho c}{m+1}\sum_{j=1}^{m} j^{-1/2} + \frac{\rho c}{m+1}(m+1)^{-1/2} \\ &= \frac{\rho c}{m+1}\sum_{j=1}^{m+1} j^{-1/2} \\ &\le \frac{2\rho c}{\sqrt{m+1}}.\end{aligned}$$

Moreover,

$$\sum_{k=m+1}^{\infty}\frac{1}{(2k+1)(k+1)} \le \frac{1}{2}\sum_{k=m+1}^{\infty}\frac{1}{k(k+1)} = \frac{1}{2(m+1)}.$$

Combining the preceding estimates, for every $N \geq m$ we obtain

$$\begin{aligned}
&\rho(2-\rho)\sum_{k=m}^{N}\|R(w^k)-R(u^k)\|^2 \\
&\leq \frac{4\rho^2c^2}{m+1}+\rho^2c^2\sum_{k=m+1}^{N}\frac{1}{(2k+1)(k+1)} \\
&\leq \frac{4\rho^2c^2}{m+1}+\frac{\rho^2c^2}{2(m+1)}=\frac{9\rho^2c^2}{2(m+1)}.
\end{aligned}$$

Therefore,

$$\sum_{k=m}^{N}\|R(w^k)-R(u^k)\|^2 \leq \frac{9\rho c^2}{2(2-\rho)(m+1)}, \quad \forall N \geq m.$$

Passing to the limit $N \to \infty$ proves

$$\sum_{k=m}^{\infty}\|R(w^k)-R(u^k)\|^2 \leq \frac{9\rho c^2}{2(2-\rho)(m+1)}. \tag{46}$$

□

Proof of Proposition 3.4. Let $\{u^i\}$ be the exact iterates from (7), and set

$$d_i := R(w^i)-R(u^i), \qquad a_\rho := 3\sqrt{\frac{2\rho}{2-\rho}}. \tag{47}$$

If $k=0$, the conclusion holds trivially since $\|R(w^0)\| \leq \|w^0-w^*\|$.

Now let $k \geq 1$ and set $m=\lfloor k/2 \rfloor$. By Lemma 3.3,

$$\sum_{j=m}^{k-1}\|d_j\|^2 \leq \frac{9\rho c^2}{2(2-\rho)(m+1)}.$$

Hence there exists $s \in \{m,\dots,k-1\}$ such that

$$\|d_s\|^2 \leq \frac{9\rho c^2}{2(2-\rho)(m+1)(k-m)}.$$

Since $s+1 \leq k$ and $k^2 \leq 4(m+1)(k-m)$ for $m=\lfloor k/2 \rfloor$, we obtain

$$(s+1)^2\|d_s\|^2 \leq \frac{9\rho c^2k^2}{2(2-\rho)k^2/4}=\frac{18\rho c^2}{2-\rho}.$$

From the definition of a_ρ (47), we therefore have

$$(s+1)\|d_s\| \leq a_\rho c. \tag{48}$$

We are now ready to compare Φ_s with the exact-iteration Lyapunov function (34),

$$\mathcal{L}_s=(s+1)\langle R(u^s),u^s-w^0\rangle+\frac{\rho s(s+1)}{2}\|R(u^s)\|^2.$$

Recall that (35) in the proof of Proposition 2.2 shows that $\mathcal{L}_s \leq 0$. Moreover, nonexpansiveness of T_ρ and $T_\rho w^*=w^*$ imply $\|u^s-w^*\| \leq \|w^0-w^*\|$. Consequently,

$$\begin{aligned}
\|u^s-w^0+\rho sR(u^s)\|^2 &= \|u^s-w^0\|^2+\frac{2\rho s}{s+1}\mathcal{L}_s \\
&\leq 2\left(\|u^s-w^*\|^2+\|w^0-w^*\|^2\right) \\
&\leq 4\|w^0-w^*\|^2.
\end{aligned} \tag{49}$$

Expanding $\Phi_s - \mathcal{L}_s$ and recalling $\Delta_s = w^s - u^s$ yields

$$\begin{aligned}\Phi_s - \mathcal{L}_s &= (s+1)\langle d_s, u^s - w^0 + \rho s R(u^s)\rangle \\ &\quad + (s+1)\langle R(u^s) + d_s, \Delta_s\rangle + \frac{\rho s(s+1)}{2}\|d_s\|^2 \\ &\le (s+1)\|d_s\|\|u^s - w^0 + \rho s R(u^s)\| \\ &\quad + (s+1)\big(\|R(u^s)\| + \|d_s\|\big)\|\Delta_s\| + \frac{\rho s(s+1)}{2}\|d_s\|^2.\end{aligned} \tag{50}$$

By Proposition 2.2 and (9),

$$(s+1)\|R(u^s)\| \le \frac{2\|w^0 - w^*\|}{\rho}, \qquad \|\Delta_s\| \le \frac{2\rho c}{\sqrt{s+1}} \le 2\rho c.$$

Combining these estimates with (48), (49), and (50) gives

$$\begin{aligned}\Phi_s - \mathcal{L}_s \le 2a_\rho c\|w^0 - w^*\| + 4c\|w^0 - w^*\| + 2\rho a_\rho c^2 \\ + \frac{\rho s(s+1)}{2}\frac{a_\rho^2 c^2}{(s+1)^2}.\end{aligned}$$

Since $s/(s+1) \le 1$ and $\mathcal{L}_s \le 0$, we obtain

$$\Phi_s \le (2a_\rho + 4)c\|w^0 - w^*\| + \left(2\rho a_\rho + \frac{\rho a_\rho^2}{2}\right)c^2. \tag{51}$$

We now telescope (12) from s to $k-1$:

$$\Phi_k \le \Phi_s + \frac{\rho^2}{2(2-\rho)}\sum_{j=s+1}^{k} j(j+1)\|e_j\|^2 + \rho\sum_{j=s+1}^{k} j\|R(w^{j-1})\|\,\|e_j\|. \tag{52}$$

Since $s \ge m$, we have $1/j \le 1/(m+1)$ for $j = s+1, \dots, k$. Moreover, since $m = \lfloor k/2 \rfloor$, it holds $k - m \le m + 1$. Therefore,

$$\sum_{j=s+1}^{k} j^{-\frac{3}{2}} \le \sum_{j=s+1}^{k}\frac{1}{j} \le \frac{k-s}{m+1} \le \frac{k-m}{m+1} \le 1. \tag{53}$$

Using $\|e_j\| \le cj^{-3/2}$, we have

$$\begin{aligned}\frac{\rho^2}{2(2-\rho)}\sum_{j=s+1}^{k} j(j+1)\|e_j\|^2 &\le \frac{\rho^2 c^2}{2(2-\rho)}\sum_{j=s+1}^{k}\left(\frac{1}{j} + \frac{1}{j^2}\right) \\ &\le \frac{\rho^2 c^2}{2(2-\rho)}\sum_{j=s+1}^{k}\frac{2}{j} \\ &\le \frac{\rho^2 c^2}{2-\rho}.\end{aligned} \tag{54}$$

For the last term in (52), using $\|e_j\| \le cj^{-3/2}$ and $R(w^{j-1}) = R(u^{j-1}) + d_{j-1}$, we have

$$\begin{aligned}&\rho\sum_{j=s+1}^{k} j\|R(w^{j-1})\|\,\|e_j\| \\ &\quad\le \rho c\sum_{j=s+1}^{k}\frac{\|R(u^{j-1})\| + \|d_{j-1}\|}{\sqrt{j}}.\end{aligned} \tag{55}$$

By Proposition 2.2 and $0 < \rho < 2$,

$$\|R(u^{j-1})\| \le \frac{2\|w^0 - w^*\|}{\rho(j-1)+2} \le \frac{2\|w^0 - w^*\|}{\rho j}.$$

Therefore,

$$\begin{aligned}
&\rho c \sum_{j=s+1}^{k} \frac{\|R(u^{j-1})\| + \|d_{j-1}\|}{\sqrt{j}} \\
&\leq 2c\|w^0 - w^*\| \sum_{j=s+1}^{k} j^{-3/2} + \rho c \sum_{j=s+1}^{k} \frac{\|d_{j-1}\|}{\sqrt{j}} \\
&\leq 2c\|w^0 - w^*\| \sum_{j=s+1}^{k} j^{-3/2} + \rho c \left(\sum_{j=s+1}^{k} \|d_{j-1}\|^2 \right)^{1/2} \left(\sum_{j=s+1}^{k} \frac{1}{j} \right)^{1/2} \\
&\leq 2c\|w^0 - w^*\| \sum_{j=s+1}^{k} j^{-3/2} + \rho c \left(\sum_{j=s+1}^{k} \|d_{j-1}\|^2 \right)^{1/2}, \qquad (56)
\end{aligned}$$

where the second inequality follows from the Cauchy–Schwarz inequality, and the last inequality follows from (53).

Using Lemma 3.3 with $m = s$ yields

$$\begin{aligned}
\sum_{j=s+1}^{k} \|d_{j-1}\|^2 = \sum_{\ell=s}^{k-1} \|d_\ell\|^2 &\leq \sum_{\ell=s}^{\infty} \|d_\ell\|^2 \leq \frac{9\rho c^2}{2(2-\rho)(s+1)} \\
&= \frac{a_\rho^2 c^2}{4(s+1)} \leq \frac{a_\rho^2 c^2}{4}. \qquad (57)
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
&\rho \sum_{j=s+1}^{k} j\|R(w^{j-1})\| \, \|e_j\| \\
&\overset{(55)}{\leq} \rho c \sum_{j=s+1}^{k} \frac{\|R(u^{j-1})\| + \|d_{j-1}\|}{\sqrt{j}} \\
&\overset{(56)}{\leq} 2c\|w^0 - w^*\| \sum_{j=s+1}^{k} j^{-3/2} + \rho c \left(\sum_{j=s+1}^{k} \|d_{j-1}\|^2 \right)^{1/2} \\
&\overset{(53)(57)}{\leq} 2c\|w^0 - w^*\| + \frac{\rho a_\rho}{2} c^2.
\end{aligned}$$

Combining this estimate with (51),(54), then (52) yields

$$\Phi_k \leq (2a_\rho + 6)c\|w^0 - w^*\| + \frac{\rho}{2}\left(5a_\rho + \frac{10}{9}a_\rho^2\right)c^2,$$

where we used $a_\rho^2 = 18\rho/(2-\rho)$. Since $5a_\rho + \frac{10}{9}a_\rho^2 \leq (2a_\rho + 6)^2$, which can be verified through a direct calculation, we obtain

$$\Phi_k \leq (2a_\rho + 6)c\|w^0 - w^*\| + \frac{\rho(2a_\rho + 6)^2 c^2}{2}. \qquad (58)$$

Combining the lower bound (13) with (58), we obtain

$$\frac{\rho k + 2}{2}\|R(w^k)\|^2 - \|w^0 - w^*\|\|R(w^k)\| \leq \frac{(2a_\rho + 6)c\|w^0 - w^*\| + \rho(2a_\rho + 6)^2 c^2/2}{k+1}.$$

Multiplying both sides by $2(\rho k + 2)$ and then adding $\|w^0 - w^*\|^2$ to both sides gives

$$\begin{aligned}
&\left((\rho k + 2)\|R(w^k)\| - \|w^0 - w^*\|\right)^2 \\
&\quad \leq \|w^0 - w^*\|^2 + \frac{\rho k + 2}{k+1}\left(2(2a_\rho + 6)c\|w^0 - w^*\| + \rho(2a_\rho + 6)^2 c^2\right).
\end{aligned}$$

Note that $\rho(k+1) \le \rho k + 2$ implies $\frac{\rho(\rho k+2)}{k+1} \le \left(\frac{\rho k+2}{k+1}\right)^2$. Therefore,

$$\big((\rho k+2)\|R(w^k)\| - \|w^0 - w^*\|\big)^2 \le \left(\|w^0 - w^*\| + \frac{\rho k+2}{k+1}(2a_\rho + 6)c\right)^2.$$

Taking square roots and substituting the expression of a_ρ (47) yields

$$\|R(w^k)\| \le \frac{2\|w^0 - w^*\|}{\rho k+2} + \frac{c}{k+1}\left(6 + 6\sqrt{\frac{2\rho}{2-\rho}}\right).$$

This completes the proof. □

Proof of Theorem 3.5. Suppose first that $p \ge 3/2$. Then

$$\|e_j\| \le cj^{-p} \le cj^{-3/2}, \qquad j \ge 1.$$

Hence, Proposition 3.4 yields

$$\|R(w^k)\| \le \frac{2\|w^0 - w^*\|}{\rho k+2} + \frac{c}{k+1}\left(6 + 6\sqrt{\frac{2\rho}{2-\rho}}\right),$$

which is the stated bound since $(\frac{3}{2} - p)_+ = 0$.

Now suppose that $0 < p < 3/2$ and fix $k \ge 0$. Since w^k depends only on $e_1, \dots, e_k$, we may set $e_j = 0$ for all $j > k$ without changing w^k. Define

$$c_k := c(k+1)^{3/2-p}.$$

For $1 \le j \le k$, we have

$$\|e_j\| \le cj^{-p} = cj^{3/2-p}j^{-3/2} \le c_k j^{-3/2},$$

while the same inequality holds trivially for $j > k$ since $e_j = 0$. Since k is fixed, c_k is a constant, and the truncated error sequence therefore satisfies the assumptions of Proposition 3.4 with $c = c_k$. Applying the proposition gives

$$\begin{aligned}\|R(w^k)\| &\le \frac{2\|w^0 - w^*\|}{\rho k+2} + \frac{c_k}{k+1}\left(6 + 6\sqrt{\frac{2\rho}{2-\rho}}\right) \\ &= \frac{2\|w^0 - w^*\|}{\rho k+2} + c\left(6 + 6\sqrt{\frac{2\rho}{2-\rho}}\right)(k+1)^{1/2-p},\end{aligned}$$

which proves the stated bound for $0 < p < 3/2$.

When $c = 0$, the error term vanishes, and the bound reduces to (10). The asserted asymptotic rates follow directly. □

A.3 Proof for Subsection 3.2

Proof of Theorem 3.6. Let $\{u^k\}$ be the exact iterates from Section 2. By Proposition 2.2 with $\rho = 2$, we have

$$\|R(u^k)\| \le \frac{\|w^0 - w^*\|}{k+1}.$$

Hence,

$$\begin{aligned}\|R(w^k)\| &\le \|R(u^k)\| + \|R(w^k) - R(u^k)\| \\ &\le \|R(u^k)\| + \|w^k - u^k\| \\ &\le \frac{\|w^0 - w^*\|}{k+1} + \frac{2}{k+1}\sum_{j=1}^{k} j\|e_j\| \\ &\le \frac{\|w^0 - w^*\| + 2c\sum_{j=1}^{k} j^{1-p}}{k+1},\end{aligned}$$

where the second inequality is due to the nonexpansiveness of R, and the third inequality follows from Lemma 2.1 with $\rho = 2$. This proves (16). The stated rates follow from the standard estimates

$$\sum_{j=1}^{k} j^{1-p} = \begin{cases} \mathcal{O}(k^{2-p}), & 0 < p < 2, \\ \mathcal{O}(\log k), & p = 2, \\ \mathcal{O}(1), & p > 2. \end{cases}$$

□

A.4 Proof for Subsection 4.1

Proof of Proposition 4.1. Fix a sufficiently large target iteration N and identify $\mathbb{R}^2$ with $\mathbb{C}$, so that $w^0 = 1$ and $w^* = 0$. Choose $\theta_N \in (0, \pi)$ and define J_N by

$$\sin^2\left(\frac{\theta_N}{2}\right) = \frac{1}{\rho(2-\rho)N}, \qquad J_N(z) := \frac{1 + e^{\mathrm{i}\theta_N}}{2} z.$$

Since $2J_N - I$ is a rotation, J_N is firmly nonexpansive and is therefore the resolvent of a maximally monotone operator A_N with $A_N^{-1}(0) = \{0\}$. Set $R = I - J_N$. The relaxed map is $T_{\rho,N}(z) = \tau_N z$, where

$$\tau_N := 1 - \frac{\rho}{2} + \frac{\rho}{2} e^{\mathrm{i}\theta_N} = |\tau_N| e^{\mathrm{i}\psi_N}.$$

Choose the errors as

$$e_j := c j^{-p} e^{\mathrm{i}j\psi_N}, \qquad j \geq 1,$$

so that $\|e_j\| = c j^{-p}$. A direct calculation gives

$$|\tau_N|^2 = 1 - \frac{1}{N}, \qquad |1 - \tau_N| = \sqrt{\frac{\rho}{(2-\rho)N}}. \tag{59}$$

The iteration (2) becomes

$$(j+1) w^j = w^0 + j\tau_N w^{j-1} + \rho j e_j, \qquad 1 \leq j \leq N.$$

Iterating this identity yields

$$(N+1) w^N = \left(\sum_{j=0}^{N} \tau_N^j \right) w^0 + \rho \sum_{j=1}^{N} j \tau_N^{N-j} e_j. \tag{60}$$

Using $R = (I - T_{\rho,N})/\rho$, it follows

$$\begin{aligned} R(w^N) &= \frac{1 - \tau_N}{\rho(N+1)} \left(\sum_{j=0}^{N} \tau_N^j \right) w^0 + \frac{1 - \tau_N}{N+1} \sum_{j=1}^{N} j \tau_N^{N-j} e_j \\ &= \frac{1 - \tau_N^{N+1}}{\rho(N+1)} w^0 + \frac{1 - \tau_N}{N+1} \sum_{j=1}^{N} j \tau_N^{N-j} e_j. \end{aligned}$$

Moreover,

$$\tau_N^{N-j} e_j = c j^{-p} |\tau_N|^{N-j} e^{\mathrm{i}N\psi_N}, \qquad 1 \leq j \leq N.$$

Thus, the propagated errors all have the same direction. The triangle inequality gives

$$\|R(w^N)\| \geq \frac{c|1 - \tau_N|}{N+1} \sum_{j=1}^{N} j^{1-p} |\tau_N|^{N-j} - \frac{|1 - \tau_N^{N+1}|}{\rho(N+1)}. \tag{61}$$

For the first term of (61), using (59), we have

$$\begin{aligned}
&N^{p-\frac{1}{2}} \frac{c|1-\tau_N|}{N+1} \sum_{j=1}^{N} j^{1-p} |\tau_N|^{N-j} \\
&= c\sqrt{\frac{\rho}{2-\rho}} \frac{N^{p-1}}{N+1} \sum_{j=1}^{N} j^{1-p} \left(1-\frac{1}{N}\right)^{(N-j)/2} \\
&= c\sqrt{\frac{\rho}{2-\rho}} \frac{N}{N+1} \left[\frac{1}{N} \sum_{j=1}^{N} \left(\frac{j}{N}\right)^{1-p} \left(1-\frac{1}{N}\right)^{(N-j)/2}\right] \\
&:= c\sqrt{\frac{\rho}{2-\rho}} \frac{N}{N+1} \left[\frac{1}{N} \sum_{j=1}^{N} f_N\left(\frac{j}{N}\right)\right],
\end{aligned}$$

where $f_N(t) := t^{1-p}\left(1-\frac{1}{N}\right)^{N(1-t)/2}$. We claim that the expression in brackets converges to

$$\int_0^1 t^{1-p} \exp\left(-\frac{1-t}{2}\right) dt.$$

Indeed, let $f(t) := t^{1-p}\exp(-\frac{1-t}{2})$, we have $f_N \to f$ uniformly on $[0,1]$ for $0 < p \leq 1$, and therefore the above claim follows from the convergence of the Riemann sums. Then, if $1 < p < 3/2$, the integrand has an integrable singularity at $t = 0$, we thus split the summation into two parts: from $j = 1$ to $\lfloor \delta N \rfloor$ and from $j = \lfloor \delta N \rfloor + 1$ to N, where $0 < \delta < 1$. On $[\delta, 1]$, the same uniform-convergence argument gives convergence to

$$\int_\delta^1 t^{1-p} e^{-(1-t)/2}\, dt.$$

For the remaining part on $[0, \delta)$, since t^{1-p} is a nonincreasing function, it follows

$$\begin{aligned}
&\frac{1}{N} \sum_{j=1}^{\lfloor \delta N \rfloor} \left(\frac{j}{N}\right)^{1-p} \left(1-\frac{1}{N}\right)^{(N-j)/2} \\
\leq &\frac{1}{N} \sum_{j=1}^{\lfloor \delta N \rfloor} \left(\frac{j}{N}\right)^{1-p} \\
\leq &\sum_{j=1}^{\lfloor \delta N \rfloor} \int_{\frac{j-1}{N}}^{\frac{j}{N}} t^{1-p}\, dt \leq \int_0^\delta t^{1-p}\, dt \\
= &\frac{\delta^{2-p}}{2-p},
\end{aligned}$$

which tends to zero as $\delta \downarrow 0$. Thus, letting first $N \to \infty$ and then $\delta \downarrow 0$, we obtain

$$\frac{1}{N} \sum_{j=1}^{N} f_N\left(\frac{j}{N}\right) \longrightarrow \int_0^1 t^{1-p} e^{-(1-t)/2}\, dt.$$

For the second term of (61), since $p < 3/2$, we have

$$N^{p-\frac{1}{2}} \frac{|1-\tau_N^{N+1}|}{\rho(N+1)} \leq \frac{2N^{p-\frac{1}{2}}}{\rho(N+1)} \longrightarrow 0, \quad N \to \infty.$$

As a result, (61) implies

$$\liminf_{N\to\infty} N^{p-\frac{1}{2}} \sup_{A,\{e_j\}} \|R(w^N)\| \geq c\sqrt{\frac{\rho}{2-\rho}} \int_0^1 t^{1-p} \exp\left(-\frac{1-t}{2}\right) dt > 0.$$

We may take the constant as

$$\xi := \frac{1}{2}\sqrt{\frac{\rho}{2-\rho}} \int_0^1 t^{1-p} \exp\left(-\frac{1-t}{2}\right) dt$$

to finish the proof. □

A.5 Proof for Subsection 4.2

Proof of Proposition 4.2. Take $\mathcal{H} = \mathbb{R}$, $A = N_{\{0\}}$, $w^* = 0$, and any $w^0 > 0$, where $N_{\{0\}}$ denotes the normal cone to $\{0\}$. Set $\rho = 2$ and $e_j = (-1)^j c j^{-p}$ for every $j \geq 1$. Then $J_A = 0$, $R = I$, and $T_2 = -I$. Hence the implementable Halpern iteration (2) becomes

$$w^{k+1} = -\frac{k+1}{k+2} w^k + \frac{2(k+1)}{k+2} e_{k+1} + \frac{1}{k+2} w^0,$$

or equivalently,

$$(k+2)w^{k+1} = -(k+1)w^k + 2(k+1)e_{k+1} + w^0.$$

Iterating this identity yields

$$(k+1)w^k = w^0 \sum_{i=0}^{k} (-1)^i + 2c(-1)^k \sum_{j=1}^{k} j^{1-p}. \tag{62}$$

For every even k,

$$\sum_{i=0}^{k} (-1)^i = 1, \qquad (-1)^k = 1.$$

Since $R = I$ and $w^* = 0$, we obtain

$$(k+1)\|R(w^k)\| = \|w^0 - w^*\| + 2c \sum_{j=1}^{k} j^{1-p}, \tag{63}$$

which proves (17). Thus equality holds in (16) for every even k.

Note that,

$$\sum_{j=1}^{k} j^{1-p} = \begin{cases} \dfrac{k^{2-p}}{2-p} + o(k^{2-p}), & 0 < p < 2, \\ \log k + \mathcal{O}(1), & p = 2, \\ \zeta(p-1) + o(1), & p > 2, \end{cases} \tag{64}$$

where $\zeta(\cdot)$ denotes the Riemann zeta function $\zeta(s) := \sum_{j=1}^{\infty} j^{-s},\ s > 1$. Combining (63) and (64) gives (18). Since (16) is a uniform upper bound and the same admissible instance attains equality for every even k, the leading constant is optimal.

□

A.6 Proof for Subsection 5.1

Proof of Corollary 5.1. Define the negative dual objective, extended to the whole space, by

$$g(y) := \begin{cases} -\inf\limits_{x \in \mathcal{X}} \{f_0(x) + \langle y, h(x) \rangle\}, & y \in \mathbb{R}^m_+, \\ +\infty, & y \notin \mathbb{R}^m_+. \end{cases}$$

The saddle-point assumption implies that g is proper, closed, and convex, and that y^* minimizes g. Hence $A := \sigma \partial g$ is maximally monotone and $y^* \in A^{-1}(0)$. The augmented-Lagrangian proximal-point identity and its error estimate, recalled in Zhang et al. (2025a, Corollary 4.1), give

$$\begin{aligned} \|\widetilde{y}^{k+1} - J_A(y^k)\|^2 &\leq 2\sigma \left(L_\sigma(\widetilde{x}^{k+1}, y^k) - \inf_{x \in \mathcal{X}} L_\sigma(x, y^k) \right) \\ &\leq \frac{c^2}{(k+1)^3}, \end{aligned}$$

where the last inequality follows from (23). Thus (21) is an instance of (2) with $\rho = 1$ and $e_{k+1} := \widetilde{y}^{k+1} - J_A(y^k)$ satisfying $\|e_{k+1}\| \leq c(k+1)^{-3/2}$. Proposition 3.4 yields $\|R(y^k)\| = \mathcal{O}((k+1)^{-1})$, and consequently

$$\|\widetilde{y}^{k+1} - y^k\| \leq \|R(y^k)\| + \|e_{k+1}\| = \mathcal{O}((k+1)^{-1}). \tag{65}$$

Moreover, let $\{u^k\}$ denote the corresponding exact Halpern iterates with $u^0 = y^0$, we have

$$\begin{aligned}\|y^k - y^*\| &\le \|y^k - u^k\| + \|u^k - y^*\| \\ &\le \frac{1}{k+1}\sum_{j=1}^{k} j\|e_j\| + \|y^0 - y^*\| \\ &\le \frac{c}{k+1}\sum_{j=1}^{k} j^{-1/2} + \|y^0 - y^*\| \\ &\le \frac{2c}{\sqrt{k+1}} + \|y^0 - y^*\|,\end{aligned}$$

where the second inequality follows from Lemma 2.1 with $\rho = 1$ and the nonexpansiveness of exact Halpern iteration. From the above equations, both multipliers $\{y^k\}$ and $\{\widetilde{y}^{k+1}\}$ are bounded.

The update preserves $y^k \ge 0$. For each component with $h_i(\widetilde{x}^{k+1}) > 0$, the multiplier update gives $\widetilde{y}_i^{k+1} - y_i^k = \sigma h_i(\widetilde{x}^{k+1})$. It follows from (65) that

$$\|[h(\widetilde{x}^{k+1})]_+\| \le \frac{1}{\sigma}\|\widetilde{y}^{k+1} - y^k\| = \mathcal{O}((k+1)^{-1}). \tag{66}$$

Since x^* is feasible and $y^k \ge 0$, we have $L_\sigma(x^*, y^k) \le f_0(x^*)$. Using (20) and (23), we obtain

$$\begin{aligned}f_0(\widetilde{x}^{k+1}) - f_0(x^*) &\le \frac{c^2}{2\sigma(k+1)^3} + \frac{\|y^k\|^2 - \|\widetilde{y}^{k+1}\|^2}{2\sigma} \\ &= \frac{c^2}{2\sigma(k+1)^3} + \frac{\big(\|y^k\| - \|\widetilde{y}^{k+1}\|\big)\big(\|y^k\| + \|\widetilde{y}^{k+1}\|\big)}{2\sigma} \\ &\le \frac{c^2}{2\sigma(k+1)^3} + \frac{\|y^k\| + \|\widetilde{y}^{k+1}\|}{2\sigma}\|y^k - \widetilde{y}^{k+1}\|.\end{aligned}$$

The right-hand side is $\mathcal{O}((k+1)^{-1})$ by boundedness of the multipliers and (65).

Conversely, from the saddle-point inequality, we have

$$L(x^*, y^*) \le L(x, y^*), \quad \forall x \in \mathcal{X}, \tag{67}$$

where $L(x, y) := f_0(x) + \langle y, h(x)\rangle$ denotes the Lagrangian function. Besides, complementary slackness gives $\langle y^*, h(x^*)\rangle = 0$, and thus (67) with $x = \widetilde{x}^{k+1}$ implies

$$f_0(\widetilde{x}^{k+1}) - f_0(x^*) \ge -\langle y^*, h(\widetilde{x}^{k+1})\rangle \ge -\langle y^*, [h(\widetilde{x}^{k+1})]_+\rangle \ge -\|y^*\|\,\|[h(\widetilde{x}^{k+1})]_+\|,$$

where the second inequality holds due to $y^* \ge 0$. The right-hand side is also $\mathcal{O}((k+1)^{-1})$ from (66).

Consequently, both $\|[h(\widetilde{x}^{k+1})]_+\|$ and $|f_0(\widetilde{x}^{k+1}) - f_0(x^*)|$ are of order $\mathcal{O}((k+1)^{-1})$. This proves (24). □

A.7 Proof for Subsection 5.2

Proof of Corollary 5.2. The projection optimality condition implies that $G(w) = 0$ if and only if $0 \in F(w) + N_C(w)$, which characterizes the saddle points of (25). Since G is $1/L$-cocoercive, both G/L and $T := I - G/L$ are firmly nonexpansive on $\mathbb{R}^{n+m}$. By the resolvent characterization of firmly nonexpansive mappings, $A := T^{-1} - I$ is maximally monotone and $J_A = T$. Any saddle point w^* belongs to $A^{-1}(0)$, and $R(w) = G(w)/L$.

Define

$$e_{k+1} := -\frac{1}{L}\big(\widetilde{G}^k - G(w^k)\big).$$

Then (27) and the assumed bound on γ_k give

$$\|e_{k+1}\| \le \frac{\gamma_k}{L} \le \frac{c}{L}(k+1)^{-3/2}, \qquad w^k - \frac{1}{L}\widetilde{G}^k = J_A(w^k) + e_{k+1}.$$

Thus (27) is exactly (2) with $\rho = 1$, and the errors satisfy (4) with $p = 3/2$. Applying Proposition 3.4 proves (28). □

A.8 Proof for Subsection 5.3

Proof of Corollary 5.3. Since F is monotone and continuous on the whole Euclidean space and G is maximally monotone, $F+G$ is maximally monotone. Thus $A=\eta(F+G)$ is maximally monotone, and the nonempty solution set of (29) ensures that Assumption 1 holds. Fix $w^*\in A^{-1}(0)$. Nonexpansiveness of J_A and R, together with (31), implies inductively that the iterates and errors have finite second moments.

Let Φ_k be defined by (11) with $\rho=1$, and set

$$H_k := \mathbb{E}\Phi_k + \|w^0-w^*\|^2.$$

Using the lower bound (13) and taking expectation on both sides, we obtain

$$\begin{aligned}
H_k &\geq \frac{(k+1)(k+2)}{2}\mathbb{E}\|R(w^k)\|^2 - (k+1)\|w^0-w^*\|\mathbb{E}\|R(w^k)\| + \|w^0-w^*\|^2 \\
&\geq \frac{(k+1)(k+2)}{2}\mathbb{E}\|R(w^k)\|^2 - \frac{(k+1)(k+2)}{4}\mathbb{E}\|R(w^k)\|^2 \\
&\quad - \frac{k+1}{k+2}\|w^0-w^*\|^2 + \|w^0-w^*\|^2 \\
&= \frac{(k+1)(k+2)}{4}\mathbb{E}\|R(w^k)\|^2 + \frac{\|w^0-w^*\|^2}{k+2} \\
&\geq \frac{(k+1)(k+2)}{4}\mathbb{E}\|R(w^k)\|^2 \geq 0,
\end{aligned} \tag{68}$$

where the second inequality follows from Young's inequality:

$$(k+1)\|w^0-w^*\|\|R(w^k)\| \leq \frac{(k+1)(k+2)}{4}\|R(w^k)\|^2 + \frac{k+1}{k+2}\|w^0-w^*\|^2.$$

Conditional Cauchy–Schwarz and (31) yield $\mathbb{E}_k\|e_{k+1}\| \leq \|R(w^k)\|/(k+2)$. Taking conditional expectations in (12) therefore gives

$$\begin{aligned}
\mathbb{E}_k[\Phi_{k+1}-\Phi_k] &\leq \frac{(k+1)(k+2)}{2}\mathbb{E}_k\|e_{k+1}\|^2 + (k+1)\|R(w^k)\|\mathbb{E}_k\|e_{k+1}\| \\
&\leq \frac{3(k+1)}{2(k+2)}\|R(w^k)\|^2.
\end{aligned}$$

Taking expectations and using the lower bound on H_k (68), we obtain

$$\begin{aligned}
H_{k+1}-H_k = \mathbb{E}[\Phi_{k+1}-\Phi_k] &\leq \frac{3(k+1)}{2(k+2)}\mathbb{E}\|R(w^k)\|^2 \\
&\leq \frac{3(k+1)}{2(k+2)}\frac{4}{(k+1)(k+2)}H_k \\
&= \frac{6}{(k+2)^2}H_k.
\end{aligned}$$

Equivalently,

$$H_{k+1} \leq \left(1+\frac{6}{(k+2)^2}\right)H_k.$$

Since $\Phi_0=0$, it follows from $1+t\leq e^t$ that

$$H_k \leq \|w^0-w^*\|^2\prod_{j=0}^{k-1}\left(1+\frac{6}{(j+2)^2}\right) \leq \|w^0-w^*\|^2\exp\left(\sum_{j=0}^{\infty}\frac{6}{(j+2)^2}\right) < \infty.$$

Thus $\{H_k\}$ is uniformly bounded. Together with the lower bound (68), this implies $\mathbb{E}\|R(w^k)\|^2 = \mathcal{O}((k+1)^{-2})$. Finally, Jensen's inequality proves the result

$$\mathbb{E}\|R(w^k)\| \leq \left(\mathbb{E}\|R(w^k)\|^2\right)^{1/2} = \mathcal{O}((k+1)^{-1}).$$

□