Time-Reversal and Anchoring Algorithms: Unified Optimal Acceleration for Minimax and Fixed-Point Problems

Chris Junchi Li[◇]

Department of Electrical Engineering and Computer Sciences[◇]
University of California, Berkeley

September 25, 2024

Abstract

In this paper, we introduce a novel class of accelerated algorithms for minimax optimization and fixed-point problems by leveraging *time-reversal ordinary differential equations* (time-reversal ODEs). This new one-parameter family of time-reversal ODEs, when discretized, leads to algorithms that achieve the same optimal convergence rates as existing anchor-based methods while employing materially different acceleration mechanisms. These algorithms, dual to traditional anchoring techniques, provide a unified framework that extends beyond existing approaches. Our results not only confirm the theoretical optimality of these new methods with matching complexity lower bounds but also reveal that the optimal acceleration mechanism is not unique. These findings offer new insights and opportunities for developing more efficient and stable algorithms, particularly for machine learning and large-scale optimization applications.

Keywords: Time-Reversal ODEs; Minimax Optimization; Fixed-Point Problems; Optimal Acceleration; Anchoring Mechanism; Convergence Rates; Maximally Monotone Operator

1 Introduction

Minimax optimization and fixed-point problems are fundamental to various fields, including machine learning, economics, and game theory. These problems are often computationally challenging, particularly when seeking solutions with optimal convergence rates. Over the past decade, accelerated algorithms using the *anchoring mechanism*—such as the Extra Anchored Gradient (EAG) for minimax problems and the Optimal Halpern Method (OHM) for fixed-point problems—have been the state-of-the-art, matching theoretical lower bounds on convergence rates and minimizing computational complexity.

However, while anchoring-based methods are celebrated for their optimality, they rely on specific acceleration mechanisms that limit exploration of alternative, potentially more robust approaches. In this paper, we introduce a new family of *time-reversal algorithms*, based on time-reversal ordinary differential equations (ODEs), which achieve the same optimal convergence rates as their anchoring-based counterparts but with materially different acceleration mechanisms.

This surprising discovery reveals that the optimal acceleration mechanism is not unique. Our time-reversal approach generalizes existing frameworks and reframes the study of optimal acceleration as an exploration of a broader family of mechanisms. The resulting algorithms maintain worst-case convergence guarantees while exhibiting distinct empirical characteristics, offering more flexibility for diverse applications. These findings open up new directions for theoretical and practical advancements in solving minimax and fixed-point problems, particularly in machine learning, game theory, and large-scale optimization.

Mathematically, let operator $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be a set-valued function such that $\mathbb{A}(x) \subseteq \mathbb{R}^d$ for $x \in \mathbb{R}^d$. The (time-forward) *Anchoring-ODE* [RYY19, SPR23] is

$$\ddot{X}(t) + \frac{2}{t}\dot{X}(t) + \frac{1}{t}\mathbb{A}(X(t)) + \frac{d}{dt}\mathbb{A}(X(t)) = 0$$

where $X(0) = X_0$ and $\dot{X}(0) = -\frac{1}{2}\mathbb{A}(X_0)$ are the initial conditions. Anchoring-ODE exhibits the rate

$$\|\mathbb{A}(X(t))\|^2 \le \frac{4\|X_0 - X_\star\|^2}{t^2}$$

for $t > 0$, where $X_\star$ is a solution or zero of $\mathbb{A}$.

In this paper we concentrate on the new *RevAn-ODE*: prescribing $T > 0$ as the *terminal time*, for $t \in (0, T)$ finite interval

$$\ddot{X}(t) + \frac{1}{T-t}\dot{X}(t) + \frac{d}{dt}\mathbb{A}(X(t)) = 0 \qquad \text{(RevAn-ODE)}$$

where $X(0) = X_0$ and $\dot{X}(0) = -\mathbb{A}(X_0)$ are initial conditions. In comparison with the above time-forward ODE it has one summand less and enjoys a cleaner form. We will be showing that (RevAn-ODE) exhibits the rate

$$\|\mathbb{A}(X(T))\|^2 \le \frac{4\|X_0 - X_\star\|^2}{T^2}$$

This rate exactly coincides with the rate of Anchoring-ODE for $t = T$.

1.1 Accelerated algorithms for minimax optimization and fixed-point problems

In this subsection, we present novel accelerated algorithms for both smooth minimax problem and fixed-point problem. For each setup, we first review the existing *time-forward* algorithm using the anchor acceleration mechanism and then show its *time-reversal* counterpart with identical rates but using a materially different acceleration mechanism.[1] The introduced algorithms for fixed-point problems and minimax optimization can be unified to, when taking proper continuous limit, the same continuous-time ODE.

Example 1: Smooth convex-concave minimax optimization. We consider the minimax optimization problem (1), which solves

$$\underset{u\in\mathbb{R}^n}{\text{minimize}}\ \underset{v\in\mathbb{R}^m}{\text{maximize}}\ \ \mathbf{L}(u, v) \tag{1}$$

where function $\mathbf{L}\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is convex-concave and L-smooth. $\|\nabla\mathbf{L}(u_k, v_k)\|^2$ is one natural performance measure under the assumption of convex-concavity, where [YR21] provided a first *optimal* accelerated $\|\nabla\mathbf{L}(u_k, v_k)\|^2 = \mathcal{O}(1/k^2)$ rate via the Extra Anchored Gradient (EAG) algorithm, along with an $\Omega(1/k^2)$ complexity lower bound.

[1] Throughout the paper, we write $N \ge 1$ to denote the pre-specified iteration count of the algorithm.

The operator in consideration is $\mathbb{A} = (\nabla_u \mathbf{L}, -\nabla_v \mathbf{L})$ for minimax problem (1) The (time-forward) *Fast Extragradient* (FEG) algorithm [LK21] has the update rule[2]

$$
\begin{aligned}
x_{k+\frac{1}{2}} &= x_k + \frac{1}{k+1}(x_0 - x_k) - \frac{k}{k+1}\alpha \mathbb{A} x_k \\
x_{k+1} &= x_k + \frac{1}{k+1}(x_0 - x_k) - \alpha \mathbb{A} x_{k+\frac{1}{2}}
\end{aligned} \tag{FEG}
$$

for $k = 0, 1, \dots$. If $0 < \alpha \le \frac{1}{L}$, (FEG) exhibits the rate

$$
\|\nabla \mathbf{L}(x_k)\|^2 = \|\mathbb{A} x_k\|^2 \le \frac{4\|x_0 - x_\star\|^2}{\alpha^2 k^2}
$$

for $k = 1, 2, \dots$ and a saddle point (solution) $x_\star = (u_\star, v_\star)$. As shown in [YR21], there exists L-smooth convex-concave $\mathbf{L}\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ with $m = n \ge 3N + 2$ such that

$$
\|\nabla \mathbf{L}(x_N)\|^2 \ge \frac{L^2 \|x_0 - x_\star\|^2}{(2\lfloor N/2 \rfloor + 1)^2}
$$

for any deterministic N-step first-order algorithm. Such a convergence rate result with $\alpha = \frac{1}{L}$ improved the convergence upper bound of [YR21] by a constant factor and is *optimal* up to a prefactor of 4. To the best of our knowledge, it achieves the best-known constant prefactor up to date.

We present the new method, *Reversal Fast Extragradient* (Rev-FEG):

$$
\begin{aligned}
x_{k+\frac{1}{2}} &= x_k - \alpha z_k - \alpha \mathbb{A} x_k \\
x_{k+1} &= x_{k+\frac{1}{2}} - \frac{N-k-1}{N-k}\alpha \left(\mathbb{A} x_{k+\frac{1}{2}} - \mathbb{A} x_k \right) \\
z_{k+1} &= \frac{N-k-1}{N-k} z_k - \frac{1}{N-k} \mathbb{A} x_{k+\frac{1}{2}}
\end{aligned} \tag{Rev-FEG}
$$

for $k = 0, 1, \dots, N-1$, where $z_0 = 0$. For $0 < \alpha \le \frac{1}{L}$, (Rev-FEG) exhibits the rate

$$
\|\nabla \mathbf{L}(x_N)\|^2 = \|\mathbb{A} x_N\|^2 \le \frac{4\|x_0 - x_\star\|^2}{\alpha^2 N^2}
$$

This rate exactly coincides with the rate of (FEG) for $k = N$, and hence (Rev-FEG) is also optimal up to a prefactor of 4.

Example 2: Fixed-point problems. The operator in consideration is $\mathbb{A} = 2(\mathbb{I} + \mathbb{T})^{-1} - \mathbb{I}$. Consider the fixed-point problem (2) where $\mathbb{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ is nonexpansive. A fixed-point problem solves

$$
\underset{y \in \mathbb{R}^d}{\text{find}} \quad y = \mathbb{T} y \tag{2}
$$

for $\mathbb{T}\colon \mathbb{R}^d \to \mathbb{R}^d$. We denote $\operatorname{Fix} \mathbb{T} = \{y \in \mathbb{R}^d \,|\, y = \mathbb{T} y\}$ and assume $\operatorname{Fix} \mathbb{T} \ne \emptyset$. A natural performance measure is the magnitude of the fixed-point residual $y_k - \mathbb{T} y_k$, where [SS17] first achieved the rate $\|y_k - \mathbb{T} y_k\|^2 = \mathcal{O}(1/k^2)$ through the Sequential Averaging Method.

[2](FEG) was designed primarily for weakly nonconvex-nonconcave problems, but we consider its application to the special case of convex-concave problems.

Lieder [Lie21] showed that Halpern iteration with specific parameters, which we call the (time-forward) *Optimal Halpern Method* (OHM) [Hal67, Lie21],[3] can be written as

$$y_{k+1} = \frac{k+1}{k+2}\mathbb{T}y_k + \frac{1}{k+2}y_0 \tag{OHM}$$

for $k = 0, 1, \ldots$. Equivalently, we can write

$$y_{k+1} = y_k - \frac{1}{k+2}(y_k - \mathbb{T}y_k) + \frac{k}{k+2}\left(\mathbb{T}y_k - \mathbb{T}y_{k-1}\right)$$

where we define $\mathbb{T}y_{-1} = y_0$. (OHM) exhibits the rate

$$\|y_{k-1} - \mathbb{T}y_{k-1}\|^2 \le \frac{4\|y_0 - y_\star\|^2}{k^2}$$

for $k = 1, 2, \ldots$ and $y_\star \in \operatorname{Fix}\mathbb{T}$ [Lie21]. Such a rate improves upon the rate of [SS17] by a factor of 16. Furthermore, [PR22] showed that the rate of [Lie21] is *exactly* optimal by providing a matching complexity lower bound,[4] in the sense that there exists a nonexpansive operator $\mathbb{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ with $d \ge 2N-2$ such that

$$\|y_{N-1} - \mathbb{T}y_{N-1}\|^2 \ge \frac{4\|y_0 - y_\star\|^2}{N^2}$$

for any deterministic algorithm using $N-1$ evaluations of $\mathbb{T}$.

We present the new method, *Reversal Optimal Halpern Method* (Rev-OHM):

$$y_{k+1} = y_k + \frac{N-k-1}{N-k}\left(\mathbb{T}y_k - \mathbb{T}y_{k-1}\right) \tag{Rev-OHM}$$

for $k = 0, 1, \ldots, N-2$, where we define $\mathbb{T}y_{-1} = y_0$. Equivalently,

$$\begin{aligned} z_{k+1} &= \frac{N-k-1}{N-k}z_k - \frac{1}{N-k}\left(y_k - \mathbb{T}y_k\right) \\ y_{k+1} &= \mathbb{T}y_k - z_{k+1} \end{aligned} \tag{3}$$

for $k = 0, 1, \ldots, N-2$, where $z_0 = 0$. (Rev-OHM) exhibits the rate

$$\|y_{N-1} - \mathbb{T}y_{N-1}\|^2 \le \frac{4\|y_0 - y_\star\|^2}{N^2}$$

for $y_\star \in \operatorname{Fix}\mathbb{T}$. This rate exactly coincides with the rate of (OHM) for $k = N$, hence as is shown in the previous argument (OHM) is exactly optimal.

Relationship between convex-concave minimax and fixed-point problems. Convex-concave minimax problems are closely related to fixed-point problems, and the anchoring mechanism of (OHM) for accelerating fixed-point algorithms has been used to accelerate algorithms for minimax problems [YR21, LK21]. *Reversal Fast Extragradient* (Rev-FEG) can be viewed as the minimax counterpart of *Reversal Optimal Halpern Method* (Rev-OHM) for fixed-point problems.

More importantly, we will show in §3.2 that

[3]Some prior work referred to this method as the "Optimized" Halpern Method, but we now know the method is (exactly) optimal as [PR22] provided a matching lower bound.

[4]This means it cannot be improved, not even by a constant factor, in terms of worst-case performance.

Proposition 1 (Informal)**.** (RevAn-ODE) *is the common continuous-time model for both*

$$\begin{aligned} x_{k+\frac{1}{2}} &= x_k - \alpha z_k - \alpha \mathbb{A} x_k \\ x_{k+1} &= x_{k+\frac{1}{2}} - \frac{N-k-1}{N-k}\alpha \left(\mathbb{A} x_{k+\frac{1}{2}} - \mathbb{A} x_k\right) \\ z_{k+1} &= \frac{N-k-1}{N-k} z_k - \frac{1}{N-k}\mathbb{A} x_{k+\frac{1}{2}} \end{aligned} \tag{Rev-FEG}$$

for minimax problem and

$$y_{k+1} = y_k + \frac{N-k-1}{N-k}\left(\mathbb{T}y_k - \mathbb{T}y_{k-1}\right) \tag{Rev-OHM}$$

for fixed-point problem.

In studying (RevAn-ODE) we also provide existence, uniqueness and convergence analysis upon therein.

1.2 Related work

Continuous-time analyses. Taking the continuous-time limit of an iterative algorithm results in an ordinary differential equation (ODE), and they often more easily reveal the structure of convergence analysis. The ODE models for OG and EG were respectively studied by [CMT19, Lu22]. The continuous-time analysis of accelerated algorithms was initiated by [SBC14, KBB15], and the anchor acceleration ODE for fixed-point and minimax problems was first introduced by [RYY19] and then generalized to a broader family of differential inclusion with more rigorous treatment by [SPR23]. [BCN23] studied a different family of ODE that achieves acceleration asymptotically. An ODE involving the coefficient of the form $\frac{1}{T-t}$ with fixed terminal time T was first presented in [SRR22], as continuous-time model of the OGM-G algorithm [KF21].

Fixed-point algorithms. Iterative algorithms for solving fixed-point problems (2) have been extensively studied over a long period, and among them, Picard iteration, Krasnosel'skii–Mann (KM) Iteration and Halpern Iteration stand out as representative classes, each defined by:

$$y_{k+1} = \mathbb{T}y_k \tag{Picard}$$

$$y_{k+1} = \lambda_{k+1} y_k + (1-\lambda_{k+1})\mathbb{T}y_k \tag{KM}$$

$$y_{k+1} = \lambda_{k+1} y_0 + (1-\lambda_{k+1})\mathbb{T}y_k \tag{Halpern}$$

The formal study of Picard iteration dates up to [Ban22]. The class KM generalizes the works of [Kra55] and [Man53]; early works [Ish76, BRS92] focused on its asymptotic convergence $\|y_k - \mathbb{T}y_k\|^2 \to 0$, while quantitative rates of $\mathcal{O}(1/k)$ to $o(1/k)$ were respectively demonstrated by [CSV14, LFP16, BC18] and [BB92, DY16, Mat17]. For the Halpern class, devised by [Hal67], [Wit92, Xu02] established asymptotic convergence. Concerning quantitative convergence, [Leu07] provided the rate $\|y_k - \mathbb{T}y_k\|^2 = \mathcal{O}(1/(\log k)^2)$, which was later improved to $\mathcal{O}(1/k)$ [Koh11], and to $\mathcal{O}(1/k^2)$, first by [SS17], and then by [Lie21] with a tighter constant, using the choice $\lambda_k = \frac{1}{k+1}$. [PR22] established the exact optimality of the convergence rate from [Lie21] (and the independently discovered equivalent result of [Kim21]) by constructing a matching complexity lower bound. Some notable acceleration results for fixed-point problems not covered by the above list include Anderson-type acceleration [And65, WN11, ZOB20] and inertial-type acceleration [Mai08, DYCR18, She18, RTCVL21].

Minimax optimization algorithms. Smooth convex-concave minimax optimization is a classical problem, whose investigation dates up to [Kor76, Pop80], which respectively first developed the Extragradient (EG) and optimistic gradient (OG) whose variants have been studied extensively [SS99, Nem04, Nes07, RS13, DISZ18] across the optimization and machine learning communities. Recently there has been rapid development in the theory of reducing $\|\nabla \mathbf{L}(\cdot)\|^2$; while EG and OG converge at the rate of $\|\nabla \mathbf{L}(u_k, v_k)\|^2 = \mathcal{O}(1/k)$ [GLG22, COZ22], using the anchoring mechanism [RYY19] motivated by the optimal Halpern iteration [Hal67, Kim21, Lie21], [Dia20] achieved the near-optimal rate $\widetilde{\mathcal{O}}(1/k^2)$, which was finally accelerated to $\mathcal{O}(1/k^2)$ via Extra Anchored Gradient (EAG) algorithm [YR21]. [LK21] provided a constant factor improvement over EAG while generalizing the acceleration to weakly nonconvex-nonconcave problems, and [TDL21, CZ23] presented single-call versions of the acceleration. [BCN23] achieved asymptotic $o(1/k^2)$ convergence based on the continuous-time perspective. The conceptual connection between smooth minimax optimization and fixed-point problems (proximal algorithms) have been formally studied in [MOP20, YR22].

Performance estimation problem (PEP) technique. At a high level, the PEP methodology [DT14, THG17, TB19, RTBG20, DGVPR24] provides a computer-assisted framework for finding tight convergence proofs and optimizing the step-sizes of optimization of algorithms, and many efficient algorithms and novel proofs have recently been discovered with the aid of this methodology [KF16, Lie21, Kim21, KF21, PR24, PR22, GLG22, TD23, JGR23, PR23].

Specifically relevant to our work are the prior, concurrent work of [Kim21] presenting APPM and of [Lie21] presenting OHM. It was later shown that APPM and OHM are equivalent, and they represent one element within the family of exact optimal algorithms that we present in this work. We find it somewhat surprising that the two authors independently found the same algorithm when there is an infinitude of answers. Therefore, we personally asked Kim and Lieder to understand how their processes led to their discovery and not any other choice.

Family of exact optimal algorithms. In the setup of minimizing a non-smooth convex function f whose subgradient magnitude is bounded by M (so the objective function is M-Lipschitz continuous), there are many different algorithms achieving the exact optimal complexity. Suppose that $\|x_0 - x_\star\| \le R$, where $x_\star$ is a minimizer of f. The subgradient method of [Nes18], given by $x_{k+1} = x_k - \frac{R}{\sqrt{N+1}} \frac{g_k}{\|g_k\|}$ where $g_k \in \partial f(x_k)$, exhibits the rate $f(x_N) - f_\star \le \frac{MR}{\sqrt{N+1}}$ where $f_\star = f(x_\star)$. The exact same rate is achieved by the algorithm of [DT16] which is a variant of the cutting-plane method, a fixed step-size algorithm of [DT20] without gradient normalization, and its line-search variant. As there is a matching lower bound $f(x_N) - f_\star \ge \frac{MR}{\sqrt{N+1}}$ [DT16], all these methods are exactly optimal in terms of the worst-case complexity.

In the fixed-point setup (or the equivalent monotone inclusion setup), a Halpern-type algorithm by [SPR23] that uses adaptive interpolation (anchoring) coefficients also achieves the exact optimal complexity (same rate as (OHM)). Hence, in a strict sense, our work is not the first discovery that there exist a family of exact optimal algorithms instead of a single one. However, to the best of our knowledge, our work is the first to identify distinct *fixed step-size* algorithms sharing the same exact optimal complexity for fixed-point problems. Additionally, while the adaptive algorithm of [SPR23] relies on a mechanism that is similar to (OHM), our time-reversal algorithms use an arguably different mechanism from that underlying (OHM).

Contributions. The key contributions of this work are as follows:

(i) We introduce a novel class of *time-reversal ODE-based algorithms* for minimax optimization and fixed-point problems that achieve optimal convergence rates.

(ii) These algorithms employ a distinct acceleration mechanism, providing an alternative to anchoring-based methods while retaining optimality, and offering a new perspective on acceleration as a *one-parameter family* of mechanisms.

(iii) We provide a comprehensive theoretical analysis, including complexity lower bounds, along with a detailed comparison to existing algorithms, demonstrating the advantages of our approach.

Organization. The remainder of this paper is organized as follows. §2 reviews the necessary preliminaries and background on existing anchoring-based methods for minimax and fixed-point problems. §3 presents the continuous-time analysis of the time-reversal ODE framework and derives the novel algorithms from these continuous-time counterparts, while §4 generalizes this analysis to the case of differential inclusions. §5 provides a detailed convergence analysis, comparing our time-reversal algorithms to existing methods. §6 concludes the paper with remarks on future research directions. The appendix presents a continuous family of exact optimal fixed-point algorithms.

2 Preliminaries

We use standard notation for set-valued operators [BC17, RY22]. Recall that an *operator* $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is a set-valued function (so $\mathbb{A}(x) \subseteq \mathbb{R}^d$ for $x \in \mathbb{R}^d$). For simplicity, we sometimes write $\mathbb{A}x = \mathbb{A}(x)$. The *graph* of $\mathbb{A}$ is defined and denoted as $\operatorname{Gra}\mathbb{A} = \{(x, y) \,|\, x \in \mathbb{R}^d, y \in \mathbb{A}x\}$. The *inverse* of $\mathbb{A}$ is defined by $\mathbb{A}^{-1}y = \{x \in \mathbb{R}^d \,|\, y \in \mathbb{A}x\}$. Scalar multiples and sums of operators are defined in the Minkowski sense. If $\mathbb{T}x$ is a singleton for all $x \in \mathbb{R}^d$, we write $\mathbb{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ and treat it as a function. An operator $\mathbb{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ is *L-Lipschitz* for some $L > 0$ if $\|\mathbb{T}x - \mathbb{T}y\| \le L\,\|x - y\|$ for all $x, y \in \mathbb{R}^d$. We say $\mathbb{T}$ is *nonexpansive* if it is 1-Lipschitz.

A function $\mathbf{L}\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is *convex-concave* if $\mathbf{L}(u, v)$ is convex in u for all fixed $v \in \mathbb{R}^m$ and concave in v for all fixed $u \in \mathbb{R}^n$. If $\mathbf{L}(u_\star, v) \le \mathbf{L}(u_\star, v_\star) \le \mathbf{L}(u, v_\star)$ for all $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$, then $(u_\star, v_\star)$ is a *saddle point* of $\mathbf{L}$. For $L > 0$, if $\mathbf{L}$ is differentiable and $\nabla\mathbf{L}$ is L-Lipschitz on $\mathbb{R}^n \times \mathbb{R}^m$, we say $\mathbf{L}$ is *L-smooth*. In this case, we define the *saddle operator* of $\mathbf{L}$ by $\nabla_\pm\mathbf{L}(u, v) = (\nabla_u\mathbf{L}(u, v), -\nabla_v\mathbf{L}(u, v))$. In most of the cases, we use the joint variable notation $x = (u, v)$ and concisely write $\nabla_\pm\mathbf{L}(x)$ in place of $\nabla_\pm\mathbf{L}(u, v)$.

Till the rest of this section, we express our analysis using the language of monotone operators. We quickly set up the notation and review the connections between fixed-point problems and monotone operators.

Monotone operators. A set-valued (non-linear) operator $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is *monotone* if $\langle g - g', x - x'\rangle \ge 0$ for all $x, x' \in \mathbb{R}^d$, $g \in \mathbb{A}x$, and $g' \in \mathbb{A}x'$. If $\mathbb{A}$ is monotone and there is no monotone operator $\mathbb{A}'$ for which $\operatorname{Gra}\mathbb{A} \subset \operatorname{Gra}\mathbb{A}'$ properly, then $\mathbb{A}$ is *maximally monotone.* If $\mathbb{A}$ is maximally monotone, then its *resolvent* $\mathbf{J}_{\mathbb{A}} := (\mathbb{I} + \mathbb{A})^{-1}\colon \mathbb{R}^d \to \mathbb{R}^d$ is a well-defined single-valued operator.

Minimax optimization and monotone operators. The minimax problem (1) can be recast as a monotone inclusion problem. Precisely, for L-smooth convex-concave $\mathbf{L}$, its saddle operator $\nabla_\pm\mathbf{L}$ is monotone and L-Lipschitz. In this case, $x_\star = (u_\star, v_\star)$ is a minimax solution for $\mathbf{L}$ if and only if $\nabla_\pm\mathbf{L}(x_\star) = 0$.

Fixed-point problems are also monotone inclusion problems. There exists a natural correspondence between the classes of nonexpansive operators and maximally monotone operators in the following sense

Proposition 2. *[EB92, Theorem 2] If $\mathbb{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ is a nonexpansive operator, then $\mathbb{A} = 2(\mathbb{I} + \mathbb{T})^{-1} - \mathbb{I}$ is maximally monotone. Conversely, if $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is maximally monotone, then $\mathbb{T} = 2\mathbb{J}_{\mathbb{A}} - \mathbb{I}$ is nonexpansive.*

When $\mathbb{T} = 2\mathbb{J}_{\mathbb{A}} - \mathbb{I}$, we have $x = \mathbb{T}x \iff 0 \in \mathbb{A}x$, i.e., $\operatorname{Fix}\mathbb{T} = \operatorname{Zer}\mathbb{A} := \{x \in \mathbb{R}^d \,|\, 0 \in \mathbb{A}x\}$. Therefore, Proposition 2 induces a one-to-one correspondence between nonexpansive fixed-point problems and monotone inclusion problems.

Fixed-point residual norm and operator norm. Given $y \in \mathbb{R}^d$, its accuracy as an approximate fixed-point solution is often measured by $\|y - \mathbb{T}y\|$, the norm of fixed-point residual. Let $\mathbb{A}$ be the maximal monotone operator satisfying $\mathbb{T} = 2\mathbb{J}_{\mathbb{A}} - \mathbb{I}$, and let $x = \mathbb{J}_{\mathbb{A}}y$. Then we see that

$$y \in (\mathbb{I} + \mathbb{A})(x) = x + \mathbb{A}x \iff y - x \in \mathbb{A}x$$

Denote $\widetilde{\mathbb{A}}x = y - x \in \mathbb{A}x$. Then

$$y - \mathbb{T}y = y - (2\mathbb{J}_{\mathbb{A}} - \mathbb{I})(y) = 2(y - \mathbb{J}_{\mathbb{A}}y) = 2\widetilde{\mathbb{A}}x$$

Therefore, $\|y - \mathbb{T}y\| = 2\|\widetilde{\mathbb{A}}x\|$.

Finally, we quickly state a handy lemma used in the convergence analyses throughout the paper.

Lemma 1. *Let $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be monotone and let $x, y \in \mathbb{R}^d$, $\widetilde{\mathbb{A}}x \in \mathbb{A}x$. Suppose, for some $\rho > 0$, $\rho \left\|\widetilde{\mathbb{A}}x\right\|^2 + \left\langle \widetilde{\mathbb{A}}x, x - y\right\rangle \le 0$ holds. Then, for $x_\star \in \operatorname{Zer}\mathbb{A}$,*

$$\left\|\widetilde{\mathbb{A}}x\right\|^2 \le \frac{\|y - x_\star\|^2}{\rho^2}$$

Proof. By monotonicity of $\mathbb{A}$ and Young's inequality,

$$0 \ge \rho \left\|\widetilde{\mathbb{A}}x\right\|^2 + \left\langle \widetilde{\mathbb{A}}x, x - y\right\rangle \ge \rho \left\|\widetilde{\mathbb{A}}x\right\|^2 + \left\langle \widetilde{\mathbb{A}}x, x_\star - y\right\rangle \ge \frac{\rho}{2}\left\|\widetilde{\mathbb{A}}x\right\|^2 - \frac{1}{2\rho}\|x_\star - y\|^2$$

which completes the proof. □

3 Continuous-time analysis of reversal-anchoring

In this section, we outline the analysis of the (RevAn-ODE). In special we provide rigorous discussion on existence, uniqueness of the solutions and almost everywhere differentiability of the involved quantities.

3.1 Equivalent forms of (RevAn-ODE)

We repeat the *second-order form* as

$$\ddot{X}(t) + \frac{1}{T-t}\dot{X}(t) + \frac{d}{dt}\mathbb{A}(X(t)) = 0 \tag{4}$$

where $X(0) = X_0$ and $\dot{X}(0) = -\mathbb{A}(X_0)$. We claim it has an equivalent *first-order form*:

Proposition 3. *Let $\mathbb{A}$ be Lipschitz continuous. Then* (4) *has the following equivalent* first-order form (5)

$$\begin{aligned}\dot{X}(t) &= -Z(t) - \mathbb{A}(X(t)) \\ \dot{Z}(t) &= -\frac{1}{T-t}Z(t) - \frac{1}{T-t}\mathbb{A}(X(t))\end{aligned} \tag{5}$$

where $X(0) = X_0$ and $Z(0) = 0$, in the sense that the solutions give equal $X(t)$'s in the almost everywhere sense.

The above result holds given a minimal assumption that $\mathbb{A}$ is Lipschitz continuous, with the equalities holding for almost every t if differentiability is not assumed.

Proof of Proposition 3. We prove the result in both ways:

First-order form $\Longrightarrow$ Second-order form. Let $\begin{pmatrix} X \\ Z \end{pmatrix} : [0,T) \to \mathbb{R}^d \times \mathbb{R}^d$ be the solution to (5) with initial conditions $X(0) = X_0$ and $Z(0) = 0$. Plugging $t = 0$ into the first line of (RevAn-ODE) gives $\dot{X}(0) = 0 - \mathbb{A}(X_0)$, which is the initial velocity condition for (4). Now observe that

$$\dot{Z}(t) = \frac{1}{T-t}\left(-Z(t) - \mathbb{A}(X(t))\right) = \frac{1}{T-t}\dot{X}(t)$$

where the last equality comes from the first line of (5). Differentiating the first line of (5) and plugging in the above identity we obtain *second-order form*:

$$0 = \ddot{X}(t) + \dot{Z}(t) + \frac{d}{dt}\mathbb{A}(X(t)) = \ddot{X}(t) + \frac{1}{T-t}\dot{X}(t) + \frac{d}{dt}\mathbb{A}(X(t))$$

Second-order form $\Longrightarrow$ First-order form. Suppose $X\colon [0,T) \to \mathbb{R}^d$ is the solution to (4) with initial conditions $X(0) = X_0$ and $\dot{X}(0) = -\mathbb{A}(X_0)$. Define $Z\colon [0,T) \to \mathbb{R}^d$ by $Z(t) = -\dot{X}(t) - \mathbb{A}(X(t))$. Then $\dot{X}(t) = -Z(t) - \mathbb{A}(X(t))$ by definition (this is the first line of (5)). Also note that $Z(0) = -\dot{X}(0) - \mathbb{A}(X_0) = 0$, which is the Z-initial condition for (5). Now differentiating Z we have

$$\dot{Z}(t) = -\ddot{X}(t) - \frac{d}{dt}\mathbb{A}(X(t)) = \frac{1}{T-t}\dot{X}(t) = -\frac{1}{T-t}\left(Z(t) + \mathbb{A}(X(t))\right)$$

The second equality directly follows from the defining equation (4). This shows that $\begin{pmatrix} X \\ Z \end{pmatrix}$ is the solution of (5).

$\square$

3.2 Derivation of ODE as continuous-time limit

In this subsection, we derive the RevAn-ODE as continuous-time limit of (Rev-OHM) and (Rev-FEG). The derivation is in the weak sense instead of strong, since we do not rigorously derive the sup-norm convergence $\lim_{\alpha\to 0^+}\sup_{t\in[0,T]}\left\|X(t)-x_{\lfloor t/\alpha\rfloor}\right\|=0$. Similar approaches have been studied in [SBC14] for Nesterov's acceleration.

(Rev-OHM) **to RevAn-ODE.** Assume $\mathbb{A}$ is continuous monotone operator and let $\alpha>0$. Consider (Rev-OHM) with $\mathbb{T}=2\mathbb{J}_{\frac{\alpha}{2}\mathbb{A}}-\mathbb{I}$. Note that this differs from the *usual identification* $\mathbb{T}=2\mathbb{J}_{\mathbb{A}}-\mathbb{I}$ by scale, because here we need a step-size α with respect to which we take the limit. Let $x_{k+1}=\mathbb{J}_{\frac{\alpha}{2}\mathbb{A}}y_k$, so that $x_{k+1}+\frac{\alpha}{2}\mathbb{A}x_{k+1}=y_k$. Then

$$\mathbb{T}y_k=2x_{k+1}-y_k=y_k-\alpha\mathbb{A}x_{k+1}$$

and we can rewrite the z-form of (Rev-OHM) (3) as

$$\begin{aligned}\frac{\frac{z_{k+1}}{\alpha}-\frac{z_k}{\alpha}}{\alpha}&=-\frac{1}{\alpha(N-k)}\frac{z_k}{\alpha}-\frac{1}{\alpha(N-k)}\mathbb{A}x_{k+1}\\ \frac{y_{k+1}-y_k}{\alpha}&=-\frac{z_{k+1}}{\alpha}-\mathbb{A}x_{k+1}\end{aligned}$$

Identifying $\alpha k=t$, $\alpha N=T$, $y_k=Y(t)$, $\frac{z_k}{\alpha}=Z(t)$ and $x_k=X(t)$, we obtain

$$\begin{aligned}\dot{Z}(t)&=-\frac{1}{T-t}Z(t)-\frac{1}{T-t}\mathbb{A}(X(t+\alpha))\\ \dot{Y}(t)&=-Z(t+\alpha)-\mathbb{A}(X(t+\alpha))\end{aligned}$$

Assuming differentiability of X,Y,Z and taking the limit $\alpha\to 0$ gives

$$\begin{aligned}\dot{Z}(t)&=-\frac{1}{T-t}Z(t)-\frac{1}{T-t}\mathbb{A}(X(t))\\ \dot{Y}(t)&=-Z(t)-\mathbb{A}(X(t))\end{aligned}\tag{6}$$

Once we show $\dot{Y}(t)=\dot{X}(t)$ in the limit $\alpha\to 0$, (6) becomes (RevAn-ODE). From $x_{k+1}+\frac{\alpha}{2}\mathbb{A}x_{k+1}=y_k$, we have

$$\dot{X}(t)+\mathcal{O}(\alpha)=\frac{x_{k+1}-x_k}{\alpha}=\frac{y_k-y_{k-1}}{\alpha}-\frac{1}{2}\left(\mathbb{A}x_{k+1}-\mathbb{A}x_k\right)=\dot{Y}(t)+\mathcal{O}(\alpha)-\frac{1}{2}\left(\mathbb{A}(X(t+\alpha))-\mathbb{A}(X(t))\right)$$

which shows that indeed, $\dot{X}(t)=\dot{Y}(t)$ in the limit $\alpha\to 0$.

(Rev-FEG) **to RevAn-ODE.** From the definition of (Rev-FEG), we have

$$\begin{aligned}x_{k+1}&=x_{k+\frac{1}{2}}-\frac{N-k-1}{N-k}\alpha\left(\mathbb{A}x_{k+\frac{1}{2}}-\mathbb{A}x_k\right)\\ &=x_k-\alpha z_k-\alpha\mathbb{A}x_k-\frac{N-k-1}{N-k}\alpha\left(\mathbb{A}x_{k+\frac{1}{2}}-\mathbb{A}x_k\right)\end{aligned}\tag{7}$$

where $\mathbb{A}=\nabla_{\pm}\mathbf{L}$. Therefore, we can write

$$\frac{x_{k+1}-x_k}{\alpha}=-z_k-\mathbb{A}x_k-\frac{\alpha(N-k-1)}{\alpha(N-k)}\left(\mathbb{A}x_{k+\frac{1}{2}}-\mathbb{A}x_k\right)$$

$$\frac{z_{k+1} - z_k}{\alpha} = -\frac{1}{\alpha(N-k)} z_k - \frac{1}{\alpha(N-k)} \mathbb{A} x_{k+\frac{1}{2}}$$

Now identify $\alpha k = t$, $\alpha N = T$, $x_k = X(t)$ and $z_k = Z(t)$ to obtain

$$\begin{aligned} \dot{X}(t) &= -Z(t) - \mathbb{A}(X(t)) - \frac{T-t-\alpha}{T-t} \left(\mathbb{A}(X(t) + \mathcal{O}(\alpha)) - \mathbb{A}(X(t))\right) \\ \dot{Z}(t) &= -\frac{1}{T-t} Z(t) - \frac{1}{T-t} \mathbb{A}(X(t) + \mathcal{O}(\alpha)) \end{aligned}$$

where we use $x_{k+\frac{1}{2}} = x_k + \mathcal{O}(\alpha)$. Thus, under the limit $\alpha \to 0$, we obtain

$$\begin{aligned} \dot{X}(t) &= -Z(t) - \mathbb{A}(X(t)) \\ \dot{Z}(t) &= -\frac{1}{T-t} Z(t) - \frac{1}{T-t} \mathbb{A}(X(t)) \end{aligned}$$

which is (RevAn-ODE).

3.3 Convergence analysis of (RevAn-ODE)

In this subsection we state the results regarding the existence, uniqueness and Lyapunov analysis of (RevAn-ODE). Define the Lyapunov function $V \colon [0, T) \to \mathbb{R}$ by

$$V(t) = -\left\|Z(t) + \mathbb{A}(X(T))\right\|^2 + \frac{2}{T-t} \left\langle Z(t) + \mathbb{A}(X(T)), X(t) - X(T) \right\rangle$$

In §3.4, we show that

$$\dot{V}(t) = -\frac{2 \left\langle X(t) - X(T), \mathbb{A}(X(t)) - \mathbb{A}(X(T)) \right\rangle}{(T-t)^2} \leq 0$$

Lemma 3 shows $\lim_{t \to T^-} V(t) = 0$. From $Z(0) = 0$,

$$V(0) = -\left\|\mathbb{A}(X(T))\right\|^2 - \frac{2}{T} \left\langle \mathbb{A}(X(T)), X(T) - X_0 \right\rangle$$

Dividing both sides of $0 = \lim_{t \to T^-} V(t) \leq V(0)$ by $\frac{2}{T}$ and applying Lemma 1 we get the desired inequality. We state the following Theorem 1:

Theorem 1. *Let $\mathbb{A} \colon \mathbb{R}^d \to \mathbb{R}^d$ be Lipschitz continuous and monotone. For $T > 0$, the solution $X \colon [0, T) \to \mathbb{R}^d$ of the* (RevAn-ODE) *with initial conditions $X(0) = X_0, Z(0) = 0$ uniquely exists, and $X(T) = \lim_{t \to T^-} X(t)$ satisfies*

$$\left\|\mathbb{A}(X(T))\right\|^2 \leq \frac{4 \left\|X_0 - X_\star\right\|^2}{T^2}$$

where $X_\star \in \operatorname{Zer} \mathbb{A}$.

Detailed proof is deferred to §3.4.

3.4 Proof of Theorem 1

Before the formal proof we first present the well-posedness of the dynamics as well as the regularity at the terminal time $t = T$.

Well-posedness of the dynamics. We have

Proposition 4. *Suppose $\mathbb{A}\colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous and let $T > 0$. Consider the differential equation*

$$\begin{pmatrix} \dot{X}(t) \\ \dot{Z}(t) \end{pmatrix} = -\begin{pmatrix} Z(t) + \mathbb{A}(X(t)) \\ \frac{1}{T-t}Z(t) + \frac{1}{T-t}\mathbb{A}(X(t)) \end{pmatrix} \tag{8}$$

for $t \in (0,T)$, with initial conditions $X(0) = X_0$ and $Z(0) = 0$. Then there is a unique solution $\begin{pmatrix} X \\ Z \end{pmatrix} \in \mathcal{C}^1([0,T), \mathbb{R}^{2d})$ that satisfies (8) for all $t \in (0,T)$.

Proof. The right hand side of (8), as a function of t and $\begin{pmatrix} X \\ Z \end{pmatrix}$, can be rewritten as

$$F\left(t, \begin{pmatrix} X \\ Z \end{pmatrix}\right) = -\begin{pmatrix} Z + \mathbb{A}(X) \\ \frac{1}{T-t}Z + \frac{1}{T-t}\mathbb{A}(X) \end{pmatrix} = -\begin{pmatrix} I & 0 \\ 0 & \frac{1}{T-t}I \end{pmatrix}\begin{pmatrix} Z + \mathbb{A}(X) \\ Z + \mathbb{A}(X) \end{pmatrix}$$

Let $L > 0$ be the Lipschitz continuity parameter for $\mathbb{A}$. Fix $\bar{t} \in (0,T)$; then for $t \in [0,\bar{t}]$, F is Lipschitz continuous with parameter

$$\sqrt{2\left(1 + \frac{1}{(T-\bar{t})^2}\right)} \max\{1, L\}$$

in $\begin{pmatrix} X \\ Z \end{pmatrix}$. By classical ODE theory, Lipschitz continuity of F on $[0,\bar{t}]$ implies that the solution $\begin{pmatrix} X \\ Z \end{pmatrix} \in \mathcal{C}^1([0,\bar{t}], \mathbb{R}^{2d})$ uniquely exists. Since $\bar{t} \in (0,T)$ is arbitrary, the solution uniquely exists on $[0,T)$ and the proof is complete. $\square$

Corollary 1. *Suppose $\mathbb{A}\colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous and let $T > 0$. The solution $\begin{pmatrix} X \\ Z \end{pmatrix}\colon [0,T) \to \mathbb{R}^{2d}$ to (8) satisfies the differential equation*

$$\ddot{X}(t) + \frac{1}{T-t}\dot{X}(t) + \frac{d}{dt}\mathbb{A}(X(t)) = 0 \tag{9}$$

almost everywhere in $t \in [0,T)$. Conversely, if $X \in \mathcal{C}^1([0,T), \mathbb{R}^d)$ satisfies (9) almost everywhere in $t \in [0,T)$, then with $Z\colon [0,T) \to \mathbb{R}^d$ defined as $Z(t) = -\dot{X}(t) - \mathbb{A}(X(t))$, $\begin{pmatrix} X \\ Z \end{pmatrix}$ is the solution of (8).

Proof. Let L be the Lipschitz continuity parameter of $\mathbb{A}$ and take $\bar{t} \in (0,T)$. As $t \mapsto \dot{X}(t)$ is continuous on the compact interval $[0,\bar{t}]$, we have $M = \max_{t\in[0,\bar{t}]} \left\|\dot{X}(t)\right\| < \infty$. Therefore $t \mapsto X(t)$ is M-Lipschitz, which implies that $t \mapsto \mathbb{A}(X(t))$ is LM-Lipschitz; hence $\mathbb{A}(X(t))$ is differentiable almost everywhere in t. Then $\dot{X}(t) = -Z(t) - \mathbb{A}(X(t))$ is absolutely continuous in t[5] Now the equivalence between (8) and (9) can be proved via same arguments as in §3.1, with all equalities holding almost everywhere in t. $\square$

[5]RHS is the sum of $\mathcal{C}^1$ function Z and Lipschitz continuous $\mathbb{A}(X(t))$), and hence differentiable almost everywhere.

Regularity at the terminal time $t = T$. Next, we show the solution on $[0,T)$ can be continuously extended to the terminal time $t = T$ due to its favorable properties. First, we need:

Lemma 2. *Suppose $\mathbb{A}\colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous and monotone. Let $\begin{pmatrix} X \\ Z \end{pmatrix}$ be the solution to* (8) *on $[0,T)$. Then*

$$\Psi(t) = \frac{1}{(T-t)^2} \left\|\dot{X}(t)\right\|^2$$

is nonincreasing in $t \in [0,T)$. In particular, this implies

$$\left\|\dot{Z}(t)\right\|^2 = \frac{1}{(T-t)^2} \left\|\dot{X}(t)\right\|^2 \leq \frac{1}{T^2} \left\|\mathbb{A}(X_0)\right\|^2$$

Proof. Take the inner product with $\frac{1}{(T-t)^2}\dot{X}(t)$ to both sides of (9), we have for almost every $t \in [0,T)$,

$$\begin{aligned} 0 &= \frac{1}{(T-t)^2}\left\langle \ddot{X}(t), \dot{X}(t)\right\rangle + \frac{1}{(T-t)^3}\left\|\dot{X}(t)\right\|^2 + \frac{1}{(T-t)^2}\left\langle \frac{d}{dt}\mathbb{A}(X(t)), \dot{X}(t)\right\rangle \\ &= \frac{d}{dt}\left(\frac{1}{2(T-t)^2}\left\|\dot{X}(t)\right\|^2\right) + \frac{1}{(T-t)^2}\left\langle \frac{d}{dt}\mathbb{A}(X(t)), \dot{X}(t)\right\rangle \end{aligned}$$

Note that because $\dot{X}(t)$ is absolutely continuous, $\Psi(t)$ is also absolutely continuous. Integrating from 0 to t and reorganizing, we obtain the following "conservation law":

$$E \equiv \frac{1}{2T^2}\left\|\dot{X}(0)\right\|^2 = \underbrace{\frac{1}{2(T-t)^2}\left\|\dot{X}(t)\right\|^2}_{\frac{\Psi(t)}{2}} + \int_0^t \frac{1}{(T-s)^2}\left\langle \frac{d}{ds}\mathbb{A}(X(s)), \dot{X}(s)\right\rangle ds$$

(E is a constant). Note that by monotonicity of $\mathbb{A}$,

$$\left\langle \frac{d}{dt}\mathbb{A}(X(t)), \dot{X}(t)\right\rangle = \lim_{h\to 0} \frac{\langle \mathbb{A}(X(t+h)) - \mathbb{A}(X(t)), X(t+h) - X(t)\rangle}{h^2} \geq 0$$

Therefore for almost every $t \in (0,T)$,

$$\dot{\Psi}(t) = -\frac{2}{(T-t)^2}\left\langle \frac{d}{dt}A(X(t)), \dot{X}(t)\right\rangle \leq 0$$

from which we conclude that $\Psi(t)$ is nonincreasing. Finally, (8) gives $\dot{Z}(t) = \frac{1}{T-t}\dot{X}(t)$ and $\dot{X}(0) = -Z(0) - \mathbb{A}(X_0) = -\mathbb{A}(X_0)$, so we conclude that

$$\left\|\dot{Z}(t)\right\|^2 = \frac{1}{(T-t)^2}\left\|\dot{X}(t)\right\|^2 = \Psi(t) \leq \Psi(0) = \frac{1}{T^2}\left\|\mathbb{A}(X_0)\right\|^2$$

□

Lemma 3. *Suppose $\mathbb{A}\colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous and monotone. Then, given the solution $\begin{pmatrix} X \\ Z \end{pmatrix}$ to* (8) *on $[0,T)$, one can continuously extend $X(t)$ to $t = T$, and the extension $X\colon [0,T] \to \mathbb{R}^d$ satisfies*

$$\lim_{t\to T^-} \dot{X}(t) = \lim_{t\to T^-} \frac{X(t) - X(T)}{t - T} = 0$$

Proof. Let $t \in [0, T)$ and let $h > 0$ satisfy $t + h < T$. Then by Lemma 2,

$$\|X(t+h) - X(t)\| = \left\| \int_t^{t+h} \dot{X}(s) ds \right\| \le \int_t^{t+h} \left\| \dot{X}(s) \right\| ds \le \int_t^{t+h} \frac{T-s}{T} \|\mathbb{A}(X_0)\| \, ds \le h \|\mathbb{A}(X_0)\|$$

This shows that $X\colon [0, T) \to \mathbb{R}^d$ is Lipschitz continuous, which implies any sequence $\{X(t_j)\}$ with $0 < t_j \nearrow T$ is a Cauchy sequence, and such sequential limit is unique. Setting $X(T)$ to this unique limit, we see that $X\colon [0, T] \to \mathbb{R}^d$ is a (unique) continuous function extending X. Additionally, Lemma 2 gives

$$\lim_{t \to T^-} \left\| \dot{X}(t) \right\| \le \lim_{t \to T^-} \frac{T-t}{T} \|\mathbb{A}(X_0)\| = 0$$

Therefore,

$$\lim_{t \to T^-} \left\| \frac{X(t) - X(T)}{t - T} \right\| = \lim_{t \to T^-} \frac{1}{T-t} \left\| \int_t^T \dot{X}(s) ds \right\| \le \lim_{t \to T^-} \frac{1}{T-t} \int_t^T \left\| \dot{X}(s) \right\| ds = \lim_{t \to T^-} \left\| \dot{X}(t) \right\| = 0$$

where the second-to-last equality uses L'Hôpital's rule. □

We are ready to present the

Proof of Theorem 1. Recall that we define $V\colon [0, T) \to \mathbb{R}$ by

$$V(t) = -\|Z(t) + \mathbb{A}(X(T))\|^2 + 2 \left\langle Z(t) + \mathbb{A}(X(T)), \frac{1}{T-t} (X(t) - X(T)) \right\rangle$$

Differentiate V and apply the substitutions $Z = -(\dot{X} + \mathbb{A}(X))$ and $\dot{Z} = \frac{1}{T-t}\dot{X}$ to obtain

$$\begin{aligned}
\dot{V}(t) &= -2 \left\langle \dot{Z}(t), Z(t) + \mathbb{A}(X(T)) \right\rangle + 2 \left\langle \dot{Z}(t), \frac{1}{T-t} (X(t) - X(T)) \right\rangle \\
&\quad + 2 \left\langle Z(t) + \mathbb{A}(X(T)), \frac{d}{dt} \left(\frac{1}{T-t} (X(t) - X(T)) \right) \right\rangle \\
&= 2 \left\langle \frac{1}{T-t} \dot{X}(t), \dot{X}(t) + \mathbb{A}(X(t)) - \mathbb{A}(X(T)) + \frac{1}{T-t} (X(t) - X(T)) \right\rangle \\
&\quad - 2 \left\langle \dot{X}(t) + \mathbb{A}(X(t)) - \mathbb{A}(X(T)), \frac{1}{(T-t)^2} (X(t) - X(T)) + \frac{1}{T-t} \dot{X}(t) \right\rangle \\
&= -\frac{2}{(T-t)^2} \langle X(t) - X(T), \mathbb{A}(X(t)) - \mathbb{A}(X(T)) \rangle \le 0
\end{aligned}$$

On the other hand, by Lemma 3, we have $\lim_{t \to T^-} \dot{X}(t) = 0$, which implies

$$\lim_{t \to T^-} Z(t) = \lim_{t \to T^-} \left(-\dot{X}(t) - \mathbb{A}(X(t)) \right) = -\mathbb{A}(X(T))$$

and thus,

$$\lim_{t \to T^-} V(t) = \lim_{t \to T^-} \left(-\|Z(t) + \mathbb{A}(X(T))\|^2 \right) + \lim_{t \to T^-} 2 \left\langle Z(t) + \mathbb{A}(X(T)), \frac{X(t) - X(T)}{T-t} \right\rangle = 0$$

where the last equality holds because both $Z(t)+\mathbb{A}(X(T))$ and $\frac{X(t)-X(T)}{t-T}$ converges to 0 as $t \to T^-$. Therefore,

$$0 = \lim_{t \to T^-} V(t) \leq V(0) = -\left\|\mathbb{A}(X(T))\right\|^2 - \frac{2}{T}\left\langle \mathbb{A}(X(T)), X(T) - X_0 \right\rangle$$

where the last equality uses $Z(0) = 0$. Now multiplying $-\frac{T}{2}$ to both sides and applying Lemma 1, we conclude

$$\left\|\mathbb{A}(X(T))\right\|^2 \leq \frac{4\left\|X_0 - X_\star\right\|^2}{T^2}$$

proving the theorem. □

4 Generalization to differential inclusion

So far, we have analyzed the (RevAn-ODE) with respect to Lipschitz continuous $\mathbb{A}$. In specific, Theorem 1 establishes the existence of a solution and the convergence analysis under the simple assumption of Lipschitz continuity of $\mathbb{A}$.

In this section, we deal with its extension to *differential inclusion*, which is a generalized continuous-time model covering the case of general (possibly set-valued) maximally monotone operator $\mathbb{A}$:[6]

$$\begin{pmatrix} \dot{X}(t) \\ \dot{Z}(t) \end{pmatrix} \in -\begin{pmatrix} Z(t) + \mathbb{A}(X(t)) \\ \frac{1}{T-t}Z(t) + \frac{1}{T-t}\mathbb{A}(X(t)) \end{pmatrix} \tag{10}$$

We say that $(X(t), Z(t))$ is a solution to this differential inclusion if it satisfies (10) almost everywhere in t.[7] Although (10) is technically not an ODE, with a slight abuse of notation, we will often refer to it as (generalized) RevAn-ODE throughout the section.

4.1 Existence of a solution to the generalized RevAn-ODE

Proposition 5. *Let $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be maximally monotone and let $T > 0$. Given the initial conditions $X(0) = X_0 \in \operatorname{dom} \mathbb{A}$ and $Z(0) = 0$, there exists a solution to the generalized RevAn-ODE, i.e., an absolutely continuous curve $\begin{pmatrix} X \\ Z \end{pmatrix} : [0, T) \to \mathbb{R}^{2d}$ that satisfies (10) for $t \in (0, T)$ almost everywhere.*

We proceed with similar ideas as in [AC84, Chapter 3] and [SPR23, Appendix B.2], i.e., we construct a sequence $\{(X_\delta(t), Z_\delta(t))\}_{\delta>0}$ of solutions to ODEs approximating (10), with $\mathbb{A}$ replaced by Yosida approximations $\mathbb{A}_\delta$ (which are much better-behaved, being single-valued and Lipschitz continuous). Then we show that the limit of $(X_\delta(t), Z_\delta(t))$ in $\delta \to 0$ is a solution to the original inclusion (10).

Below we present some well-known facts needed to rigorously establish the approximation argument.

[6] Such a methodology for modeling nonsmooth first-order methods was earlier studied by [AAD18, FRV18, YZLS19, FRV23, Li24].

[7] Unlike in ODEs, we do not require X, Z to be differentiable everywhere, but only require absolute continuity, which implies differentiability almost everywhere.

Lemma 4. *Let $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be maximally monotone. Then for $\delta > 0$, the* Yosida approximation operators

$$\mathbb{A}_\delta = \frac{1}{\delta}\left(\mathbb{I} - \mathbb{J}_{\delta\mathbb{A}}\right) = \frac{1}{\delta}\left(\mathbb{I} - (\mathbb{I} + \delta\mathbb{A})^{-1}\right)$$

satisfy the followings:

(i) $\forall x \in \mathbb{R}^d$, $\mathbb{A}_\delta(x) \in \mathbb{A}(\mathbb{J}_{\delta\mathbb{A}} x)$.

(ii) $\mathbb{A}_\delta\colon \mathbb{R}^d \to \mathbb{R}^d$ *is singled-valued, $\frac{1}{\delta}$-Lipschitz continuous and maximally monotone.*

(iii) $\|\mathbb{A}_\delta(x)\| \le \|m(\mathbb{A}(x))\|$.

(iv) $\lim_{\delta\to 0+} \mathbb{A}_\delta(x) = m(\mathbb{A}(x)) := \arg\min_{u\in\mathbb{A}(x)} \|u\|$.[8]

Proof. See [AC84, Chapter 3.1, Theorem 2]. □

Lemma 5. *Let $\{x_n(\cdot)\}_n$ be a sequence of absolutely continuous functions $x_n\colon I \to \mathbb{R}^d$, where $I \subset \mathbb{R}$ is an interval of finite length. Suppose that*

(i) $\forall t \in I$, *the set* $\{x_n(t)\}_n$ *is bounded.*

(ii) There exists $M > 0$ such that $\|\dot{x}_n(t)\| \le M$ almost everywhere in $t \in I$.

Then there exists a subsequence $x_{n_k}(\cdot)$ such that

(i) $x_{n_k}(\cdot)$ uniformly converges to $x(\cdot)$ over compact subsets of I.

(ii) $x(\cdot)$ is absolutely continuous, so $\dot{x}$ exists almost everywhere.

(iii) $\dot{x}_{n_k}(\cdot)$ converges weakly to $\dot{x}(\cdot)$ in $L^2(I, \mathbb{R}^d)$.

Proof. This is a simplified version of [AC84, Chapter 0.3, Theorem 4]. For completeness, we outline the proof below.

The family $\{x_n(\cdot)\}_n$ is pointwisely bounded (condition (i)) and equicontinuous as

$$\|x_n(t) - x_n(s)\| = \left\|\int_s^t \dot{x}_n(\tau)\, d\tau\right\| \le M|t-s| \qquad \forall s, t \in I$$

By Arzelà-Ascoli theorem, there exists a subsequence $x_{n_k}(\cdot)$ such that $x_{n_k} \to x$ uniformly over compact subsets of I, where $x\colon I \to \mathbb{R}^d$ is continuous (so far we have not shown absolute continuity). Now observe that $\{\dot{x}_{n_k}(\cdot)\}_k$ is a bounded set within $L^\infty(I, \mathbb{R}^d) = L^1(I, \mathbb{R}^d)^*$. Therefore, by Banach-Alaoglu theorem, $\dot{x}_{n_k}(\cdot)$ again has a subsequence that converges in the weak-$*$ sense. By replacing n_k with this subsequence if necessary, assume without loss of generality that $\dot{x}_{n_k} \xrightarrow{\text{w}*} y \in L^\infty(I, \mathbb{R}^d)$. This means that for any $c \in L^1(I, \mathbb{R}^d)$,

$$\lim_{k\to\infty} \int_I \langle \dot{x}_{n_k}(\tau), c(\tau)\rangle\, d\tau = \int_I \langle y(\tau), c(\tau)\rangle\, d\tau \tag{11}$$

For any $s, t \in I$ and $\mathbf{a} \in \mathbb{R}^d$, taking $c(\tau) = 1_{[s,t]}(\tau)\mathbf{a}$ in particular, we have

$$\lim_{k\to\infty} \int_s^t \langle \dot{x}_{n_k}(\tau), \mathbf{a}\rangle\, d\tau = \int_s^t \langle y(\tau), \mathbf{a}\rangle\, d\tau = \left\langle \int_s^t y(\tau)\, d\tau, \mathbf{a}\right\rangle$$

[8] The minimum norm element of the set $\mathbb{A}(x)$ is well-defined because $\mathbb{A}(x)$ is a closed and convex set due to maximal monotonicity of $\mathbb{A}$.

On the other hand, the left hand side equals $\lim_{k\to\infty} \langle x_{n_k}(t) - x_{n_k}(s), \mathbf{a}\rangle = \langle x(t) - x(s), \mathbf{a}\rangle$. Since $\mathbf{a} \in \mathbb{R}^d$ is arbitrary, this shows

$$x(t) - x(s) = \int_s^t y(\tau)\, d\tau \qquad y \in L^\infty(I, \mathbb{R}^d)$$

so x is absolutely continuous (in fact, Lipschitz continuous) and $\dot{x} = y$ almost everywhere. Finally, note that $L^2(I, \mathbb{R}^d) \subset L^1(I, \mathbb{R}^d)$ (as I is of finite length), so we can take $c \in L^2(I, \mathbb{R}^d)$ in (11), showing that $\dot{x}_{n_k} \to y$ weakly in $L^2(I, \mathbb{R}^d)$.

□

Using the above results, we can now prove Theorem 5. For $\delta > 0$, the ODE

$$\begin{pmatrix} \dot{X}_\delta(t) \\ \dot{Z}_\delta(t) \end{pmatrix} = -\begin{pmatrix} Z_\delta(t) + \mathbb{A}_\delta(X_\delta(t)) \\ \frac{1}{T-t} Z_\delta(t) + \frac{1}{T-t}\mathbb{A}_\delta(X_\delta(t)) \end{pmatrix} \tag{12}$$

with initial conditions $X_\delta(0) = X_0 \in \operatorname{dom} \mathbb{A}$ and $Z_\delta(0) = 0$ has a unique $\mathcal{C}^1$ solution $(X_\delta(t), Z_\delta(t))$ for $t \in [0, T]$ by Theorem 4 and Lemma 3. Additionally by Lemma 2, X_δ and Z_δ satisfy

$$\left\|\dot{X}_\delta(t)\right\| \le \frac{T-t}{T} \|m(\mathbb{A}(X_0))\| \le \|m(\mathbb{A}(X_0))\| \qquad \left\|\dot{Z}_\delta(t)\right\| \le \frac{1}{T} \|m(\mathbb{A}(X_0))\| \tag{13}$$

Therefore, we can apply Lemma 5 to obtain a positive sequence $\delta_n \to 0$ such that

- $\begin{pmatrix} X_{\delta_n} \\ Z_{\delta_n} \end{pmatrix} \to \begin{pmatrix} X \\ Z \end{pmatrix}$ uniformly on $[0, T]$, where $\begin{pmatrix} X \\ Z \end{pmatrix}$ is absolutely continuous,
- $\begin{pmatrix} \dot{X}_{\delta_n} \\ \dot{Z}_{\delta_n} \end{pmatrix} \to \begin{pmatrix} \dot{X} \\ \dot{Z} \end{pmatrix}$ weakly in $L^2([0, T], \mathbb{R}^{2d})$.

Now, we wish to show that $(X(t), Z(t))$ is a solution to (10). Define $G_\delta, G \colon [0, T] \to \mathbb{R}^d$ as

$$G_\delta(t) = Z_\delta(t) + \dot{X}_\delta(t) \qquad G(t) = Z(t) + \dot{X}(t)$$

By construction,

$$G_{\delta_n}(\cdot) \to G(\cdot) \quad \textit{weakly} \text{ in } L^2([0, T], \mathbb{R}^d)$$

Note that because (X_δ, Z_δ) solves (12) and Lemma 4 (i) holds, we have

$$G_{\delta_n}(t) = -\mathbb{A}_\delta(X_\delta(t)) \in -\mathbb{A}(\mathbb{J}_{\delta_n \mathbb{A}}(X_{\delta_n}(t)))$$

On the other hand, because $Z_{\delta_n}(0) = 0$, by (13) we have

$$\|Z_{\delta_n}(t)\| = \left\| \int_0^t \dot{Z}_{\delta_n}(s) ds \right\| \le \int_0^t \left\|\dot{Z}_{\delta_n}(s)\right\| ds \le \int_0^t \frac{\|\mathbb{A}_{\delta_n}(X_0)\|}{T} ds \le \|\mathbb{A}_{\delta_n}(X_0)\| \le \|m(\mathbb{A}(X_0))\|$$

Together with the norm bound on $\dot{X}$ in (13), this implies that for all $t \in [0, T]$,

$$\begin{aligned} &\|X_{\delta_n}(t) - \mathbb{J}_{\delta_n \mathbb{A}}(X_{\delta_n}(t))\| = \delta_n \|\mathbb{A}_{\delta_n}(X_{\delta_n}(t))\| \\ &= \delta_n \left\| Z_{\delta_n}(t) + \dot{X}_{\delta_n}(t) \right\| \le \delta_n \left(\|Z_{\delta_n}(t)\| + \|\dot{X}_{\delta_n}(t)\| \right) \le 2\delta_n \|m(\mathbb{A}(X_0))\| \end{aligned}$$

As $X_{\delta_n}(\cdot)$ converges uniformly to $X(\cdot)$, the above result shows that $\mathbb{J}_{\delta_n \mathbb{A}}(X_{\delta_n}(\cdot))$ uniformly converges to $X(\cdot)$ as well. In particular,

$$\mathbb{J}_{\delta_n \mathbb{A}}(X_{\delta_n}(\cdot)) \to X(\cdot) \quad \textit{strongly} \text{ in } L^2([0,T],\mathbb{R}^d)$$

Now, define $\mathcal{A}\colon L^2([0,T],\mathbb{R}^d) \to L^2([0,T],\mathbb{R}^d)$ by

$$\mathcal{A}(y) = \left\{ u \in L^2([0,T],\mathbb{R}^d) \mid u(t) \in \mathbb{A}(y(t)) \text{ for a.e. } t \in [0,T] \right\}$$

$\mathcal{A}$ is monotone because $\mathbb{A}$ is monotone: if $u \in \mathcal{A}(y), v \in \mathcal{A}(z)$ then

$$\langle u - v, y - z \rangle_{L^2([0,T],\mathbb{R}^d)} = \int_0^T \langle u(t) - v(t), y(t) - z(t) \rangle \, dt \geq 0$$

since $u(t) \in \mathbb{A}(y(t))$, $v(t) \in \mathbb{A}(z(t))$ a.e. in $t \in [0,T]$. If $\mathcal{I}\colon L^2([0,T],\mathbb{R}^d) \to L^2([0,T],\mathbb{R}^d)$ is the identity operator, then $\mathcal{I} + \mathcal{A}$ is surjective: for any $u \in L^2([0,T],\mathbb{R}^d)$, we have $y(t) = \mathbb{J}_{\mathbb{A}}(u(t)) \in L^2([0,T],\mathbb{R}^d)$ because $\mathbb{J}_{\mathbb{A}}$ is nonexpansive, and then $u \in \mathcal{A}(y)$ by construction. By Minty's surjectivity theorem [Min62], this implies that $\mathcal{A}$ is maximally monotone. Now because $\mathbb{J}_{\delta_n \mathbb{A}}(X_{\delta_n})$ converges to X strongly and $-G_{\delta_n} \in \mathcal{A}(\mathbb{J}_{\delta_n \mathbb{A}}(X_{\delta_n}))$ converges to $-G$ weakly, and the graph of a maximally monotone operator is closed under the strong-weak topology [BC17, Proposition 20.38], we conclude that $-G \in \mathcal{A}(X)$, i.e.,

$$G(t) = Z(t) + \dot{X}(t) \in -\mathbb{A}(X(t)) \iff \dot{X}(t) \in -\left(Z(t) + \mathbb{A}(X(t))\right) \quad \text{a.e. in } t \in [0,T]$$

Finally, because $(T-t)\dot{Z}_\delta(t) = \dot{X}_\delta(t)$ for all $\delta > 0$ we have

$$(T-t)\dot{Z}_{\delta_n} = \dot{X}_{\delta_n} \xrightarrow{\text{w}} \dot{X} \quad \text{in } L^2([0,T],\mathbb{R}^d)$$

On the other hand, we had $\dot{Z}_{\delta_n} \xrightarrow{\text{w}} \dot{Z}$ in $L^2([0,T],\mathbb{R}^d)$, which implies

$$(T-t)\dot{Z}_{\delta_n} \xrightarrow{\text{w}} (T-t)\dot{Z}$$

because for any $c(t) \in L^2([0,T],\mathbb{R}^d)$, we have $(T-t)c(t) \in L^2([0,T],\mathbb{R}^d)$ as well. By uniqueness of the (weak) limit, we have $\dot{X} = (T-t)\dot{Z}$ a.e., so

$$\dot{Z}(t) = \frac{1}{T-t}\dot{X}(t) \in -\frac{1}{T-t}\left(Z(t) + \mathbb{A}(X(t))\right) \qquad \text{a.e. in } t \in [0,T]$$

This shows that $\begin{pmatrix} X \\ Z \end{pmatrix}$ is indeed a solution of (10).

4.2 Behavior at the terminal time $t = T$

So far, we have successfully constructed an absolutely continuous solution $\begin{pmatrix} X \\ Z \end{pmatrix} : [0,T] \to \mathbb{R}^{2d}$. In this section we show that $X(t), Z(t)$ have two very favorable properties at the terminal time: X has left derivative 0, and $Z(T) \in -\mathbb{A}(X(T))$. To show this, we need:

Lemma 6. *Let $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be maximally monotone and let $\begin{pmatrix} X \\ Z \end{pmatrix}$ be the solution of* (10) *constructed in Theorem 5. Then for almost every $t \in [0,T)$,*

$$\frac{1}{(T-t)^2}\left\|\dot{X}(t)\right\|^2 \leq \frac{1}{T^2}\left\|m(\mathbb{A}(X_0))\right\|^2$$

Proof. Let $\delta_n > 0$ be the sequence taken in the proof of Theorem 5, for which the solutions $(X_{\delta_n}, Z_{\delta_n})$ to (12) with $\delta = \delta_n$ converge uniformly to (X, Z) and $(\dot{X}_{\delta_n}, \dot{Z}_{\delta_n})$ converge weakly to $(\dot{X}, \dot{Z})$. We have shown $\left\|\dot{X}_{\delta_n}(t)\right\| \leq \frac{T-t}{T}\left\|m(\mathbb{A}(X_0))\right\|$ in (13). Thus, the proof is done once we show

$$\left\|\dot{X}(t)\right\| \leq \limsup_{n\to\infty}\left\|\dot{X}_{\delta_n}(t)\right\|$$

for almost every $t \in [0, T]$. In fact, the above display is straightforward to show if $\dot{X}$ were the pointwise limit of $\dot{X}_{\delta_n}$, but it is not the case; it is the weak limit. We need a careful argument, as presented below.

Let D be any measurable subset of $[0, T]$. Since $\dot{X}_{\delta_{n_k}} \to \dot{X}$ weakly in $L^2([0, T], \mathbb{R}^d)$ and $\chi_D\dot{X} \in L^2([0, T], \mathbb{R}^d)$, we have

$$\begin{aligned}\int_D \left\|\dot{X}(t)\right\|^2 dt &= \int_0^T \left\langle \dot{X}(t), \chi_D(t)\dot{X}(t)\right\rangle dt = \lim_{n\to\infty}\int_0^T \left\langle \dot{X}_{\delta_n}(t), \chi_D(t)\dot{X}(t)\right\rangle dt \\ &= \lim_{n\to\infty}\int_D \left\langle \dot{X}_{\delta_n}(t), \dot{X}(t)\right\rangle dt \leq \limsup_{n\to\infty}\int_D \left\|\dot{X}_{\delta_n}(t)\right\|\left\|\dot{X}(t)\right\| dt\end{aligned} \tag{14}$$

Now from $\left\|\dot{X}_{\delta_n}(\cdot)\right\| \leq \left\|m(\mathbb{A}(X_0))\right\|$ and $\left\|\dot{X}(\cdot)\right\| \in L^2([0, T], \mathbb{R}^d) \subset L^1([0, T], \mathbb{R}^d)$, we obtain

$$\left\|\dot{X}_{\delta_n}(\cdot)\right\|\left\|\dot{X}(\cdot)\right\| \leq \left\|m(\mathbb{A}(X_0))\right\|\left\|\dot{X}(\cdot)\right\| \in L^1([0, T], \mathbb{R}^d)$$

Thus by reverse Fatou Lemma,

$$\limsup_{n\to\infty}\int_D \left\|\dot{X}_{\delta_n}(t)\right\|\left\|\dot{X}(t)\right\| dt \leq \int_D \limsup_{n\to\infty}\left\|\dot{X}_{\delta_n}(t)\right\|\left\|\dot{X}(t)\right\| dt$$

Combining the above inequality with (14) we obtain

$$\int_D \left(\limsup_{n\to\infty}\left\|\dot{X}_{\delta_n}(t)\right\| - \left\|\dot{X}(t)\right\|\right)\left\|\dot{X}(t)\right\| dt \geq 0$$

As D was an arbitrary measurable subset of $[0, T]$, we conclude that for almost every $t \in [0, T]$,

$$\left(\limsup_{n\to\infty}\left\|\dot{X}_{\delta_n}(t)\right\| - \left\|\dot{X}(t)\right\|\right)\left\|\dot{X}(t)\right\| \geq 0 \quad\Longrightarrow\quad \left\|\dot{X}(t)\right\| \leq \limsup_{n\to\infty}\left\|\dot{X}_{\delta_n}(t)\right\|$$

□

Proposition 6. *Let $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be maximally monotone and let $\begin{pmatrix} X \\ Z \end{pmatrix}$ be the solution of (10) constructed in Theorem 5. Then the following holds true:*

$$\lim_{t\to T^-}\frac{X(t) - X(T)}{t - T} = 0 \qquad -Z(T) \in \mathbb{A}(X(T))$$

Proof. Using Lemma 6, we obtain

$$\lim_{t\to T^-}\left\|\frac{X(t) - X(T)}{t - T}\right\| \leq \lim_{t\to T^-}\int_t^T \frac{\left\|\dot{X}(s)\right\|}{T - t} ds \leq \lim_{t\to T^-}\int_t^T \frac{\left\|\dot{X}(s)\right\|}{T - s} ds \leq \lim_{t\to T^-}\int_t^T \frac{\left\|m(\mathbb{A}(X_0))\right\|}{T} ds = 0$$

which proves the first property.

Next, because $(X(t), Z(t))$ satisfies (10) a.e. and $\|\dot{X}(t)\| \le \frac{T-t}{T} \|m(\mathbb{A}(X_0))\|$ a.e. (by Lemma 6), we can take a sequence $t_k \in (0, T)$ such that $\lim_{k\to\infty} t_k = T$ and

$$-\left(\dot{X}(t_k) + Z(t_k)\right) \in \mathbb{A}(X(t_k)) \qquad \left\|\dot{X}(t_k)\right\| \le \frac{T - t_k}{T} \|\mathbb{A}(X_0)\|$$

Then we have $\lim_{k\to\infty} \left\|\dot{X}(t_k)\right\| \le \lim_{k\to\infty} \frac{T-t_k}{T} \|\mathbb{A}(X_0)\| = 0$. On the other hand, $X(t_k) \to X(T)$ and $Z(t_k) \to Z(T)$ because X, Z are continuous. Finally, because the graph of $\mathbb{A}$ is closed in $\mathbb{R}^d \times \mathbb{R}^d$ [BC17, Proposition 20.38], we conclude

$$-Z(T) = -\lim_{k\to\infty} \left(\dot{X}(t_k) + Z(t_k)\right) \in \mathbb{A}\left(\lim_{k\to\infty} X(t_k)\right) = \mathbb{A}(X(T))$$

proving the proposition.

□

4.3 Convergence analysis

Based on the previous analyses, we can prove that the constructed solution has a Lyapunov function similar to that of Theorem 1 for the case of Lipschitz continuous operators

Theorem 2. *Let $\mathbb{A}\colon \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be maximally monotone and* $\operatorname{Zer} \mathbb{A} \neq \emptyset$. *Let $\begin{pmatrix} X \\ Z \end{pmatrix}$ be the solution of* (10) *constructed in Theorem 5. Then the function $V\colon [0, T) \to \mathbb{R}$ defined by*

$$V(t) = -\|Z(t) - Z(T)\|^2 + \frac{2}{T-t}\langle Z(t) - Z(T), X(t) - X(T)\rangle$$

is nonincreasing, and $\lim_{t\to T^-} V(t) = 0$. *Furthermore, for $X_\star \in \operatorname{Zer} \mathbb{A}$,*

$$\|m(\mathbb{A}(X(T)))\|^2 \le \|Z(T)\|^2 \le \frac{4\|X_0 - X_\star\|^2}{T^2}$$

where $m(\mathbb{A}(X(T)))$ is the minimum norm element of $\mathbb{A}(X(T))$.

Proof of Theorem 2. Let $\delta_n > 0$ be the sequence taken in the proof of Theorem 5. Let

$$V_{\delta_n}(t) = -\|Z_{\delta_n}(t) - Z_{\delta_n}(T)\|^2 + \frac{2}{T-t}\langle Z_{\delta_n}(t) - Z_{\delta_n}(T), X_{\delta_n}(t) - X_{\delta_n}(T)\rangle$$

for $t \in [0, T)$. As $\begin{pmatrix} X_{\delta_n} \\ Z_{\delta_n} \end{pmatrix}$ converges uniformly on $[0, T]$, for all $t \in [0, T)$ we have

$$V(t) = \lim_{n\to\infty} V_{\delta_n}(t)$$

Observe that $V_{\delta_n}(\cdot)$ is nonincreasing on $[0, T)$ for each n (by the result of Theorem 1 and $Z_{\delta_n}(T) = -\mathbb{A}(X_{\delta_n}(T))$, which follows from Lemma 6 and uniqueness of the solution in the Lipschitz case). Therefore, for any $s, t \in [0, T)$ such that $s < t$, we have

$$V(s) = \lim_{n\to\infty} V_{\delta_n}(s) \ge \lim_{n\to\infty} V_{\delta_n}(t) = V(t)$$

which shows that V is nonincreasing. Furthermore, from $\lim_{t\to T^-} Z(t) = Z(T)$ (continuity of Z) and Lemma 6, we have

$$\lim_{t\to T^-} V(t) = -\left\|\lim_{t\to T^-} Z(t) - Z(T)\right\|^2 - 2\left\langle \lim_{t\to T^-} Z(t) - Z(T), \lim_{t\to T^-} \frac{X(t)-X(T)}{t-T}\right\rangle = 0$$

Therefore

$$0 = \lim_{t\to T^-} V(t) \le V(0) = -\|Z(T)\|^2 - \frac{2}{T}\langle Z(T), X_0 - X(T)\rangle$$

and Lemma 1 gives

$$\|Z(T)\|^2 \le \frac{4\|X_0 - X_\star\|^2}{T^2}$$

Finally, because $-Z(T) \in \mathbb{A}(X(T))$ by Lemma 6, the left hand side is lower bounded by $\|m(\mathbb{A}(X(T)))\|^2$, which gives the desired convergence rate.

□

5 Discrete-time analyses of reversal algorithms

In this section we present discrete-time analyses of reversal algorithms. In §5.1 we clarify all essential correspondence between discrete and continuous-time quantities. In §5.2, we present the convergence analysis of (Rev-FEG). In §5.3, we present the convergence analysis of (Rev-OHM), showing that it is another exact optimal algorithm for solving nonexpansive fixed-point problems.

5.1 Translating our continuous-time analysis to discrete-time analysis

In this subsection, we overview how the continuous-time analysis presented above corresponds to respective analyses of (Rev-OHM) and (Rev-FEG). Precisely, we verify that under the identifcation of terms from §3.2, the following holds true in the asymptotic sense:

(i) The discrete Lyapunov function V_k corresponds to $V(t)$

(ii) The consecutive difference $V_{k+1} - V_k$ corresponds to $\dot{V}(t)$

Analysis of (Rev-FEG)**.** As above, we use the identification $\alpha k = t$, $\alpha N = T$, $x_k = X(t)$ and $z_k = Z(t)$.

(i) $\frac{V_k}{\alpha} \longleftrightarrow V(t)$.

The Lyapunov function of (Rev-FEG) is

$$V_k = -\alpha\|z_k + \mathbb{A}x_N\|^2 + \frac{2}{N-k}\langle z_k + \mathbb{A}x_N, x_k - x_N\rangle$$

Dividing both sides by α and applying the identifcations, we have

$$\begin{aligned}\frac{V_k}{\alpha} &= -\|z_k + \mathbb{A}x_N\|^2 + \frac{2}{\alpha(N-k)}\langle z_k + \mathbb{A}x_N, x_k - x_N\rangle \\ &= -\|Z(t) + \mathbb{A}(X(T))\|^2 + \frac{2}{T-t}\langle Z(t) + \mathbb{A}(X(T)), X(t) - X(T)\rangle\end{aligned}$$

we get the desired correspondence.

(ii) $\frac{\frac{1}{\alpha}V_{k+1}-\frac{1}{\alpha}V_k}{\alpha} \longleftrightarrow \dot{V}(t)$.

The Lyapunov analysis of (Rev-FEG) establishes

$$V_{k+1}-V_k = -\frac{2}{(N-k)(N-k-1)}\left\langle \mathbb{A}x_N - \mathbb{A}x_{k+\frac{1}{2}}, x_N - x_{k+\frac{1}{2}}\right\rangle \\ -\frac{1}{\alpha(N-k)^2}\left(\|x_{k+\frac{1}{2}}-x_k\|^2 - \alpha^2\|\mathbb{A}x_{k+\frac{1}{2}}-\mathbb{A}x_k\|^2\right)$$

Because $x_{k+\frac{1}{2}} = x_k + \mathcal{O}(\alpha)$, dividing the both sides by α^2, we obtain

$$\begin{aligned}\frac{\frac{1}{\alpha}V_{k+1}-\frac{1}{\alpha}V_k}{\alpha} &= -\frac{2}{\alpha^2(N-k)(N-k-1)}\left\langle \mathbb{A}x_N - \mathbb{A}x_{k+\frac{1}{2}}, x_N - x_{k+\frac{1}{2}}\right\rangle \\ &\quad -\frac{1}{\alpha^2(N-k)^2}\left(\frac{1}{\alpha}\|x_{k+\frac{1}{2}}-x_k\|^2 - \alpha\|\mathbb{A}x_{k+\frac{1}{2}}-\mathbb{A}x_k\|^2\right) \\ &= -\frac{2}{(T-t)^2+\mathcal{O}(\alpha)}\left\langle \mathbb{A}(X(T)) - \mathbb{A}(X(t)+\mathcal{O}(\alpha)), X(T)-X(t)+\mathcal{O}(\alpha)\right\rangle \\ &\quad \underbrace{-\frac{1}{(T-t)^2}\left(\frac{1}{\alpha}\mathcal{O}(\alpha^2) - \alpha\|\mathbb{A}(X(t)+\mathcal{O}(\alpha)) - \mathbb{A}(X(t))\|^2\right)}_{\mathcal{O}(\alpha)}\end{aligned}$$

Taking $\alpha \to 0$, we have

$$\text{(RHS)} = -\frac{2}{(T-t)^2}\left\langle X(t)-X(T), \mathbb{A}(X(t)) - \mathbb{A}(X(T))\right\rangle = \dot{V}(t)$$

which gives the desired correspondence.

Analysis of (Rev-OHM)**.** In §3.2 we consider (Rev-OHM) with $\mathbb{T} = 2\mathbb{J}_{\frac{\alpha}{2}\mathbb{A}} - \mathbb{I}$ for continuous monotone $\mathbb{A}$ to derive the continuous-time limit, and we have the identities $x_{k+1} = \mathbb{J}_{\frac{\alpha}{2}\mathbb{A}}y_k$ and $x_{k+1} + \frac{\alpha}{2}\mathbb{A}x_{k+1} = y_k$. Therefore, the corresponding Lyapunov analysis of (Rev-OHM) should use $g_k = \frac{\alpha}{2}\mathbb{A}x_k$ (instead of $g_k = \widetilde{\mathbb{A}}x_k$). Recall that we identify $\alpha k = t$, $\alpha N = T$, $x_k = X(t)$, $y_k = Y(t)$ and $\frac{z_k}{\alpha} = Z(t)$.

(i) $\frac{V_k}{\alpha^2} \longleftrightarrow V(t)$.

Replacing $g_N = \widetilde{\mathbb{A}}x_N$ with $\frac{\alpha}{2}\mathbb{A}x_N$ in the Lyapunov function of (Rev-OHM) gives

$$V_k = -\frac{N-k-1}{N-k}\|z_k + \alpha\mathbb{A}x_N\|^2 + \frac{2}{N-k}\langle z_k + \alpha\mathbb{A}x_N, y_k - y_{N-1}\rangle$$

Therefore

$$\begin{aligned}\frac{V_k}{\alpha^2} &= -\left(1 - \frac{\alpha}{\alpha(N-k)}\right)\left\|\frac{z_k}{\alpha} + \mathbb{A}x_N\right\|^2 + \frac{2}{\alpha(N-k)}\left\langle \frac{z_k}{\alpha} + \mathbb{A}x_N, y_k - y_{N-1}\right\rangle \\ &= -\left(1+\mathcal{O}(\alpha)\right)\|Z(t) + \mathbb{A}(X(T))\|^2 + \frac{2}{T-t}\langle Z(t) + \mathbb{A}(X(T)), Y(t) - Y(T-\alpha)\rangle\end{aligned}$$

From the identity $x_{k+1} + \frac{\alpha}{2}\mathbb{A}x_{k+1} = y_k$ we have $Y(t) = X(t+\alpha) + \mathcal{O}(\alpha)$, so in the limit $\alpha \to 0$, $Y(t) = X(t)$. Then the above equation establishes the desired correspondence, as

$$\text{(RHS)} = -\|Z(t) + \mathbb{A}(X(T))\|^2 + \frac{2}{T-t}\langle Z(t) + \mathbb{A}(X(T)), X(t) - X(T)\rangle$$

(ii) $\frac{\frac{1}{\alpha^2}V_{k+1}-\frac{1}{\alpha^2}V_k}{\alpha} \longleftrightarrow \dot{V}(t)$.

The Lyapunov analysis of (Rev-OHM) establishes

$$V_{k+1} - V_k = -\frac{4}{(N-k)(N-k-1)} \langle x_{k+1} - x_N, g_{k+1} - g_N \rangle$$

Replacing g_k with $\frac{\alpha}{2}\mathbb{A}x_k$, dividing both sides by α^3 and applying the identifications, we obtain

$$\begin{aligned}\frac{\frac{1}{\alpha^2}V_{k+1} - \frac{1}{\alpha^2}V_k}{\alpha} &= -\frac{4}{\alpha^3(N-k)(N-k-1)} \left\langle x_{k+1} - x_N, \frac{\alpha}{2}\mathbb{A}x_{k+1} - \frac{\alpha}{2}\mathbb{A}x_N \right\rangle \\ &= -\frac{2}{(\alpha N - \alpha k)(\alpha N - \alpha k - \alpha)} \langle x_{k+1} - x_N, \mathbb{A}(X(t+\alpha) - \mathbb{A}(X(T)) \rangle \\ &= -\frac{2}{(T-t)^2 + \mathcal{O}(\alpha)} \langle X(t+\alpha) - X(T), \mathbb{A}(X(t+\alpha) - \mathbb{A}(X(T)) \rangle\end{aligned}$$

Taking $\alpha \to 0$, we have

$$\text{(RHS)} = -\frac{2}{(T-t)^2} \langle X(t) - X(T), \mathbb{A}(X(t)) - \mathbb{A}(X(T)) \rangle = \dot{V}(t)$$

which gives the desired correspondence.

5.2 Analysis of (Rev-FEG) for minimax problems

In this subsection, we present the convergence analysis of (Rev-FEG). Algorithms such as (OHM) and (Rev-OHM) are sometimes referred to as *implicit methods* since they can be expressed using resolvents, as discussed in §5.3.2. The results of this section show that (Rev-OHM) has an *explicit* counterpart (Rev-FEG), which uses direct gradient evaluations instead.

First, we formally state the convergence result.

Theorem 3. *Let* $\mathbf{L}\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *be convex-concave and* L*-smooth. For* $N \geq 1$ *fixed, if* $0 < \alpha \leq \frac{1}{L}$, (Rev-FEG) *exhibits the rate*

$$\|\nabla \mathbf{L}(x_N)\|^2 = \|\nabla_{\pm} \mathbf{L}(x_N)\|^2 \leq \frac{4\|x_0 - x_\star\|^2}{\alpha^2 N^2}$$

where $x_\star \in \operatorname{Zer} \nabla_{\pm}\mathbf{L}$ *is a solution.*

Proof outline. Define $g_N = \nabla_{\pm}\mathbf{L}(x_N)$. In §5.2.1, we show that

$$V_k = -\alpha \|z_k + g_N\|^2 + \frac{2}{N-k} \langle z_k + g_N, x_k - x_N \rangle$$

is nonincreasing in $k = 0, 1, \ldots, N-1$, and $V_{N-1} \geq 0$. This implies $0 \leq V_0 = -\alpha \|g_N\|^2 + \frac{2}{N} \langle g_N, x_0 - x_N \rangle$. Finally, divide both sides by $\frac{2}{N}$ and apply Lemma 1. □

Comparison of time-forward and time-reversal algorithms. We observe that the reversal algorithms require N or T to be specified in advance, while the time-forward ones do not. In this respect, time-forward algorithms are advantageous over reversal algorithms. On the other hand, we observe that for problems involving strongly monotone operators, reversal algorithms exhibit much faster convergence than time-forward algorithms in the earlier iterations. As the time-forward and reversal algorithms share the same worst-case guarantee but display distinct characteristics, determining the best choice of algorithm may require considering other criterion that depend on the particular application scenario.

5.2.1 Lyapunov analysis of (Rev-FEG)

In this section, we denote $\mathbb{A}x = \nabla_{\pm}\mathbf{L}(x)$ for simplicity. Recall the update rules of (Rev-FEG):

$$
\begin{aligned}
x_{k+\frac{1}{2}} &= x_k - \alpha z_k - \alpha \mathbb{A}x_k \\
x_{k+1} &= x_{k+\frac{1}{2}} - \frac{N-k-1}{N-k}\alpha\left(\mathbb{A}x_{k+\frac{1}{2}} - \mathbb{A}x_k\right) \\
z_{k+1} &= \frac{N-k-1}{N-k} z_k - \frac{1}{N-k}\mathbb{A}x_{k+\frac{1}{2}}
\end{aligned}
\tag{Rev-FEG}
$$

and the Lyapunov function

$$
V_k = -\alpha \left\|z_k + \mathbb{A}x_N\right\|^2 + \frac{2}{N-k}\left\langle z_k + \mathbb{A}x_N, x_k - x_N\right\rangle
$$

for $k = 1, \dots, N-1$. We prove that

$$
\begin{aligned}
V_k - V_{k+1} = &\frac{2}{(N-k)(N-k-1)} \underbrace{\left\langle \mathbb{A}x_N - \mathbb{A}x_{k+\frac{1}{2}}, x_N - x_{k+\frac{1}{2}}\right\rangle}_{\mathrm{MI}_k} \\
&+ \frac{1}{\alpha(N-k)^2}\underbrace{\left(\|x_{k+\frac{1}{2}} - x_k\|^2 - \alpha^2\|\mathbb{A}x_{k+\frac{1}{2}} - \mathbb{A}x_k\|^2\right)}_{\mathrm{LI}_k}
\end{aligned}
$$

for $k = 1, \dots, N-2$. This implies $V_k \geq V_{k+1}$ because $\mathrm{MI}_k \geq 0$ by monotonicity of $\mathbb{A}$ and $\mathrm{LI}_k \geq 0$ from L-Lipschitz continuity of $\mathbb{A}$ (i.e., L-smoothness of $\mathbf{L}$) and $0 < \alpha \leq \frac{1}{L}$.

We first decompose V_k as following:

$$
V_k = -\alpha\|z_k\|^2 - \alpha\|\mathbb{A}x_N\|^2 + \underbrace{\left(\frac{2}{N-k}\langle \mathbb{A}x_N, x_k - x_N\rangle - 2\alpha\langle \mathbb{A}x_N, z_k\rangle\right)}_{:=V_k^{(1)}} + \underbrace{\frac{2}{N-k}\langle z_k, x_k - x_N\rangle}_{:=V_k^{(2)}}
$$

Observe that $x_{k+\frac{1}{2}}$ can be written in the following two ways:

$$
\begin{aligned}
x_{k+\frac{1}{2}} &= x_k - \alpha z_k - \alpha\mathbb{A}x_k \\
x_{k+\frac{1}{2}} &= x_{k+1} - \frac{N-k-1}{N-k}\alpha(\mathbb{A}x_k - \mathbb{A}x_{k+\frac{1}{2}})
\end{aligned}
\tag{15}
$$

Appropriately plugging (15) into MI_k we can rewrite it as

$$
\begin{aligned}
\mathrm{MI}_k &= (N-k)\left\langle \mathbb{A}x_N, x_N - x_{k+\frac{1}{2}}\right\rangle - (N-k-1)\left\langle \mathbb{A}x_N, x_N - x_{k+\frac{1}{2}}\right\rangle - \left\langle \mathbb{A}x_{k+\frac{1}{2}}, x_N - x_{k+\frac{1}{2}}\right\rangle \\
&= (N-k)\left\langle \mathbb{A}x_N, x_N - x_{k+1} + \frac{N-k-1}{N-k}\alpha(\mathbb{A}x_k - \mathbb{A}x_{k+\frac{1}{2}})\right\rangle \\
&\quad - (N-k-1)\langle \mathbb{A}x_N, x_N - x_k + \alpha z_k + \alpha\mathbb{A}x_k\rangle - \left\langle \mathbb{A}x_{k+\frac{1}{2}}, x_N - x_{k+\frac{1}{2}}\right\rangle \\
&= (N-k)\langle \mathbb{A}x_N, x_N - x_{k+1}\rangle - (N-k-1)\langle \mathbb{A}x_N, x_N - x_k\rangle \\
&\quad - \alpha(N-k-1)\left\langle \mathbb{A}x_N, \mathbb{A}x_{k+\frac{1}{2}} + z_k\right\rangle - \left\langle \mathbb{A}x_{k+\frac{1}{2}}, x_N - x_{k+\frac{1}{2}}\right\rangle
\end{aligned}
$$

Now multiply $\frac{2}{(N-k)(N-k-1)}$ to MI_k:

$$
\begin{aligned}
&\frac{2}{(N-k)(N-k-1)}\mathrm{MI}_k \\
&= \frac{2}{N-k-1}\left\langle \mathbb{A}x_N, x_N - x_{k+1}\right\rangle - \frac{2}{N-k}\left\langle \mathbb{A}x_N, x_N - x_k\right\rangle \\
&\quad - 2\alpha\left\langle \mathbb{A}x_N, \frac{1}{N-k}\mathbb{A}x_{k+\frac{1}{2}} + \frac{1}{N-k}z_k\right\rangle - \frac{2}{N-k-1}\left\langle \frac{1}{N-k}\mathbb{A}x_{k+\frac{1}{2}}, x_N - x_{k+\frac{1}{2}}\right\rangle
\end{aligned} \tag{16}
$$

and plug the following identity, which is the third line of (Rev-FEG):

$$
\frac{1}{N-k}\mathbb{A}x_{k+\frac{1}{2}} = \frac{N-k-1}{N-k}z_k - z_{k+1} \tag{17}
$$

into (16) to obtain

$$
\begin{aligned}
&\frac{2}{(N-k)(N-k-1)}\mathrm{MI}_k \\
&= \frac{2}{N-k-1}\left\langle \mathbb{A}x_N, x_N - x_{k+1}\right\rangle - \frac{2}{N-k}\left\langle \mathbb{A}x_N, x_N - x_k\right\rangle \\
&\quad - 2\alpha\left\langle \mathbb{A}x_N, z_k - z_{k+1}\right\rangle - \frac{2}{N-k-1}\left\langle \frac{N-k-1}{N-k}z_k - z_{k+1}, x_N - x_{k+\frac{1}{2}}\right\rangle \\
&= \frac{2}{N-k}\left\langle \mathbb{A}x_N, x_k - x_N\right\rangle - 2\alpha\left\langle \mathbb{A}x_N, z_k\right\rangle - \left(\frac{2}{N-k-1}\left\langle \mathbb{A}x_N, x_{k+1} - x_N\right\rangle - 2\alpha\left\langle \mathbb{A}x_N, z_{k+1}\right\rangle\right) \\
&\quad + \underbrace{2\left\langle \frac{1}{N-k}z_k - \frac{1}{N-k-1}z_{k+1}, x_{k+\frac{1}{2}} - x_N\right\rangle}_{:=R_k} \\
&= V_k^{(1)} - V_{k+1}^{(1)} + R_k
\end{aligned} \tag{18}
$$

We rewrite R_k as following, using (15):

$$
\begin{aligned}
R_k &= 2\left\langle \frac{1}{N-k}z_k, x_{k+\frac{1}{2}} - x_N\right\rangle - 2\left\langle \frac{1}{N-k-1}z_{k+1}, x_{k+\frac{1}{2}} - x_N\right\rangle \\
&= 2\left\langle \frac{1}{N-k}z_k, x_k - x_N - \alpha z_k - \alpha\mathbb{A}x_k\right\rangle - 2\left\langle \frac{1}{N-k-1}z_{k+1}, x_{k+1} - x_N - \frac{N-k-1}{N-k}\alpha(\mathbb{A}x_k - \mathbb{A}x_{k+\frac{1}{2}})\right\rangle \\
&= \frac{2}{N-k}\left\langle z_k, x_k - x_N\right\rangle - \frac{2}{N-k-1}\left\langle z_{k+1}, x_{k+1} - x_N\right\rangle \\
&\quad + \frac{2\alpha}{N-k}\left\langle \mathbb{A}x_k, z_{k+1} - z_k\right\rangle - \frac{2\alpha}{N-k}\|z_k\|^2 - 2\alpha\left\langle z_{k+1}, \frac{1}{N-k}\mathbb{A}x_{k+\frac{1}{2}}\right\rangle \\
&= V_k^{(2)} - V_{k+1}^{(2)} + \alpha\left(\frac{2}{N-k}\left\langle \mathbb{A}x_k, z_{k+1} - z_k\right\rangle - \frac{2}{N-k}\|z_k\|^2 - \frac{2(N-k-1)}{N-k}\left\langle z_{k+1}, z_k\right\rangle + 2\left\|z_{k+1}\right\|^2\right)
\end{aligned} \tag{19}
$$

where the last equality uses (17). Now multiplying $\frac{1}{\alpha(N-k)^2}$ to LI_k and applying the identities (15), (17) we obtain

$$
\frac{1}{\alpha(N-k)^2}(\mathrm{LI}_k)
$$

$$
\begin{aligned}
&= \frac{1}{\alpha(N-k)^2}\left(\|x_{k+\frac{1}{2}} - x_k\|^2 - \alpha^2\|\mathbb{A}x_{k+\frac{1}{2}} - \mathbb{A}x_k\|^2\right) \\
&= \frac{1}{(N-k)^2}\left(\alpha\,\|z_k + \mathbb{A}x_k\|^2 - \alpha\|(N-k-1)z_k - (N-k)z_{k+1} - \mathbb{A}x_k\|^2\right) \\
&= \alpha\left(\left\|\frac{1}{N-k}z_k + \frac{1}{N-k}\mathbb{A}x_k\right\|^2 - \left\|\frac{1}{N-k}z_k + \frac{1}{N-k}\mathbb{A}x_k + (z_{k+1} - z_k)\right\|^2\right) \\
&= -\alpha\left\langle z_{k+1} - \left(1 - \frac{2}{N-k}\right)z_k + \frac{2}{N-k}\mathbb{A}x_k, z_{k+1} - z_k\right\rangle \\
&= -\alpha\left(\|z_{k+1}\|^2 + \left(1 - \frac{2}{N-k}\right)\|z_k\|^2 - \frac{2(N-k-1)}{N-k}\langle z_k, z_{k+1}\rangle + \frac{2}{N-k}\langle \mathbb{A}x_k, z_{k+1} - z_k\rangle\right)
\end{aligned} \tag{20}
$$

holds. Now, we add (18) with (20), plug in (19) and simplify to obtain:

$$
\begin{aligned}
\frac{2}{(N-k)(N-k-1)}(\mathrm{MI}_k) + \frac{1}{\alpha(N-k)^2}(\mathrm{LI}_k) &= (V_k^{(1)} + V_k^{(2)} - \alpha\|z_k\|^2) - (V_{k+1}^{(1)} + V_{k+1}^{(2)} - \alpha\|z_{k+1}\|^2) \\
&= V_k - V_{k+1}
\end{aligned}
$$

It remains to show $V_{N-1} \geq 0$. When $k = N-1$, we have

$$
x_N = x_{N-\frac{1}{2}} = x_{N-1} - \alpha z_{N-1} - \alpha\mathbb{A}x_{N-1} \quad \Longleftrightarrow \quad z_{N-1} + \mathbb{A}x_{N-1} = -\frac{1}{\alpha}(x_N - x_{N-1})
$$

Therefore,

$$
\begin{aligned}
V_{N-1} &= -\alpha\,\|z_{N-1} + \mathbb{A}x_N\|^2 + 2\langle z_{N-1} + \mathbb{A}x_N, x_{N-1} - x_N\rangle \\
&= -\alpha\,\|z_{N-1} + \mathbb{A}x_{N-1} + (\mathbb{A}x_N - \mathbb{A}x_{N-1})\|^2 + 2\langle z_{N-1} + \mathbb{A}x_{N-1}, x_{N-1} - x_N\rangle \\
&= -\alpha\left\|-\frac{1}{\alpha}(x_N - x_{N-1}) + (\mathbb{A}x_N - \mathbb{A}x_{N-1})\right\|^2 + \frac{2}{\alpha}\|x_N - x_{N-1}\|^2 \\
&= \frac{1}{\alpha}\left(\|x_N - x_{N-1}\|^2 - \alpha^2\|\mathbb{A}x_N - \mathbb{A}x_{N-1}\|^2\right) + 2\langle x_N - x_{N-1}, \mathbb{A}x_N - \mathbb{A}x_{N-1}\rangle \geq 0
\end{aligned}
$$

which concludes the proof.

5.3 Analysis of (Rev-OHM) for fixed-point problems

In this subsection, we present the convergence analysis of (Rev-OHM), showing that it is another exact optimal algorithm for solving nonexpansive fixed-point problems.

5.3.1 Convergence analyses of (OHM) and (Rev-OHM)

We formally state the convergence result of (Rev-OHM) and outline its proof.

Theorem 4. *Let $\mathbb{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ be nonexpansive and $y_\star \in \operatorname{Fix}\mathbb{T}$. For $N \geq 1$,* (Rev-OHM) *exhibits the rate*

$$
\|y_{N-1} - \mathbb{T}y_{N-1}\|^2 \leq \frac{4\|y_0 - y_\star\|^2}{N^2}
$$

Proof outline. Let $\mathbb{A}$ be the unique maximally monotone operator such that $\mathbb{T} = 2\mathbb{J}_{\mathbb{A}} - \mathbb{I}$ (defined as in Proposition 2). Let $x_{k+1} = \mathbb{J}_{\mathbb{A}}(y_k)$ for $k = 0, 1, \ldots$, so that $\widetilde{\mathbb{A}}x_{k+1} = y_k - x_{k+1} \in \mathbb{A}x_{k+1}$. Recall the alternative form (3) of (Rev-OHM). Define

$$V_k = -\frac{N-k-1}{N-k}\left\|z_k + 2\widetilde{\mathbb{A}}x_N\right\|^2 + \frac{2}{N-k}\left\langle z_k + 2\widetilde{\mathbb{A}}x_N, y_k - y_{N-1}\right\rangle$$

for $k = 0, 1, \ldots, N-1$. We show in §5.3.3 that

$$V_k - V_{k+1} = \frac{4}{(N-k)(N-k-1)}\left\langle x_N - x_{k+1}, \widetilde{\mathbb{A}}x_N - \widetilde{\mathbb{A}}x_{k+1}\right\rangle$$

i.e., $V_k \geq V_{k+1}$ for $k = 0, 1, \ldots, N-2$. Observe that $V_{N-1} = 0$ and because $z_0 = 0$,

$$V_0 = -\frac{4(N-1)}{N}\left\|\widetilde{\mathbb{A}}x_N\right\|^2 + \frac{4}{N}\left\langle \widetilde{\mathbb{A}}x_N, y_0 - y_{N-1}\right\rangle = -4\left\|\widetilde{\mathbb{A}}x_N\right\|^2 + \frac{4}{N}\left\langle \widetilde{\mathbb{A}}x_N, y_0 - x_N\right\rangle$$

where the second line uses $x_N = y_{N-1} - \widetilde{\mathbb{A}}x_N$. Finally, divide both sides of $V_0 \geq \cdots \geq V_{N-1} = 0$ by $\frac{4}{N}$, apply Lemma 1 and the identity $y_{N-1} - \mathbb{T}y_{N-1} = 2\widetilde{\mathbb{A}}x_N$:

$$\|y_{N-1} - \mathbb{T}y_{N-1}\|^2 = 4\left\|\widetilde{\mathbb{A}}x_N\right\|^2 \leq \frac{4\|y_0 - y_\star\|^2}{N^2}$$

□

We point out that the convergence analysis for (OHM) can be done in a similar style [Lie21]. Define $\mathbb{A}$, x_{k+1} and $\widetilde{\mathbb{A}}x_{k+1}$ as above. Define $U_0 = 0$ and

$$U_k = k^2\left\|\widetilde{\mathbb{A}}x_k\right\|^2 + k\left\langle \widetilde{\mathbb{A}}x_k, x_k - y_0\right\rangle$$

for $k = 1, 2, \ldots$. It can be shown that [RY22]

$$U_j - U_{j+1} = j(j+1)\left\langle x_{j+1} - x_j, \widetilde{\mathbb{A}}x_{j+1} - \widetilde{\mathbb{A}}x_j\right\rangle \geq 0$$

for $j = 0, 1, \ldots$. Then $0 = U_0 \geq \cdots \geq U_k$, and dividing both sides by k and applying Lemma 1 gives the rate

$$\|\widetilde{\mathbb{A}}x_k\|^2 \leq \frac{\|y_0 - x_\star\|^2}{k^2}$$

5.3.2 Proximal forms of (OHM) and (Rev-OHM)

It is known that (OHM) can be equivalently written as

$$\begin{aligned} x_{k+1} &= \mathbb{J}_{\mathbb{A}}(y_k) \\ y_{k+1} &= x_{k+1} + \frac{k}{k+2}(x_{k+1} - x_k) - \frac{k}{k+2}(x_k - y_{k-1}) \end{aligned}$$

for $k = 0, 1, \ldots$, where $x_0 = y_0$. This proximal form is called Accelerated Proximal Point Method (APPM) [Kim21]. Likewise, (Rev-OHM) can be equivalently written as

$$\begin{aligned} x_{k+1} &= \mathbb{J}_{\mathbb{A}}(y_k) \\ y_{k+1} &= x_{k+1} + \frac{N-k-1}{N-k}(x_{k+1} - x_k) - \frac{N-k-1}{N-k}(x_k - y_{k-1}) - \frac{1}{N-k}(x_{k+1} - y_k) \end{aligned}$$

for $k = 0, 1, \ldots, N-2$, where $x_{-1} = y_0 = x_0$. Limited by space we forgo proving the equivalence.

5.3.3 Lyapunov analysis of (Rev-OHM)

Recall the following form of (Rev-OHM):

$$z_{k+1} = \frac{N-k-1}{N-k} z_k - \frac{1}{N-k}\left(y_k - \mathbb{T} y_k\right) \tag{21}$$

$$y_{k+1} = \mathbb{T} y_k - z_{k+1} \tag{22}$$

Substituting (21) into (22) we get

$$y_{k+1} = \mathbb{T} y_k - \frac{N-k-1}{N-k} z_k + \frac{1}{N-k}\left(y_k - \mathbb{T} y_k\right) = \frac{1}{N-k} y_k + \frac{N-k-1}{N-k} \mathbb{T} y_k - \frac{N-k-1}{N-k} z_k \tag{23}$$

With the substitution $\mathbb{T} = 2\mathbb{J}_{\mathbb{A}} - \mathbb{I}$ and $x_{k+1} = \mathbb{J}_{\mathbb{A}} y_k$ we can write (21) and (23) as

$$z_{k+1} = \frac{N-k-1}{N-k} z_k - \frac{2}{N-k} \widetilde{\mathbb{A}} x_{k+1} \tag{24}$$

$$\begin{aligned} y_{k+1} &= \frac{1}{N-k} y_k + \frac{N-k-1}{N-k}\left(y_k - 2\widetilde{\mathbb{A}} x_{k+1}\right) - \frac{N-k-1}{N-k} z_k \\ &= y_k - \frac{2(N-k-1)}{N-k} \widetilde{\mathbb{A}} x_{k+1} - \frac{N-k-1}{N-k} z_k \end{aligned} \tag{25}$$

where we have used $\mathbb{T} y_k = 2\mathbb{J}_{\mathbb{A}} y_k - y_k = 2x_{k+1} - y_k = y_k - 2(y_k - x_{k+1}) = y_k - 2\widetilde{\mathbb{A}} x_{k+1}$. For simplicity, write $g_j = \widetilde{\mathbb{A}} x_j$ for $j = 1, \ldots, N$. To complete the proof, it remains to show that for

$$V_k = \underbrace{-\frac{N-k-1}{N-k} \left\| z_k + 2g_N \right\|^2}_{:=V_k^{(1)}} + \underbrace{\frac{2}{N-k} \left\langle z_k + 2g_N, y_k - y_{N-1} \right\rangle}_{:=V_k^{(2)}}$$

the following holds:

$$V_k - V_{k+1} = \frac{4}{(N-k)(N-k-1)} \left\langle x_{k+1} - x_N, g_{k+1} - g_N \right\rangle$$

Rewriting the right hand side, we have

$$\begin{aligned} &\frac{4}{(N-k)(N-k-1)} \left\langle x_{k+1} - x_N, g_{k+1} - g_N \right\rangle \\ &= \frac{4}{(N-k)(N-k-1)} \left\langle y_k - y_{N-1} - (g_{k+1} - g_N), g_{k+1} - g_N \right\rangle \\ &= \frac{2}{N-k-1} \left\langle y_k - y_{N-1}, \frac{2}{N-k} g_{k+1} - \frac{2}{N-k} g_N \right\rangle \\ &\quad - \frac{4}{(N-k)(N-k-1)} \left\| g_{k+1} - g_N \right\|^2 \\ &\overset{(24)}{=} \frac{2}{N-k-1} \left\langle y_k - y_{N-1}, \frac{N-k-1}{N-k}(z_k + 2g_N) - (z_{k+1} + 2g_N) \right\rangle \\ &\quad - \frac{4}{(N-k)(N-k-1)} \left\| g_{k+1} - g_N \right\|^2 \end{aligned}$$

$$
\begin{aligned}
&= \frac{2}{N-k}\left\langle y_k - y_{N-1}, z_k + 2g_N \right\rangle - \frac{2}{N-k-1}\left\langle y_k - y_{N-1}, z_{k+1} + 2g_N \right\rangle \\
&\quad - \frac{4}{(N-k)(N-k-1)}\left\| g_{k+1} - g_N \right\|^2 \\
&\overset{(25)}{=} V_k^{(2)} - \frac{2}{N-k-1}\left\langle y_{k+1} - y_{N-1} + \frac{2(N-k-1)}{N-k} g_{k+1} + \frac{N-k-1}{N-k} z_k, z_{k+1} + 2g_N \right\rangle \\
&\quad - \frac{4}{(N-k)(N-k-1)}\left\| g_{k+1} - g_N \right\|^2 \\
&= V_k^{(2)} - V_{k+1}^{(2)} \underbrace{- \frac{2}{N-k}\left\langle 2g_{k+1} + z_k, z_{k+1} + 2g_N \right\rangle - \frac{1}{(N-k)(N-k-1)}\left\| 2g_{k+1} - 2g_N \right\|^2}_{:=R_k}
\end{aligned}
$$

and the proof is done once we show $R_k = V_k^{(1)} - V_{k+1}^{(1)}$. From (24) we have $2g_{k+1} = (N-k-1)z_k - (N-k)z_{k+1}$, and plugging this into R_k we obtain

$$
\begin{aligned}
R_k &= -2\left\langle z_{k+1} - z_k, z_{k+1} + 2g_N \right\rangle - \frac{1}{(N-k)(N-k-1)}\left\| (N-k-1)z_k - (N-k)z_{k+1} - 2g_N \right\|^2 \\
&= 2\left\langle (z_{k+1} + 2g_N) - (z_k + 2g_N), z_{k+1} + 2g_N \right\rangle \\
&\quad - \frac{1}{(N-k)(N-k-1)}\left\| (N-k-1)(z_k + 2g_N) - (N-k)(z_{k+1} + 2g_N) \right\|^2 \\
&= 2\left\| z_{k+1} + 2g_N \right\|^2 - \frac{N-k-1}{N-k}\left\| z_k + 2g_N \right\|^2 - \frac{N-k}{N-k-1}\left\| z_{k+1} + 2g_N \right\|^2 \\
&= -\frac{N-k-1}{N-k}\left\| z_k + 2g_N \right\|^2 + \frac{N-k-2}{N-k-1}\left\| z_{k+1} + 2g_N \right\|^2 \\
&= V_k^{(1)} - V_{k+1}^{(1)}
\end{aligned}
$$

which proves that indeed, $R_k = V_k^{(1)} - V_{k+1}^{(1)}$.

6 Conclusion and Outlook

In this paper, we introduced a novel class of time-reversal ODEs, which, when discretized, yield new algorithms that achieve the same optimal accelerated convergence rates as existing anchor-based methods. These algorithms use materially different acceleration mechanisms, revealing that the optimal acceleration mechanism for minimax optimization and fixed-point problems is not unique. This opens a new avenue of research, as we now have a family of methods that empirically exhibit varied characteristics while maintaining the same optimal worst-case guarantees.

The implications of this work challenge traditional understandings of acceleration in optimization. Future research could explore detailed empirical comparisons between these new algorithms and existing methods in large-scale machine learning, signal processing, and other applications. Additionally, further investigation into why different mechanisms achieve the same convergence rates could deepen our understanding of optimization landscapes.

In conclusion, our work advances both the theoretical framework and practical innovations in accelerated optimization algorithms. These insights are poised to inspire continued research and development in the field.

References

[AAD18] Vassilis Apidopoulos, Jean-François Aujol, and Charles Dossal. The differential inclusion modeling FISTA algorithm and optimality of convergence rate in the case $b \leq 3$. *SIAM Journal on Optimization*, 28(1):551–574, 2018.

[AC84] Jean-Pierre Aubin and Arrigo Cellina. *Differential Inclusions*. Springer, 1984.

[And65] Donald G Anderson. Iterative procedures for nonlinear integral equations. *Journal of the ACM (JACM)*, 12(4):547–560, 1965.

[Ban22] Stefan Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922.

[BB92] Jean-Bernard Baillon and Ronald E Bruck. Optimal rates of asymptotic regularity for averaged nonexpansive mappings. *Fixed Point Theory and Applications*, 128:27–66, 1992.

[BC17] Heinz H Bauschke and Patrick L Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2nd edition, 2017.

[BC18] Mario Bravo and Roberto Cominetti. Sharp convergence rates for averaged nonexpansive maps. *Israel Journal of Mathematics*, 227:163–188, 2018.

[BCN23] Radu Ioan Boţ, Ernö Robert Csetnek, and Dang-Khoa Nguyen. Fast optimistic gradient descent ascent (OGDA) method in continuous and discrete time. *Foundations of Computational Mathematics*, pages 1–60, 2023.

[BRS92] Jonathan Borwein, Simeon Reich, and Itai Shafrir. Krasnoselski-Mann iterations in normed spaces. *Canadian Mathematical Bulletin*, 35(1):21–28, 1992.

[CMT19] Ernö Robert Csetnek, Yura Malitsky, and Matthew K Tam. Shadow Douglas–Rachford splitting for monotone inclusions. *Applied Mathematics & Optimization*, 80:665–678, 2019.

[COZ22] Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Finite-time last-iterate convergence for learning in multi-player games. *Advances in Neural Information Processing Systems*, 35:33904–33919, 2022.

[CSV14] Roberto Cominetti, José A Soto, and José Vaisman. On the rate of convergence of Krasnosel'skiĭ-Mann iterations and their connection with sums of Bernoullis. *Israel Journal of Mathematics*, 199:757–772, 2014.

[CZ23] Yang Cai and Weiqiang Zheng. Accelerated single-call methods for constrained min-max optimization. In *The Eleventh International Conference on Learning Representations*, 2023.

[DGVPR24] Shuvomoy Das Gupta, Bart PG Van Parys, and Ernest K Ryu. Branch-and-bound performance estimation programming: A unified methodology for constructing optimal optimization methods. *Mathematical Programming*, 204:567–639, 2024.

[Dia20] Jelena Diakonikolas. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. In *Conference on Learning Theory*, pages 1428–1451. 2020.

[DISZ18] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with optimism. In *International Conference on Learning Representations*, 2018.

[DT14] Yoel Drori and Marc Teboulle. Performance of first-order methods for smooth convex minimization: A novel approach. *Mathematical Programming*, 145:451–482, 2014.

[DT16] Yoel Drori and Marc Teboulle. An optimal variant of Kelley's cutting-plane method. *Mathematical Programming*, 160:321–351, 2016.

[DT20] Yoel Drori and Adrien B Taylor. Efficient first-order methods for convex minimization: A constructive approach. *Mathematical Programming*, 184:183–220, 2020.

[DY16] Damek Davis and Wotao Yin. Convergence rate analysis of several splitting schemes. In Roland Glowinski, Stanley J Osher, and Wotao Yin, editors, *Splitting Methods in Communication, Imaging, Science, and Engineering*, pages 115–163. Springer, Cham, 2016.

[DYCR18] QL Dong, HB Yuan, YJ Cho, and Th M Rassias. Modified inertial Mann algorithm and inertial CQ-algorithm for nonexpansive mappings. *Optimization Letters*, 12:87–102, 2018.

[EB92] Jonathan Eckstein and Dimitri P Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[FRV18] Guilherme Franca, Daniel Robinson, and Rene Vidal. ADMM and accelerated ADMM as continuous dynamical systems. In *International Conference on Machine Learning*, pages 1559–1567. PMLR, 2018.

[FRV23] Guilherme França, Daniel P Robinson, and René Vidal. A nonsmooth dynamical systems perspective on accelerated extensions of ADMM. *IEEE Transactions on Automatic Control*, 68(5):2966–2978, 2023.

[GLG22] Eduard Gorbunov, Nicolas Loizou, and Gauthier Gidel. Extragradient method: $O(1/K)$ last-iterate convergence for monotone variational inequalities and connections with cocoercivity. In *International Conference on Artificial Intelligence and Statistics*, pages 366–402. PMLR, 2022.

[Hal67] Benjamin Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73(6):957–961, 1967.

[Ish76] Shiro Ishikawa. Fixed points and iteration of a nonexpansive mapping in a Banach space. *Proceedings of the American Mathematical Society*, 59(1):65–71, 1976.

[JGR23] Uijeong Jang, Shuvomoy Das Gupta, and Ernest K Ryu. Computer-assisted design of accelerated composite optimization methods: OptISTA. *arXiv preprint arXiv:2305.15704*, 2023.

[KBB15] Walid Krichene, Alexandre Bayen, and Peter L Bartlett. Accelerated mirror descent in continuous and discrete time. *Advances in Neural Information Processing Systems*, 28, 2015.

[KF16] Donghwan Kim and Jeffrey A Fessler. Optimized first-order methods for smooth convex minimization. *Mathematical Programming*, 159:81–107, 2016.

[KF21] Donghwan Kim and Jeffrey A Fessler. Optimizing the efficiency of first-order methods for decreasing the gradient of smooth convex functions. *Journal of Optimization Theory and Applications*, 188:192–219, 2021.

[Kim21] Donghwan Kim. Accelerated proximal point method for maximally monotone operators. *Mathematical Programming*, 190(1):57–87, 2021.

[Koh11] Ulrich Kohlenbach. On quantitative versions of theorems due to F.E. Browder and R. Wittmann. *Advances in Mathematics*, 226(3):2764–2795, 2011.

[Kor76] Galina M Korpelevich. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematicheskie Metody*, 12:747–756, 1976.

[Kra55] Mark Aleksandrovich Krasnosel'skii. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10(1):123–127, 1955.

[Leu07] Laurentiu Leustean. Rates of asymptotic regularity for Halpern iterations of nonexpansive mappings. *Journal of Universal Computer Science*, 13(11):1680–1691, 2007.

[LFP16] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Convergence rates with inexact non-expansive operators. *Mathematical Programming*, 159:403–434, 2016.

[Li24] Chris Junchi Li. A general continuous-time formulation of stochastic ADMM and its variants. *arXiv preprint arXiv:2404.14358*, 2024.

[Lie21] Felix Lieder. On the convergence rate of the Halpern-iteration. *Optimization Letters*, 15(2):405–418, 2021.

[LK21] Sucheol Lee and Donghwan Kim. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. *Advances in Neural Information Processing Systems*, 34:22588–22600, 2021.

[Lu22] Haihao Lu. An $O(s^r)$-resolution ODE framework for understanding discrete-time algorithms and applications to the linear convergence of minimax problems. *Mathematical Programming*, 194:1061–1112, 2022.

[Mai08] Paul-Emile Maingé. Convergence theorems for inertial KM-type algorithms. *Journal of Computational and Applied Mathematics*, 219(1):223–236, 2008.

[Man53] William Robert Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4(3):506–510, 1953.

[Mat17] Shin-Ya Matsushita. On the convergence rate of the Krasnosel'skiĭ–Mann iteration. *Bulletin of the Australian Mathematical Society*, 96(1):162–170, 2017.

[Min62] George J Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962.

[MOP20] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In *International Conference on Artificial Intelligence and Statistics*, pages 1497–1507. PMLR, 2020.

[Nem04] Arkadi Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

[Nes07] Yurii Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, 109(2):319–344, 2007.

[Nes18] Yurii Nesterov. *Lectures on Convex Optimization*, volume 137. Springer, 2018.

[Pop80] LD Popov. A modification of the Arrow–Hurwicz method for search of saddle points. *Matematicheskie Zametki*, 28(5):777–784, 1980.

[PR22] Jisun Park and Ernest K Ryu. Exact optimal accelerated complexity for fixed-point iterations. In *International Conference on Machine Learning*, pages 17420–17457. PMLR, 2022.

[PR23] Jisun Park and Ernest K Ryu. Accelerated infeasibility detection of constrained optimization and fixed-point iterations. In *International Conference on Machine Learning*, pages 27294–27345. PMLR, 2023.

[PR24] Chanwoo Park and Ernest K Ryu. Optimal first-order algorithms as a function of inequalities. *Journal of Machine Learning Research*, 25(51):1–66, 2024.

[RS13] Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019. 2013.

[RTBG20] Ernest K Ryu, Adrien B Taylor, Carolina Bergeling, and Pontus Giselsson. Operator splitting performance estimation: Tight contraction factors and optimal parameter selection. *SIAM Journal on Optimization*, 30(3):2251–2271, 2020.

[RTCVL21] Simeon Reich, Duong Viet Thong, Prasit Cholamjiak, and Luong Van Long. Inertial projection-type methods for solving pseudomonotone variational inequality problems in Hilbert space. *Numerical Algorithms*, 88:813–835, 2021.

[RY22] Ernest K Ryu and Wotao Yin. *Large-Scale Convex Optimization*. Cambridge University Press, 2022.

[RYY19] Ernest K Ryu, Kun Yuan, and Wotao Yin. ODE analysis of stochastic gradient methods with optimism and anchoring for minimax problems. *arXiv preprint arXiv:1905.10899*, 2019.

[SBC14] Weijie Su, Stephen Boyd, and Emmanuel J Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Advances in Neural Information Processing Systems*, 27, 2014.

[She18] Yekini Shehu. Convergence rate analysis of inertial Krasnoselskii–Mann type iteration with applications. *Numerical Functional Analysis and Optimization*, 39(10):1077–1091, 2018.

[SPR23] Jaewook J Suh, Jisun Park, and Ernest K Ryu. Continuous-time analysis of anchor acceleration. 36:32782–32866, 2023.

[SRR22] Jaewook J Suh, Gyumin Roh, and Ernest K Ryu. Continuous-time analysis of accelerated gradient methods via conservation laws in dilated coordinate systems. In *International Conference on Machine Learning*, pages 20640–20667. PMLR, 2022.

[SS99] MV Solodov and BF Svaiter. A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*, 7(4):323–345, 1999.

[SS17] Shoham Sabach and Shimrit Shtern. A first order method for solving convex bilevel optimization problems. *SIAM Journal on Optimization*, 27(2):640–660, 2017.

[TB19] Adrien Taylor and Francis Bach. Stochastic first-order methods: Non-asymptotic and computer-aided analyses via potential functions. In *Conference on Learning Theory*, pages 2934–2992. PMLR, 2019.

[TD23] Adrien Taylor and Yoel Drori. An optimal gradient method for smooth strongly convex minimization. *Mathematical Programming*, 199:557–594, 2023.

[TDL21] Quoc Tran-Dinh and Yang Luo. Halpern-type accelerated and splitting algorithms for monotone inclusions. *arXiv preprint arXiv:2110.08150*, 2021.

[THG17] Adrien B Taylor, Julien M Hendrickx, and François Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161:307–345, 2017.

[Wit92] Rainer Wittmann. Approximation of fixed points of nonexpansive mappings. *Archiv der Mathematik*, 58:486–491, 1992.

[WN11] Homer F Walker and Peng Ni. Anderson acceleration for fixed-point iterations. *SIAM Journal on Numerical Analysis*, 49(4):1715–1735, 2011.

[Xu02] Hong-Kun Xu. Iterative algorithms for nonlinear operators. *Journal of the London Mathematical Society*, 66(1):240–256, 2002.

[YR21] Taeho Yoon and Ernest K Ryu. Accelerated algorithms for smooth convex-concave minimax problems with $\mathcal{O}(1/k^2)$ rate on squared gradient norm. In *International Conference on Machine Learning*, pages 12098–12109. PMLR, 2021.

[YR22] TaeHo Yoon and Ernest K Ryu. Accelerated minimax algorithms flock together. *arXiv preprint arXiv:2205.11093*, 2022.

[YZLS19] Huizhuo Yuan, Yuren Zhou, Chris Junchi Li, and Qingyun Sun. Differential inclusions for modeling nonsmooth ADMM variants: A continuous limit theory. In *International Conference on Machine Learning*, pages 7232–7241. PMLR, 2019.

[ZOB20] Junzi Zhang, Brendan O'Donoghue, and Stephen Boyd. Globally convergent type-I Anderson acceleration for nonsmooth fixed-point iterations. *SIAM Journal on Optimization*, 30(4):3170–3197, 2020.