A Continuous-Time Perspective on Monotone Equation Problems

Tianyi Lin\textsuperscript{\circ}  Michael I. Jordan\textsuperscript{\circ,\dag}

Department of Electrical Engineering and Computer Sciences\textsuperscript{\circ}
Department of Statistics\textsuperscript{\dag}
University of California, Berkeley

June 13, 2022

Abstract

We study rescaled gradient dynamical systems in a Hilbert space $\mathcal{H}$, where the implicit discretization in a finite-dimensional Euclidean space leads to high-order methods for solving monotone equations (MEs). Our framework generalizes the celebrated dual extrapolation method [Nesterov, 2007] from first order to high order via appeal to the regularization toolbox of optimization theory [Nesterov, 2021a, b]. We establish the existence and uniqueness of a global solution and analyze the convergence properties of solution trajectories. We also present discrete-time counterparts of our high-order continuous-time methods, and we show that the $p$th-order method achieves an ergodic rate of $O(k^{-(p+1)/2})$ in terms of a restricted merit function and a pointwise rate of $O(k^{-p/2})$ in terms of a residue function. Under regularity conditions, the restarted version of $p$th-order methods achieves local convergence with the order $p \geq 2$.

1 Introduction

Monotone equations (MEs) are a large class of problems that include unconstrained convex optimization problems, unconstrained convex-concave saddle-point problems and various equilibrium computation problems. Let $\mathcal{H}$ be a real Hilbert space and let $F(x) : \mathcal{H} \mapsto \mathcal{H}$ be a continuous operator that is monotone: $\langle F(x) - F(y), x - y \rangle \geq 0$ for any $x, y \in \mathcal{H}$. The ME problem is to find a point $x^* \in \mathcal{H}$ such that

$$F(x^*) = 0 \in \mathcal{H}. \tag{1.1}$$

Given our assumptions, the solution to Eq. (1.1) can also be written as a solution to a variational inequality (VI) corresponding to $F$ and $\mathcal{H}$ [Facchinei and Pang, 2007]. Indeed, a point $x^* \in \mathcal{H}$ is the solution to Eq. (1.1) if and only if we have

$$\langle F(x^*), x - x^* \rangle \geq 0, \ \forall x \in \mathcal{H}.$$ 

Note also that the solution set of Eq. (1.1) is also equivalent to the solution set of an alternative VI, $\langle F(x), x - x^* \rangle \geq 0$ for all $x \in \mathcal{H}$. This equivalence holds given the assumptions that $F$ is continuous and monotone.

The ME problem is a fundamental problem in applied mathematics, scientific computing and optimization [Muller, 1956, Ostiowski, 1960, Traub, 1982, Kelley, 1995, Dennis Jr and Schnabel, 1996, Ortega and Rheinboldt, 2000, Nocedal and Wright, 2006]. Major applications can be found in a wide range of fields, including economics [Morgenstern and Von Neumann, 1953, Osborne, 2004], interval
arithmetic [Moore, 1979, Hong and Stahl, 1994], kinematics [Morgan and Sommese, 1987], chemical engineering [Meintjes and Morgan, 1990], neurophysiology [Verschelde et al., 1994] and combustion [Morgan, 2009]. For an overview of recent progress on solving nonlinear equations and many more applications, we refer to Heath [2018] and references therein. Recently, this line of research has been brought into contact with machine learning (ML), in the guise of optimality conditions for saddle-point problems, with applications including robust prediction and regression [Xu et al., 2009, Esfahani and Kuhn, 2018], adversarial learning [Goodfellow et al., 2014], online learning with multiple agents [Cesa-Bianchi and Lugosi, 2006] and distributed computing [Shamma, 2008, Mateos et al., 2010]. Emerging applications of machine learning often involve multi-agent systems, multi-way markets, or social context, and this is driving increasing interest in equilibrium and dynamical formulations that can be cast as ME problems.

The state-of-the-art method for solving the ME problem is the Newton method [Kelley, 2003]:

\[ x_{k+1} = x_k - (\nabla F(x_k))^{-1} F(x_k). \]

The method is well defined only when \( \nabla F \) is invertible and it attains local quadratic convergence to a solution \( x^* \) when \( F \) is strongly monotone and \( \nabla F \) is Lipschitz continuous around \( x^* \). Higher-order methods have also been proposed that improve upon Newton—they are guaranteed to achieve local superquadratic convergence under the high-order smoothness of \( F \) [Homeier, 2004, Frontini and Sormani, 2004, Darvishi and Barati, 2007, Cordero and Torregrosa, 2007, McDougal and Womerspoon, 2014, Potra, 2017]. As for global convergence, while it is possible to give an asymptotic guarantee using a line search scheme, the global convergence rate for these methods remains unknown.

Since ME problems can be reformulated as unconstrained monotone VIs, the extragradient (EG) [Korpelevich, 1976, Antipin, 1978, Nemirovski, 2004] and optimistic gradient (OG) [Popov, 1980, Daskalakis et al., 2018] methods can be applied. These have the following forms:

\[
\text{(EG)} \quad v_{k+1} = x_k - \lambda F(x_k), \quad x_{k+1} = x_k - \lambda F(v_{k+1}).
\]

\[
\text{(OG)} \quad x_{k+1} = x_k - \beta (F(x_k) + \lambda (F(x_k) - F(x_{k-1}))).
\]

Mokhtari et al. [2020a,b] have emphasized that both schemes can be viewed as inexact proximal point methods [Moreau, 1965, Martinet, 1970, Rockafellar, 1976] and have established an ergodic rate of \( O(1/k) \) (in terms of a gap function) for smooth and convex-concave saddle-point problems. It is straightforward to extend their results to a restricted gap function. An alternative to EG and OG is the dual extrapolation (DE) [Nesterov, 2007] method which can be written as follows:

\[
\text{(DE)} \quad v_{k+1} = x_0 + \beta s_k, \quad x_{k+1} = v_{k+1} - \beta F(v_{k+1}), \quad s_{k+1} = s_k - \lambda F(x_{k+1}).
\]

The DE method exhibits the same convergence properties as EG and OB under a proper choices of parameters; i.e., it has an ergodic convergence rate of \( O(1/k) \) in terms of a restricted merit function. However, this line of work has been restricted to first-order methods and does not show how to obtain global acceleration via appeal to the high-order smoothness of \( F \). Indeed, as noted by Monteiro and Svaiter [2012], a gap exists in our understanding of global convergence rate of high-order methods; this drives much of the recent work on high-order generalizations of EG and OG methods [Bullins and Lai, 2020, Jiang and Mokhtari, 2022]. In particular, Lin and Jordan [2021b] studied a closed-loop damping formulation, obtaining a new class of high-order VI methods which include the high-order EG method in Bullins and Lai [2020]. All of these methods can be applied to solve the ME problems with a global rate of \( O(k^{-\alpha}/\log(k)) \), but they require a nontrivial binary search procedure at each iteration. Such a procedure is generally unsatisfactory from a practical viewpoint [Nesterov, 2018]. The problem
of finding simple high-order methods with a global rate of $O(k^{-(p+1)/2})$ and a local rate with the order $p$ remains open.

In this paper, we present a continuous-time perspective that unifies first-order and higher-order dual extrapolation methods for solving VIs. Our work builds on a two-decade trend that investigates the connection between continuous-time and discrete-time perspectives on dynamical systems for variational inequalities and monotone inclusions [Attouch and Redont, 2001, Attouch and Svaiter, 2011, Maingé, 2013, Abbas et al., 2014, Attouch et al., 2013, 2016a, Attouch and Peyrouquet, 2019, Attouch and Cabot, 2020, Attouch and László, 2020, 2021, Csetnek, 2020, Adly et al., 2021, Labarre and Maingé, 2021, Lin and Jordan, 2021b]. As in these papers, we make use of Lyapunov functions that transfer asymptotic behavior and rates of convergence between continuous time and discrete time. In particular, we propose and analyze a class of rescaled gradient dynamical systems for ME problems that model the dual extrapolation step as an integral. Formally, our system can be written as follows:

$$\dot{s}(t) = -\frac{F(x(t))}{\|F(x(t))\|^{1-1/p}}, \quad v(t) = x_0 + s(t), \quad x(t) - v(t) + \frac{F(x(t))}{\|F(x(t))\|^{1-1/p}} = 0, \quad (1.2)$$

where $\theta > 0$ and $p \in \{1, 2, \ldots\}$ and the initial condition is $s(0) = 0$ and $x(0) = x_0 \in \Omega = \{x \in \mathcal{H} \mid F(x) \neq 0\}$. Note that this condition is not restrictive since $F(x) = 0$ implies that $x \in \mathcal{H}$ is a weak solution. Throughout the paper, unless otherwise indicated, we assume that $F : \mathcal{H} \rightarrow \mathcal{H}$ is continuous and monotone and the ME in Eq. (1.1) has at least one solution.

**Our contributions.** We summarize the main contributions of this paper as follows:

1. We study the rescaled gradient systems in Eq. (1.2) and prove the existence and uniqueness of a global solution (see Theorem 2.2). We analyze the convergence of solution trajectories via appeal to Lyapunov arguments. Our results include asymptotic weak and strong convergence (see Theorem 2.4) and we obtain a nonasymptotic convergence rate in terms of a restricted merit function (ergodic) and a residue function (nonergodic) (see Theorem 2.6).

2. We provide an algorithmic framework via an implicit discretization of the system in Eq. (1.2) in finite-dimensional Euclidean space. Combining this framework with an approximate tensor subroutine, we obtain a new suite of simple $p^{th}$-order methods for solving the ME problem. These resulting methods include not only the classical method [Nesterov, 2007] (for the case of $p = 1$) but yield high-order methods (for the case of $p \geq 2$).

3. We present a unified discrete-time convergence analysis for first-order and high-order methods, including an ergodic rate of $O(k^{-(p+1)/2})$ in terms of a restricted merit function and a pointwise rate of $O(k^{-p/2})$ in terms of a residue function (see Theorem 4.1). We also prove that the restarted version of our $p^{th}$-order method achieve local convergence with the order $p \geq 2$.

Concurrently appearing on arXiv, Lin and Jordan [2022] and Adil et al. [2022] proposed some new high-order methods for solving smooth VIs without requiring any binary search. The way that operator and mixing steps are coupled is clever yet not well understood. Our results are derived independently and complement Lin and Jordan [2022] by showing that a class of simple high-order methods can be developed systematically by appeal to the discretization of rescaled gradient dynamical systems.
Further related work. A flurry of recent work in convex optimization has focused on uncovering a general principle underlying Nesterov’s accelerated gradient method (NAG) [Nesterov, 1983, 2013, Beck and Teboulle, 2009], with a particular focus on interpreting the acceleration phenomenon using a temporal discretization of a continuous-time dynamical system with damping [Su et al., 2016, Attouch and Peypouquet, 2016, Attouch and Cabot, 2017, Attouch et al., 2016b, 2018, 2019a,b, 2020, 2021, Vassilis et al., 2018, Muehlebach and Jordan, 2019, 2021, Diakonikolas and Orecchia, 2019, Sebbouh et al., 2020, Shi et al., 2021]. These papers show that the key to acceleration in continuous-time dynamical systems comes from asymptotically vanishing damping [Su et al., 2016, Attouch et al., 2018, 2019a] and Hessian-driven damping [Alvarez and Pérez, 1998, Alvarez et al., 2002, Attouch et al., 2012, 2016b, 2020, 2021, Shi et al., 2021]. These approaches are, however, not suited for understanding and designing accelerated high-order algorithms in convex smooth optimization [Monteiro and Svaiter, 2013, Gasnikov et al., 2019, Alves, 2022]. Recently, Lin and Jordan [2021a] presented an initial foray into the analysis of continuous-time dynamics of high-order algorithms using closed-loop control. Their approach can systematically generate discrete-time accelerated high-order algorithms, simplifying and generalizing the analysis in Monteiro and Svaiter [2013] using a unified Lyapunov function.

Extending the continuous-time dynamics and Lyapunov analysis from convex optimization to monotone variational inequality and inclusion problems has been an active research area during the last two decades [Attouch and Redont, 2001, Attouch and Svaiter, 2011, Maingé, 2013, Abbas et al., 2014, Attouch et al., 2013, 2016a, Attouch and László, 2020, 2021, Csetnek, 2020, Adly et al., 2021, Labarre and Maingé, 2021, Lin and Jordan, 2021b]. Notably, Attouch et al. [2016a] proposed a second-order method for solving monotone inclusion problems but only obtained a convergence rate estimate for convex optimization. Recently, Lin and Jordan [2021b] have proposed a closed-loop damping approach which generalizes Attouch et al. [2016a] and leads to a suite of high-order algorithms with convergence rate estimation for monotone inclusion problems.

The continuous-time framework has also been productive in generating discrete-time algorithms via numerical integration strategies [Zhang et al., 2018, O’Donoghue and Maddison, 2019, Wilson et al., 2019, França et al., 2020]. In particular, Zhang et al. [2018] applied Runge-Kutta integration to an inertial gradient system without Hessian-driven damping [Wibisono et al., 2016]. The resulting algorithm is faster than NAG when the objective function is sufficiently smooth and when the order of the integrator is sufficiently large. O’Donoghue and Maddison [2019] and França et al. [2020] studied conformal Hamiltonian systems and proved that the resulting algorithms achieve fast convergence under certain conditions. The closest work to ours is that of Wilson et al. [2019] who derived first-order algorithms via the discretization of accelerated rescaled gradient dynamical systems and proved that their algorithms achieve the same convergence rate as accelerated high-order algorithms [Monteiro and Svaiter, 2013, Gasnikov et al., 2019, Alves, 2022]. However, Wilson et al. [2019] did not establish global existence or uniqueness for their systems and the strong smoothness assumptions that they invoked limits the range of possible applications. In contrast, the high-order algorithms developed in this paper are applicable to general monotone and smooth VI problems.

Organization. The remainder of this paper is organized as follows. In Section 2, we prove the existence and uniqueness of a global solution of the rescaled gradient dynamical systems in Eq. (1.2) and analyze the convergence properties of solution trajectories. In Section 3, we propose an algorithmic framework by implicit discretization of the systems in Eq. (1.2). This together with an approximate tensor subroutine leads to a suite of $p^{th}$-order methods. In Section 4, we conduct the discrete-time convergence analysis, obtaining a global convergence rate in terms of a restricted merit function (ergodic)
and a residue function (nonergodic) when \( F \) is \((p - 1)\)th-order smooth. In Section 5, we conclude our work with a brief discussion of future directions.

**Notation.** We use bold lower-case letters such as \( x \) to denote vectors, and upper-case letters such as \( X \) to denote tensors. We let \( \mathcal{H} \) be a real Hilbert space which is endowed with the scalar product \( \langle \cdot , \cdot \rangle \). For \( x \in \mathcal{H} \), we let \( \|x\| \) denote its norm induced by \( \langle \cdot , \cdot \rangle \). If \( \mathcal{H} = \mathbb{R}^d \) is a real Euclidean space, \( \|x\| \) reduces to the \( \ell_2 \)-norm of \( x \). Then, for \( X \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_p} \), we define

\[
X[z^1, \cdots, z^p] = \sum_{1 \leq i_1 \leq d_1, \cdots, 1 \leq i_p \leq p} (X_{i_1, \cdots, i_p})z_{i_1}^1 \cdots z_{i_p}^p,
\]

and let \( \|X\|_{op} = \max_{\|z\| = 1, 1 \leq j \leq p} X[z^1, \cdots, z^p] \). Fixing \( p \geq 1 \), we define \( G^p_L(\mathbb{R}^d) \) as a class of monotone and \((p - 1)\)th-order smooth operators \( F : \mathbb{R}^d \to \mathbb{R}^d \) satisfying that

\[
\|\nabla^{(p-1)} F(x') - \nabla^{(p-1)} F(x)\|_{op} \leq L\|x' - x\|, \quad \text{for all } x, x' \in \mathbb{R}^d.
\]

where \( \nabla^{(k)} F(x) \) stands for the \( k \)th-order derivative of \( F \) at \( x \in \mathbb{R}^d \) and \( \nabla^{(0)} F = F \). In other words, for \( \{z_1, z_2, \ldots, z_k\} \subseteq \mathbb{R}^d \), we have

\[
\nabla^{(k)} F(x)[z^1, \cdots, z^k] = \sum_{1 \leq i_1, \cdots, i_k \leq d} \left[ \frac{\partial F_{i_1}}{\partial x_{i_2} \cdots \partial x_{i_k}}(x) \right] z_{i_1}^1 \cdots z_{i_k}^k.
\]

Lastly, the notation \( a = O(b(k)) \) stands for an upper bound \( a \leq C \cdot b(k) \), where \( C > 0 \) is independent of the iteration count \( k \in \{1, 2, \ldots\} \) and \( a = \tilde{O}(b(k)) \) indicates the same inequality where \( C > 0 \) depends on logarithmic factors of \( k \).

## 2 Rescaled Gradient Dynamical System

We reformulate the rescaled gradient system in Eq. (1.2) as a closed-loop control system. We use dynamical systems concepts to establish the existence and uniqueness of a global solution of the system in Eq. (1.2) and to prove asymptotic weak and strong convergence of solution trajectories. We propose a novel Lyapunov function, which we employ to derive nonasymptotic convergence rates in terms of a restricted merit function (ergodic) and a residue function (nonergodic).

### 2.1 Reformulation via a closed-loop control

We rewrite the rescaled gradient system in Eq. (1.2) as follows:

\[
\dot{s}(t) = -\frac{F(x(t))}{\|F(x(t))\|^{1 - 1/p}}, \quad v(t) = x_0 + s(t), \quad x(t) - v(t) + \frac{F(x(t))}{\|F(x(t))\|^{1 - 1/p}} = 0.
\]

Introducing the function \( \lambda(t) = \frac{1}{\|F(x(t))\|^{1 - 1/p}} \), we have

\[
\dot{s}(t) = -\lambda(t)F(x(t)), \quad v(t) = x_0 + s(t), \quad x(t) - v(t) + \lambda(t)F(x(t)) = 0. \tag{2.1}
\]

Since \( s(t) = v(t) - x_0 \) and \( s(0) = 0 \), we have \( v(0) = x_0 \) and

\[
\dot{v}(t) = \dot{s}(t) = -\lambda(t)F(x(t)) = x(t) - v(t).
\]
In addition, we have \( x(t) = (I + \lambda(t)F)^{-1} v(t) \). Thus, we can rewrite Eq. (2.1) as

\[
\begin{cases}
\dot{v}(t) = (I + \lambda(t)F)^{-1} v(t) - v(t), \\
s(t) = v(t) - x_0, \\
x(t) = (I + \lambda(t)F)^{-1} v(t).
\end{cases}
\]

This implies that \((x, s)\) depend on the variables \((v, \lambda)\) explicitly and can be eliminated from the system. By the definition of \(\lambda(\cdot)\), we have

\[
1 = (\lambda(t))^p \| F(x(t)) \|^{p-1} = \lambda(t) \| \lambda(t) F(x(t)) \|^{p-1} = \lambda(t) \| \dot{v}(t) \|^{p-1}.
\]

Summarizing, the rescaled gradient system in Eq. (1.2) can be equivalently reformulated as a closed-loop control system [cf. Attouch et al., 2016a, Lin and Jordan, 2021b] as follows:

\[
\begin{cases}
\dot{v}(t) = (I + \lambda(t)F)^{-1} v(t) - v(t), \\
\lambda(t) \| \dot{v}(t) \|^{p-1} = 1, \\
v(0) = x_0 \in \Omega = \{x \in \mathcal{H} \mid F(x) \neq 0\}.
\end{cases}
\]

**Remark 2.1** The aforementioned relationship between the rescaled gradient systems in Eq. (1.2) and the systems in Eq. (2.2) pave the way for the existence and uniqueness of a global solution and the asymptotic weak and strong convergence of solution trajectories via appeal to results for a more general closed-loop control system studied by Lin and Jordan [2021b, Theorem 2.7, 3.1 and 3.6]. The closed-loop formulation provides a pathway to establishing convergence results that is distinct from those employed in other rescaled gradient dynamical systems associated with convex optimization algorithms [Cortés, 2006, Wibisono et al., 2016, Wilson et al., 2019, Romero and Benosman, 2020]. For further discussion of dynamical systems approaches to optimization we refer to Romero and Benosman [2020] and references therein.

### 2.2 Main results

We state our main theorem on the existence and uniqueness of a global solution.

**Theorem 2.2** The rescaled gradient dynamical system in Eq. (1.2) has a unique global solution, \((x, v, s) : [0, +\infty) \mapsto \mathcal{H} \times \mathcal{H} \times \mathcal{H}\). In addition, \(x(\cdot)\) is continuous, \(v(\cdot)\) and \(s(\cdot)\) are continuously differentiable, and \(\| F(x(t)) \|^{1/p-1}\) is nondecreasing as a function of \(t\). For the case of \(p \geq 2\), we have

\[
\| F(x(t)) \| \geq \| F(x_0) \| e^{-\frac{pt}{p-1}}, \quad \text{for all } t \geq 0.
\]

**Proof.** By Lin and Jordan [2021b, Theorem 2.7 and Lemma 2.11], the closed-loop control system in Eq. (2.2) has a unique global solution, \((v, \lambda) : [0, +\infty) \mapsto \mathcal{H} \times (0, +\infty)\). In addition, \(v(\cdot)\) is continuously differentiable and \(\lambda(\cdot)\) is locally Lipschitz continuous and nondecreasing. If \(p \geq 2\), we have

\[
\| v(t) - (I + \lambda(t)F)^{-1} v(t) \| \geq \| v(0) - (I + \lambda(0)F)^{-1} v(0) \| e^{-\frac{pt}{p-1}}, \quad \text{for all } t \geq 0.
\]

(2.3)

Recall that the rescaled gradient system in Eq. (1.2) can be reformulated as the system in Eq. (2.2) with the following transformations:

\[
s(t) = v(t) - x_0, \quad x(t) = (I + \lambda(t)F)^{-1} v(t), \quad \lambda(t) = \frac{1}{\| F(x(t)) \|^{1-1/p}}.
\]
Since $v(\cdot)$ is continuously differentiable, we have $s(\cdot)$ is continuously differentiable. Since $\lambda(\cdot)$ is locally Lipschitz continuous and nondecreasing, we obtain from the definition of $\lambda(t)$ that $\|F(x(t))\|^{1/p-1}$ is nondecreasing as a function of $t$. As for the relationship between $x(\cdot)$ and $v(\cdot)$, we have

$$v(t) = x(t) + \lambda(t)F(x(t)) = x(t) + \frac{F(x(t))}{\|F(x(t))\|^{1-1/p}},$$

which implies that $\|x(t) - v(t)\| = \|F(x(t))\|^{1/p}$ and further we have

$$F(x(t)) + \|x(t) - v(t)\|^{p-1}(x(t) - v(t)) = 0, \quad \text{for all } t \geq 0. \quad (2.4)$$

We claim that $x(\cdot)$ is uniquely determined by $v(\cdot)$ and that $x(\cdot)$ is continuous. Indeed, let $x_1(t)$ and $x_2(t)$ both satisfy Eq. (2.4) and let $x_1(t_0) \neq x_2(t_0)$ for some $t_0 > 0$. For simplicity, we assume that $x_1 = x_1(t_0)$, $x_2 = x_2(t_0)$, $v = v(t_0)$ and $\tilde{F}(\cdot) = F(\cdot) + \|\cdot - v\|^{p-1}(\cdot - v)$. Then, we have

$$\tilde{F}(x_1) = \tilde{F}(x_2) = 0.$$  

Define $f(\cdot) = \frac{1}{p+1}\|\cdot - v\|^{p+1}$, we have $\tilde{F}(\cdot) = F(\cdot) + \nabla f(\cdot)$. Since $f$ is strictly convex and $F$ is monotone, we have that $\tilde{F}$ is strictly monotone. Putting these pieces together yields that $x_1 = x_2$. By the definition of $x_1$ and $x_2$, this contradicts $x_1(t_0) \neq x_2(t_0)$. Thus we conclude that $x(\cdot)$ is uniquely determined by $v(\cdot)$.

We next prove the continuity of $x(\cdot)$ by contradiction. Assume that there exists $t_0 > 0$ such that $x(t_n) \to \tilde{x} \neq x(t_0)$ for some sequence $t_n \to t_0$. Note that Eq. (2.4) implies that

$$F(x(t_n)) + \|x(t_n) - v(t_n)\|^{p-1}(x(t_n) - v(t_n)) = 0.$$  

Taking the limit of this equation with respect to the sequence $t_n$ and using the continuity of $F$ and $v(\cdot)$, we have

$$F(\tilde{x}) + \|\tilde{x} - v(t_0)\|^{p-1}(\tilde{x} - v(t_0)) = 0.$$  

In addition, we have

$$F(x(t_0)) + \|x(t_0) - v(t_0)\|^{p-1}(x(t_0) - v(t_0)) = 0.$$  

Combining these two equations with the previous argument implies that $\tilde{x} = x(t_0)$ which is a contradiction. Thus, we conclude that $x(\cdot)$ is continuous.

Finally, we have

$$v(t) - (I + \lambda(t)F)^{-1}v(t) = \lambda(t)F(x(t)) = \frac{F(x(t))}{\|F(x(t))\|^{1-1/p}}. \quad (2.5)$$

Combining Eq. (2.3) and Eq. (2.5) implies that

$$\|F(x(t))\|^{1/p} \geq \|F(x_0)\|^{1/p}e^{-\frac{pt}{p}}, \quad \text{for all } t \geq 0.$$  

which is equivalent to

$$\|F(x(t))\| \geq \|F(x_0)\|e^{-\frac{pt}{p}}, \quad \text{for all } t \geq 0.$$  

This completes the proof. 

□
Remark 2.3 Theorem 2.2 demonstrates that $F(x(t)) \neq 0$ for all $t \geq 0$ which is equivalent to the assertion that the orbit $x(t)$ stays in $\Omega = \{x \in \mathcal{H} \mid F(x) \neq 0\}$. In other words, if $x(0) = x_0 \in \Omega$ and $s(0) = 0$ hold, the rescaled gradient dynamical system in Eq. (1.2) can not be stabilized in finite time, which helps clarify the asymptotic convergence behavior of the discrete-time dual extrapolation methods to a weak solution of variational inequality problems; see Nesterov [2007] for an example. In contrast, many other rescaled gradient dynamical systems associated with convex optimization algorithms [Cortés, 2006, Romero and Benosman, 2020] can exhibit finite-time convergence to an optimal solution when the objective function satisfies the generalized Polyak-Łojasiewicz (PL) inequality.

We present our results on the asymptotic weak convergence of solution trajectories and further establish asymptotic strong convergence results under additional conditions.

Theorem 2.4 Suppose that $(x, v, s) : [0, +\infty) \rightarrow \mathcal{H} \times \mathcal{H} \times \mathcal{H}$ is a global solution of the rescaled gradient dynamical system in Eq. (1.2). Then, there exists some $\bar{x} \in \{x \in \mathcal{H} \mid F(x) = 0\}$ such that the trajectories $x(t)$ and $v(t)$ both weakly converge to $\bar{x}$ as $t \rightarrow +\infty$. In addition, strong convergence results can be established if either of the following conditions holds true:

1. $F = \nabla f$ where $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex, differentiable and inf-compact.

2. $\{x \in \mathcal{H} \mid F(x) = 0\}$ has a nonempty interior.

Proof. By Lin and Jordan [2021b, Theorem 3.1 and 3.6], the unique global solution $(v, \lambda)$ of the closed-loop control system in Eq. (2.2) satisfies the condition that there exists some $\bar{x} \in \{x \in \mathcal{H} \mid F(x) = 0\}$ such that the trajectory $v(t)$ weakly converges to $\bar{x}$ and $\|\dot{v}(t)\| \rightarrow 0$ as $t \rightarrow +\infty$. Since the rescaled gradient system in Eq. (1.2) can be reformulated as the system in Eq. (2.2) with the same $v(t)$, there exists $\bar{x} \in \{x \in \mathcal{H} \mid F(x) = 0\}$ such that the trajectory $v(t)$ weakly converges to $\bar{x}$ as $t \rightarrow +\infty$ and the strong convergence results hold under the additional conditions stated in the theorem. Further, we see from Eq. (2.2) that $\|x(t) - v(t)\| = \|\dot{v}(t)\| \rightarrow 0$. This implies that the trajectory $x(t)$ weakly converges to the same $\bar{x}$ as $t \rightarrow +\infty$ with strong convergence under the same additional conditions. \hfill \Box

Remark 2.5 In an infinite-dimensional setting, the weak convergence of $x(\cdot)$ and $v(\cdot)$ to some $\bar{x} \in \{x \in \mathcal{H} \mid F(x) = 0\}$ in Theorem 2.4 is the best possible result we can expect without any additional conditions. Nonetheless, the strong convergence is more desirable since it guarantees that the distance between the trajectories and $\bar{x}$ converges to zero in terms of a norm [Bauschke and Combettes, 2001]. Indeed, Güler [1991] provided an example showing the importance of strong convergence where the sequence $\{f(x_k)\}_{k \geq 0}$ converges faster if $\{x_k\}_{k \geq 0}$ achieves strong convergence. In addition, the conditions imposed by Theorem 2.4 are verifiable and satisfied by many application problems.

Before stating the main theorems on the convergence rate, we define the restricted merit function and the residue function for the ME problems. The restricted merit function was introduced by Nesterov [2007] and plays a pivotal role in the VI literature [Facchinei and Pang, 2007]. Recall that $x_0 \in \Omega = \{x \in \mathcal{H} \mid F(x) \neq 0\}$ is an initial point and we define $D > 0$ such that there exists $\bar{x} \in \{x \in \mathcal{H} \mid F(x) = 0\}$ satisfying $\|x_0 - \bar{x}\| \leq D$. Then, the restricted merit function is defined by

$$\text{MERIT}(x) = \sup_{z \in \mathcal{H}} \{\langle F(z), x - z \rangle \mid \|z - x_0\| \leq D\}. \quad (2.6)$$
Clearly, $\text{merit}(\cdot)$ is well defined, closed and convex on $\mathcal{H}$. We also know from Nesterov [2007, Lemma 1] that $\text{merit}(x) \geq 0$ for all $x \in \mathcal{H}$ with equality if and only if $F(x) = 0$ holds. In addition, the residue function is directly derived from the ME problem as follows,

$$\text{res}(x) = \|F(x)\|. \tag{2.7}$$

We are now ready to present our main results on the global convergence rate estimation in terms of the restricted merit function in Eq. (2.6) and the residue function in Eq. (2.7).

**Theorem 2.6** Suppose that $(x, v, s): [0, +\infty) \mapsto \mathcal{H} \times \mathcal{H} \times \mathcal{H}$ is a global solution of the rescaled gradient dynamical system in Eq. (1.2). Then we have

$$\text{merit}(\tilde{x}(t)) = O(t^{-\frac{p+1}{2}}),$$

and

$$\text{res}(x(t)) = O(t^{-\frac{p}{2}}),$$

where $\tilde{x}(\cdot)$ is uniquely determined by $x(\cdot)$ as follows,

$$\tilde{x}(t) = \frac{1}{\int_0^t \|F(x(s))\|^{-\frac{1}{p}} \, ds} \left( \int_0^t \|F(x(s))\|^{-\frac{1}{p}} x(s) \, ds \right). \tag{2.9}$$

**Remark 2.7** Theorem 2.6 is new and extends known results for discrete-time dual extrapolation; indeed, the discrete-time version of our results have been obtained by the dual extrapolation method for $p = 1$ [Nesterov, 2007]. The idea of averaging for convex optimization and monotone VIs is not new and goes back to at least the mid-seventies [Bruck Jr, 1977, Lions, 1978, Nemirovski and Yudin, 1978, Nemirovski, 1981]. Its advantage was recently justified by proving that lower bounds for averaged iterate and last iterate are different [Ouyang and Xu, 2021]. Our results are in the same vein and demonstrate the importance of averaging in continuous-time dynamics of dual extrapolation methods by showing that the ergodic convergence rate can be faster for all $p \geq 1$.

We define a Lyapunov function for the system in Eq. (1.2) as follows:

$$E(t) = \|s(t)\|^2. \tag{2.8}$$

Note that the function in Eq. (2.8) is the squared norm of $s(t)$ which can be seen as a continuous-time version of a dual variable in the dual extrapolation methods. This function is simpler than that used for analyzing the convergence of Newton-like inertial systems [Attouch and Svaiter, 2011, Attouch et al., 2013, Abbas et al., 2014, Bot and Csetnek, 2016, Attouch and László, 2020, 2021] and closed-loop control systems [Attouch et al., 2016a, Lin and Jordan, 2021b]. We also note that our discrete-time analysis for high-order methods largely depends on the discrete-time version of the Lyapunov function in Eq. (2.8) with a small modification to handle the constraint sets (see Section 4).

**Proof of Theorem 2.6:** Using the definition of $E(\cdot)$ in Eq. (2.8), we have

$$\frac{dE(t)}{dt} = 2\langle \dot{s}(t), s(t) \rangle = -\frac{2\langle F(x(t)), v(t) - x_0 \rangle}{\|F(x(t))\|^{1-1/p}}. \tag{2.9}$$

9
Expanding this equality with any $z \in \mathcal{H}$, we have
\[
\frac{d\mathcal{E}(t)}{dt} = 2\langle F(x(t)), x_0 - z \rangle + \langle F(x(t)), z - x(t) \rangle + \langle F(x(t)), x(t) - v(t) \rangle
\]
\[
\|F(x(t))\|^{1-1/p}
\]
Since $\dot{s}(t) = -\frac{F(x(t))}{\|F(x(t))\|^{1-1/p}}$, we have
\[
\frac{\langle F(x(t)), x_0 - z \rangle}{\|F(x(t))\|^{1-1/p}} = \langle \dot{s}(t), x_0 - z \rangle.
\]
Since $F$ is monotone, we have
\[
\langle F(x(t)), z - x(t) \rangle \leq \langle F(z), z - x(t) \rangle.
\]
Since $x(t) - v(t) + \frac{F(x(t))}{\|F(x(t))\|^{1-1/p}} = 0$, we have
\[
\frac{\langle F(x(t)), x(t) - v(t) \rangle}{\|F(x(t))\|^{1-1/p}} = -\|F(x(t))\|^{\frac{2}{p}}.
\]
Plugging these pieces together in Eq. (2.9) yields that, for any $z \in \mathcal{H}$, we have
\[
\frac{d\mathcal{E}(t)}{dt} \leq 2\langle \dot{s}(t), x_0 - z \rangle - 2\|F(x(t))\|^{\frac{2}{p}} \langle F(z), x(t) - z \rangle - 2\|F(x(t))\|^\frac{2}{p}. \tag{2.10}
\]
Integrating this inequality over $[0, t]$ and using the fact that $\|F(x(t))\| \geq 0$ yields:
\[
\int_0^t \|F(x(s))\|^{\frac{1}{p}} \langle F(z), x(s) - z \rangle \, ds \leq \frac{1}{2} (\mathcal{E}(0) - \mathcal{E}(t) + \langle s(t), x_0 - z \rangle), \text{ for all } t \geq 0.
\]
Note that $\mathcal{E}(0) = \|s_0\|^2 = 0$. Then, we have
\[
\langle F(z), \tilde{x}(t) - z \rangle \leq \frac{1}{2 \left( \int_0^t \|F(x(s))\|^{\frac{1}{p}} \, ds \right)} \left( \langle s(t), x_0 - z \rangle - \|s(t)\|^2 \right), \text{ for all } t \geq 0.
\]
Young’s inequality implies that $\langle s(t), x_0 - z \rangle - \|s(t)\|^2 \leq \frac{1}{4} \|x_0 - z\|^2$. By the definition of $\text{MERIT}(\cdot)$, we have
\[
\text{MERIT}(\tilde{x}(t)) \leq \frac{D^2}{8 \left( \int_0^t \|F(x(s))\|^{\frac{1}{p}} \, ds \right)} = O \left( \frac{1}{\left( \int_0^t \|F(x(s))\|^{\frac{1}{p}} \, ds \right)} \right). \tag{2.11}
\]
Letting $z = \bar{x}$ be a solution satisfying $F(\bar{x}) = 0$ and $\|x_0 - \bar{x}\| \leq D$ in Eq. (2.10), we have
\[
\frac{d\mathcal{E}(t)}{dt} \leq 2\langle \dot{s}(t), x_0 - \bar{x} \rangle - 2\|F(x(t))\|^\frac{2}{p}.
\]
Integrating this inequality over $[0, t]$ and using $\mathcal{E}(0) = 0$, we have
\[
\int_0^t \|F(x(s))\|^{\frac{2}{p}} \, ds \leq \frac{1}{2} \left( \langle s(t), x_0 - z \rangle - \|s(t)\|^2 \right) \leq \frac{D^2}{8}, \text{ for all } t \geq 0.
\]
We claim that \( t \mapsto \|F(x(t))\| \) is nonincreasing. Indeed, for the case of \( p = 1 \), we have
\[
\dot{s}(t) = -F(x(t)), \quad v(t) = x_0 + s(t), \quad x(t) - v(t) + F(x(t)) = 0.
\]
In this case, we can write \( x(t) = (I + F)^{-1}v(t) \). Since \( v(\cdot) \) is continuously differentiable, we have \( x(\cdot) \) is also continuously differentiable. Define the function \( g(t) = \frac{1}{2}\|v(t) - x(t)\|^2 \). Then, we have
\[
\dot{g}(t) = \langle \dot{v}(t) - \dot{x}(t), v(t) - x(t) \rangle = -\langle \dot{v}(t) - \dot{x}(t), \dot{v}(t) \rangle = -\|\dot{v}(t) - \dot{x}(t)\|^2 - \langle \dot{v}(t) - \dot{x}(t), \dot{x}(t) \rangle.
\]
Since \( \dot{v}(t) - \dot{x}(t) = F(\dot{x}(t)) \) and \( F \) is monotone, we have
\[
\langle \dot{v}(t) - \dot{x}(t), \dot{x}(t) \rangle \geq 0.
\]
This implies that \( \dot{g}(t) \leq 0 \) and hence \( g(t) \) is nonincreasing. For the case of \( p \geq 2 \), Theorem 2.2 implies that \( \|F(x(t))\|^{1/p-1} \) is nondecreasing. Since \( 1/p - 1 < 0 \), we have \( t \mapsto \|F(x(t))\| \) is nonincreasing as desired. Putting these pieces together yields that
\[
\text{res}(x(t)) = \|F(x(t))\| \leq \left( \frac{D^2}{8t} \right)^{\frac{p}{2}} = O(t^{-\frac{p}{2}}).
\]
Further, we have
\[
\int_0^t \|F(x(s))\|^{\frac{p}{2}} ds \geq \int_0^t \left( \frac{D^2}{8s} \right)^{\frac{1-p}{2}} ds = \Omega(t^{\frac{p+1}{2}}).
\]
Plugging the above inequality into Eq. (2.11) yields the desired result.

2.3 Discussion

We compare the rescaled gradient system in Eq. (1.2) to related dynamical systems for monotone variational inequality and inclusion problems. We also give an overview of the existing rescaled gradient dynamical systems for convex optimization and comment on weak versus strong convergence.

Existing systems for variational inequality and inclusion problems. For convex optimization problems with a potential function, \( \Phi : H \mapsto \mathbb{R} \), Polyak [1964] initiated a line of work on damping approaches to speed up steepest descent methods. A decisive step to obtain a fast rate was taken by Su et al. [2016] who considered using asymptotically vanishing damping. Another key ingredient for obtaining acceleration is Hessian-driven damping which originated from a variational characterization of general regularization algorithms [Alvarez and P´erez, 1998] and have been extensively studied [Attouch and Redont, 2001, Alvarez et al., 2002, Attouch and Svaiter, 2011]. Subsequently, Attouch et al. [2016b] connected Nesterov acceleration with Hessian-driven damping and asymptotically vanishing damping and Shi et al. [2021] interpreted Nesterov acceleration via multiscale limits that distinguish it from the heavy ball method. Recently, Lin and Jordan [2021a] have introduced closed-loop control into the dynamical system with Hessian-driven damping and asymptotically vanishing damping, providing a variational characterization of optimal high-order algorithms for convex optimization.

These techniques have been translated to variational inequality and inclusion problems; indeed, Attouch and Svaiter [2011] considered the following regularized systems for inclusion problems:
\[
v(t) \in Ax(t), \quad \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0.
\]
If $A$ is maximally monotone and $\lambda(\cdot)$ satisfies minimal conditions, this system is well posed and the solution trajectory converges weakly to $A^{-1}(0)$. Note that various first-order algorithms have been obtained by implicit discretization of this system and several variants [Attouch et al., 2013, Abbas et al., 2014].

Further, if $A$ is a point-to-point cocoercive operator, inertial systems taking the following form have been considered [Attouch and Maingé, 2011, Maingé, 2013, Bot and Csetnek, 2016]:

$$\ddot{x}(t) + \alpha\dot{x}(t) + A(x(t)) = 0.$$  

Cocoercivity is crucial for guaranteeing the existence of a global solutions, weak asymptotic stabilization, and nonasymptotic convergence rate. Note that $A_{\lambda} = \frac{1}{\lambda}(I - (I + \lambda A)^{-1})$ is $\lambda$-cocoercive for some $\lambda > 0$ and $A_{\lambda}^{-1}(0) = A^{-1}(0)$. This motivate us to consider the following inertial system:

$$\ddot{x}(t) + \alpha\dot{x}(t) + A_{\lambda}(x(t)) = 0,$$

and the following variant:

$$\ddot{x}(t) + \alpha(t)\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0. \quad (2.12)$$

The discretization of the system in Eq. (2.12) gives a class of relaxed inertial proximal algorithms with theoretical guarantees [Attouch and Cabot, 2018, 2020, Attouch and Peypouquet, 2019]. Recently, Attouch and László [2020, 2021] proposed and analyzed Newton-like inertial dynamics for monotone inclusion problems with the following general formulation:

$$\ddot{x}(t) + \alpha(t)\dot{x}(t) + \beta(t)\frac{d}{dt}(A_{\lambda(t)}x(t)) + b(t)A_{\lambda(t)}x(t) = 0.$$  

Note that the term $\frac{d}{dt}(A_{\lambda(t)}x(t))$ generalizes Hessian-driven damping and gives a well-posed system with weak convergence of trajectories to $A^{-1}(0)$. However, convergence rates were obtained only for solving convex optimization where $A$ is the subdifferential of a convex function.

Most of the dynamical systems have been developed for modeling first-order algorithms for variational inequality and inclusion problems. The only exception we are aware of is Lin and Jordan [2016a, Lin and Jordan, 2021a] to monotone inclusion problems. However, the closed-loop control system in Lin and Jordan [2021b] fails to model the dual extrapolation method [Nesterov, 2007] and it remains unknown how the dual extrapolation step can exploit high-order smoothness.

**Rescaled gradient dynamical systems.** In continuous-time optimization, a dynamical system is designed to be computable under the oracles of a convex and differentiable function $f : \mathcal{H} \to \mathbb{R}$ such that the solution trajectory $x(t)$ converges to a minimizer of $f$. A classical example is the gradient flow in the form of $\dot{x}(t) + \nabla f(x(t)) = 0$ which has been long studied due to its ability to yield that $x(t)$ converges to a minimizer of $f$ as $t \to +\infty$ [Hadamard, 1908, Courant, 1943, Ambrosio et al., 2005]. Subsequently, Schropp [1995] and Schropp and Singer [2000] explored links between nonlinear dynamical systems and gradient-based optimization, including nonlinear constraints.

In his seminal work, Cortés [2006] proposed two discontinuous normalized modifications of gradient flows to attain finite-time convergence. Specifically, his dynamical systems are

$$\dot{x}(t) + \nabla f(x(t))\frac{1}{\|\nabla f(x(t))\|} = 0,$$  

$$(2.13)$$
and
\[ \dot{x}(t) + \text{sign}(\nabla f(x(t))) = 0. \] (2.14)

If \( f \) is twice continuously differentiable and strongly convex in an open neighborhood \( D \subseteq \mathbb{R}^n \) of \( x^* \), he proved that the maximal solutions to the dynamical systems in Eq. (2.13) and (2.14) (in the sense of Filippov [Filippov, 1964, Paden and Sastry, 1987, Arscott and Filippov, 1988]) exist and converge in finite time to \( x^* \), given that \( x_0 \) is in some positively invariant compact subset \( S \subseteq D \). The convergence times are upper bounded by

\[ t^* \leq \begin{cases} \frac{\|\nabla f(x_0)\|}{\min_{x \in S} \lambda_{\min} (\nabla^2 f(x))}, & \text{for Eq. (2.13)}, \\ \frac{\min_{x \in S} \lambda_{\min} (\nabla^2 f(x))}{\lambda_{\min} (\nabla^2 f(x))_1}, & \text{for Eq. (2.14)}. \end{cases} \]

Recently, Romero and Benosman [2020] have extended the finite-time convergence results to the rescaled gradient flow proposed by Wibisono et al. [2016] as follows:

\[ \dot{x}(t) + \frac{\nabla f(x(t))}{\|\nabla f(x(t))\|^{1-1/p}} = 0, \] (2.15)

and considered a new rescaled gradient flow as follows,

\[ \dot{x}(t) + \|\nabla f(x(t))\|^{1/p} \text{sign}(\nabla f(x(t))) = 0. \] (2.16)

In particular, if \( f \) is continuously differentiable and \( \mu \)-gradient-dominated of order \( q \in (1, p + 1) \) near a strict local minimizer \( x^* \), the maximal solutions to the dynamical systems in Eq. (2.15) and (2.16) (in the sense of Filippov) exist and converge in finite time to \( x^* \), given that \( \|x_0 - x^*\| > 0 \) is sufficiently small. The upper bounds for convergence times are summarized in Romero and Benosman [2020, Theorem 1].

In the context of dynamical systems associated with convex optimization algorithms, there exist other simple yet powerful techniques to strengthen the convergence properties of trajectories, including time scaling [Attouch et al., 2019a,c, 2021] and dry friction [Adly and Attouch, 2020, 2021].

However, the aforementioned works are restricted to the study of dynamical systems associated with convex optimization algorithms. It has been unknown if these methodologies can be extended to variational inequality and inclusion problems and further lead to accelerated first-order and high-order algorithms with convergence guarantees.

**Weak versus strong convergence.** We consider a Hilbert space with a generalized steepest descent dynamical system derived from a convex potential function \( \Phi \):

\[ -\dot{x}(t) \in \partial \Phi(x(t)), \quad x(0) = x_0. \]

It is well known that the solution trajectory of this system converges to a point \( \bar{x} \in \{ x : f(x) = \inf_{x \in H} f(x) \} \neq \emptyset \) [Brézis, 1973, 1978]. However, it remains challenging to characterize the relationship between \( \bar{x} \) and the choice of \( x_0 \) [Lemaire, 1996]; indeed, one counterexample shows that the trajectories of the above system converge weakly but not strongly [Baillon, 1978]. Accordingly, this issue is likely to hold for our rescaled gradient dynamical systems in Eq. (1.2) and we consider it an interesting open problem to find a counterexample (weak versus strong convergence).

Moreover, the convergence properties of solution trajectories are important aspects of the analysis, especially in an infinite-dimensional setting [Attouch and Svaiter, 2011, Attouch et al., 2013, 2016a,
Abbas et al., 2014, Bot and Csetnek, 2016, Attouch and Peypouquet, 2019, Attouch and Cabot, 2020, Attouch and László, 2020, 2021]. A few results are valid for weak convergence in general and become true for strong convergence only under additional conditions. Some results are valid in the ergodic sense with some particular metrics, e.g., the faster rate of $O(t^{-(p+1)/2})$ in this paper.

3 Implicit Discretization and High-Order Methods

We propose an algorithmic framework that arises via the implicit discretization of the system in Eq. (1.2). Our approach highlights the importance of the dual extrapolation step, interpreting it as the implicit discretization of a continuous-time integral. With an approximate tensor subroutine, we obtain a suite of simple $p^{th}$-order methods for solving monotone equation (ME) problems. Notably, our new methods do not require any line search schemes as are necessary for the existing $p^{th}$-order methods [Bullins and Lai, 2020, Lin and Jordan, 2021b, Jiang and Mokhtari, 2022].

3.1 Conceptual algorithmic framework

We begin by studying an algorithm which is derived by implicit discretization of the rescaled gradient dynamical system in Eq. (1.2) in a Euclidean setting. We have

$$\dot{s}(t) = -\frac{F(x(t))}{\|F(x(t))\|^{1/p}}, \quad v(t) = x_0 + s(t), \quad x(t) - v(t) + \frac{F(x(t))}{\|F(x(t))\|^{1-1/p}} = 0.$$  

By introducing $\lambda(t) = \|F(x(t))\|^{1/p-1}$, we obtain

$$\begin{cases}
\dot{s}(t) + \lambda(t)F(x(t)) = 0, \\
v(t) = x_0 + s(t), \\
x(t) - v(t) + \frac{F(x(t))}{\|F(x(t))\|^{1-1/p}} = 0, \\
(\lambda(t))^{p}\|F(x(t))\|^{p-1} = 1.
\end{cases}$$

Note that $\|x(t) - v(t)\| = \|F(x(t))\|^{1/p} = \|\lambda(t)F(x(t))\|$. Thus, we can rewrite the system as

$$\begin{cases}
\dot{s}(t) + \lambda(t)F(x(t)) = 0, \\
-s(t) + v(t) - x_0 = 0, \\
F(x(t)) + \|x(t) - v(t)\|^{p-1}(x(t) - v(t)) = 0, \\
\lambda(t)\|x(t) - v(t)\|^{p-1} = 1.
\end{cases} \tag{3.1}$$

Let us define a discrete-time sequence, $\{(x_k, v_{k+1}, s_k, \lambda_{k+1})\}_{k \geq 0}$, that corresponds to the continuous-time trajectory $\{(x(t), v(t), s(t), \lambda(t))\}_{t \geq 0}$. By using an implicit discretization of the dynamical system in Eq. (3.1), we have

$$\begin{cases}
s_{k+1} - s_k + \lambda_{k+1}F(x_{k+1}) = 0, \\
-s_k + v_{k+1} - x_0 = 0, \\
F(x_{k+1}) + \|x_{k+1} - v_{k+1}\|^{p-1}(x_{k+1} - v_{k+1}) = 0, \\
\lambda_{k+1}\|x_{k+1} - v_{k+1}\|^{p-1} = 1, \\
x_0 \in \{x \in \mathbb{R}^d \mid F(x) \neq 0\} \text{ and } s_0 = 0.
\end{cases} \tag{3.2}$$

We relax the equation $\lambda_{k+1}\|x_{k+1} - v_{k+1}\|^{p-1} = 1$ using $\lambda_{k+1}\|x_{k+1} - v_{k+1}\|^{p-1} \geq \theta$, where $\theta > 0$ is a scaling parameter which enhances the flexibility of our method. Putting these pieces together yields our conceptual algorithmic framework which is summarized in Algorithm 1.
Algorithm 1 Conceptual algorithmic framework

STEP 0: Let \( x_0 \in \mathcal{H} \), \( s_0 = 0 \in \mathbb{R}^d \), \( \theta > 0 \) and set \( k = 0 \).

STEP 1: If \( x_k \in \mathcal{X} \) is a solution of the ME problem in Eq. (1.1), then stop.

STEP 2: Compute \( v_{k+1} = x_k + s_k \).

STEP 3: Compute \( x_{k+1} \in \mathbb{R}^d \) such that \( F(x_{k+1}) + \|x_{k+1} - v_{k+1}\|^{p-1}(x_{k+1} - v_{k+1}) = 0 \).

STEP 4: Compute \( \lambda_{k+1} > 0 \) such that \( \lambda_{k+1}\|x_{k+1} - v_{k+1}\|^{p-1} \geq \theta \).

STEP 5: Compute \( s_{k+1} = s_k - \lambda_{k+1}F(x_{k+1}) \).

Note that our framework is intrinsically different from the framework of Monteiro and Svaiter [2012] and its high-order extension [Bullins and Lai, 2020, Lin and Jordan, 2021b, Jiang and Mokhtari, 2022] in that computing \( x_{k+1} \) in Algorithm 1 does not require any knowledge of \( \lambda_{k+1} \). In contrast, existing methods need to compute a pair of \( x_{k+1} \in \mathbb{R}^d \) and \( \lambda_{k+1} > 0 \) jointly since the computation of \( x_{k+1} \) is based on \( \lambda_{k+1} \) and the computation of \( \lambda_{k+1} \) is based on \( x_{k+1} \). This is why the nontrivial line search scheme is necessary. Note also that Algorithm 1 resembles a generalization of the classical first-order dual extrapolation method [Nesterov, 2007]; see more discussion in the next subsection.

3.2 High-order regularization method

By instantiating Algorithm 1 with an approximate tensor subroutine [Nesterov, 2021a], we obtain a new suite of simple \( p \)-th order methods for ME problems in which the operator \( F \in \mathcal{G}^p_L(\mathbb{R}^d) \) satisfies that \((p-1)\)-th order derivative is Lipschitz continuous.

In Algorithm 1, the following step plays a pivotal role:

Find \( x_{k+1} \in \mathcal{X} \) such that \( F(x_{k+1}) + \|x_{k+1} - v_{k+1}\|^{p-1}(x_{k+1} - v_{k+1}) = 0 \),

which involves solving a strictly monotone equation problem and which may be challenging from a computational viewpoint. Fortunately, we have \( F \in \mathcal{G}^p_L(\mathbb{R}^d) \) and can instead solve another subproblem with the \((p-1)\)-th order Taylor expansion of \( F \) at a point \( v \in \mathbb{R}^d \). We define

\[
F_v(x) = F(v) + \langle \nabla F(v), x - v \rangle + \ldots + \frac{1}{(p-1)!} \nabla^{(p-1)} F(v) [x - v]^{p-1} + \frac{2L}{(p-1)!} \|x - v\|^{p-1}(x - v),
\]

and let the \( x \)-subproblem be given by

Find \( x_{k+1} \in \mathcal{X} \) such that \( F_{v_{k+1}}(x_{k+1}) = 0 \). (3.4)

We summarize our new \( p \)-th order methods in Algorithm 2 which combines Algorithm 1 with an approximate tensor subroutine. For simplicity, we assume an access to an oracle which returns an exact solution of \( x \)-subproblem at each iteration.

We now show that the ME problem in Eq. (3.4) is relatively strongly monotone with respect to a reference function and thus be solved by a variant of Bregman first-order methods [Lu et al., 2018] with a linear rate of convergence. Indeed, we have \( \nabla F_{v_{k+1}}(x) = L \cdot I_{d \times d} \) for all \( x \in \mathbb{R}^d \) where \( I_{d \times d} \in \mathbb{R}^{d \times d} \) is an identity matrix if \( p = 1 \). Otherwise, we have

\[
\nabla F_{v_{k+1}}(x) = \nabla F(v_{k+1}) + \ldots + \frac{1}{(p-2)!} \nabla^{(p-1)} F(v_{k+1}) [x - v_{k+1}]^{p-2}
+ \frac{2L}{(p-1)!} \|x - v_{k+1}\|^{p-1} I_{d \times d} + \frac{2L}{(p-2)!} \|x - v_{k+1}\|^{p-2}(x - v_{k+1})(x - v_{k+1})^\top.
\]
Algorithm 2 A class of $p$-th order methods

**Input:** order $p$, initial point $x_0 \in \mathbb{R}^d$, parameter $L$ and iteration number $T$.
**Initialization:** set $s_0 = 0_d \in \mathbb{R}^d$.

for $k = 0, 1, 2, \ldots, T$ do

**STEP 1:** If $x_k \in X$ is a solution of the ME problem in Eq. (1.1), then **stop**.

**STEP 2:** Compute $v_{k+1} = x_0 + s_k$.

**STEP 3:** Compute $x_{k+1} \in \mathbb{R}^d$ such that $F_{v_{k+1}}(x_{k+1}) = 0$ holds.

**STEP 4:** Compute $\lambda_{k+1} > 0$ such that $\frac{1}{12p-6} \leq \frac{\lambda_{k+1}L\|x_{k+1} - v_{k+1}\|^{p-1}}{p!} \leq \frac{1}{4p+2}$.

**STEP 5:** Compute $s_{k+1} = s_k - \lambda_{k+1}F(x_{k+1})$.
end for

**Output:** $\bar{x}_T = \frac{1}{\sum_{k=1}^{T} \lambda_k} \sum_{k=1}^{T} \lambda_k x_k$.

Since $F \in G^p_L(\mathbb{R}^d)$ satisfies that $(p-1)^{th}$-order derivative is Lipschitz continuous with a constant $L > 0$, we have

$$
\|\nabla F(x) - (\nabla F(v_{k+1}) + \ldots + \frac{1}{(p-2)!}\nabla^{(p-1)}F(v_{k+1})[x - v_{k+1}]^{p-2}\| \leq \frac{L}{(p-1)!}\|x - v_{k+1}\|^{p-1}.
$$

This implies that

$$
\nabla F_{v_{k+1}}(x) = \nabla F(x) + \frac{L}{(p-1)!}\|x - v_{k+1}\|^{p-1}I_d \times d + \frac{2L}{(p-2)!}\|x - v_{k+1}\|^{p-2}(x - v_{k+1})(x - v_{k+1})^T.
$$

Since $F$ is monotone, we have that $\nabla F(x)$ is positive semidefinite for all $x \in \mathbb{R}^d$. As a consequence, we have that $F_{v_{k+1}}$ is relatively strongly monotone.

In practice, we often work with the case of $p = 2$ and hope to find $x_{k+1} \in \mathbb{R}^d$ such that

$$
F_{v_{k+1}}(x_{k+1}) = 0. \tag{3.5}
$$

For the optimization setting, with $F = \nabla f$ where $f$ is convex and smooth function and has a Lipschitz Hessian, solving the ME in Eq. (3.5) is equivalent to solving the subproblem of the cubic regularization of Newton’s method [Nesterov and Polyak, 2006, Nesterov, 2008]:

$$
x_{k+1} = \arg\min_{x \in \mathbb{R}^d} (\nabla f(v_{k+1}), x - v_{k+1}) + \frac{1}{2}(x - v_{k+1}, \nabla^2 f(v_{k+1})(x - v_{k+1}) + \frac{2L}{3}\|x - v_{k+1}\|^3.
$$

Concrete examples of cubic subsolvers include the generalized conjugate gradient methods with the Lanczos process [Gould et al., 1999, 2010] and a simple variant of gradient descent [Carmon and Duchi, 2019]. There are also several easy-to-satisfy approximation conditions that allow the subproblem to be solved inexactly in high-order methods [Birgin et al., 2017, Bellavia et al., 2019, Jiang et al., 2020, Yao et al., 2021, Grapiglia and Nesterov, 2021]. Can we generalize these subsolvers and approximation conditions to handle the ME in Eq. (3.5), similar to what has been accomplished in the optimization setting? This topic is beyond the scope of this paper and we leave it to future work.

3.3 Restarted scheme

We present the restarted version of Algorithm 2 in Algorithm 3, which combines Algorithm 2 with the classical restart scheme; [see, e.g., Nemirovskii and Nesterov, 1985, Nesterov, 2013, 2018].
Algorithm 3 The restart version of Algorithm 2

**Input:** order $p$, initial point $x_0 \in \mathbb{R}^d$, parameter $L$ and iteration number $T$.

**for** $k = 0, 1, 2, \ldots, T$ **do**

**STEP 1:** If $x_k \in \mathbb{R}^d$ is a solution of the ME problem in Eq. (1.1), then **stop**.

**STEP 2:** Compute $x_{k+1}$ as an output of Algorithm 2 with the input $(p, x_k, L, 1)$, i.e., one iteration of Algorithm 2 but with $x_k$ as an initial point.

**end for**

**Output:** $x_{T+1}$.

The restart scheme stops an algorithm when a criterion is satisfied and then restarts the algorithm with a new input. It has been an important algorithmic component for speeding up first-order methods when the objective function is strongly convex [Nesterov, 2013]. Assuming that $F$ is strongly monotone around the solution $x^*$, it is thus natural to consider Algorithm 3 which is expected to achieve the linear convergence for $p = 1$ (first-order methods) and local superlinear convergence for $p \geq 2$ (high-order methods). Concerning the scheme of Algorithm 3, we run one iteration of Algorithm 2 with $x_k$ as an initial point at each iteration; indeed, we restart Algorithm 2 every iteration.

Many adaptive restart schemes have been proposed in recent years to speed up convergence of first-order methods [O’donoghue and Candes, 2015] and some of them have been analyzed theoretically when the objective function is smooth and has Hölderian growth [Roulet and d’Aspremont, 2017, Fercoq and Qu, 2019]. However, these schemes depend on learning appropriately accurate approximations of problem parameters. To alleviate this issue, Renegar and Grimmer [2022] proposed a simple adaptive restart scheme which only require the information that would be available in practice. To that end, it is promising to either combine our method with adaptive restart schemes, or prove the local convergence under other weaker conditions than the strong monotonicity of $F$.

### 3.4 Discussion

We provide an overview of Lyapunov analysis techniques for analyzing the existing dynamical systems associated with convex optimization and monotone inclusion problems. This sheds light on our discrete-time convergence analysis in Section 4. We also comment on high-order algorithms obtained via time discretization of dynamical systems in some generality.

**Lyapunov analysis.** The interplay between continuous-time dynamics and discrete-time algorithms has been a particular milestone in optimization theory. Typical examples include (1) the interpretation of steepest descent, heavy ball and proximal algorithms as explicit and implicit discretization of gradient-like dissipative systems [Polyak, 1987, Antipin, 1994, Attouch and Cominetti, 1996, Alvarez, 2000, Attouch et al., 2000, Alvarez and Attouch, 2001]; and (2) the explicit discretization of Newton-like and Levenberg-Marquardt regularized systems [Alvarez and Pérez, 1998, Attouch and Redont, 2001, Alvarez et al., 2002, Attouch and Svaiter, 2011, Attouch et al., 2012, Maingé, 2013, Attouch et al., 2013, Abbas et al., 2014, Attouch et al., 2016a, Attouch and Lásló, 2020, 2021]. One particularly salient way that these connections have spurred research is via the Lyapunov functions which transfer asymptotic behavior and rates of convergence between continuous time and discrete time.

Starting with work of [Su et al., 2016], there has been a line of research on Lyapunov analysis for accelerated first-order optimization methods. In particular, Wilson et al. [2021] found an unified time-
dependent Lyapunov function and showed that their Lyapunov analysis is equivalent to Nesterov’s estimate sequence technique in a number of cases, including quasi-monotone subgradient, accelerated gradient descent and conditional gradient. Moreover, the Lyapunov analysis for variational inequality and inclusion problems become simple since the associated dynamical systems do not contain any damping terms [Attouch and Svaiter, 2011, Attouch et al., 2013, Abbas et al., 2014, Attouch et al., 2016a, Lin and Jordan, 2021b] when the operator is monotone without any additional structure (e.g., cocoercivity). Notably, the Lyapunov function has an intuitive interpretation with a common choice being the distance between the solution trajectory and an optimal solution.

**From dynamical systems to high-order algorithms.** Most of the high-order algorithms obtained via discretization of dynamical systems focus on convex optimization [Wibisono et al., 2016, Song et al., 2021, Lin and Jordan, 2021a], with an exception being recent work of Lin and Jordan [2021b] on monotone inclusion problems. In particular, Wibisono et al. [2016] showed that the cubic regularization of Newton’s methods and their variants [Nesterov and Polyak, 2006, Nesterov, 2008, 2021a] can be derived from the implicit discretization of the following open-loop system without Hessian-driven damping:

\[ \ddot{x}(t) + \alpha(t) \dot{x}(t) + C(p+1)^2 \dot{x}(t) + C(p+1)^2 \nabla \Phi(x(t)) = 0. \]

The convergence rate of their algorithms is \( O(k^{-(p+1)}) \) in terms of objective function gap. Another contribution is due to Song et al. [2021], who proposed an alternative open-loop dynamical system (we consider the simplified form in a Euclidean setting):

\[ \ddot{x}(t) + \left( \frac{2 \dot{a}(t)}{a(t)} - \frac{\ddot{a}(t)}{a(t)} \right) \dot{x}(t) + \left( \frac{\dot{a}(t)^2}{a(t)} \right) \nabla \Phi(x(t)) = 0. \]

Recently, Lin and Jordan [2021a] have provided a control-theoretic perspective on accelerated high-order algorithms where they considered a closed-loop system with Hessian-driven damping:

\[ \ddot{x}(t) + \alpha(t) \dot{x}(t) + \beta(t) \nabla^2 \Phi(x(t)) \dot{x}(t) + b(t) \nabla \Phi(x(t)) = 0, \]

where \( (\alpha, \beta, b) \) are defined by

\[
\begin{align*}
\alpha(t) &= \frac{2 \dot{a}(t)}{a(t)} - \frac{\ddot{a}(t)}{a(t)}, \\
\beta(t) &= \frac{\dot{a}(t)^2}{a(t)}, \\
b(t) &= \frac{\dot{a}(t)(\dot{a}(t) + \ddot{a}(t))}{a(t)}, \\
a(t) &= \frac{1}{4} \int_0^t \sqrt{\lambda(s)} ds, \\
(\lambda(t))^{p-1} \nabla \Phi(x(t)) ||^{p-1} = \theta.
\end{align*}
\]

They recovered a class of best possible \( p \)-order algorithms [Monteiro and Svaiter, 2013, Gasnikov et al., 2019, Alves, 2022] from implicit discretization of the above system and also proved the convergence rate of \( O(t^{-(3p+1)/2}) \) via appeal to a novel Lyapunov function.

There is comparatively little work on the development of high-order tensor algorithms for variational inequality and inclusion problems; indeed, we are only aware of Lin and Jordan [2021b] who obtained high-order extragradient methods [Bullins and Lai, 2020] via the discretization of closed-loop systems. High-order optimistic gradient methods have also been proposed [Jiang and Mokhtari, 2022] but the derivation does not flow from a single underlying principle and involves case-specific algebra. It is interesting to investigate the continuous-time dynamics and Lyapunov analysis for these algorithms.
4 Discrete-Time Convergence Analysis

We provide the discrete-time convergence analysis for our new $p$th-order methods from Algorithm 2. More specifically, we prove the global rate of $O(k^{-(p+1)/2})$ in terms of a restricted merit function (ergodic) and $O(k^{-p/2})$ in terms of a residue function (nonergodic).

4.1 Global convergence rate estimation

We present an ergodic and a pointwise estimate of the global convergence rate for Algorithm 2. Our analysis is motivated by the previous continuous-time analysis and our results are expressed in terms of discrete-time versions of the restricted merit function in Eq. (2.6) and the residue function in Eq. (2.7).

We start with the presentation of our main results for Algorithm 2, which generalizes the theoretical results in Nesterov [2007, Theorem 2] in terms of a restricted merit function. To facilitate the presentation, we define a restricted merit function as follows:

$$\text{merit}(x) = \sup_{z \in \mathbb{R}^d} \{(F(z), x - z) \mid \|z - x_0\| \leq D\},$$

where $D > 0$ is defined such that there exists $\bar{x} \in \{x \in \mathbb{R}^d \mid F(x) = 0\}$. This definition is commonly used for measuring an approximate weak solution of variational inequalities (VIs) in the literature [Nesterov, 2007, Facchinei and Pang, 2007] and is clearly the discrete-time counterpart of the function in Eq. (2.6). However, such an optimality criterion may be mainly of theoretical interest since $D > 0$ is unknown in practice. We are also aware of other optimality criteria for characterizing the approximate solution of VIs when the constraint set is unbounded [Monteiro and Svaiter, 2010]. Our analysis can be extended using their modified optimality criterion but the proof becomes longer and its link with continuous-time analysis becomes unclear. Accordingly, we focus on the restricted merit function for simplicity.

We also present theoretical analysis in terms of a residue function. More specifically, we define a residue function as follows:

$$\text{res}(x) = \|F(x)\|.$$

We summarize our results in the following theorem.

**Theorem 4.1** Let $k \geq 1$ be an integer and let $F \in \mathcal{G}_p^f(\mathbb{R}^d)$ satisfy that the $(p - 1)$th-order derivative is Lipschitz continuous. For the iterates $\{x_k\}_{k \geq 0}$ generated by Algorithm 2, we have

$$\text{merit}(\tilde{x}_k) = O(k^{-\frac{p+1}{2}}),$$

and

$$\inf_{0 \leq i \leq k} \text{res}(x_i) = O(k^{-\frac{p}{2}}),$$

where the ergodic iterates are defined by $\tilde{x}_k = \frac{1}{\sum_{i=1}^k \lambda_i} (\sum_{i=1}^k \lambda_i x_i)$ for all $k \geq 1$.

**Remark 4.2** The ergodic convergence result in Theorem 4.1 was established by Nesterov [2007, Theorem 2] for the case of $p = 1$ and the same result was derived by Nemirovski [2004] and Monteiro and Svaiter [2010] for the extragradient method (EG) and its variants. For $p \geq 2$ in general, the convergence rates of $O(k^{-\frac{p+1}{2}} \log(k))$ and $O(k^{-\frac{p}{2}} \log(k))$ were obtained for high-order EG methods [Bullins and Lai, 2020, Lin and Jordan, 2021b] and high-order optimistic gradient methods [Jiang and Mokhtari, 2022].
We construct a discrete-time Lyapunov function for Algorithm 2 as follows:

\[ \mathcal{E}_k = \frac{1}{2}\|s_k\|^2, \quad (4.1) \]

which will be used to prove technical results that pertain to Algorithm 2.

**Lemma 4.3** For every integer \( T \geq 1 \), we have

\[
\sum_{k=1}^{T} \lambda_k \langle F(x_k), x_k - x \rangle \leq \mathcal{E}_0 - \mathcal{E}_T + \langle s_T, x - x_0 \rangle - \frac{1}{10} \left( \sum_{k=1}^{T} \|x_k - v_k\|^2 \right), \quad \text{for all } x \in \mathcal{X}.
\]

**Proof.** We have

\[
\mathcal{E}_{k+1} - \mathcal{E}_k = \langle s_{k+1} - s_k, s_{k+1} - s_k \rangle - \frac{1}{2} \|s_{k+1} - s_k\|^2.
\]

Combining this equation with the definition of \( v_{k+1} \), we have

\[
\mathcal{E}_{k+1} - \mathcal{E}_k = \lambda_{k+1} \langle F(x_{k+1}), x_0 - v_{k+2} \rangle - \frac{1}{2} \|v_{k+2} - v_{k+1}\|^2.
\]

Letting \( x \in \mathbb{R}^d \) be any point, we have

\[
\mathcal{E}_{k+1} - \mathcal{E}_k \leq \lambda_{k+1} \langle F(x_{k+1}), x_0 - x \rangle + \lambda_{k+1} \langle F(x_{k+1}), x - x_{k+1} \rangle + \lambda_{k+1} \langle F(x_{k+1}), x_{k+1} - v_{k+2} \rangle - \frac{1}{2} \|v_{k+2} - v_{k+1}\|^2.
\]

Summing up this inequality over \( k = 0, 1, \ldots, T - 1 \) and changing the counter \( k + 1 \) to \( k \) yields

\[
\sum_{k=1}^{T} \lambda_k \langle F(x_k), x_k - x \rangle \leq \mathcal{E}_0 - \mathcal{E}_T + \underbrace{\sum_{k=1}^{T} \lambda_k \langle F(x_k), x_0 - x \rangle}_{I} \quad \underbrace{+ \sum_{k=1}^{T} \lambda_k \langle F(x_k), x_k - v_{k+1} \rangle - \frac{1}{2} \|v_{k+2} - v_{k+1}\|^2}_{II}. \quad (4.2)
\]

Using the update formula \( s_{k+1} = s_k - \lambda_{k+1} F(x_{k+1}) \) and \( s_0 = 0 \in \mathbb{R}^d \), we have

\[
I = \left\langle \sum_{k=1}^{T} \lambda_k F(x_k), x_0 - x \right\rangle = \langle s_0 - s_T, x_0 - x \rangle = \langle s_T, x - x_0 \rangle. \quad (4.3)
\]

We rewrite the update formula of \( x_k \) as follows:

\[
F_{v_k}(x_k) = 0, \quad (4.4)
\]

where \( F_v(x) : \mathbb{R}^d \to \mathbb{R}^d \) is defined with any fixed \( v \in \mathcal{X} \) as follows,

\[
F_v(x) = F(v) + \langle \nabla F(v), x - v \rangle + \ldots + \frac{1}{(p-1)!} \nabla^{(p-1)} F(v) [x - v]^{p-1} + \frac{2L}{(p-1)!} \|x - v\|^{p-1} (x - v).
\]

Since \( F \in \mathcal{G}_L^p(\mathbb{R}^d) \), we have

\[
\|F(x_k) - F_{v_k}(x_k) + \frac{2L}{(p-1)!} \|x_k - v_k\|^{p-1} (x_k - v_k)\| \leq \frac{L}{p!} \|x_k - v_k\|^p. \quad (4.5)
\]
We perform a decomposition of \( \langle F(x_k), x_k - v_{k+1} \rangle \) and derive from Eq. (4.4) and Eq. (4.5) that
\[
\langle F(x_k), x_k - v_{k+1} \rangle = \langle F(x_k) - Fv_k(x_k), x_k - v_{k+1} \rangle + \langle Fv_k(x_k), x_k - v_{k+1} \rangle - \frac{2L}{(p-1)!} \| x_k - v_k \|^{p-1} \langle x_k - v_k, x_k - v_{k+1} \rangle \\
\leq \| F(x_k) - Fv_k(x_k) \| \| x_k - v_k \|^{p-1} \| x_k - v_{k+1} \| + \langle Fv_k(x_k), x_k - v_{k+1} \rangle - \frac{2L}{(p-1)!} \| x_k - v_k \|^{p-1} \langle x_k - v_k, x_k - v_{k+1} \rangle \\
\leq \frac{L}{p!} \| x_k - v_k \|^{p} \| x_k - v_{k+1} \| + \frac{2L}{(p-1)!} \| x_k - v_k \|^{p-1} \langle x_k - v_k, x_k - v_{k+1} \rangle \\
\leq \frac{L}{p!} \| x_k - v_k \|^{p+1} + \frac{2L}{p!} \| x_k - v_k \|^{p} \| v_k - v_{k+1} \| - \frac{2L}{(p-1)!} \| x_k - v_k \|^{p-1} \langle x_k - v_k, x_k - v_{k+1} \rangle.
\]
Note that we have
\[
\langle x_k - v_k, x_k - v_{k+1} \rangle = \| x_k - v_k \|^{2} + \langle x_k - v_k, x_k - v_{k+1} \rangle \geq \| x_k - v_k \|^{2} - \| x_k - v_k \| \| v_k - v_{k+1} \|.
\]
Putting these pieces together yields that
\[
\langle F(x_k), x_k - v_{k+1} \rangle \leq \frac{(2p+1)L}{p!} \| x_k - v_k \|^{p} \| v_k - v_{k+1} \| - \frac{(2p-1)\lambda L}{p!} \| x_k - v_k \|^{p+1}.
\]
Since \( \frac{1}{12p-6} \leq \frac{\lambda \| x_k - v_k \|^{p-1}}{p!} \leq \frac{1}{4p+2} \) for all \( k \geq 1 \), we have
\[
\| \| \leq \sum_{k=1}^{T} \left( \frac{(2p+1)\lambda L}{p!} \| x_k - v_k \|^{p} \| v_k - v_{k+1} \| - \frac{1}{2} \| v_k - v_{k+1} \|^{2} - \frac{1}{6} \| x_k - v_k \|^{2} \right) \\
\leq \sum_{k=1}^{T} \left( \frac{1}{2} \| x_k - v_k \|^{2} - \frac{1}{6} \| x_k - v_k \|^{2} \right) \\
\leq \sum_{k=1}^{T} \left( \max_{\eta \geq 0} \left\{ \frac{1}{2} \| x_k - v_k \|^{2} - \frac{1}{6} \| x_k - v_k \|^{2} \right\} \right) \\
= -\frac{1}{24} \left( \sum_{k=1}^{T} \| x_k - v_k \|^{2} \right).
\]
Plugging Eq. (4.3) and Eq. (4.6) into Eq. (4.2) yields that
\[
\sum_{k=1}^{T} \lambda_k \langle F(x_k), x_k - x \rangle \leq \mathcal{E}_0 - \mathcal{E}_T + \langle s_T, x - x_0 \rangle - \frac{1}{24} \left( \sum_{k=1}^{T} \| x_k - v_k \|^{2} \right).
\]
This completes the proof. \( \Box \)

**Lemma 4.4** For every integer \( T \geq 1 \) and let \( x \in \mathbb{R}^d \), we have
\[
\sum_{k=1}^{T} \lambda_k \langle F(x_k), x_k - x \rangle \leq \frac{1}{2} \| x - x_0 \|^{2}, \quad \sum_{k=1}^{T} \| x_k - v_k \|^{2} \leq 12 \| x^* - x_0 \|^{2},
\]
where \( x^* \in \mathbb{R}^d \) is a solution of the ME in Eq. (1.1) satisfying that \( F(x^*) = 0 \).
Proof. For any \( x \in \mathbb{R}^d \), we have
\[
\mathcal{E}_0 - \mathcal{E}_T + \langle s_T, x - x_0 \rangle = \mathcal{E}_0 - \frac{1}{2} \| s_T \|^2 + \langle s_T, x - x_0 \rangle.
\]
Since \( s_0 = 0 \), we have \( \mathcal{E}_0 = 0 \). By Young’s inequality, we have
\[
\mathcal{E}_0 - \mathcal{E}_T + \langle s_T, x - x_0 \rangle \leq -\frac{1}{2} \| s_T \|^2 + \frac{1}{2} \| s_T \|^2 + \frac{1}{2} \| x - x_0 \|^2 = \frac{1}{2} \| x - x_0 \|^2.
\]
This together with Lemma 4.3 yields that
\[
\sum_{k=1}^{T} \lambda_k \langle F(x_k), x_k - x \rangle + \frac{1}{2} \left( \sum_{k=1}^{T} \| x_k - v_k \|^2 \right) \leq \frac{1}{2} \| x - x_0 \|^2, \quad \text{for all } x \in \mathbb{R}^d,
\]
which implies the first inequality. Letting \( x = x^* \) be a solution of the ME in Eq. (1.1) satisfying that \( F(x^*) = 0 \) in the above inequality yields the second inequality.
\[\square\]

We provide a technical lemma on a lower bound for \( \sum_{k=1}^{T} \lambda_k \). The analysis is motivated by continuous-time analysis for the system in Eq. (1.2).

**Lemma 4.5** For \( p \geq 1 \) and every integer \( k \geq 1 \), we have
\[
\sum_{k=1}^{T} \lambda_k \geq \frac{p}{(12p - 6)L} \left( \frac{1}{12 \| x^* - x_0 \|^2} \right)^{\frac{p-1}{p+1}} T^{\frac{p+1}{p}}
\]
where \( x^* \in \mathbb{R}^d \) is a solution of the ME in Eq. (1.1) satisfying that \( F(x^*) = 0 \).

*Proof.* For \( p = 1 \), we have \( \lambda_k = \frac{1}{6L} \) for all \( k \geq 1 \). For \( p \geq 2 \), we have
\[
\sum_{k=1}^{T} (\lambda_k)^{-\frac{2}{p+1}} \left( \frac{p}{(12p - 6)L} \right)^{\frac{2}{p+1}} \leq \sum_{k=1}^{T} (\lambda_k)^{-\frac{2}{p+1}} \lambda_k \| x_k - v_k \|^{p-1} \frac{2}{p+1} \| x_k - v_k \|^2 \leq \sum_{k=1}^{T} \| x_k - v_k \|^2 \leq 12 \| x - x_0 \|^2.
\]

By Hölder’s inequality, we have
\[
\sum_{k=1}^{T} 1 = \sum_{k=1}^{T} \left( (\lambda_k)^{-\frac{2}{p+1}} (\lambda_k)^{\frac{2}{p+1}} \right) \leq \left( \sum_{k=1}^{T} (\lambda_k)^{-\frac{2}{p+1}} \right)^{-\frac{1}{p+1}} \left( \sum_{k=1}^{T} (\lambda_k)^{\frac{2}{p+1}} \right)^{\frac{1}{p+1}}.
\]

Putting these pieces together yields that
\[
T \leq (12 \| x^* - x_0 \|^2)^{\frac{p-1}{p+1}} \left( \frac{(12p - 6)L}{p} \right)^{\frac{2}{p+1}} \left( \sum_{k=1}^{T} \lambda_k \right)^{\frac{1}{p+1}},
\]
which implies that
\[
\sum_{k=1}^{T} \lambda_k \geq \frac{p}{(12p - 6)L} \left( \frac{1}{12 \| x^* - x_0 \|^2} \right)^{\frac{p-1}{p+1}} T^{\frac{p+1}{p}}.
\]
This completes the proof. \[\square\]
Proof of Theorem 4.1: For simplicity, we let \( x^* \in \mathbb{R}^d \) be a solution of the ME in Eq. (1.1) satisfying \( F(x^*) = 0 \). Then, we can derive from the first inequality in Lemma 4.4 and the monotonicity of \( F \) that

\[
\langle F(z), \tilde{x}_k - z \rangle \leq \frac{1}{\sum_{i=1}^k \lambda_i} \left( \sum_{i=1}^k \lambda_i \langle F(x_i), x_i - z \rangle \right) = \frac{\|z - x_0\|^2}{2(\sum_{i=1}^k \lambda_i)}, \quad \text{for all } z \in \mathbb{R}^d.
\]

By the definition of a restricted merit function, we have

\[
\operatorname{MERIT}(&\tilde{x}_k) = \sup_{z \in \mathbb{R}^d} \{ \langle F(z), \tilde{x}_k - z \rangle \mid \|z - x_0\| \leq D \} \leq \frac{D^2}{2(\sum_{i=1}^k \lambda_i)}.
\]

By Lemma 4.5, we have

\[
\operatorname{MERIT}(&\tilde{x}_k) \leq \frac{(6p-3)L}{p!} \left( 12\|x^* - x_0\|^2 \right)^{p-1} D^2 k^{-\frac{p+2}{2}} = O(k^{-\frac{p+2}{2}}).
\]

The second inequality in Lemma 4.4 implies that

\[
\inf_{1 \leq i \leq k} \|x_i - v_i\|^2 \leq \frac{L}{p!} \sum_{i=1}^k \|x_i - v_i\|^2 \leq \frac{12\|x^* - x_0\|^2}{k}.
\]

In addition, we recall that the update formula of \( x_k \) is

\[
F_{v_k}(x_k) = 0.
\]

Recall that \( F \in G^p_L(\mathbb{R}^d) \) and \( F_v(x) : \mathbb{R}^d \to \mathbb{R}^d \) is defined by

\[
F_v(x) = F(v) + \langle \nabla F(v), x - v \rangle + \ldots + \frac{1}{(p-1)!} \nabla^{(p-1)} F(v)[x - v]^{p-1} + \frac{2L}{(p-1)!} \|x - v\|^{p-1} (x - v).
\]

Then, we have

\[
\|F(x_k) - F_{v_k}(x_k)\| + \frac{2L}{(p-1)!} \|x_k - v_k\|^{p-1} (x_k - v_k) \leq \frac{L}{p!} \|x_k - v_k\|^p,
\]

and

\[
\|F(x_k)\| \leq \frac{L}{p!} \|x_k - v_k\|^p + \|F_{v_k}(x_k)\| + \frac{2L}{(p-1)!} \|x_k - v_k\|^{p-1} \|x_k - v_k\|^p \leq \frac{(2p+1)L}{p!} \|x_k - v_k\|^p.
\]

By the definition of \( \operatorname{RES}(\cdot) \), we have

\[
\inf_{1 \leq i \leq k} \operatorname{RES}(x_i) = \inf_{1 \leq i \leq k} \|F(x_i)\| \leq \left( \inf_{1 \leq i \leq k} \|x_i - v_i\|^p \right) \left( \frac{(2p+1)L}{p!} \right).
\]

Plugging Eq. (4.7) into Eq. (4.8) yields that

\[
\inf_{1 \leq i \leq k} \operatorname{RES}(x_i) = O(k^{-\frac{p}{2}}).
\]

This completes the proof.

Remark 4.6 The discrete-time analysis in this subsection is based on the discrete-time Lyapunov function in Eq. (4.1), which is closely related to the continuous-time counterpart in Eq. (2.8). It is also worth mentioning that the proofs of these technical results follow the same path for the analysis in Theorem 2.6.
4.2 Local convergence rate estimation

We present an estimate of the local convergence rate for Algorithm 3 when \( p \geq 2 \). Our analysis is based on the analysis in the last subsection and our results are expressed in terms of the distance function \( \| x_k - x^* \| \), where \( x^* \in \mathbb{R}^d \) is a solution of the ME in Eq. (1.1) satisfying \( F(x^*) = 0 \).

We say that \( F \) is \( \mu \)-strongly monotone if we have

\[
\langle F(x) - F(x'), x - x' \rangle \geq \mu \| x - x' \|^2.
\]

We summarize our results in the following theorem.

**Theorem 4.7** Let \( k \geq 1 \) be an integer and let \( F \in \mathcal{G}_p^\theta(\mathbb{R}^d) \) satisfy that \((p - 1)\)th-order derivative is Lipschitz continuous and be \( \mu \)-strongly monotone. For the iterates \( \{ x_k \}_{k \geq 0} \) generated by Algorithm 3, we have

\[
\| x_{k+1} - x^* \| \leq \left( \frac{4p(2p+1)\kappa}{L \mu} \right)^{\frac{1}{p+1}} \| x_k - x^* \| \kappa, \quad \text{for all } k \geq 0,
\]

where \( \kappa = L/\mu > 0 \) and \( x^* \in \mathbb{R}^d \) is the solution of the ME. As a consequence, if \( p \geq 2 \) and the following condition holds true,

\[
\| x_0 - x^* \| \leq \frac{1}{2} \left( \frac{4p(2p+1)\kappa}{L \mu} \right)^{\frac{1}{p+1}},
\]

the iterates \( \{ x_k \}_{k \geq 0} \) will converge to \( x^* \in \mathbb{R}^d \) with a rate at least of order \( p \geq 2 \).

**Remark 4.8** The local convergence results in Theorem 4.7 are derived for the second-order and high-order methods (i.e., the case of \( p \geq 2 \)) and is posited as an advantage over first-order method if we hope to pursue high-accurate solutions. In this context, Jiang and Mokhtari [2022] provided a local convergence guarantee for their generalized optimistic gradient methods without counting the complexity bound of binary search procedure. However, the order of \( \frac{p+1}{2} \) in their results is worse since they allow the subproblem to be solved inexacty. Thus, our results complement those in Jiang and Mokhtari [2022]. It is also interesting to ask if we can improve the order of local convergence rate using some other easy-to-satisfy approximation conditions.

**Proof of Theorem 4.7:** We derive from Eq. (4.7) and Eq. (4.8) in the proof of Theorem 4.1 that one iteration of Algorithm 2 with an input \( x_0 \) satisfies that

\[
\| F(x_1) \| \leq \frac{(2p+1)L}{p!} (12 \| x_0 - x^* \|^2)^{\frac{p}{2}} \leq \frac{4p(2p+1)L}{p} \| x_0 - x^* \|^p.
\]

By abuse of notation, we let the iterates \( \{ x_k \}_{k \geq 0} \) be generated by Algorithm 3. Then, \( x_{k+1} \) is an output of Algorithm 2 with the input \((p, x_k, L, 1)\), i.e., one iteration of Algorithm 2 with an input \( x_k \). This implies that

\[
\| F(x_{k+1}) \| \leq \frac{4p(2p+1)L}{p} \| x_k - x^* \|^p.
\]

Since \( F \) is \( \mu \)-strongly monotone, we have

\[
\| x_{k+1} - x^* \|^2 \leq \frac{1}{\mu} \langle F(x_{k+1}) - F(x^*), x_{k+1} - x^* \rangle \leq \frac{1}{\mu} \langle F(x_{k+1}), x_{k+1} - x^* \rangle \leq \frac{1}{\mu} \| F(x_{k+1}) \| \| x_{k+1} - x^* \|.
\]

Putting these pieces together yields that

\[
\| x_{k+1} - x^* \| \leq \frac{4p(2p+1) L}{p} \frac{\| x_k - x^* \|^p}{\mu},
\]

24
which implies that

\[ \|x_{k+1} - x^*\| \leq \left( \frac{4^p(2p+1)\kappa}{p!} \right) \|x_k - x^*\|^p. \]  

(4.9)

For the case of \( p \geq 2 \), we have \( p - 1 \geq 1 \). If the following condition holds true,

\[ \|x_0 - x^*\| \leq \frac{1}{2} \left( \frac{p!}{4^p(2p+1)\kappa} \right)^{\frac{1}{p-1}}. \]

Then, we have

\[
\left( \frac{4^p(2p+1)\kappa}{p!} \right)^{\frac{1}{p-1}} \|x_{k+1} - x^*\| \leq \left( \frac{4^p(2p+1)\kappa}{p!} \right)^{\frac{p}{p-1}} \|x_k - x^*\|^p
\]

\[
= \left( \frac{4^p(2p+1)\kappa}{p!} \right)^{\frac{1}{p-1}} \|x_k - x^*\|^p \leq \left( \frac{4^p(2p+1)\kappa}{p!} \right)^{\frac{1}{p-1}} \|x_0 - x^*\| \leq \left( \frac{1}{2} \right)^{p+1}.
\]

This completes the proof.

5 Conclusions

We have presented a new class of rescaled gradient dynamical systems for deriving high-order methods for solving monotone equation (ME) problems. Our analysis of these systems is based on continuous-time and discrete-time Lyapunov functions which yield convergence characterizations for both the solution trajectories and their discretization. Our framework provides a systematic way to derive and analyze \( p \)th-order methods for solving ME problems.

There are several avenues for future research. In particular, it is of interest to bring our continuous-time perspective for understanding ME methods into register with the Lagrangian and Hamiltonian frameworks that have proved productive in recent works \[ \text{[Wibisono et al., 2016, Diakonikolas and Jordan, 2021, Muehlebach and Jordan, 2021, França et al., 2021]} \]. Moreover, it is worth noting that neither second-order methods nor high-order methods are fully characterized for MEs and lower bounds have only been established for convex optimization \[ \text{[Arjevani et al., 2019, Kornowski and Shamir, 2020, Nesterov, 2021a]} \]. This is in contrast to their first-order counterparts where lower bounds for solving monotone and strongly monotone MEs are known \[ \text{[Ouyang and Xu, 2021, Zhang et al., 2021]} \] and the optimal methods have been derived. ME problems generalize unconstrained saddle-point problems and lower bounds for the latter class can be applied to the former class. As such, it is promising to either improve the convergence rate of \( p \)th-order methods for solving the ME problems, or show that this is impossible by establishing a lower bound based on an appropriate oracle model.

Acknowledgments

This work was supported in part by the Mathematical Data Science program of the Office of Naval Research under grant number N00014-18-1-2764 and by the Vannevar Bush Faculty Fellowship program under grant number N00014-21-1-2941.
References

B. Abbas, H. Attouch, and B. F. Svaiter. Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces. *Journal of Optimization Theory and Applications*, 161(2):331–360, 2014. (Cited on pages 3, 4, 9, 12, 14, 17, and 18.)

D. Adil, B. Bullins, A. Jambulapati, and S. Sachdeva. Line search-free methods for higher-order smooth monotone variational inequalities. *ArXiv Preprint: 2205.06167*, 2022. (Cited on page 3.)

S. Adly and H. Attouch. Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping. *SIAM Journal on Optimization*, 30(3):2134–2162, 2020. (Cited on page 13.)

S. Adly and H. Attouch. First-order inertial algorithms involving dry friction damping. *Mathematical Programming*, pages 1–41, 2021. (Cited on page 13.)

S. Adly, H. Attouch, and V. N. Vo. Asymptotic behavior of Newton-like inertial dynamics involving the sum of potential and nonpotential terms. *Fixed Point Theory and Algorithms for Sciences and Engineering*, 2021(1):1–30, 2021. (Cited on pages 3 and 4.)

F. Alvarez. On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM Journal on Control and Optimization*, 38(4):1102–1119, 2000. (Cited on page 17.)

F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Analysis*, 9(1):3–11, 2001. (Cited on page 17.)

F. Alvarez and J. M. Pérez. A dynamical system associated with Newton’s method for parametric approximations of convex minimization problems. *Applied Mathematics and Optimization*, 38:193–217, 1998. (Cited on pages 4, 11, and 17.)

F. Alvarez, H. Attouch, J. Bolte, and P. Redont. A second-order gradient-like dissipative dynamical system with Hessian-driven damping: Application to optimization and mechanics. *Journal de mathématiques pures et appliquées*, 81(8):747–779, 2002. (Cited on pages 4, 11, and 17.)

M. M. Alves. Variants of the A-HPE and large-step A-HPE algorithms for strongly convex problems with applications to accelerated high-order tensor methods. *Optimization Methods and Software*, 0(0):1–31, 2022. (Cited on pages 4 and 18.)

L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: in Metric Spaces and in the Space of Probability Measures*. Springer Science & Business Media, 2005. (Cited on page 12.)

A. S. Antipin. Method of convex programming using a symmetric modification of Lagrange function. *Matekon*, 14(2):23–38, 1978. (Cited on page 2.)

A. S. Antipin. Minimization of convex functions on convex sets by means of differential equations. *Differential Equations*, 30(9):1365–1375, 1994. (Cited on page 17.)

Y. Arjevani, O. Shamir, and R. Shiff. Oracle complexity of second-order methods for smooth convex optimization. *Mathematical Programming*, 178(1):327–360, 2019. (Cited on page 25.)
F. M. Arscott and A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides: Control Systems*. Springer Netherlands, 1988. (Cited on page 13.)

H. Attouch and A. Cabot. Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. *Journal of Differential Equations*, 263(9):5412–5458, 2017. (Cited on page 4.)

H. Attouch and A. Cabot. Convergence of damped inertial dynamics governed by regularized maximally monotone operators. *Journal of Differential Equations*, 264(12):7138–7182, 2018. (Cited on page 12.)

H. Attouch and A. Cabot. Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. *Mathematical Programming*, 184(1):243–287, 2020. (Cited on pages 3, 12, and 14.)

H. Attouch and R. Cominetti. A dynamical approach to convex minimization coupling approximation with the steepest descent method. *Journal of Differential Equations*, 128(2):519–540, 1996. (Cited on page 17.)

H. Attouch and S. C. László. Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. *SIAM Journal on Optimization*, 30(4):3252–3283, 2020. (Cited on pages 3, 4, 9, 12, 14, and 17.)

H. Attouch and S. C. László. Continuous Newton-like inertial dynamics for monotone inclusions. *Set-Valued and Variational Analysis*, 29(3):555–581, 2021. (Cited on pages 3, 4, 9, 12, 14, and 17.)

H. Attouch and P-E. Maingé. Asymptotic behavior of second-order dissipative evolution equations combining potential with non-potential effects. *ESAIM: Control, Optimisation and Calculus of Variations*, 17(3):836–857, 2011. (Cited on page 12.)

H. Attouch and J. Peypouquet. The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than $1/k^2$. *SIAM Journal on Optimization*, 26(3):1824–1834, 2016. (Cited on page 4.)

H. Attouch and J. Peypouquet. Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. *Mathematical Programming*, 174(1):391–432, 2019. (Cited on pages 3, 12, and 14.)

H. Attouch and P. Redont. The second-order in time continuous Newton method. In *Approximation, optimization and mathematical economics*, pages 25–36. Springer, 2001. (Cited on pages 3, 4, 11, and 17.)

H. Attouch and B. F. Svaiter. A continuous dynamical Newton-like approach to solving monotone inclusions. *SIAM Journal on Control and Optimization*, 49(2):574–598, 2011. (Cited on pages 3, 4, 9, 11, 13, 17, and 18.)

H. Attouch, X. Goudou, and P. Redont. The heavy ball with friction method, I. the continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. *Communications in Contemporary Mathematics*, 2(01):1–34, 2000. (Cited on page 17.)

H. Attouch, P-E. Maingé, and P. Redont. A second-order differential system with Hessian-driven damping: Application to non-elastic shock laws. *Differential Equations and Applications*, 4(1):27–65, 2012. (Cited on pages 4 and 17.)
H. Attouch, P. Redont, and B. F. Svaiter. Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces. *Journal of Optimization Theory and Applications*, 157(3):624–650, 2013. (Cited on pages 3, 4, 9, 12, 13, 17, and 18.)

H. Attouch, M. M. Alves, and B. F. Svaiter. A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert spaces, with complexity $o(1/n^2)$. *Journal of Convex Analysis*, 23(1):139–180, 2016a. (Cited on pages 3, 4, 6, 9, 12, 13, 17, and 18.)

H. Attouch, J. Peypouquet, and P. Redont. Fast convex optimization via inertial dynamics with Hessian driven damping. *Journal of Differential Equations*, 261(10):5734–5783, 2016b. (Cited on pages 4 and 11.)

H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, 168(1-2):123–175, 2018. (Cited on page 4.)

H. Attouch, Z. Chbani, and H. Riahi. Fast convex optimization via time scaling of damped inertial gradient dynamics. *Pure and Applied Functional Analysis*, To appear, 2019a. (Cited on pages 4 and 13.)

H. Attouch, Z. Chbani, and H. Riahi. Rate of convergence of the nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:2, 2019b. (Cited on page 4.)

H. Attouch, Z. Chbani, and H. Riahi. Fast proximal methods via time scaling of damped inertial dynamics. *SIAM Journal on Optimization*, 29(3):2227–2256, 2019c. (Cited on page 13.)

H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. First-order optimization algorithms via inertial systems with Hessian driven damping. *Mathematical Programming*, pages 1–43, 2020. (Cited on page 4.)

H. Attouch, A. Balhag, Z. Chbani, and H. Riahi. Fast convex optimization via inertial dynamics combining viscous and Hessian-driven damping with time rescaling. *Evolution Equations & Control Theory*, To appear, 2021. (Cited on pages 4 and 13.)

J. B. Baillon. Un exemple concernant le comportement asymptotique de la solution du problème du \( dt + \partial \vartheta (\mu) \in 0 \). *Journal of Functional Analysis*, 28(3):369–376, 1978. (Cited on page 13.)

H. H. Bauschke and P. L. Combettes. A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. *Mathematics of Operations Research*, 26(2):248–264, 2001. (Cited on page 8.)

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. (Cited on page 4.)

S. Bellavia, G. Gurioli, B. Morini, and P. L. Toint. Adaptive regularization algorithms with inexact evaluations for nonconvex optimization. *SIAM Journal on Optimization*, 29(4):2881–2915, 2019. (Cited on page 16.)

E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos, and P. L. Toint. Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models. *Mathematical Programming*, 163(1):359–368, 2017. (Cited on page 16.)
R. I. Bot and E. R. Csetnek. Second order forward-backward dynamical systems for monotone inclusion problems. *SIAM Journal on Control and Optimization*, 54(3):1423–1443, 2016. (Cited on pages 9, 12, and 14.)

H. Brézis. *Opérateurs Maximaux Monotones: Et semi-groupes de contractions dans les espaces de Hilbert*. Elsevier, 1973. (Cited on page 13.)

H. Brézis. Asymptotic behavior of some evolution systems. In *Nonlinear Evolution Equations*, pages 141–154. Elsevier, 1978. (Cited on page 13.)

R. E. Bruck Jr. On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space. *Journal of Mathematical Analysis and Applications*, 61(1):159–164, 1977. (Cited on page 9.)

B. Bullins and K. A. Lai. Higher-order methods for convex-concave min-max optimization and monotone variational inequalities. *ArXiv Preprint: 2007.04528*, 2020. (Cited on pages 2, 14, 15, 18, and 19.)

Y. Carmon and J. Duchi. Gradient descent finds the cubic-regularized nonconvex Newton step. *SIAM Journal on Optimization*, 29(3):2146–2178, 2019. (Cited on page 16.)

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006. (Cited on page 2.)

A. Cordero and J. R. Torregrosa. Variants of Newton’s method using fifth-order quadrature formulas. *Applied Mathematics and Computation*, 190(1):686–698, 2007. (Cited on page 2.)

J. Cortés. Finite-time convergent gradient flows with applications to network consensus. *Automatica*, 42(11):1993–2000, 2006. (Cited on pages 6, 8, and 12.)

R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49(1):1–23, 1943. (Cited on page 12.)

E. R. Csetnek. Continuous dynamics related to monotone inclusions and non-smooth optimization problems. *Set-Valued and Variational Analysis*, 28(4):611–642, 2020. (Cited on pages 3 and 4.)

M. T. Darvishi and A. Barati. A third-order Newton-type method to solve systems of nonlinear equations. *Applied Mathematics and Computation*, 187(2):630–635, 2007. (Cited on page 2.)

C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng. Training GANs with optimism. In *ICLR*, 2018. URL https://openreview.net/forum?id=SJJySbbAZ. (Cited on page 2.)

J. E. Dennis Jr and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM, 1996. (Cited on page 1.)

J. Diakonikolas and M. I. Jordan. Generalized momentum-based methods: A Hamiltonian perspective. *SIAM Journal on Optimization*, 31(1):915–944, 2021. (Cited on page 25.)

J. Diakonikolas and L. Orecchia. The approximate duality gap technique: A unified theory of first-order methods. *SIAM Journal on Optimization*, 29(1):660–689, 2019. (Cited on page 4.)
P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1): 115–166, 2018. (Cited on page 2.)

Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Science & Business Media, 2007. (Cited on pages 1, 8, and 19.)

O. Fercoq and Z. Qu. Adaptive restart of accelerated gradient methods under local quadratic growth condition. *IMA Journal of Numerical Analysis*, 39(4):2069–2095, 2019. (Cited on page 17.)

A. F. Filippov. Differential equations with discontinuous right-hand side. *American Mathematical Society Translations*, 42:199–231, 1964. (Cited on page 13.)

G. França, J. Sulam, D. P. Robinson, and R. Vidal. Conformal symplectic and relativistic optimization. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(12):124008, 2020. (Cited on page 4.)

G. França, M. I. Jordan, and R. Vidal. On dissipative symplectic integration with applications to gradient-based optimization. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(4):043402, 2021. (Cited on page 25.)

M. Frontini and E. Sormani. Third-order methods from quadrature formulae for solving systems of nonlinear equations. *Applied Mathematics and Computation*, 149(3):771–782, 2004. (Cited on page 2.)

A. Gasnikov, P. Dvurechensky, E. Gorbunov, E. Vorontsova, D. Selikhanovych, C. A. Uribe, B. Jiang, H. Wang, S. Zhang, S. Bubeck, Q. Jiang, Y. T. Lee, Y. Li, and A. Sidford. Near optimal methods for minimizing convex functions with Lipschitz p-th derivatives. In *COLT*, pages 1392–1393. PMLR, 2019. (Cited on pages 4 and 18.)

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NeurIPS*, pages 2672–2680, 2014. (Cited on page 2.)

N. I. M Gould, S. Lucidi, M. Roma, and P. L. Toint. Solving the trust-region subproblem using the Lanczos method. *SIAM Journal on Optimization*, 9(2):504–525, 1999. (Cited on page 16.)

N. I. M. Gould, D. P. Robinson, and H. S. Thorne. On solving trust-region and other regularised subproblems in optimization. *Mathematical Programming Computation*, 2(1):21–57, 2010. (Cited on page 16.)

G. N. Grapiglia and Y. Nesterov. On inexact solution of auxiliary problems in tensor methods for convex optimization. *Optimization Methods and Software*, 36(1):145–170, 2021. (Cited on page 16.)

O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419, 1991. (Cited on page 8.)

J. Hadamard. *Mémoire sur le problème d’analyse relatif à l’équilibre des plaques élastiques encastrées*, volume 33. Imprimerie Nationale, 1908. (Cited on page 12.)

M. T. Heath. *Scientific Computing: An Introductory Survey, Revised Second Edition*. SIAM, 2018. (Cited on page 2.)
H. H. H. Homeier. A modified Newton method with cubic convergence: the multivariate case. *Journal of Computational and Applied Mathematics*, 169(1):161–169, 2004. (Cited on page 2.)

H. Hong and V. Stahl. Safe starting regions by fixed points and tightening. *Computing*, 53(3):323–335, 1994. (Cited on page 2.)

B. Jiang, T. Lin, and S. Zhang. A unified adaptive tensor approximation scheme to accelerate composite convex optimization. *SIAM Journal on Optimization*, 30(4):2897–2926, 2020. (Cited on page 16.)

R. Jiang and A. Mokhtari. Generalized optimistic methods for convex-concave saddle point problems. *ArXiv Preprint*: 2202.09674, 2022. (Cited on pages 2, 14, 15, 18, 19, and 24.)

C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. SIAM, 1995. (Cited on page 1.)

C. T. Kelley. *Solving nonlinear equations with Newton’s method*. SIAM, 2003. (Cited on page 2.)

G. Kornowski and O. Shamir. High-order oracle complexity of smooth and strongly convex optimization. *ArXiv Preprint*: 2010.06642, 2020. (Cited on page 25.)

G. M. Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976. (Cited on page 2.)

F. Labarre and P-E. Maingé. First-order frameworks for continuous Newton-like dynamics governed by maximally monotone operators. *Set-Valued and Variational Analysis*, pages 1–27, 2021. (Cited on pages 3 and 4.)

B. Lemaire. An asymptotical variational principle associated with the steepest descent method for a convex function. *Journal of Convex Analysis*, 3:63–70, 1996. (Cited on page 13.)

T. Lin and M. I. Jordan. A control-theoretic perspective on optimal high-order optimization. *Mathematical Programming*, pages 1–47, 2021a. (Cited on pages 4, 11, 12, and 18.)

T. Lin and M. I. Jordan. On monotone inclusions, acceleration and closed-loop control. *ArXiv Preprint*: 2111.08093, 2021b. (Cited on pages 2, 3, 4, 6, 8, 9, 12, 14, 15, 18, and 19.)

T. Lin and M. I. Jordan. Perseus: A simple high-order regularization method for variational inequalities. *ArXiv Preprint*: 2205.03202, 2022. (Cited on page 3.)

P. L. Lions. Une méthode itérative de résolution d’une inéquation variationnelle. *Israel Journal of Mathematics*, 31(2):204–208, 1978. (Cited on page 9.)

H. Lu, R. M. Freund, and Y. Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIAM Journal on Optimization*, 28(1):333–354, 2018. (Cited on page 15.)

P-E. Maingé. First-order continuous Newton-like systems for monotone inclusions. *SIAM Journal on Control and Optimization*, 51(2):1615–1638, 2013. (Cited on pages 3, 4, 12, and 17.)

B. Martinet. Régularisation d’inéquations variationnelles par approximations successives. rev. française informat. *Recherche Opérationnelle*, 4:154–158, 1970. (Cited on page 2.)

G. Mateos, J. A. Bazerque, and G. B. Giannakis. Distributed sparse linear regression. *IEEE Transactions on Signal Processing*, 58(10):5262–5276, 2010. (Cited on page 2.)
T. J. McDougall and S. J. Wotherspoon. A simple modification of Newton’s method to achieve convergence of order $1+\sqrt{2}$. *Applied Mathematics Letters*, 29:20–25, 2014. (Cited on page 2.)

K. Meintjes and A. P. Morgan. Chemical equilibrium systems as numerical test problems. *ACM Transactions on Mathematical Software (TOMS)*, 16(2):143–151, 1990. (Cited on page 2.)

A. Mokhtari, A. Ozdaglar, and S. Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In *AISTATS*, pages 1497–1507. PMLR, 2020a. (Cited on page 2.)

A. Mokhtari, A. E. Ozdaglar, and S. Pattathil. Convergence rate of $o(1/k)$ for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 30(4):3230–3251, 2020b. (Cited on page 2.)

R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20(6):2755–2787, 2010. (Cited on page 19.)

R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM Journal on Optimization*, 22(3):1092–1125, 2013. (Cited on pages 4 and 18.)

R. E. Moore. *Methods and Applications of Interval Analysis*. SIAM, 1979. (Cited on page 2.)

J-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société mathématique de France*, 93:273–299, 1965. (Cited on page 2.)

A. Morgan. *Solving polynomial systems using continuation for engineering and scientific problems*. SIAM, 2009. (Cited on page 2.)

A. Morgan and A. Sommese. Computing all solutions to polynomial systems using homotopy continuation. *Applied Mathematics and Computation*, 24(2):115–138, 1987. (Cited on page 2.)

O. Morgenstern and J. Von Neumann. *Theory of Games and Economic Behavior*. Princeton University Press, 1953. (Cited on page 1.)

M. Muehlebach and M. I. Jordan. A dynamical systems perspective on Nesterov acceleration. In *ICML*, pages 4656–4662, 2019. (Cited on page 4.)

M. Muehlebach and M. I. Jordan. Optimization with momentum: Dynamical, control-theoretic, and symplectic perspectives. *Journal of Machine Learning Research*, 22(73):1–50, 2021. (Cited on pages 4 and 25.)

D. E. Muller. A method for solving algebraic equations using an automatic computer. *Mathematical Tables and Other Aids to Computation*, 10(56):208–215, 1956. (Cited on page 1.)
A. 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. (Cited on pages 2 and 19.)

A. S. Nemirovski. Effective iterative methods for solving equations with monotone operators. *Ekonom. i Mat. Metody*, 17(2):344–359, 1981. (Cited on page 9.)

A. S. Nemirovski and D. B. Yudin. Cesari convergence of the gradient method of approximating saddle points of convex-concave functions. *Doklady Akademii Nauk*, 239(5):1056–1059, 1978. (Cited on page 9.)

A. S. Nemirovskii and Y. E. Nesterov. Optimal methods of smooth convex minimization. *USSR Computational Mathematics and Mathematical Physics*, 25(2):21–30, 1985. (Cited on page 16.)

Y. Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, 109(2):319–344, 2007. (Cited on pages 1, 2, 3, 8, 9, 12, 15, and 19.)

Y. Nesterov. Accelerating the cubic regularization of Newton’s method on convex problems. *Mathematical Programming*, 112(1):159–181, 2008. (Cited on pages 16 and 18.)

Y. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013. (Cited on pages 4, 16, and 17.)

Y. Nesterov. *Lectures on Convex Optimization*, volume 137. Springer, 2018. (Cited on pages 2 and 16.)

Y. Nesterov. Implementable tensor methods in unconstrained convex optimization. *Mathematical Programming*, 186(1):157–183, 2021a. (Cited on pages 1, 15, 18, and 25.)

Y. Nesterov. Inexact accelerated high-order proximal-point methods. *Mathematical Programming*, pages 1–26, 2021b. (Cited on page 1.)

Y. Nesterov and B. T. Polyak. Cubic regularization of Newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006. (Cited on pages 16 and 18.)

Y. E. Nesterov. A method for solving the convex programming problem with convergence rate $o(1/k^2)$. In *Dokl. akad. nauk Sssr*, volume 269, pages 543–547, 1983. (Cited on page 4.)

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2006. (Cited on page 1.)

B. O’Donoghue and C. J. Maddison. Hamiltonian descent for composite objectives. In *NeurIPS*, pages 14470–14480, 2019. (Cited on page 4.)

J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. SIAM, 2000. (Cited on page 1.)

M. J. Osborne. *An Introduction to Game Theory*. Oxford University Press, New York, 2004. (Cited on page 1.)

A. M. Ostiowsk. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1960. (Cited on page 1.)

Y. Ouyang and Y. Xu. Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. *Mathematical Programming*, 185(1):1–35, 2021. (Cited on pages 9 and 25.)

33
B. O’donoghue and E. Candes. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15(3):715–732, 2015. (Cited on page 17.)

B. Paden and S. Sastry. A calculus for computing Filippov’s differential inclusion with application to the variable structure control of robot manipulators. *IEEE Transactions on Circuits and Systems*, 34(1):73–82, 1987. (Cited on page 13.)

B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *Ussr Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964. (Cited on page 11.)

B. T. Polyak. *Introduction to Optimization*. Optimization Software, Inc, New York, 1987. (Cited on page 17.)

L. D. Popov. A modification of the Arrow-Hurwicz method for search of saddle points. *Mathematical notes of the Academy of Sciences of the USSR*, 28(5):845–848, 1980. (Cited on page 2.)

F. A. Potra. A superquadratic variant of Newton’s method. *SIAM Journal on Numerical Analysis*, 55(6):2863–2884, 2017. (Cited on page 2.)

J. Renegar and B. Grimmer. A simple nearly optimal restart scheme for speeding up first-order methods. *Foundations of Computational Mathematics*, 22(1):211–256, 2022. (Cited on page 17.)

R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976. (Cited on page 2.)

O. Romero and M. Benosman. Finite-time convergence in continuous-time optimization. In *ICML*, pages 8200–8209. PMLR, 2020. (Cited on pages 6, 8, and 13.)

V. Roulet and A. d’Aspremont. Sharpness, restart and acceleration. In *NeurIPS*, pages 1119–1129, 2017. (Cited on page 17.)

J. Schropp. Using dynamical systems methods to solve minimization problems. *Applied Numerical Mathematics*, 18(1-3):321–335, 1995. (Cited on page 12.)

J. Schropp and I. Singer. A dynamical systems approach to constrained minimization. *Numerical Functional Analysis and Optimization*, 21(3-4):537–551, 2000. (Cited on page 12.)

O. Sebbouh, C. Dossal, and A. Rondepierre. Convergence rates of damped inertial dynamics under geometric conditions and perturbations. *SIAM Journal on Optimization*, 30(3):1850–1877, 2020. (Cited on page 4.)

J. Shamma. *Cooperative Control of Distributed Multi-agent Systems*. John Wiley & Sons, 2008. (Cited on page 2.)

B. Shi, S. S. Du, M. I. Jordan, and W. J. Su. Understanding the acceleration phenomenon via high-resolution differential equations. *Mathematical Programming*, pages 1–70, 2021. (Cited on pages 4 and 11.)

C. Song, Y. Jiang, and Y. Ma. Unified acceleration of high-order algorithms under general hölder continuity. *SIAM Journal on Optimization*, 31(3):1797–1826, 2021. (Cited on page 18.)
W. Su, S. Boyd, and E. J. Candès. A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *The Journal of Machine Learning Research*, 17(1):5312–5354, 2016. (Cited on pages 4, 11, and 17.)

J. F. Traub. *Iterative Methods for the Solution of Equations*, volume 312. American Mathematical Soc., 1982. (Cited on page 1.)

A. Vassilis, A. Jean-François, and D. Charles. 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. (Cited on page 4.)

J. Verschelde, P. Verlinden, and R. Cools. Homotopies exploiting Newton polytopes for solving sparse polynomial systems. *SIAM Journal on Numerical Analysis*, 31(3):915–930, 1994. (Cited on page 2.)

A. Wibisono, A. C. Wilson, and M. I. Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47):E7351–E7358, 2016. (Cited on pages 4, 6, 13, 18, and 25.)

A. C. Wilson, L. Mackey, and A. Wibisono. Accelerating rescaled gradient descent: Fast optimization of smooth functions. In *NeurIPS*, pages 13555–13565, 2019. (Cited on pages 4 and 6.)

A. C. Wilson, B. Recht, and M. I. Jordan. A Lyapunov analysis of accelerated methods in optimization. *Journal of Machine Learning Research*, 22(113):1–34, 2021. (Cited on page 17.)

H. Xu, C. Caramanis, and S. Mannor. Robustness and regularization of support vector machines. *Journal of Machine Learning Research*, 10(Jul):1485–1510, 2009. (Cited on page 2.)

Z. Yao, P. Xu, F. Roosta, and M. W. Mahoney. Inexact nonconvex Newton-type methods. *INFORMS Journal on Optimization*, 3(2):154–182, 2021. (Cited on page 16.)

J. Zhang, A. Mokhtari, S. Sra, and A. Jadbabaie. Direct Runge-Kutta discretization achieves acceleration. In *NeurIPS*, pages 3900–3909, 2018. (Cited on page 4.)

J. Zhang, M. Hong, and S. Zhang. On lower iteration complexity bounds for the convex concave saddle point problems. *Mathematical Programming*, pages 1–35, 2021. (Cited on page 25.)