Discovery of Dynamics using Linear Multistep Methods

Rachael Keller* and Qiang Du*

Abstract.
Linear multistep methods (LMMs) are popular time discretization techniques for the numerical solution of differential equations. Traditionally they are applied to solve for the state given the dynamics (the forward problem), but here we consider their application for learning the dynamics given the state (the inverse problem). This repurposing of LMMs is largely motivated by growing interest in data-driven modeling of dynamics, but the behavior and analysis of LMMs for discovery turn out to be significantly different from the well-known, existing theory for the forward problem. Assuming the highly idealized setting of being given the exact state, we establish for the first time a rigorous framework based on refined notions of consistency and stability to yield convergence using LMMs for discovery. When applying these concepts to three popular $M-$step LMMs, the Adams-Bashforth, Adams-Moulton, and Backwards Differentiation Formula schemes, with $M \in \mathbb{N}$, the new theory suggests that Adams-Bashforth for $1 \leq M \leq 6$, Adams-Moulton for $M = 0$ and $M = 1$, and Backwards Differentiation Formula for all M are convergent, and, otherwise, the methods are not convergent in general. In addition, we provide numerical experiments to both motivate and substantiate our theoretical analysis.

Key words. discovery of dynamics, data-driven modeling, linear multistep methods, stability and convergence, root condition, learning dynamics, artificial intelligence

AMS subject classifications. 65L06, 65L09, 65L20, 65P99, 68T99

1. Introduction.

Understanding nature has been a human endeavor since early existence. Indeed, the ancient Greek Thales of Miletus correctly predicted a solar eclipse occurring on May 28th, 585 BC, a date the great Isaac Asimov suggests to be the birth of science [32, 21]. In this work, we focus on the *discovery* of dynamical systems with given states, where finitely many discrete measurements are used to approximately recover the unknown dynamical system – a *data-driven* discovery of dynamics [3, 29, 16]. A number of techniques for modeling nonlinear dynamical systems for artificial intelligence have arisen in recent years with breakthroughs in symbolic regression [31] and deep learning [15], studies in sparse regression and compressed sensing [3], and an outgrowing of new work in physics-informed neural networks [28, 25]. Meanwhile, advancements in the fields of machine learning [17, 13] and data science have witnessed renewed vigor into study of modeling of complex systems. Furthermore, the massive availability and accessibility of software packages to build neural networks in the past decade have enabled expansive new modeling and understanding of complex data [23, 3].

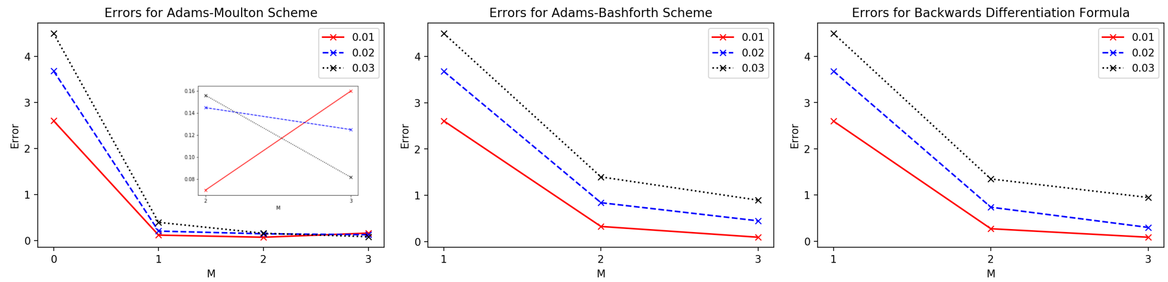
The mathematical modeling of dynamical systems formally dates back to the mid-1600s when Newton developed differential equations to study Kepler's law of planetary motion. In his work, Newton used analytical methods to calculate the motion of the earth around the sun, the famous two-body problem. Extending the method to solve even the three-body problem proved impossible, and in the late 1800s Poincaré focused instead on asymptotic and geometric features of celestial mechanics, from which the modern subjects of dynamical systems and chaos arose [34]. Since then, these fields have been broadened to gain insight into

*Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY R.K. rachael.keller@columbia.edu, Q.D. qd2125@columbia.edu)

problems ranging from general relativity to quantum mechanics.

The mathematical field of numerical analysis, on the other hand, focuses on the approximation of mathematical systems and in particular the discretization of continuum mechanics [22]. Dating back to the most fundamental technique, Euler's method, the subject has grown to incorporate the study of scientific computing, including round-off error and matrix inversion problems [10]. Many of the methods developed in numerical analysis combine physics with computing capabilities to yield systems for numerical simulations. Thus, the subject is a lens from which one can study the approximation of dynamical systems.

1.1. Motivation: Data-driven discovery of dynamical systems. Data-driven discovery of dynamical systems is experiencing a renaissance as costs of sensors, data storage, and computational resources has decreased [29]. As observational data on the state is increasingly available, and the underlying dynamic patterns of the data remain unknown, renewed vigor into modeling dynamics has naturally arisen. In comparison to celebrated scientific discovery in the human history, as an analogy, instead of Newton's learning the positions (states) of the earth and sun given Kepler's model of planetary motion (the dynamics), we seek to learn Kepler's law given the positions of the earth and sun; i.e., we solve the inverse problem. Recent works in data-driven discovery include [16, 26, 28, 27, 14, 25, 3, 29, 20, 38, 24, 30, 35, 36, 19, 11, 37]. In this work, we consider using linear multistep methods (LMMs) to discover the dynamics given the state at equidistant time steps and contribute to the fundamental theory of using LMMs for data-driven discovery.



(a) AM (b) AB (c) BDF

Figure 1: Absolute ℓ_2-errors for the 2D Cubic System (6.1) of the first coordinate on $t \in [0, 5]$ with varying time mesh $h = 0.01, 0.02, 0.03$, using a single 256-hidden layer neural network with a tanh activation function, as used in [26].

Coined "LMNet," LMMs are combined with neural networks for discovery of dynamics in [26, 38, 36]. Figure 1 shows the absolute errors associated with learning $\boldsymbol{f}$ for a nonlinearly-damped, 2D cubic oscillator (6.1) using neural networks with three representative schemes of LMMs – Adams-Moulton (AM), Adams-Bashforth (AB), and Backwards Differentiation Formula (BDF). These results are generated using the code repository built for [26]; reported are the errors of the dynamics rather than the integrated dynamics, which are shown in [26]. For solving differential equations with smooth solutions, increasing M corresponds to higher accuracy if the scheme is also stable. The AM scheme is an example of such a method; hence, the perplexing behavior in the errors of AM, shown in Figure 1 and observed in [26, 38, 36], warrants further investigation. It is the goal of this paper to investigate these findings and provide a theoretical explanation of the phenomena.

To do so, we introduce a systematic framework for the numerical analysis of discovery of dynamics using LMMs. Our new framework is rooted in the classical theory for LMMs and their applications to the numerical solution of differential equations, but it adopts new stability and convergence criteria due to the inverse nature of using time integrators for dynamics discovery. Consequently, it draws different conclusions regarding convergence in stark contrast to the conventional wisdom. The stability properties of particular schemes depart from the traditional numerical differential equation (forward problem) viewpoint, and some methods that are stable for the forward problem do not retain the property for the inverse problem. Our theory is able to explain the unusual phenomena as reported in Figure 1 and lays a rigorous foundation for elucidating the effect of neural networks on dynamics discovery via LMMs through follow-up studies. Therefore, this helps the scientific community broadly in our goal of making machine learning more transparent, explainable, stable and trustworthy.

1.2. Summary of Results. We present a framework in Section 3 consisting of nuanced notions of consistency and stability to handle unique challenges presented by using LMMs for discovery. These concepts are then combined to prove convergence. With this foundation, in Theorems 4.1 and 4.2 we outline consistency and stability properties of the Adams-Bashforth, Adams-Moulton, and Backwards Differentiation Formula schemes, and consequentially, Corollary 4.3, their convergence guarantees.

1.3. Outline. This paper is organized as follows. In Section 2 we briefly review LMMs and their theory for solving ordinary differential equations, including the standard notions in numerical analysis of truncation error, consistency, stability, and convergence, along with an algebraic root condition for stability. In Section 3 we frame the problem of discovery using LMMs and develop nuanced versions of consistency and stability for discovery. In particular, in Section 3.2, we discuss how truncation error for discovery is inherited from the forward problem and introduce a stronger notion of consistency; in Section 3.3 we refine the traditional definition of stability and the algebraic root condition, and we show equivalent theorems connecting the root conditions and the refined notions of stability. In Section 4, the discovery framework of Section 3 is applied to characterize convergence properties of the Adams-Bashforth, Adams-Moulton, and Backwards Differentiation Formula schemes. Some discussions on the long time dynamics discovery are made in section 5. Then, in Section 6, we show results of numerical experiments. Finally, in Section 7, we summarize the results and discuss future directions.

2. LMMs: Notation and a Quick Review.

In this section, we introduce notation used throughout this work and briefly review theory of LMMs as time integrators. While LMMs are well-documented in standard textbooks for solving ordinary differential equations (see [9, 22, 2, 12]), we include the salient points to facilitate direct comparison with the new theory for the discovery of unknown dynamics developed in the next section.

2.1. Notation. Consider the ordinary differential equation

$$\frac{d}{dt}\mathbf{x}(t) = \boldsymbol{f}(\mathbf{x}(t)),\ a \leq t \leq b,\ \mathbf{x}(t_0) = \mathbf{x}_0, \tag{2.1}$$

where $\mathbf{x} \in C^{\infty}(0,\infty)^d$. Discretizing the model problem (2.1), we assume a grid on the interval $[\,a,b\,]$ defined to be a set of points: $a = t_0 < t_1 < \cdots < t_N = b$ with equidistant mesh $t_{n+1} - t_n = h = (b-a)/N$, $n \in \{0,1,\ldots,N-1\}$. Let $[\,a,b\,]_h$ denote this ordered set, and define $\Gamma_h[\,a,b\,] = \{\mathbf{z}\,|\,\mathbf{z} \in \mathbb{R}^{(N+1\times d}, \ \mathbf{z}_n = \mathbf{z}(t_n) \in \mathbb{R}^d, t_n \in [\,a,b\,]_h\}$. $\Gamma_h[\,a,b\,]$ is called the set of grid functions on $[\,a,b\,]_h$ [9]. For $\mathbf{z} \in \Gamma_h[\,a,b\,]$, we let $\mathbf{z}_n = \mathbf{z}(t_n)$ for all $n = 0,1,\ldots,N$. Furthermore, we define the ℓ^p norm of the grid as $\|\mathbf{z}\|_p = \left(h\sum_{i=0}^N |\mathbf{z}_i|^p\right)^{1/p}$ for $p < \infty$ and the ℓ^∞ norm as $\|\mathbf{z}\| = \max_{0\leq n\leq N} |\mathbf{z}_i|$ [18].

Lastly, to deal with subtleties that arise from discovery, we find it helpful to introduce the notation $\|\cdot\|_W$ to denote the operator norm $\|\cdot\|_{\ell_W\to\ell^\infty}$, where $W \in \{1,\infty\}$. In later sections, we use these norms for a nuanced discussion of the interplay between consistency and stability to guarantee convergence for discovery.

2.2. LMMs: A Quick Review. A linear $M-$multistep method approximates the n^{th} value $\mathbf{x}_n = \mathbf{x}(t_n)$ in terms of the previous M time steps [9, 22, 2, 12]. An $M-$step linear multistep method is given by, $\alpha_0 \neq 0$,

$$\sum_{m=0}^{M} \alpha_m \mathbf{x}_{n-m} = h\sum_{m=0}^{M} \beta_m \boldsymbol{f}(\mathbf{x}_{n-m}),\ n = M, M+1, \ldots, N, \tag{2.2}$$

where $\mathbf{x} \in \Gamma_h[\,a,b\,]$ and the coefficients $\alpha_m, \beta_m \in \mathbb{R}$ for $m = 0,1,\ldots,M$. The function $\boldsymbol{f}$ is assumed to be given and Lipschitz, and the LMM scheme (2.2) defines an iterative procedure stepping forward in the independent variable $t \in [\,a,b\,]$ to solve for $\mathbf{x}(t)$ at the gridpoints.

For the numerical integration of differential equations, the method (2.2) is called explicit if $\beta_0 = 0$ and implicit otherwise [9, 22, 2]. Implicit methods require a nonlinear solver to the generated system of equations, whereas explicit methods do not. Existence and uniqueness of solutions in the case of implicit schemes is shown in [9, 12]. For both implicit and explicit methods, a kickstarting method for initial M values must be chosen, and as such a critical component of analyzing any multistep method scheme is to understand how much errors in initial values pollute the subsequent calculations [9]. This aspect of numerical methods is called numerical stability [2].

Remark 2.1. To fix ideas, we use the hat notation $\hat{}$ to mark grid functions generated by the discretization (2.2). In the forward problem, the state $\mathbf{x}(t)$ is iteratively produced by LMMs, and hence we study $\hat{\mathbf{x}}$, whereas for dynamics discovery, we study $\hat{\boldsymbol{f}}$, see Section 3.

2.3. The Adams Family and BDF. Adams-Bashforth (AB), Adams-Moulton (AM), and the Backwards Differentiation Formula (BDF) are three popular multistep method schemes that arise from a Lagrange interpolating polynomial of the state or dynamics at time t_n using M previous time steps. Without loss of generality, we consider the scalar model problem in this section; for higher dimensions, the theory need only be applied in each dimension. The Lagrange interpolating polynomial of a function $u : \mathbb{R} \to \mathbb{R}$ over the set $\Lambda = \{-M+1, -M+$

$2, \dots, -1, 0\}$ is the polynomial of degree $(M-1)$ obtained from the linear combination of basis functions

$$\ell_{k,n}(t;\Lambda) = \prod_{i \in \Lambda \setminus \{k\}} \frac{t - t_{n+i}}{t_{n+k} - t_{n+i}}, \; k \in \Lambda, \tag{2.3}$$

with $u(t_{n+k})$ as the coefficient of (2.3). Adams-Bashforth and Adams-Moulton, or the Adams family, arise from interpolating the dynamics $f(x(t_n)) \equiv f(t_n)$ by Lagrange interpolating polynomials on different sets $\Lambda_0 = \Lambda$ and $\Lambda_1 = \cup_{k \in \Lambda_0} k+1$, and applying the fundamental theorem of calculus on the model problem (2.1).

$$x(t_n) \approx x(t_{n-1}) + \int_{t_{n-1}}^{t_n} \sum_{k \in \tilde{\Lambda}} f(t_{n+k}) \ell_k(t; \tilde{\Lambda}) ds. \tag{2.4}$$

AB arises from using only the M previous time steps to approximate the current time step, i.e. $\tilde{\Lambda} = \Lambda_0$, whereas AM includes the current step as well, i.e. $\tilde{\Lambda} = \Lambda_1$.

BDF, on the other hand, is derived from interpolating the state $\{x(t_n)\}$ in (2.1) directly, but it shares the interpolating lattice Λ_1, so that

$$\sum_{k \in \Lambda_1} x(t_{n+k}) \frac{d\ell_{k,n}}{dt}(t_n; \Lambda_1) \approx \frac{d}{dt} x(t_n) = f(x(t_n)). \tag{2.5}$$

For Adams family and BDF methods, the coefficients in (2.2), are computed using (2.4) and (2.5) respectively. For solving the differential equation (2.1), Λ_1 corresponds to implicit methods, while Λ_0 determines the coefficients of explicit methods.

2.4. Truncation Error and Consistency. How accurately the discretization (2.2) approximates the solution to (2.1) is measured by the truncation error. We introduce the residual operator defined [9] for $n = M, M+1, \dots, N$ to be:

$$(R_h \hat{\mathbf{x}})_n := \frac{1}{h} \sum_{m=0}^{M} \alpha_m \hat{\mathbf{x}}_{n-m} - \sum_{m=0}^{M} \beta_m \boldsymbol{f}(\hat{\mathbf{x}}_{n-m}),$$

where $\hat{\mathbf{x}} \in \Gamma_h[\,a, b\,]$. The local truncation error is obtained by substituting the exact functions into the numerical scheme. Specifically, $(\boldsymbol{\tau}_h)_n = (R_h \mathbf{x})_n$ where $\mathbf{x} \in \Gamma_h[\,a, b\,]$ is the exact grid function satisfying the initial value problem.

Definition 2.2 (Local Truncation Error [22, 2, 12, 9]). Assuming $\mathbf{x}$ and $\boldsymbol{f}$ are smooth functions, the local truncation error of an $M-$step linear multistep method is given by

$$(\boldsymbol{\tau}_h)_n = \sum_{m=0}^{\infty} C_m h^{m-1} \nabla_{\mathbf{x}}^m \mathbf{x}(t_n), \quad \text{for } \; n = M, M+1, \dots, N, \tag{2.6}$$

where

$$C_0 = \sum_{k=0}^{M} \alpha_k, \; C_m = (-1)^m \left[\frac{1}{m!} \sum_{k=1}^{M} k^m \alpha_k + \frac{1}{(m-1)!} \sum_{k=0}^{M} k^{m-1} \beta_k \right], \; m = 1, 2, \dots.$$

Now, we proceed to define order of error and the notion of consistency.

Definition 2.3 (Order of Error [9]). A linear multistep method has error order of p if $\|\boldsymbol{\tau}_h\|_\infty = \mathcal{O}(h^p)$ as $h \to 0$ and admits a *principal error function* $\boldsymbol{e}(t) \in C[\,a,b\,]$ provided

$$\boldsymbol{e}(t) \neq \mathbf{0} \text{ and } (\boldsymbol{\tau}_h)_n = \boldsymbol{e}(t_n)h^p + \mathcal{O}(h^{p+1}) \text{ as } h \to 0,$$

or simply, $\|\boldsymbol{\tau}_h - h^p\boldsymbol{e}\|_\infty = \mathcal{O}(h^{p+1})$.

Definition 2.4 (Consistency [9]). A linear multistep method is consistent with the differential equation provided $\|\boldsymbol{\tau}_h\|_\infty \to 0$ as $h \to 0$.

The Adams family and BDF are consistent in the sense of Definition 2.4. Moreover, the local truncation error associated with the $M-$step AB scheme to be $\mathcal{O}(h^M)$, whereas for the $M-$step AM and BDF methods, the local truncation error is $\mathcal{O}(h^{M+1})$ [22, 2].

2.5. Stability and the Root Condition. In this section, we review definitions of stability and the root condition for LMMs. Stability is defined as follows.

Definition 2.5 (Stability [9]). A linear $M-$step method for the ordinary differential equation $\dot{\mathbf{x}} = \boldsymbol{f}(t, \mathbf{x}(t))$ is called stable on $[\,a,b\,]$ provided there exists a constant K not depending on h such that, for any two grid functions $\boldsymbol{u}, \boldsymbol{v} \in \Gamma_h[\,a,b\,]$, we have for all h sufficiently small

$$\|\boldsymbol{u} - \boldsymbol{v}\|_\infty \leq K \left(\max_{0\leq i\leq M-1} \|\boldsymbol{u}_i - \boldsymbol{v}_i\| + \|R_h\boldsymbol{u} - R_h\boldsymbol{v}\|_\infty \right).$$

Now we introduce the characteristic polynomials of a linear multistep method. For a linear $M-$step method, the first and second characteristic polynomials, respectively, are given by

$$\rho(z) = \sum_{m=0}^{M} \alpha_{M-m} z^m, \quad \text{and} \quad \sigma(z) = \sum_{m=0}^{M} \beta_{M-m} z^m, \tag{2.7}$$

where it is assumed that $\alpha_0 \neq 0$ [22]. These polynomials may be used to determine the stability of a linear multistep method via the root condition.

Definition 2.6 (Algebraic Root Condition [22, 9]). A polynomial satisfies the root condition provided the roots of the polynomial do not exceed magnitude 1, and those of magnitude 1 are simple.

The following theorem states the equivalence between the stability and the root condition.

Theorem 2.7 (Stability and the Root Condition [22, 9]). *A linear multistep method is stable if and only if its first characteristic polynomial $\rho(z)$ satisfies the algebraic root condition given by Definition 2.6.*

2.6. Convergence.

Definition 2.8 (Convergence [9]). Consider the initial value problem (2.1) and a fixed linear multistep method defined by (2.2). Let $\hat{\mathbf{x}} = \{\hat{\mathbf{x}}_n\} \in \Gamma_h[\,a,b\,]$ be the grid function obtained by applying (2.2) on a uniform, real-valued grid of $[\,a,b\,]$ with mesh size h, and let $\mathbf{x} = \{\mathbf{x}_n\} \in \Gamma_h[\,a,b\,]$ be the exact solution of (2.1) at the grid points. The linear multistep method is said to converge on $[\,a,b\,]$ if

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_\infty \to 0 \ \text{ as } \ h \to 0 \ \text{ whenever } \ \max_{0\leq k\leq M-1} \|\hat{\mathbf{x}}_k(h) - \mathbf{x}_0\|_\infty \to 0.$$

With Definition 2.8, one can obtain the Dahlquist Equivalence Theorem, Theorem 2.9 [22].

Theorem 2.9 (Equivalence Theorem [9]). *The multistep method (2.2) converges in the sense of Definition 2.8 for all Lipschitz $\boldsymbol{f}$ if and only if it is consistent and stable.*

In this work, we develop an analogous theory for multistep methods modifying these theorems to deal with the discovery of dynamics rather than solving the differential equation. In particular, we show how the second characteristic polynomial is determinant of stability for discovery and whether the Adams family and BDF are stable or not.

3. Discovery of Dynamics. In this study, we consider a *data-driven* technique to solve for the dynamics $\boldsymbol{f}$ given information on the state $\mathbf{x}$ at equidistant time steps [26]. First, we introduce the problem and then discuss notions of consistency, stability, and convergence. We now proceed to define the problem of LMMs for discovery.

3.1. Problem Definition. Following earlier discussions, we are concerned with the initial value problem (2.1). Our task is to use multistep methods to produce a function $\hat{\boldsymbol{f}} \in \Gamma_h[\,a, b\,]$ such that $\hat{\boldsymbol{f}}_n \approx \boldsymbol{f}(\mathbf{x}(t_n))$ using an $M-$step multistep method where $M \in \mathbb{N}$. Note that it is an idealized situation to assume that all of the exact states $\{\mathbf{x}(t_n)\}$, along with suitably approximated initial dynamics, are given. In practice, one often encounters situations with only partial (incomplete) data and data containing observation errors and uncertainties; these complications are typical for inverse problems. However, the idealized setting is a first step towards the understanding of the mathematical and computational issues, which are the focuses of the current work. The findings made here shed light on future studies of similar issues under more realistic conditions.

The procedure is defined as follows. Given $\mathbf{x}_s$ for $0 \le s \le N$ and $\hat{\boldsymbol{f}}_s$ as suitable approximations of $\boldsymbol{f}(\mathbf{x}_s)$ for $0 \le s \le M-1$,

$$\sum_{m=0}^{M} \beta_m \hat{\boldsymbol{f}}_{n-m} = \frac{1}{h} \sum_{m=0}^{M} \alpha_m \mathbf{x}_{n-m}, \quad n = M, M+1, \dots, N. \tag{3.1}$$

Indeed, (3.1) is simply (2.2) rewritten for learning the dynamics rather than the state. Recall that from Section 2 an LMM for solving the ODE is explicit if $\beta_0 = 0$ and implicit otherwise; this distinction arises from the unknown state $\mathbf{x}_n$ appearing in both sides of (2.2) with $\boldsymbol{f}$ possibly nonlinear. For discovery, the system (2.2) is linear in $\boldsymbol{f}$, and so in (3.1) we have no implicit or explicit distinction for any linear multistep method in the solver. However, choosing an implicit or explicit method changes the size of the system. In particular, for explicit methods $\beta_0 = 0$, so that (3.1) effectively begins at $n = M+1$ rather than at $n = M$. In either case, the procedures for discovery terminate at $n = N - M + 1$. We let N_M denote the number of grid points on $[\,a, b\,]$ for which a fixed $M-$step scheme generates coordinates of a grid function.

The equation (3.1) can be viewed as a linear recursive relation for determining the unknown $\{\hat{\boldsymbol{f}}_n\}$ as shown. It can also be formulated in a compact matrix-vector form:

$$B\hat{\boldsymbol{f}} = h^{-1} A\mathbf{x} - \hat{\boldsymbol{g}}, \tag{3.2}$$

where A is the $N_M \times N+1$ matrix of coefficients for $\{\alpha_m\}_{m=0}^{M}$ corresponding to $\mathbf{x}_{n-m}$ in (3.1), B is the $N_M \times N_M$ banded, lower-triangular matrix of $\{\beta_m\}_{m=0}^{M}$ corresponding to the unknown values $\hat{\boldsymbol{f}}_{n-m}$, and $\hat{\boldsymbol{g}} \in \mathbb{R}^{N_M}$ is a vector generated by some assumed, suitably approximated starting values given by

$$\hat{\boldsymbol{g}}_k = \begin{cases} \sum_{m=k+1}^{M} \beta_m \hat{\boldsymbol{f}}_{n-m}, & k = 0, 1, \ldots, M-1 \\ 0, & \text{otherwise}. \end{cases}$$

As presented in the next sections, the theory for discovering differential equations inherits some aspects of the theory for solving the differential equation as noted in Section 2 but also exhibits some differences due to the nature of the inverse problem.

3.2. Truncation Error and Consistency. LMMs for discovery inherit the truncation error of solving ordinary differential equations with LMMs. Indeed, truncation error is specific to the discretization of the continuous problem; therefore, the truncation error $\boldsymbol{\tau}_h$ of a scheme for dynamics discovery remains the same as that for solving an ordinary differential equation for the state. However, in addition to inheriting the same concept of consistency from Section 2, Definition 2.4, we also introduce some strengthened notions of consistency for dynamics discovery. We complement these concepts later on with refined notions of stability for a more nuanced discussion of convergence for discovery using LMMs. Consistency and its strengthened forms are defined below.

Definition 3.1 (Consistency for Dynamics Discovery). A linear multistep method is consistent with the differential equation for dynamics discovery provided $\|\boldsymbol{\tau}_h\|_\infty \to 0$ as $h \to 0$, and it is strongly consistent if $\|\boldsymbol{\tau}_h\|_1 \to 0$ as $h \to 0$. Furthermore, a method is consistent of degree k, for $k \geq 1$, provided $N^k \|\boldsymbol{\tau}_h\|_\infty \to 0$ as $h \to 0$.

Remark 3.2. With the Definition 3.1, all LMMs having at least $(k+1)-$order truncation error are consistent of degree at least k. Moreover, since

$$\|\boldsymbol{\tau}_h\|_1 = \sum_{n=M}^{N} |(\boldsymbol{\tau}_h)_n| \leq N \|\boldsymbol{\tau}_h\|_\infty ,$$

LMMs having at least 2nd order truncation are automatically strongly consistent. The notion of consistency of degree k is given here to make the discussion of this paper more general, though for the special LMMs considered later (AB, AM and BDF), we only need the notions of conventional consistency (i.e. $k = 0$) and strong consistency with $k = 1$.

3.3. Stability and the Root Condition for Discovery. In this section we develop stability in a similar spirit as in Section 2 but also introduce more refined notions of stability for convergence analysis. For discovery, the main distinction from theory for solving the forward problem is that now we consider perturbations to the recovered dynamics as opposed to the integrated states for the numerical solution of the differential equation. To begin we introduce a linear operator given by

$$(\hat{R}_h \hat{\boldsymbol{f}})_n := -\sum_{m=0}^{M} \beta_m \hat{\boldsymbol{f}}_{n-m}. \tag{3.3}$$

Definition 3.3 (Stability for Dynamics Discovery). A linear $M-$step method for the dynamics discovery is called stable on $[\,a,b\,]$ provided there exists a constant $K<\infty$, not depending on N, such that, for any two grid functions $\boldsymbol{u},\boldsymbol{v}\in\Gamma_h[\,a,b\,]$, we have

$$\|\boldsymbol{u}-\boldsymbol{v}\|_1 \leq K\left(\max_{0\leq i\leq M-1}\|\boldsymbol{u}_i-\boldsymbol{v}_i\|_\infty + \left\|\hat{R}_h(\boldsymbol{u}-\boldsymbol{v})\right\|_1\right).$$

Definition 3.4 (Strong Stability for Dynamics Discovery). A linear $M-$step method for the dynamics discovery is called strongly stable on $[\,a,b\,]$ provided there exists a constant $K<\infty$, not depending on N, such that, for any two grid functions $\boldsymbol{u},\boldsymbol{v}\in\Gamma_h[\,a,b\,]$, we have

$$\|\boldsymbol{u}-\boldsymbol{v}\|_\infty \leq K\left(\max_{0\leq i\leq M-1}\|\boldsymbol{u}_i-\boldsymbol{v}_i\|_\infty + \left\|\hat{R}_h(\boldsymbol{u}-\boldsymbol{v})\right\|_\infty\right).$$

Definition 3.5 (Stability of Degree k for Dynamics Discovery). A linear $M-$step method for the dynamics discovery is called marginally stable of degree k for $k>0$ on $[\,a,b\,]$ provided there exists a constant $K<\infty$, not depending on N, such that, for any two grid functions $\boldsymbol{u},\boldsymbol{v}\in\Gamma_h[\,a,b\,]$, we have

$$\|\boldsymbol{u}-\boldsymbol{v}\|_\infty \leq K\left(\max_{0\leq i\leq M-1}\|\boldsymbol{u}_i-\boldsymbol{v}_i\|_\infty + N^k\left\|\hat{R}_h(\boldsymbol{u}-\boldsymbol{v})\right\|_\infty\right).$$

Note that the stability notion in Definition 3.5 generalizes that in Definition 3.4 as the latter corresponds to stability of degree 0. Moreover, similar to the observation given in Remark 3.2, marginal stability of degree 1 follows from stability (Definition 3.3).

We would like to turn the property of stability into an algebraic condition as for the case of numerical solution to ODEs. For the forward problem, the algebraic root condition (Definition 2.6) serves this purpose; however, for the inverse problem, we require a more subtle treatment of the root condition to capture the nuances in stability for dynamics discovery.

Definition 3.6 (Strong Root Condition [1, 33, 7, 2]). A polynomial satisfies the strong root condition provided the roots of the polynomial have magnitude less than 1.

Likewise, we also generalize the above root conditions.

Definition 3.7 (k^{th}-multiplicity Root Condition). A polynomial satisfies the root condition of degree $k\in\mathbb{N}$ provided the roots of the polynomial do not exceed magnitude 1, and those of magnitude 1 have multiplicity no larger than k.

Remark 3.8. One may view the conventional (algebraic) root condition (Definition 2.6) and the strong root condition (Definition 3.7) as special cases of the k^{th}-multiplicity root condition of Definition 3.7 with $k=1$ and $k=0$, respectively. The strong root condition has been used in the numerical analysis, control theory, and linear recurrence relation literature for study of relative stability for LMM as time integrators and asymptotic properties associated with the linear recurrence relations [1, 33, 7, 2].

Naturally, we can see that the notions of stability for discovery for LMMs are tied to the bounds on the solutions to the linear recurrence equations determined by the coefficients $\{\beta_j\}$.

Lemma 3.9 (Stable $\iff$ Bounded). *A linear multistep method scheme for dynamics discovery is stable provided there exists some positive constant $K \in \mathbb{R}$, independent of N, such that $\|B^{-1}\|_1 \leq K$ and strongly stable provided $\|B^{-1}\|_\infty \leq K$. Furthermore, it is marginally stable of degree k if $\|B^{-1}\|_\infty \leq KN^k$.*

Proof. Fix $W \in \{1, \infty\}$. For brevity, we denote $\|\cdot\|_W$ as the norm $\ell_W \to \ell^\infty$. Suppose that $\hat{\boldsymbol{f}}, \tilde{\boldsymbol{f}} \in \Gamma_h[a, b]$, are generated by

$$\sum_{m=0}^{M} \beta_m \hat{\boldsymbol{f}}_{n-m} = -(\hat{R}_h[\hat{\boldsymbol{f}}])_n \quad \text{and} \quad \sum_{m=0}^{M} \beta_m \tilde{\boldsymbol{f}}_{n-m} = -(\hat{R}_h[\tilde{\boldsymbol{f}}])_n,$$

with different initial data. By subtracting the two, we have

$$\sum_{m=0}^{M} \beta_m (\hat{\boldsymbol{f}}_{n-m} - \tilde{\boldsymbol{f}}_{n-m}) = -(\hat{R}_h[\hat{\boldsymbol{f}} - \tilde{\boldsymbol{f}}])_n.$$

We form the matrix system

$$B(\hat{\boldsymbol{f}} - \tilde{\boldsymbol{f}}) = \hat{R}_h[\tilde{\boldsymbol{f}} - \hat{\boldsymbol{f}}] + \boldsymbol{w}_h, \tag{3.4}$$

where $\boldsymbol{w}_h$ encodes the differences in initial values

$$(\boldsymbol{w}_h)_k = \begin{cases} \sum_{m=k+1}^{M} \beta_m (\hat{\boldsymbol{f}}_{n-m} - \tilde{\boldsymbol{f}}_{n-m}), & k = 0, 1, \ldots, M-1 \\ 0, & \text{otherwise.} \end{cases}$$

Since B is invertible, (3.4) implies

$$\hat{\boldsymbol{f}} - \tilde{\boldsymbol{f}} = B^{-1}(\hat{R}_h[\tilde{\boldsymbol{f}} - \hat{\boldsymbol{f}}] + \boldsymbol{w}_h),$$

and hence

$$\left\|\hat{\boldsymbol{f}} - \tilde{\boldsymbol{f}}\right\|_\infty \leq \|B^{-1}\|_W \left(\left\|\hat{R}_h[\hat{\boldsymbol{f}} - \tilde{\boldsymbol{f}}]\right\|_W + C \max_{0 \leq s \leq M-1} \left\|\hat{\boldsymbol{f}}_s - \tilde{\boldsymbol{f}}_s\right\|_\infty \right),$$

where C is a constant independent of h for h sufficiently small. (In fact, we can take $C = 1$ if $W = \infty$ and $C = M$ if $W = 1$.) Therefore, the stability criteria are equivalent to the existence of some positive $K \in \mathbb{R}$ such that $\|B^{-1}\|_W \leq K$. In particular, if $W = 1$, we have stability in the sense of Definition 3.3, whereas if $W = \infty$, we have strong stability as defined in Definition 3.4. Finally, stability of degree k follows similarly. ■

We now relate the stability notions with the root conditions. Notice that while the stability in Theorem 2.7 for numerical integration of the given dynamics is concerned with the first characteristic polynomial $\rho(r)$, the stability in Theorem 3.10 for the discovery of dynamics is concerned with the second characteristic polynomial $\sigma(r)$. Hence, we see a fundamental difference in the two stability notions. This might be unexpected as it has not appeared in the numerical differential equation literature, but it is also not surprising given the inverse nature of using LMMs for dynamics discovery.

Theorem 3.10 (Stability for Discovery). *A linear multistep method for discovery of dynamics is stable provided the roots of the second characteristic polynomial $\sigma(r)$, defined by (2.7), satisfy the algebraic root condition (Definition 2.6). Likewise, a linear multistep method for discovery of dynamics is strongly stable provided the roots of the second characteristic polynomial $\sigma(r)$, defined by (2.7), satisfy the strong root condition (Definition 3.6). Furthermore, an LMM for discovery of dynamics is marginally stable of degree k provided the roots of the second characteristic polynomial satisfy the k^{th}-multiplicity root condition (Definition 3.7).*

Proof. We see from Lemma 3.9 that the various notions of stability are equivalent to the various bounds of B^{-1}, i.e., bounds on the solution of inhomogeneous linear recurrence relation, with $\{\beta_j\}$ be the coefficients, with respect to suitable norms of the initial data and the inhomogeneous right hand side.

By standard recurrence and linear algebra theory [1, 9], for $W = 1$, the solutions are bounded in $\|\cdot\|_\infty$ if and only if there exists some positive constant $K^* \in \mathbb{R}$ such that the companion matrix of the recurrence relation, denoted by $\mathcal{Z}$, satisfies

$$\max_{M\le n\le N} \|\mathcal{Z}^n\|_\infty \le K^* < \infty.$$

This bound is valid if and only if the root condition is satisfied. Meanwhile, for $W = \infty$, the solutions of the inhomogeneous linear recurrence relation are bounded in $\|\cdot\|_\infty$ if and only if we have the stronger condition that

$$\sum_{n=M}^{N} \|\mathcal{Z}^n\|_\infty \le K^* < \infty,$$

which is equivalent to the strong root condition. Likewise, we can argue that solutions have a polynomial growth of degree k if and only if

$$\max_{M\le n\le N} \|\mathcal{Z}^n\|_W \le K^* N^k < \infty,$$

which is equivalent to the root condition of degree k. ■

Remark 3.11. Again, as stated earlier, the notion of stability of degree k for $k > 1$ is not needed for the particular LMMs considered later, but it is included for generality.

3.4. Error Analysis and Convergence. In this section, we use the truncation error to learn the error for discovery. In particular, we prove the following theorem.

Theorem 3.12 (Error for Discovery). *Consider the ordinary differential equation (2.1) discretized by an $M-$step linear multistep method given by (3.1) where $M \in \mathbb{N}$. Let $\boldsymbol{f}, \hat{\boldsymbol{f}} \in \Gamma_h[a, b]$, where $\boldsymbol{f}$ is the exact grid function on the $N+1$ grid points and $\hat{\boldsymbol{f}}$ the approximation from the chosen scheme. Then,*

$$\left\|\boldsymbol{f} - \hat{\boldsymbol{f}}\right\| = \left\|B^{-1}\boldsymbol{\tau}_h\right\|, \tag{3.5}$$

where B is the matrix as defined in (3.2), $\|\cdot\|$ is any vector norm, and $\boldsymbol{\tau}_h$ is the local truncation error of the scheme.

Proof. Consider the discretized form of (2.1) for a given M-step linear multistep method. Denote $\boldsymbol{f} \in \Gamma_h[a,b]$ as the exact grid function on the $N+1$ grid points and $\hat{\boldsymbol{f}} \in \Gamma_h[a,b]$ as the approximation from the chosen scheme. Further, let $\boldsymbol{\tau}_h$ be the local truncation error associated with the $M-$step method and $\boldsymbol{\varepsilon}$ be the pointwise error of the approximation, i.e. $\hat{\boldsymbol{f}} = \boldsymbol{f} + \boldsymbol{\varepsilon}$. Then, subtracting the following two equations

$$\sum_{m=0}^{M} [\alpha_m \mathbf{x}_{n-m} + h\beta_m(\boldsymbol{f}_{n-m} + \boldsymbol{\varepsilon}_{n-m})] = \mathbf{0},$$

$$\sum_{m=0}^{M} [\alpha_m \mathbf{x}_{n-m} + h\beta_m \boldsymbol{f}_{n-m}] = \boldsymbol{\tau}_h.$$

we have $\sum_{m=0}^{M} \beta_m \boldsymbol{\varepsilon}_m = \boldsymbol{\tau}_h$. The matrix of non-zero coefficients acting on $\boldsymbol{\varepsilon}$ formed by this system of equations is the B matrix as defined in (3.2); it is a lower triangular, banded matrix with a non-zero, constant coefficient along the diagonal. It is full rank by construction, and thus invertible. Hence, $\|\boldsymbol{\varepsilon}\| = \left\|B^{-1}\boldsymbol{\tau}_h\right\|$, implying the result. ■

Theorem 2.9 for solving states that convergence requires both stability and consistency. The error (3.5) shows the interplay between solving the system, manifested in B^{-1}, and the truncation error, $\boldsymbol{\tau}_h$, from discretization of the differential equation. In Section 4.3, we show how the behavior of terms in B^{-1} as $h \to 0$ is connected to the second characteristic polynomial, and hence this component of the error is associated with stability. All three classes of LMMs considered in this study are consistent, which is a classical result, and hence $\boldsymbol{\tau}_h \to 0$ as $h \to 0$. Following (3.5) and the discussion of stability, we define convergence of a linear multistep method for discovery in the following way.

Definition 3.13 (Convergence for Discovery). Consider the initial value problem (2.1) and a fixed linear multistep method defined by (2.2). Let $\hat{\boldsymbol{f}} = \{\hat{\boldsymbol{f}}_n\} \in \Gamma_h[a,b]$ be the grid function obtained by applying (3.1) on a uniform, real-valued grid of $[a,b]$ with mesh size h, and let $\boldsymbol{f} = \{\boldsymbol{f}_n\} \in \Gamma_h[a,b]$ be the exact solution of (2.1) at the grid points. The linear multistep method for dynamics discovery is said to converge on $[a,b]$ if

$$\left\|\boldsymbol{f} - \hat{\boldsymbol{f}}\right\|_\infty \to 0 \text{ as } h \to 0 \text{ whenever } \max_{0 \le s \le M-1} \left\|\boldsymbol{f}_s - \hat{\boldsymbol{f}}_s\right\|_\infty = o(1) \to 0.$$

Finally, using the introduced notions of consistency and stability, we now present analogous convergence theorems for dynamics discovery.

Theorem 3.14 (Convergence Theorems for Discovery). *With the definitions of consistency, stability and convergence for the dynamics discovery problem defined above, if a linear multistep method is strongly consistent and stable, consistent and strongly stable, or marginally stable and consistent of degree k, then it is convergent.*

Proof. Let $\boldsymbol{f}, \hat{\boldsymbol{f}} \in \Gamma_h[0,T]$ where $\boldsymbol{f}$ is exact and $\hat{\boldsymbol{f}}$ is obtained from a linear multistep method that is stable in either the sense of Definition 3.3 or 3.4. Fix $W \in \{1, \infty\}$. For either case, there exists a constant $K_W < \infty$ independent of h, for h sufficiently small, such that

$$\begin{aligned}\left\|\boldsymbol{f} - \hat{\boldsymbol{f}}\right\|_W &\leq K_W \left(\max_{0 \leq i \leq M-1} \left\|\boldsymbol{f}_i - \hat{\boldsymbol{f}}_i\right\|_\infty + \left\|R_h \boldsymbol{f} - R_h \hat{\boldsymbol{f}}\right\|_W \right) \\ &\leq K_W \left(\quad o(1) \quad + \quad \|\boldsymbol{\tau}_h\|_W \quad \right)\end{aligned}$$

If the LMM is strongly stable and consistent, then the stability bound is with respect to $W = \infty$, and consistency implies $\left\|\boldsymbol{f} - \hat{\boldsymbol{f}}\right\|_\infty \to 0$ as $h \to 0$; hence, the method is convergent. If, on the other hand, the LMM is stable but not strongly stable ($W = 1$), only strong consistency guarantees convergence. As a final note, if the LMM is marginally stable and consistent of degree k, by canceling the factors of N^k and N^{-k} from stability and consistency, respectively, we also obtain convergence. Combining these three arguments, we have Theorem (3.14). ■

Remark 3.15. We can succinctly write Theorem (3.14) by stating that strong stability and consistency, or stability and strong consistency, implies convergence.

4. Application to AB, AM, and BDF.

We now apply the general theorem on LMM for the dynamics discovery to three popular special classes of methods– Adams-Bashforth (AB), Adams-Moulton (AM), and Backwards Differentiation Formula (BDF).

4.1. Consistency of AB, AM and BDF.

It is well-know that for methods like the $M-$step AB, AM and BDF, the truncation error is $\mathcal{O}(h^M)$ for explicit schemes, and $\mathcal{O}(h^{M+1})$ for implicit schemes. As a result, the three classes of LMM methods studies here remain consistent for dynamics discovery. Moreover, as a consequence of the truncation error for explicit and implicit schemes, the explicit methods are consistent of degree M, and the implicit methods are consistent of degree $M+1$, as noted in Remark 3.2. Indeed, the latter fact is crucial to the convergence of AM $M = 1$.

Theorem 4.1 (Consistency of AB, AM and BDF for Dynamics Discovery). *The linear multistep method schemes Adams-Bashforth, Adams-Moulton, and Backwards Differentiation Formula are all consistent for dynamics discovery. Furthermore, the Adams-Moulton scheme is consistent of degree* 1 *and thus strongly consistent for* $M = 1$.

4.2. Stability and Convergence of AB, AM and BDF.

Theorem 4.2. *With the notions of stability defined in Definitions 3.3 and 3.4,*
1. *BDF for all* $M \in \mathbb{N}$*, AB for* $1 \leq M \leq 6$*, and AM for* $M = 0$*, are strongly stable;*
2. *AM for* $M = 1$ *is stable and thus marginally stable of degree* 1*;*
3. *AB for* $7 \leq M \leq 10$ *and AM for* $M \geq 2$ *are unstable.*

The proof of Theorem 4.2 is given in Section 4.3.

Corollary 4.3. *BDF for all* $M \in \mathbb{N}$*, AB for* $1 \leq M \leq 6$*, and AM for* $M = 0$ *and* 1*, are convergent. On the other hand, AB for* $M > 6$ *and AM for* $M > 1$ *are not convergent.*

Proof. Corollary 4.3 follows immediately from the application of Theorems 4.1 and 4.2 in the spirit of the Dahlquist Equivalence Theorem. ■

Remark 4.4. The finite range of instability with respect to the order M for the AB scheme is due to limitation of explicit calculations. We conjecture that the scheme is unstable for all $M \geq 7$. Interestingly, that $M = 6$ is a threshold for stability of the polynomial echoes the stability criterion for the forward problem BDF [12], for which $M = 6$ is also the largest known order method that is stable. Explicit numerical calculation or Routh Arrays (see [7]) are used to show this fact [12, 5, 8]. Schur polynomials have since been used [4] to show a generalized stability argument for $M \geq 13$ [8]. We leave open a generalized stability result for $M \geq 7$ using the polynomial roots, but we have validated numerically the instability for $7 \leq M \leq 20$.

4.3. Verification of root conditions for AB, AM and BDF. We now verify, for the three classes of LMMs, the root condition holds for cases stated in Theorem 4.2.

We begin by calculating the roots of the second characteristic polynomial associated with AB and AM. The results imply specific of (in)stability for AB and AM in some range of M. For brevity, we show results for only $1 \leq M \leq 10$ with both schemes being unstable for $7 \leq M \leq 10$. We have also numerically validated instability for $11 \leq M \leq 20$ and expect instability to persist for all $M \geq 7$. For AM, a general instability result is given in Lemma 4.7; however, we rely on numerical calculation for the AB method.

Fix $M \in \mathbb{N}$ and $\Lambda \in \{\Lambda_0, \Lambda_1\}$, where we recall from Section 2,

$$\Lambda_0 = \{-M, \dots, 0\} \text{ and } \Lambda_1 = \{-M+1, \dots, 1\}.$$

Exchanging the integral and the summand in the formula for the Lagrange interpolating polynomial, one can observe that finding the roots of the second characteristic polynomial is equivalent to choosing $r \in \mathbb{C}$ satisfying a mean-zero equation. In particular,

$$\sum_{x\in\Lambda} \int_0^1 \ell_x(u) r^x du \iff \int_0^1 \sum_{x\in\Lambda} \ell_x(u) r^x du = 0.$$

Table 1: Largest Magnitude Roots

M	1	2	3	4	5
AB	–	0.3333	0.4663	0.6338	0.8075
AM	1.0000	1.7165	2.3658	2.9775	3.5639
M	6	7	8	9	10
AB	0.9829	1.1587	1.3345	1.5100	1.6852
AM	4.1312	4.6851	5.2267	5.7586	6.2820

As we see in the Table 1, which is computed symbolically by Mathematica, the profile of the roots of the characteristic polynomial associated with the different lattices varies significantly. The data in Table 1 immediately establishes the following lemma.

Lemma 4.5. *Fix $\Lambda \in \{\Lambda_0, \Lambda_1\}$, and let $\ell_x(u;\Lambda)$ be the Lagrange interpolating polynomial of order M with $M \in \mathbb{N}$ associated with node $x \in \Lambda$. Then, we can characterize the M roots*

$r \in \mathbb{C}$ of the equation

$$\int_0^1 \sum_{x\in\Lambda} \ell_x(u;\Lambda) r^x du = 0, \tag{4.1}$$

as follows.

1. *If $\Lambda = \Lambda_0$,*
   (a) *$1 \le M \le 6$, then $|r| < 1$.*
   (b) *$7 \le M \le 10$, there exists at least one root r with $|r| > 1$.*
2. *If $\Lambda = \Lambda_1$,*
   (a) *$M = 1$, then the single root satisfies $|r| = 1$.*
   (b) *$2 \le M \le 10$, there exists at least one root r with $|r| > 1$.*

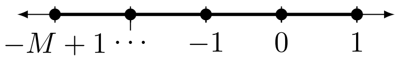


Figure 2: Interpolation Points of Λ_1 for (AM)

To further investigate stability properties of AM, we utilize the following properties of its second characteristic polynomial $\sigma(r)$ and the corresponding coefficients of its B matrix.

Lemma 4.6. *For $M \ge 2$, the coefficients $\{\beta_m\}_{m=0}^M$ of the AM method have the properties:*

1. *$\beta_1 > \beta_0 > 0$, and*
2. *$\operatorname{sign}(\beta_{m+1}) = -\operatorname{sign}(\beta_m),\ 1 \le m \le M$.*

Proof. Fix $M \in \mathbb{N}$ with $M \ge 2$. The $M-$step AM scheme has coefficients

$$\beta_m = \frac{(-1)^m}{m!(M-m)!} \int_0^1 \prod_{\substack{i=0\\ i\neq m}}^{M} (u+i-1)\ du, \tag{4.2}$$

for $m = 0, 1, \ldots, M$. The coefficients β_0 and β_1 are given by

$$\beta_0 = \frac{1}{M!} \int_0^1 \prod_{i=0}^{M-1} (u+i) du \quad \text{and} \quad \beta_1 = \frac{1}{(M-1)!} \int_0^1 (1-u) \prod_{i=1}^{M-1} (u+i) du.$$

Obviously $\beta_0 > 0$. Notice

$$\beta_1 > \beta_0 \iff \frac{M}{M+1} \int_0^1 \prod_{i=1}^{M-1} (u+i) du > \int_0^1 \prod_{i=0}^{M-1} (u+i) du. \tag{4.3}$$

We prove (4.3) by induction. As the base case, $M = 2$. For $M = 2$, we have $\beta_1 = 8/12 > \beta_0 = 5/12$. Now assume (4.3) holds up to some arbitrary $M \in \mathbb{N}$, with $M > 2$. We will show the result for $M+1$.

$$\frac{M+1}{M+2}\int_0^1 \prod_{i=1}^{M}(u+i)du = \frac{M+1}{M+2}\int_0^1 \left(u\prod_{i=1}^{M-1}(u+i) + M\prod_{i=1}^{M-1}(u+i)du\right) \tag{4.4}$$

$$\overset{(4.3)}{>}\frac{M+1}{M+2}\left(\int_0^1 \prod_{i=0}^{M-1}(u+i) + \frac{M(M+1)}{M}\prod_{i=0}^{M-1}(u+i)du\right) \tag{4.5}$$

$$= \frac{(M+1)(M+2)}{(M+2)}\int_0^1 \prod_{i=0}^{M-1}(u+i)du \tag{4.6}$$

$$> (M+1)\int_0^1 \frac{u+M}{M+1}\prod_{i=0}^{M-1}(u+i)du = \int_0^1 \prod_{i=0}^{M}(u+i)du, \tag{4.7}$$

as desired. Note we used the inductive hypothesis on the second term in (4.5). The proof by induction showing for $M \geq 2$, $\beta_1 > \beta_0$ is complete.
To prove Part 2, note that the relation of signs between coefficients follows from the sign of the Lagrange basis polynomials in the integrand of the coefficients. For $m \in \{2, 3, \dots, M\}$, the integrand of (4.2) are of the same sign, and therefore the sign of β_m depends only on the multiplier $(-1)^m$. Hence Part 2 of Lemma 4.6 follows. ■

Lemma 4.7 (General Instability of AM $M \geq 2$). *The linear multistep method formed by the Adams-Moulton scheme for $M \geq 2$ does not satisfy the root condition.*

Proof. Fix $M \geq 2$ and consider the second characteristic polynomial associated with the Adams-Moulton scheme. We write it as $\sigma(r) = \sum \beta_m r^{M-m}$. From Lemma 4.6, $\beta_1/\beta_0 > 1$. Moreover, by construction of the AM method, $(-1)^m\beta_m < 0$ for $m \geq 2$.

For $r > 0$ sufficiently large.

$$(-1)^M\sigma(-r) = (-1)^{2M}r^M\left[\beta_0 - \beta_1/r + \sum_{m=2}^{M}(-1)^m\beta_m r^{-m}\right]. \tag{4.8}$$

Taking the limit as $r \to +\infty$, we see that $(-1)^M\sigma(-\infty) = \infty$ since $\beta_0 > 0$. Meanwhile,

$$(-1)^M\sigma(-\beta_1/\beta_0) = \sum_{m\geq 2}(-1)^{-m}\beta_m(\beta_1/\beta_0)^m < 0.$$

Hence, it follows from the Intermediate Value Theorem that there is at least one root of $\sigma(r)$ that is real in $(-\infty, -\beta_1/\beta_0) \subset (-\infty, -1)$, violating the root condition. The result thus follows. ■

Theorem 4.8 (Root Condition of AB, AM, BDF). *The strong root condition for discovery is satisfied by BDF for all $M \in \mathbb{N}$, AB scheme for $1 \leq M \leq 6$, and AM for $M = 0$. The algebraic root condition, or the k^{th} root condition with $k = 1$, is satisfied for AM with $M = 1$. On the other hand, the root condition is not satisfied for the AB scheme with $7 \leq M \leq 10$ or the AM scheme with $M \geq 2$.*

Proof. The lattices of Lemma 4.5 correspond to the interpolation lattices used to construct the coefficients of the AB and AM methods' second characteristic polynomials; in particular, Λ_0 is associated with the AB scheme and Λ_1 with AM. Lemma 4.5 therefore implies the results of Theorem 4.8 for AB with $1 \leq M \leq 10$ and for AM with $1 \leq M \leq 10$. Furthermore, by Lemma 4.7, the AM scheme violates the root conditon and hence is generally unstable for $M \geq 2$. Finally, the $M-$step BDF method has second characteristic polynomial $\sigma(r) = r^{M-1}$, for all $M \in \mathbb{N}$, which has roots that are 0. Hence, the root condition is satisfied for the BDF scheme for arbitrary $M \in \mathbb{N}$. As a result, AM $M = 0$, identical to BDF 1, is satisfies the root conditon as well. ∎

Finally, Theorem 4.2 follows directly from Theorems 4.8 and 3.10.

5. Long Time Dynamics Discovery. In this section, we consider the problem of discovering dynamics of (2.1) over a variable interval $(0, T)$, with terminal time $1 \ll T \to \infty$, and a fixed mesh h. Notice by increasing T we increase the number of grid points $N = T/h$; hence we hope to relate our previous studies with variable mesh and fixed domain to this setting. For the numerical analysis of time integration, this study is reminiscent to that of asymptotic stability, which is often treated via the study of linear dynamics [9, 22, 2].

By rescaling time, $\tilde{t} = t/T$, where $0 \leq \tilde{t} \leq 1$, and defining $\tilde{\mathbf{x}}(\tilde{t}) = \mathbf{x}(t)$, we have via change of variables that the scaled dynamics $\tilde{\boldsymbol{f}}$ may be related to that of the original variables by

$$\frac{d}{d\tilde{t}}\tilde{\mathbf{x}}(\tilde{t}) = T\frac{d}{dt}\mathbf{x}(t) = T\boldsymbol{f}(\mathbf{x}(t)) = T\boldsymbol{f}(\tilde{\mathbf{x}}(\tilde{t})).$$

Then, if we define $\tilde{\boldsymbol{f}}(\tilde{\mathbf{x}}(\tilde{t})) = T\boldsymbol{f}(\tilde{\mathbf{x}}(\tilde{t})) = T\boldsymbol{f}(\mathbf{x}(t))$, the rescaled differential equation becomes

$$\frac{d}{d\tilde{t}}\tilde{\mathbf{x}}(\tilde{t}) = \tilde{\boldsymbol{f}}(\tilde{\mathbf{x}}(t)), \quad 0 \leq \tilde{t} \leq 1, \quad \tilde{\mathbf{x}}(0) = \mathbf{x}(0) = \mathbf{x}_0. \tag{5.1}$$

Now, consider applying the LMM scheme to $\tilde{\mathbf{x}}$ using the transformed model problem (5.1) with a step size $\tilde{h} = 1/N$. Under this rescaling of time, one can check directly that the leading truncation error term, the consistency error, of an LMM of order p (see Definitions 2.2 and 2.3) is

$$C_{p+1}\tilde{h}^p\frac{d^{p+1}}{d\tilde{t}^{p+1}}\tilde{\mathbf{x}}(\tilde{t}) = C_{p+1}\tilde{h}^pT^{p+1}\frac{d^{p+1}}{dt^{p+1}}\mathbf{x}(t) = C_{p+1}Th^p\frac{d^{p+1}}{dt^{p+1}}\mathbf{x}(t). \tag{5.2}$$

In light of (5.2), we can see that the truncation error of the discovered dynamics of (2.1) in the original time scale is a multiple of the truncation error of the rescaled model (5.1) by the factor T^{-1}. Meanwhile, from the analysis of Section 3.3, the error from stability is only directly dependent on $\sigma(r)$ and N, not the specific time domain.

Using these observations of the effects on consistency and stability, we can deduce the behavior of an LMM in the long-time regime. For a strongly stable $p^{th}-$order LMM, the global error behaves like $O\left(T^{-1}Th^p\right) = O(h^p)$ provided that $\max_{t\in(0,T)} |\mathbf{x}^{(p+1)}(t)|$ remains uniformly bounded as T increases. Hence, we may view strongly stable LMMs as A-stable, in the case of dynamics discovery, for fixed h as $T \to \infty$. This can be seen as another difference with the case of the forward problem of time integration, where the order of A-stable LMMs is known to be limited by 2 due to the celebrated Dahquist barrier theorems [6, 9, 22, 2]. On the other hand, for unstable methods, the exponential growth in N of the inverse matrix B^{-1} dominates over any gain in accuracy from consistency. Thus, lack of stability leads to an exponentially increasing error as T grows linearly.

As a peculiar example, the marginally stable AM-1 (AM with $M = 1$) is not asymptotically stable for dynamics discovery in the long-time regime. Recall that that AM-1 is stable of degree $k = 1$ for a fixed time interval (see Definition 3.5) while also having error of order $p = 2$ for time integration (Definition 2.3). For discovered dynamics, the global error in the long-time regime behaves like $O(NT^{-1}Th^2) = O(Th)$, where the additional factor of N arises from the stability estimate of B^{-1}. Thus, we expect AM-1 to have a linearly increasing error in T for a fixed h, which is supported by numerical experiments presented in the next section.

To recap, from the analysis in this section, for dynamics discovery, AB and BDF enjoy asymptotic stability for a fixed time step size h as T increases, while AM is not for $M \geq 1$. As shown in Figure 4, the errors from AB and BDF remain fixed across various values of T, while the AM methods yield exponential growth of error in T for $M \geq 2$ and a linear growth of error in T for $M = 1$.

6. Numerical Experiments. In this section, we discuss the matrix systems for each of the studied multistep methods specifically and show numerical evidence consistent with the theoretical findings. We limit ourselves to the idealized setting of numerically exact states considered for the theoretical analysis and to low dimensional dynamic systems for the sake of illustration and benchmarking.

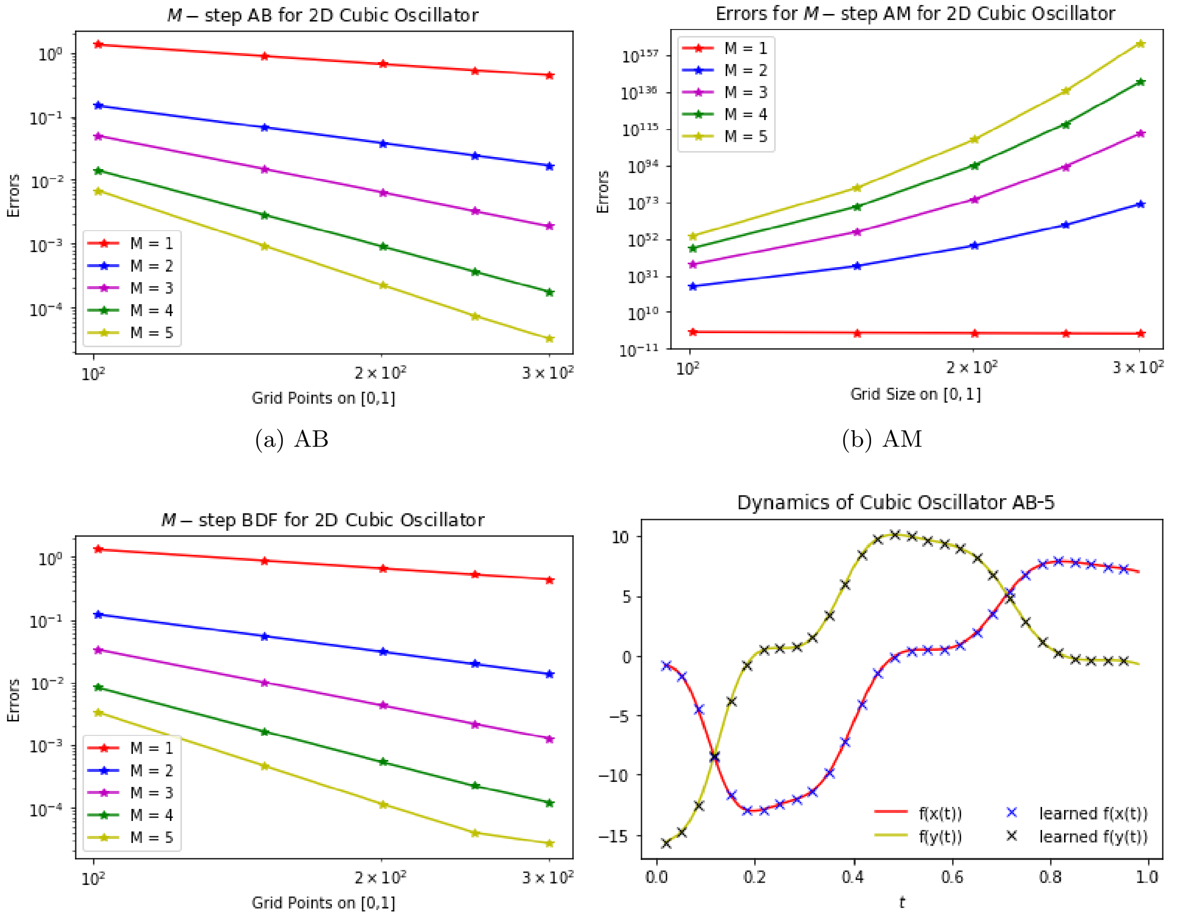
6.1. Fixed Time Domain. First we study the methods on a fixed time domain, $t \in [0, 1]$, with varying time step. For a model problem, we consider the 2D Cubic System, a nonlinearly damped oscillator, specified as in [26, 3].

$$\begin{cases} \dot{x}_1 = -0.1\ x_1^3 + 2.0\ x_2^3, \\ \dot{x}_2 = -2.0\ x_1^3 - 0.1\ x_2^3, \\ [x_1, x_2] = [2, 0]. \end{cases} \tag{6.1}$$

We show in Figure 3 the results from the Adams family and BDF methods.

The exact dynamics is computed by numerically integrating (6.1) on a very refined mesh. The errors of the discovered dynamics in the $\ell^\infty-$norm are shown in Figure 3 for different M against different number of grid points. In addition, Figure 3d shows a slice of the approximated dynamics captured over the interval versus the true dynamics using a stable and convergent method (AB-5). Clearly, the numerical results support the theoretical findings of this paper.

6.2. Long Time Behavior. In this section, we consider the problem of discovering dynamics over a changing domain $[0, T], T \gg 1.$ with fixed mesh size h. In Figure 4, we discover the dynamics of the 2D Cubic System over the specified ranges of T ($T = 10, 20, 30, 40$). For AM (Figure 4a) we use $h = 0.1$, the smallest order mesh that avoids numerically singular matrices for the specified range of T due to the instability of the methods, while for AB and BDF (Figures 4b-4c) we use $h = 0.01$. AM clearly suffers from the exponential error growth for the AM-M with $M \geq 2$ and has a linear growth for $M = 1$, just as predicted in Section



(a) AB

(b) AM

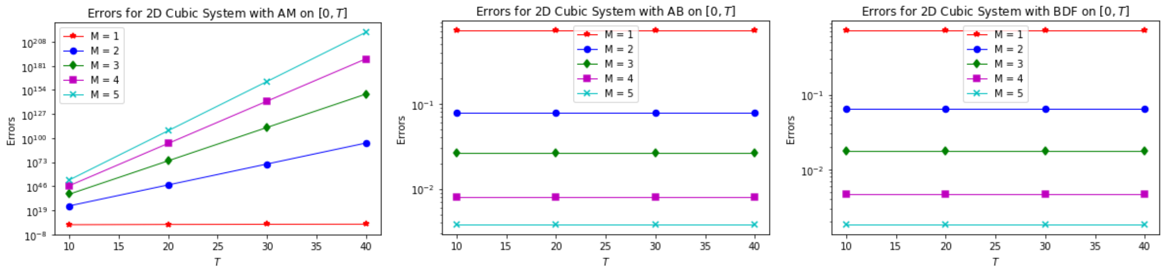
(c) BDF

(d) Captured Dynamics with AB

Figure 3: Numerical results of the three types of schemes on the 2D cubic system (6.1) on the unit time interval for different choices of M and N.

5. Meanwhile, also consistent with the analysis of Secton 5, AB and BDF are robust for the long-time dynamics discovery – yielding a constant error for fixed mesh as T increases and a decreasing error for larger M.

7. Conclusions and Future Steps. In this paper, we extended the foundational work of solving ordinary differential equations using LMMs for the discovery of dynamics. We introduced refined notions of consistency, and stability, and convergence for discovery based on classical definitions, and we showed how three prominent schemes – Adams-Bashforth, Adams-Moulton, and Backwards Differentiation Formula – may or may not be convergent numerical methods for discovery in general. To do so, we derived an explicit construction for the inverse matrix formed for discovery of the dynamics, and with that in hand, we deduced which schemes and choice of M are stable and unstable. Lastly, we presented numerical results,



(a) AM with $h = 0.1$ (b) AB with $h = 0.01$ (c) BDF with $h = 0.01$

Figure 4: Long Time Errors for Discovery of 2D Cubic System (6.1)

which show agreement with the theoretical findings. In conclusion, we found theoretically and numerically that the systems for BDF for all $M \in \mathbb{N}$, AB for $1 \leq M \leq 6$, and AM for $M = 0$ and 1 are and convergent, while AB for $7 \leq M \leq 10$ and AM for $M \geq 2$ are not.

The current study is assuming the best possible case that the exact states along with suitable approximations to the initial dynamics are all given. While this setting is highly idealized, based on the studies made here, we can speculate about the impact of incomplete data on the properties of stability and convergence caused by different choices of time discretization schemes for a more informed attempt at discovery of unknown dynamics. Furthermore, we also only considered the special class of time integrators given by LMMs and limited our investigation to be without regularization. We are now working in various directions to extend the results to inexact and noisy data, for example, and some explorations are underway on the study of

1. the effects of regularization. In the idealized setting, we can study effects of regularization for discovery, which can come from constraints on the dynamics either explicitly enforced or through a penalty formulation. In particular, we hope to study the effect of regularizing the learned functions via neural networks (LMNet) [26, 38, 24, 36] and compressed representation by promoting sparsity [3, 16]; we may also explore physics-informed neural networks [28] as another form of regularization to achieve physics-informed and data-driven discovery of the dynamics.
2. more general class of time-integrators and extending the stability framework to incorporate other multistep schemes such as predictor-corrector and multistage methods Runge-Kutta [30];
3. reduced-order models for the state or dynamics such as presented in [38];
4. the errors in numerically *integrated* states based on learned dynamics [26];
5. distributed dynamic systems such as time-dependent PDEs and examine the additional effect due to spatial discretization;
6. extending the above tasks to the study of dynamics discovery problem with incomplete and uncertain data.

To conclude, we see from this study that there are many new challenges in physics-based and data-driven modeling and simulations warranting further numerical analysis research.

8. Acknowledgments. The authors would like to thank the CM3 group at Columbia University for invigorating discussions, including Wen Ding for his helpful insight using Mathematica. We also thank the NSF for supporting this work under the NSF Graduate Research Fellowship NSF DGE-1644869, NSF-DMS-1719699, and CCF-1704833. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] R. P. Agarwal, *Difference equations and inequalities: theory, methods, and applications*, CRC Press, 2000.

[2] K. Atkinson, W. Han, and D. E. Stewart, *Numerical solution of ordinary differential equations*, vol. 108, John Wiley & Sons, 2011.

[3] S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proceedings of the National Academy of Sciences, 113 (2016), pp. 3932–3937.

[4] D. M. Creedon and J. J. Miller, *The stability properties ofq-step backward difference schemes*, BIT Numerical Mathematics, 15 (1975), pp. 244–249.

[5] C. W. Cryer, *On the instability of high order backward-difference multistep methods*, BIT Numerical Mathematics, 12 (1972), pp. 17–25.

[6] G. G. Dahlquist, *A special stability problem for linear multistep methods*, BIT Numerical Mathematics, 3 (1963), pp. 27–43.

[7] R. C. Dorf and R. H. Bishop, *Modern control systems*, Pearson, 2011.

[8] C. Fredebeul, *A-bdf: a generalization of the backward differentiation formulae*, SIAM journal on numerical analysis, 35 (1998), pp. 1917–1938.

[9] W. Gautschi, *Numerical analysis*, Springer Science & Business Media, 1997.

[10] H. H. Goldstine, *A History of Numerical Analysis from the 16th through the 19th Century*, vol. 2, Springer Science & Business Media, 2012.

[11] N. S. Gulgec, Z. Shi, N. Deshmukh, S. Pakzad, and M. Takáč, *Fd-net with auxiliary time steps: Fast prediction of pdes using hessian-free trust-region methods*, arXiv preprint arXiv:1910.12680, (2019).

[12] P. Henrici, *Discrete variable methods in ordinary differential equations*, (1962).

[13] M. I. Jordan and T. M. Mitchell, *Machine learning: Trends, perspectives, and prospects*, Science, 349 (2015), pp. 255–260.

[14] S. H. Kang, W. Liao, and Y. Liu, *Ident: Identifying differential equations with numerical time evolution*, arXiv preprint arXiv:1904.03538, (2019).

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet classification with deep convolutional neural networks*, in Advances in neural information processing systems, 2012, pp. 1097–1105.

[16] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic mode decomposition: data-driven modeling of complex systems*, SIAM, 2016.

[17] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, nature, 521 (2015), p. 436.

[18] R. J. LeVeque, *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*, vol. 98, Siam, 2007.

[19] Z. Long, Y. Lu, and B. Dong, *Pde-net 2.0: Learning pdes from data with a numeric-symbolic hybrid deep network*, Journal of Computational Physics, 399 (2019), p. 108925, https://doi.org/https://doi.org/10.1016/j.jcp.2019.108925, http://www.sciencedirect.com/science/article/pii/S0021999119306308.

[20] F. Lu, M. Zhong, S. Tang, and M. Maggioni, *Nonparametric inference of interaction laws in systems of agents from trajectory data*, Proceedings of the National Academy of Sciences, 116 (2019), pp. 14424–14433.

[21] T. Mandel, *Happy birthday to science*, Chicago Sun-Times, https://web.archive.org/web/20140611123359/http://www.highbeam.com/doc/1P2-4000661.html. [Online; archived by HighBeam Research.].

[22] D. Mayers and E. Süli, *An introduction to numerical analysis*, Cambridge university press Cambridge, 2003.
[23] C. Metz, *You don't have to be google to build an artificial brain*, https://www.wired.com/2014/09/google-artificial-brain/. [Online. Date Accessed October 28, 2019.].
[24] T. Qin, K. Wu, and D. Xiu, *Data driven governing equations approximation using deep neural networks*, Journal of Computational Physics, 395 (2019), pp. 620–635.
[25] M. Raissi, *Deep hidden physics models: Deep learning of nonlinear partial differential equations*, The Journal of Machine Learning Research, 19 (2018), pp. 932–955.
[26] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Multistep neural networks for data-driven discovery of nonlinear dynamical systems*, arXiv preprint arXiv:1801.01236, (2018).
[27] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Numerical gaussian processes for time-dependent and nonlinear partial differential equations*, SIAM Journal on Scientific Computing, 40 (2018), pp. A172–A198, https://doi.org/10.1137/17M1120762, https://doi.org/10.1137/17M1120762, https://arxiv.org/abs/https://doi.org/10.1137/17M1120762.
[28] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, Journal of Computational Physics, 378 (2019), pp. 686–707.
[29] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Data-driven discovery of partial differential equations*, Science Advances, 3 (2017), p. e1602614.
[30] S. H. Rudy, J. N. Kutz, and S. L. Brunton, *Deep learning of dynamics and signal-noise decomposition with time-stepping constraints*, Journal of Computational Physics, 396 (2019), pp. 483–506.
[31] M. Schmidt and H. Lipson, *Distilling free-form natural laws from experimental data*, science, 324 (2009), pp. 81–85.
[32] F. R. Stephenson and L. J. Fatoohi, *Thales's prediction of a solar eclipse*, Journal for the History of Astronomy, 28 (1997), pp. 279–282.
[33] S. Strelitz, *On the routh-hurwitz problem*, The American Mathematical Monthly, 84 (1977), pp. 542–544.
[34] S. H. Strogatz, *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*, CRC Press, 2018.
[35] Y. Sun, L. Zhang, and H. Schaeffer, *Neupde: Neural network based ordinary and partial differential equations for modeling time-dependent data*, arXiv preprint arXiv:1908.03190, (2019).
[36] R. Tipireddy, P. Perdikaris, P. Stinis, and A. Tartakovsky, *A comparative study of physics-informed neural network models for learning unknown dynamics and constitutive relations*, 2019, https://arxiv.org/abs/1904.04058.
[37] K. Wu and D. Xiu, *Data-driven deep learning of partial differential equations in modal space*, arXiv preprint arXiv:1910.06948, (2019).
[38] X. Xie, G. Zhang, and C. G. Webster, *Non-intrusive inference reduced order modeling of fluid dynamics using linear multistep network*, arXiv preprint arXiv:1809.07820, (2018).