Guaranteed Stable Quadratic Models and their applications in SINDy and Operator Inference

Pawan Goyal, Igor Pontes Duff, and Peter Benner

Abstract: Scientific machine learning for learning dynamical systems is a powerful tool that combines data-driven modeling models, physics-based modeling, and empirical knowledge. It plays an essential role in an engineering design cycle and digital twinning. In this work, we primarily focus on an *operator inference* methodology that builds dynamical models, preferably in low-dimension, with a prior hypothesis on the model structure, often determined by known physics or given by experts. Then, for inference, we aim to learn the operators of a model by setting up an appropriate optimization problem. One of the critical properties of dynamical systems is *stability*. However, such a property is not guaranteed by the inferred models. In this work, we propose inference formulations to learn quadratic models, which are stable by design. Precisely, we discuss the parameterization of quadratic systems that are locally and globally stable. Moreover, for quadratic systems with no stable point yet bounded (e.g., Chaotic Lorenz model), we discuss an attractive trapping region philosophy and a parameterization of such systems. Using those parameterizations, we set up inference problems, which are then solved using a gradient-based optimization method. Furthermore, to avoid numerical derivatives and still learn continuous systems, we make use of an integration form of differential equations. We present several numerical examples, illustrating the preservation of stability and discussing its comparison with the existing state-of-the-art approach to infer operators. By means of numerical examples, we also demonstrate how proposed methods are employed to discover governing equations and energy-preserving models.

Keywords: Scientific machine learning, operator inference, sparse-regression, quadratic dynamical systems, stability, Lyapunov function, Runge-Kutta scheme, energy-preserving systems.

Novelty statement:

- Inference of stability-guaranteed quadratic systems.
- Utilize locally and globally stable quadratic system parameterizations.
- Discuss an attractive trapping region for quadratic systems and their parametrization.
- Make use of an integral form for the inference of continuous systems so that we do not require estimating the derivative information numerically.
- Several numerical examples are presented, demonstrating the stability guarantee of the inferred models.

1. Introduction

In recent years, the field of learning dynamical systems from data has gained significant popularity due to the availability of vast amounts of data and its diverse range of applications, including robotics, self-driving cars, epidemiology, neuroscience, and climate modeling. Constructing models for such applications using first principles is often impractical. Leaning dynamical models using data allows us to make predictions, optimize system parameters, and system control using feedback loops. One of the most common approaches to learning dynamical systems from data is based on the Koopman operator theory, which aims at representing the underlying dynamics of a system as a linear operator acting on a high-dimensional space of observables [1], which may even of infinite dimensional. This approach is particularly useful for studying systems with complex and nonlinear dynamics using the tools from the linear theory. The identification procedure is done using the *Dynamic Mode Decomposition* (DMD), which has been initially developed in [2], and several extensions have been proposed, such as extended DMD [3], DMD with control input operator [4,5], kernel DMD [6], higher-order DMD [7], and DMD for quadratic-bilinear control systems [8]. Furthermore, In the context of nonlinear systems, the so-called Sparse Identification of Nonlinear Dynamics (SINDy) has gained prominence [9]. Its underlying principle involves selecting a few nonlinear terms from a vast library of potential candidate nonlinear basis functions to describe the underlying system's dynamics. This selection is made via sparse regression approaches using the available time-domain data. Additionally, in the context of *system identification*, various methodologies have been proposed to identify systems using input-output data in the time domain, including the eigensystem realization algorithm [10] and subspace methods [11,12].

The integration of prior knowledge into learning frameworks has become increasingly important for creating accurate and interpretable surrogate models in various applications. This is particularly relevant when we have prior knowledge, such as model hypotheses or a form of governing equations but lack information about system parameters and discretization schemes and only have access to time-domain measurements or simulated data. In such scenarios, the operator inference technique has become popular as a valuable tool for learning operators. For high-dimensional data, one can focus on learning dynamics in a low-dimensional subspace, thus yielding reduced-order models. Such a problem was initially studied in [13], aiming at identifying nonlinear polynomial systems, which has been extended to general nonlinear systems [14, 15]. Moreover, the methodology has been tailored to capture second-order or mechanical behavior [16, 17], adapted to incompressible flow problems [18], and parametric systems [19, 20].

Physical phenomena often exhibit remarkable stability, with their state variables remaining well-behaved and bounded over long periods. Thus, accurately modeling these processes requires a set of stable differential equations, which are crucial for numerical computations. Despite its importance, stability is often overlooked in frameworks for learning dynamical systems. This paper focuses on learning quadratic differential equations, which guarantee stability. Quadratic models naturally appear in discretized fluid mechanical models. Moreover, smooth nonlinear dynamical systems can be recast as quadratic dynamical systems by means of a lifting transformation [14, 21–23].

In this work, we focus on imposing three distinct stabilities on the learned quadratic models. The first focuses on local (asymptotic) stability around an equilibrium point. The necessary condition for systems to be locally asymptotic stable requires its Jacobian at the equilibrium to be a stable matrix. With this spirit, the authors in [24] discussed an inference problem and set up an optimization problem with matrix inequality constraints, which can be computationally expensive. Furthermore, it is done where the linear matrix (or Jacobian at the equilibrium point) is symmetric and negative definite. But, for many stable systems, the Jacobian doesn't need to be symmetric. To overcome these limitations, in this work, we adopt the stable matrix parametrization from [25] to the inference problem, allowing us to learn quadratic models that are guaranteed to be locally stable by design. Notably, such a parametrization has been previously used in [26] to learn linear dynamical systems.

Next, we examine global (asymptotic) stability as the second type of stability. To investigate global stability, we assume the quadratic nonlinearity is (generalized) energy-preserving [27]. It is worth noting that many dynamical systems in fluid dynamics have energy-preserving nonlinearities, as discussed in [27]. These nonlinearities are commonly seen in discretized fluid mechanical models with a wide range of boundary conditions, such as those presented in [28–31], as well as in magneto-hydrodynamics applications, as reported in [32–35]. By means of a Lyapunov function, we derive the conditions for quadratic systems with energy-preserving nonlinearities to be globally stable. Consequently, we parametrize a family of energy-preserving quadratic systems that are globally stable. Leveraging this, we then introduce an inference problem to guarantee the learned models are globally stable by construction.

Besides local asymptotic stability or global stability of a fixed equilibrium point, many nonlinear dynamical systems exhibit attractor behavior, e.g., the chaotic Lorenz example. In this case, the trajectories are bounded and converge (asymptotically) to a stable region despite not having a stable equilibrium point. Therefore, this work considers the third type of stability, involving the existence of (asymptotic) attractive trapping regions. Energy-preserving quadratic systems possessing attractive trapping regions have been initially studied in [27]. Building upon this work, the authors in [36] proposed an extension of the SINDy algorithm that promotes boundness for quadratic systems. Therein, the proposed algorithm includes energy-preserving nonlinearities as soft constraints and projects the solution at each iteration onto the space of quadratic bounded systems. In contrast, in this work, we propose a parametrization of quadratic systems possessing trapping regions. This parametrization is then leveraged in the inference problem, ensuring by design that the learned models are always energy-preserving, bounded, and possess trapping regions at each iteration without the need for projection.

The remaining paper is structured as follows. In Section 2, we briefly recall the operator inference approach [13] for learning nonlinear systems with quadratic nonlinearities from data. Then, in Section 3, we describe local asymptotic stability and introduce a parametrization of locally asymptotically stable quadratic systems. Leveraging this, we propose an inference problem so that the learned models are guaranteed to be locally asymptotically stable. Thereafter, global asymptotic stability is studied in Section 4 for quadratic systems possessing energy-preserving nonlinearity. Then, a novel parametrization of globally asymptotic stable systems is proposed, guaranteeing that learned models are globally stable. In Section 5, we study energy-preserving quadratic systems possessing bounded trajectories with attractive trapping regions and their parametrization. It is worth mentioning that all parameterizations proposed assumed that the underlying quadratic dynamical system possesses a quadratic-like Lyapunov function. Then, in Section 6, an integration scheme is reviewed so that incorporating them in learning models can avoid the requirement of computing the derivative information from data. Finally, Section 7 presents several numerical experiments illustrating the guaranteed stabilities of the learned models. By means of two examples, we also illustrate how the proposed parameterization can be used to discover governing equations in the context of SINDy that are stable and can be energy-preserving. In Section 8, we conclude the paper with a short summary and future avenues.

2. Model Inference with Quadratic Non-linearity

In the following, we provide a brief overview of the standard operator inference (`OpInf`) approach [13] by restricting ourselves to quadratic nonlinearities. This approach also closely resembles `SINDy` approach [9] when the dictionary contains only linear and quadratic terms. We begin by discussing learning models using high-dimensional data. However, the low-dimensional case is rather straightforward than the high-dimensional case, where the projection step—which we shall see shortly—is not required.

In the following, we formulate the inference problem to obtain low-dimensional dynamical models using high-dimensional data. To that end, we consider a high-fidelity dynamical system with N-degrees of freedom, and its state vector is denoted by $\mathbf{y}(t) \in \mathbb{R}^N$. Furthermore, we assume that at time steps $\{t_0, t_1, \ldots, t_{\mathcal{N}}\}$, we have access to the snapshots $\mathbf{y}(t_k)$, for $k = 1, \ldots, \mathcal{N}$. Next, let us collect these snapshots in a matrix as follows:

$$\mathbf{Y} = \left[\mathbf{y}(t_0), \ldots, \mathbf{y}(t_{\mathcal{N}})\right] \in \mathbb{R}^{N \times \mathcal{N}}. \tag{1}$$

Though the state $\mathbf{y}(t)$ is N-dimensional, so is the dynamics, it is often possible to approximate it using a lower-dimensional subspace accurately. By doing so, we can significantly simplify the inference problem. To achieve this, we first identify a low-dimensional representation of $\mathbf{y}(t)$ by computing the dominant bases $\mathbf{V} \in \mathbb{R}^{N \times n}$, which can be obtained using the SVD of the matrix $\mathbf{Y}$, followed by taking its n-most dominant left singular vectors. This allows us to compute the reduced state trajectory as follows:

$$\mathbf{X} = \mathbf{V}^\top \mathbf{Y},$$

where $\mathbf{X} := \left[\mathbf{x}(t_0), \ldots, \mathbf{x}(t_{\mathcal{N}})\right]$ with $\mathbf{x}(t_i) = \mathbf{V}^\top \mathbf{y}(t_i)$. With our quadratic model hypothesis, we then aim to learn its operators. Precisely, our goal is to learn a quadratic model of the following form:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{H}\left(\mathbf{x}(t) \otimes \mathbf{x}(t)\right) + \mathbf{B}, \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{H} \in \mathbb{R}^{n \times n^2}$, and $\mathbf{B} \in \mathbb{R}^{n \times 1}$ are the operators. The quadratic systems (2) is an important class of nonlinear systems, as a large class of smooth nonlinear systems can be written in this form, see [14, 21–23, 37–40].

Next, we cast the inference problem, which is as follows. Having the low-dimensional trajectories $\{\mathbf{x}(t_0), \ldots, \mathbf{x}(t_{\mathcal{N}})\}$, we aim to learn operators $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{B}$ in (2). At the moment, let us also assume to have the derivative information of $\mathbf{x}$ at time $\{t_0, \ldots, t_{\mathcal{N}}\}$, which are denoted by $\dot{\mathbf{x}}(t_0), \ldots, \dot{\mathbf{x}}(t_{\mathcal{N}})$. Using this derivative information, we form the following matrix:

$$\dot{\mathbf{X}} = \left[\dot{\mathbf{x}}(t_0), \ldots, \dot{\mathbf{x}}(t_{\mathcal{N}})\right]. \tag{3}$$

Then, in a naive case, determining operators boils down to solving a least-squares optimization problem, which can be given as

$$\min_{\mathbf{A},\mathbf{H},\mathbf{B}} \left\|\dot{\mathbf{X}} - \left[\mathbf{A},\ \mathbf{H},\ \mathbf{B}\right] \mathcal{D}\right\|_F, \tag{4}$$

where $\mathcal{D} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X} \tilde{\otimes} \mathbf{X} \\ \mathbf{1} \end{bmatrix}$ with $\mathbf{1}$ being a row-vector of ones, and the product $\tilde{\otimes}$ is defined as $\mathbf{G} \tilde{\otimes} \mathbf{G} = [\mathbf{g}_1 \otimes \mathbf{g}_1, \dots, \mathbf{g}_\mathcal{N} \otimes \mathbf{g}_\mathcal{N}]$ with $\mathbf{g}_i$ being i-th column of the matrix $\mathbf{G} \in \mathbb{R}^{n \times \mathcal{N}}$. Note that $\mathbf{g} \otimes \mathbf{g} = \left[\mathbf{g}_1^2, \mathbf{g}_1\mathbf{g}_2, \dots, \mathbf{g}_1\mathbf{g}_n, \dots, \mathbf{g}_n^2\right]^\top \in \mathbb{R}^{n^2}$, where $\mathbf{g} = [\mathbf{g}_1, \dots, \mathbf{g}_n]^\top$.

Although the optimization problem (4) appears straightforward, it has a couple of major drawbacks. Firstly, it requires knowledge of the state derivative. When the collected data are finely sampled and noise-free, one can estimate it using numerical methods. However, in practical scenarios, data can be scarce and noisy, thus making the computation of the derivative information challenging. As a remedy to this, one can cast the problem (4) in an integral form and infer operators. Several works have discussed such an approach, see, e.g., [41–44]. This approach enables robust learning of operators under conditions of limited and noisy data. We discuss it in more detail later in Section 6. In addition, the matrix $\mathcal{D}$ can be ill-conditioned, thus making the optimization problem (4) challenging. A way to circumvent this problem is to make use of suitable regularization schemes, and many proposals are made in this direction in the literature, see, e.g., [19, 45].

Another challenge—one of the most crucial ones, if not the most, from the dynamical system perspective—is regarding the stability of inferred models. When the optimization problem (4) is solved, then the inferred operators only aim to minimize the specific design objective. However, it does not guarantee that the resulting dynamical system will be stable; hence, they cannot ensure any type of stability. Therefore, as major contributions of the paper, we, in the next sections, propose novel inference problems that address this issue and ensure local and global asymptotic stability, and global bounded stability of the learned dynamical systems by the construction design.

Lastly, it is worth noting that when $\mathbf{y}(t)$ is a low-dimensional, we do not need to identify a low-dimensional projection. We instead can learn a quadratic model using data for $\mathbf{y}$. Furthermore, to obtain interpretable models, we can employ sparse-regression methodology [9], where we seek to identify the operators $\mathbf{A}$, $\mathbf{H}$, and $\mathbf{B}$ that are potentially sparse.

3. Inference of Local Asymptotic Stable Quadratic Models

In this section, we present a characterization of local asymptotic stable quadratic models of the form (2). This characterization allows us to parametrize the inference problem, which enables constructing local asymptotic stable quadratic models through deliberate parameterization.

3.1. Local asymptotic stabilty characterization

Assuming quadratic systems with zero as a stable equilibrium point, we can use a linear operator of them to determine the local asymptotic stability around zero, and the matrix $\mathbf{A}$ in (2) defines its linear operator. It can be shown that the quadratic system (2) is local asymptotic stable (`LAS`) if none of the eigenvalues of the matrix $\mathbf{A}$ is in the open right half complex plane. Hence, we only need to ensure the stability of the matrix $\mathbf{A}$ for `LAS`. In this regard, a recent work characterizes a stable matrix in [25, Lemma 1], where every asymptotic stable matrix $\mathbf{A}$ can be written as follows:

$$\mathbf{A} = (\mathbf{J} - \mathbf{R})\mathbf{Q} \tag{5}$$

where $\mathbf{J} = -\mathbf{J}^\top$ is a skew symmetric matrix, and $\mathbf{R} = \mathbf{R}^\top \succ 0$ and $\mathbf{Q} = \mathbf{Q}^\top \succ 0$ are symmetric positive definite (SPD) matrices. Using this parameterization for stable matrices, the work [26] learns stable linear systems. Moreover, since zero is a stable equilibrium point, this implies $\mathbf{B} = 0$. Therefore, we assume $\mathbf{B}$ to be zero in the inference of local asymptotic stable models. It is also worth noticing that the parametrization (5) encodes a Lyapunov function for the underlying dynamical system, which is stated in the following lemma.

Lemma 1. *Consider a quadratic system as in (2), where $\mathbf{A}$ takes the form given in (5), i.e., $\mathbf{A} = (\mathbf{J} - \mathbf{R})\mathbf{Q}$, where $\mathbf{J} = -\mathbf{J}^\top$, and $\mathbf{R}, \mathbf{Q}$ are SPD matrices. Then, the quadratic function*

$$\mathbf{V}(\mathbf{x}(t)) = \frac{1}{2}\mathbf{x}(t)^\top \mathbf{Q}\mathbf{x}(t),$$

is a Lyapunov function of the system, defining local asymptotic stability. Moreover, the stability radius can be given by

$$\|\mathbf{x}\|_F = \frac{\sigma_{\min}(\mathbf{L})}{2\sqrt{\|\mathbf{Q}\|_2}\|\mathbf{H}\|_2}, \tag{6}$$

where $\mathbf{L}\mathbf{L}^\top = 2\cdot\mathbf{Q}^\top\mathbf{R}\mathbf{Q}$, *and* $\sigma_{\min}(\cdot)$ *is the minimum singular value of a matrix.*

Proof. First note that $\mathbf{V}(\mathbf{x}(t)) > 0$. Moreover, we have

$$\begin{aligned}\dot{\mathbf{V}}(\mathbf{x}(t)) &= \mathbf{x}(t)^\top\mathbf{Q}\dot{\mathbf{x}}(t) = \mathbf{x}(t)^\top\mathbf{Q}^\top\dot{\mathbf{x}}(t)\\ &= \mathbf{x}(t)^\top\mathbf{Q}^\top\left(\mathbf{A}\mathbf{x}(t) + \mathbf{H}\left(\mathbf{x}(t)\otimes\mathbf{x}(t)\right)\right)\\ &= \mathbf{x}(t)^\top\mathbf{Q}^\top\left(\left(\mathbf{J}-\mathbf{R}\right)\mathbf{Q}\mathbf{x}(t) + \mathbf{H}\left(\mathbf{x}(t)\otimes\mathbf{x}(t)\right)\right)\\ &= \mathbf{x}(t)^\top\mathbf{Q}^\top\mathbf{J}\mathbf{Q}\mathbf{x}(t) - \mathbf{x}(t)^\top\mathbf{Q}\left(\mathbf{R}\mathbf{Q}\mathbf{x}(t) + \mathbf{H}\left(\mathbf{x}(t)\otimes\mathbf{x}(t)\right)\right)\\ &= -\mathbf{x}(t)^\top\mathbf{Q}^\top\mathbf{R}\mathbf{Q}\mathbf{x}(t) + \mathbf{x}(t)^\top\mathbf{Q}^\top\mathbf{H}\left(\mathbf{x}(t)\otimes\mathbf{x}(t)\right).\end{aligned}$$

For small $\mathbf{x}$, e.g., in the Frobenius-norm, the term $\mathbf{x}(t)^\top\mathbf{Q}\mathbf{R}\mathbf{Q}\mathbf{x}(t)$ is dominant almost everywhere. Additionally, since $\mathbf{Q}^\top\mathbf{R}\mathbf{Q}\succ 0$, this yields $\dot{\mathbf{V}}(\mathbf{x}(t))\leq 0$ for small $\mathbf{x}$. This implies that the system is `LAS`.

Toward determining the radius of stability, we begin by noticing that

$$\begin{aligned}\mathbf{A}^\top\mathbf{Q} + \mathbf{Q}\mathbf{A} &= \left((\mathbf{J}-\mathbf{R})\mathbf{Q}\right)^\top\mathbf{Q} + \mathbf{Q}(\mathbf{J}-\mathbf{R})\mathbf{Q}\\ &= \mathbf{Q}^\top(\mathbf{J}^\top - \mathbf{R}^\top)\mathbf{Q} + \mathbf{Q}(\mathbf{J}-\mathbf{R})\mathbf{Q}\\ &= -\mathbf{Q}^\top\mathbf{R}^\top\mathbf{Q} - \mathbf{Q}\mathbf{R}\mathbf{Q} = -2\cdot\mathbf{Q}^\top\mathbf{R}\mathbf{Q}\prec 0.\end{aligned}$$

Next, we define the Cholesky-factorization of $2\cdot\mathbf{Q}^\top\mathbf{R}\mathbf{Q} =: \mathbf{L}\mathbf{L}^\top$. Then, using [23, 24, 46], we can obtain the stability radius as follows:

$$\|\mathbf{x}\|_F = \frac{\sigma_{\min}(\mathbf{L})}{2\sqrt{\|\mathbf{Q}\|_2}\|\mathbf{H}\|_2}, \tag{7}$$

where $\sigma_{\min}(\mathbf{L})$ is the minimum singular value of $\mathbf{L}$. This concludes the proof. □

Summarizing, the parameterization (5) to construct the matrix $\mathbf{A}$ characterizes `LAS` for quadratic systems. Thus, we need not to impose constraints based on matrix inequality or eigenvalues to infer quadratic models. In contrast to [24], we can have the matrix $\mathbf{A}$ non-symmetric as well.

3.2. Local asymptotic stable model inference via suitable parameterization

Using the stable matrix parameterization, we state the problem formulation to infer `LAS` quadratic models as follows. Given the data $\mathbf{X}$, e.g., in reduced-dimension, and the derivative information $\dot{\mathbf{X}}$, we aim at inferring operators of a quadratic model using the following criterion:

$$\begin{aligned}(\mathbf{J},\mathbf{R},\mathbf{Q},\mathbf{H}) &= \underset{\widehat{\mathbf{J}},\widehat{\mathbf{R}},\widehat{\mathbf{Q}},\widehat{\mathbf{H}}}{\arg\min}\|\dot{\mathbf{X}} - (\widehat{\mathbf{J}}-\widehat{\mathbf{R}})\widehat{\mathbf{Q}}\mathbf{X} - \widehat{\mathbf{H}}\mathbf{X}^{\otimes}\|_F,\\ &\text{subject to}\ \ \widehat{\mathbf{J}} = -\widehat{\mathbf{J}}^\top, \widehat{\mathbf{R}} = \widehat{\mathbf{R}}^\top\succ 0, \widehat{\mathbf{Q}} = \tilde{\mathbf{Q}}^\top\succ 0.\end{aligned} \tag{8}$$

Once we have the optimal value for $\{\mathbf{J},\mathbf{R},\mathbf{Q}\}$, we can construct the stable matrix $\mathbf{A}$ as $(\mathbf{J}-\mathbf{R})\mathbf{Q}$, thus leading to a quadratic model of the form (2). Note that the optimization problem is a constrained one. However, as shown in, e.g., [26], we can use adequate parameterizations for these matrices so that we can have an unconstrained optimization. This is,

$$(\bar{\mathbf{J}},\bar{\mathbf{R}},\bar{\mathbf{Q}},\bar{\mathbf{H}}) = \underset{\acute{\mathbf{J}},\acute{\mathbf{Q}},\acute{\mathbf{R}},\acute{\mathbf{H}}}{\arg\min}\|\dot{\mathbf{X}} - (\acute{\mathbf{J}} - \acute{\mathbf{J}}^\top - \acute{\mathbf{R}}\acute{\mathbf{R}}^\top)\acute{\mathbf{Q}}\acute{\mathbf{Q}}^\top\mathbf{X} - \acute{\mathbf{H}}\mathbf{X}^{\otimes}\|_F. \tag{lasMI}$$

The above formulation is obtained by utilizing the Cholesky factorization of $\mathbf{Q}$ and $\mathbf{R}$, and any skew-symmetric matrix $\widehat{\mathbf{J}}$ can be written as $\widehat{\mathbf{J}} = \acute{\mathbf{J}} - \acute{\mathbf{J}}^\top$. Then, we can construct the system (2) with a matrix $\mathbf{A}$ with $\left(\bar{\mathbf{J}} - \bar{\mathbf{J}}^\top - \bar{\mathbf{R}}\bar{\mathbf{R}}^\top\right)\bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top$ and $\mathbf{H} = \bar{\mathbf{H}}$. Not surprisingly, the optimization problem (lasMI) is nonlinear and non-convex; hence, there is no analytical solution to the problem. Therefore, we utilize a gradient-based approach to obtain a solution to the problem.

Remark 1. *We highlight that we have relaxed the SPD condition on* $\mathbf{Q}$ *and* $\mathbf{R}$ *for* `LAS` *to them being symmetric semi-positive definite ones in* (lasMI)*. This slightly eases the optimization process. However, one can strictly impose the definite condition by adding an identity matrix multiplied with a small factor to* $\bar{\mathbf{Q}}\bar{\mathbf{Q}}$ *to define* $\mathbf{Q}$*, i.e.,* $\mathbf{Q} = \bar{\mathbf{Q}}\bar{\mathbf{Q}} + \alpha\mathbf{I}$*, where* $\alpha > 0$*, and similarly for* $\mathbf{R}$*, i.e.,* $\mathbf{R} = \bar{\mathbf{R}}\bar{\mathbf{R}} + \beta\mathbf{I}$*, where* $\beta > 0$*. As a result, for given* $\alpha,\beta > 0$*, the inference problem becomes as follows:*

$$(\bar{\mathbf{J}},\bar{\mathbf{R}},\bar{\mathbf{Q}},\bar{\mathbf{H}}) = \underset{\acute{\mathbf{J}},\acute{\mathbf{Q}},\acute{\mathbf{R}},\acute{\mathbf{H}}}{\arg\min}\left\|\dot{\mathbf{X}} - \left(\acute{\mathbf{J}} - \acute{\mathbf{J}}^\top - \acute{\mathbf{R}}\acute{\mathbf{R}}^\top - \beta\mathbf{I}\right)\left(\acute{\mathbf{Q}}\acute{\mathbf{Q}}^\top + \alpha\mathbf{I}\right)\mathbf{X} - \acute{\mathbf{H}}\mathbf{X}^{\otimes}\right\|_F. \tag{9}$$

4. Inference of Global Asymptotic Stable Quadratic Models

Several quadratic dynamical systems are not only locally stable but also globally stable. We cannot guarantee their global stability properties if we enforce only local stability like in the previous section. Thus, in this section, we study the problem of enforcing global asymptotic stability to quadratic systems (2). To this aim, we first define the concept of energy-preserving nonlinearities, initially presented in [27,47], which we, later on, generalize. This allows us to use the Lyapunov direct method to establish sufficient and necessary conditions for a quadratic system with generalized energy-preserving nonlinearities to be global asymptotic stable (GAS). Finally, we propose a parametrization of quadratic systems that inherently has the global stability property. We then leverage it to infer quadratic models.

4.1. Global asymptotic stability characterization

Let us consider quadratic systems of the form (2). In this section, we assume that $\mathbf{B} = 0$, so the origin is an equilibrium of the system (2). Additionally, we assume that the quadratic nonlinearity satisfies algebraic constraints, which are also discussed in [27,47]. Precisely, the $\mathbf{H}$ matrix in (2) is said to be an energy-preserving quadratic term when it satisfies as follows:

$$\mathbf{H}_{ijk} + \mathbf{H}_{ikj} + \mathbf{H}_{jik} + \mathbf{H}_{jki} + \mathbf{H}_{kij} + \mathbf{H}_{kji} = 0, \tag{10}$$

where $\{i, j, k\} \in \{1, \dots n\}$, $\mathbf{H}_{ijk} := \mathbf{e}_i^\top \mathbf{H}(\mathbf{e}_j \otimes \mathbf{e}_k)$, and $\mathbf{e}_i \in \mathbb{R}^n$ denotes the vector whose entries are all zero besides the i-th entry whose value is one. The condition (10), in terms of the Kronecker-product notation, can be written as follows (see Theorem 2 in Appendix A):

$$\mathbf{x}^\top \mathbf{H}(\mathbf{x} \otimes \mathbf{x}) = 0, \quad \text{for every } \mathbf{x} \in \mathbb{R}^n. \tag{11}$$

For quadratic systems possessing energy-preserving quadratic terms, we can characterize monotonically globally asymptotically stability, i.e., to establish the conditions under which the energy of the state vector $\mathbf{E}(\mathbf{x}(t)) := \frac{1}{2}\mathbf{x}^\top(t)\mathbf{x}(t) = \frac{1}{2}\|\mathbf{x}(t)\|_2^2$ is strictly monotonically decreasing through all trajectories of the quadratic system (2).

Theorem 1. *Consider a quadratic system (2) whose the matrix $\mathbf{H}$ is an energy preserving quadratic term, meaning it satisfies (11). Then, the following statements are equivalent:*

(a) The quadratic system (2) is monotonically global asymptotic stable.

(b) The matrix $\mathbf{A}_s = \frac{1}{2}\left(\mathbf{A} + \mathbf{A}^\top\right)$ is strictly stable, i.e., the eigenvalues of the matrix $\mathbf{A}_s$ are all strictly negative real numbers.

Proof. $(b) \Rightarrow (a)$: Let us consider the state energy of the system as $\mathbf{E}(\mathbf{x}(t)) = \frac{1}{2}\mathbf{x}^\top(t)\mathbf{x}(t)$, where $\mathbf{x}(t)$ is a trajectory of the considered quadratic system. The derivative of the energy function $\mathbf{E}$ with respect to time t at $\mathbf{x}(t)$ is given by

$$\begin{aligned}
\frac{d}{dt}\mathbf{E}(\mathbf{x}(t)) &= \frac{1}{2}\left(\frac{d}{dt}\mathbf{x}(t)\right)^\top \mathbf{x}(t) + \frac{1}{2}\mathbf{x}(t)^\top \left(\frac{d}{dt}\mathbf{x}(t)\right) \\
&= \frac{1}{2}\left(\mathbf{A}\mathbf{x}(t) + \mathbf{H}(\mathbf{x}(t) \otimes \mathbf{x}(t))\right)^\top \mathbf{x}(t) + \frac{1}{2}\mathbf{x}(t)^\top \left(\mathbf{A}\mathbf{x}(t) + \mathbf{H}(\mathbf{x}(t) \otimes \mathbf{x}(t))\right) \\
&= \frac{1}{2}\mathbf{x}(t)^\top(\mathbf{A} + \mathbf{A}^\top)\mathbf{x}(t) + \cancelto{0}{\mathbf{x}(t)^\top \mathbf{H}(\mathbf{x}(t) \otimes \mathbf{x}(t))} \\
&= \mathbf{x}(t)^\top \mathbf{A}_s \mathbf{x}(t) < 0.
\end{aligned}$$

Hence, $\mathbf{E}(\mathbf{x}(t))$ is a strict Lyapunov function, and the quadratic system is, thus, monotonically global asymptotic stable.

$(a) \Rightarrow (b)$: If the system in (2) is monotonically global asymptotic stable, then $\frac{d}{dt}\mathbf{E}(\mathbf{x}(t)) < 0$. Moreover, since the system has an energy preserving quadratic term $\frac{d}{dt}\mathbf{E}(\mathbf{x}(t)) = \mathbf{x}(t)^\top \mathbf{A}_s \mathbf{x}(t)$. Hence, the condition $\mathbf{x}(t)^\top \mathbf{A}_s \mathbf{x}(t) < 0$ hold if all the eigenvalues of $\mathbf{A}_s$ are strictly negative. □

Notice that Theorem 1 characterizes monotonically global asymptotic stability for quadratic systems, providing the quadratic term is energy-preserving. However, it does not cover global asymptotic stable (GAS) quadratic systems for which $\mathbf{E}(\mathbf{x}) = \|\mathbf{x}\|_2^2$ is not a Lyapunov function. Indeed, a sufficient condition for a system to be GAS is the existence of a global strict Lyapunov function. The following remark illustrates by means of an example that Theorem 1 is not enough to determine global asymptotic stability.

Remark 2. *Let us consider the quadratic system of the form* (2), *where*

$$\tilde{\mathbf{A}} = \begin{bmatrix} -4 & -4 \\ 1 & 0 \end{bmatrix} \quad \textit{and} \quad \tilde{\mathbf{H}} = \begin{bmatrix} 2 & 5 & 4 & 10 \\ -1 & -2 & -2 & -4 \end{bmatrix}.$$

Notice that the matrix $\frac{1}{2}(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top)$ *is not stable; additionally, the quadratic term represented by* $\tilde{\mathbf{H}}$ *is not energy-preserving, meaning it does not satisfy* (11). *Hence, according to Theorem 1, this system would not be monotonically globally stable. However, the system is* `GAS` *since*

$$\mathbf{V}(\mathbf{x}(t)) = \mathbf{x}^\top(t)\tilde{\mathbf{Q}}\mathbf{x}(t) \quad \textit{with} \quad \tilde{\mathbf{Q}} = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \succ 0.$$

is a global strict Lyapunov function.

The above remark motives us to derive more general quadratic Lyapunov functions for `GAS` of the form, namely, $\mathbf{V}(\mathbf{x}(t)) = \mathbf{x}^\top(t)\mathbf{Q}\mathbf{x}(t)$ for a given SPD matrix $\mathbf{Q}$. To that aim, we introduce a notion of a generalized energy-preserving quadratic term with respect to $\mathbf{Q}$. For a given SPD matrix $\mathbf{Q}$, the $\mathbf{H}$ matrix in (2) corresponds to a generalized energy-preserving quadratic term when

$$\mathbf{x}^\top\mathbf{Q}\mathbf{H}(\mathbf{x} \otimes \mathbf{x}) = 0, \quad \text{for every } \mathbf{x} \in \mathbb{R}^n. \tag{12}$$

Next, we present conditions under which the function $\mathbf{V}(\mathbf{x}(t)) = \mathbf{x}^\top(t)\mathbf{Q}\mathbf{x}(t)$ would be a strict Lyapunov function for the underlying quadratic system. As a result, due to the direct Lyapunov method, the quadratic system is `GAS`.

Corollary 1. *Consider a quadratic system* (2) *which has the matrix* $\mathbf{H}$ *so that it is an energy-preserving with respect to a SPD matrix* $\mathbf{Q}$. *Then, the following statements are equivalent:*

(a) $\mathbf{V}(\mathbf{x}(t)) = \mathbf{x}^\top(t)\mathbf{Q}\mathbf{x}(t)$ *is a global strict Lyapunov function for the quadratic system* (2).

(b) The matrix $(\mathbf{Q}\mathbf{A})_s := \frac{1}{2}\left(\mathbf{Q}\mathbf{A} + \mathbf{A}^\top\mathbf{Q}\right)$ *is strictly stable, i.e, the eigenvalues* $(\mathbf{Q}\mathbf{A})_s$ *are all strictly negative real numbers.*

Moreover, if these conditions hold, the quadratic system (2) *is* `GAS`.

Proof. Notice that the above result coincides with Theorem 1 if $\mathbf{Q} = \mathbf{I}$. Whenever $\mathbf{Q} \neq \mathbf{I}$, the result is a consequence of Theorem 1 when the linear change of coordinates as $\tilde{\mathbf{x}} := \mathbf{Q}^{\frac{1}{2}}\mathbf{x}$ is applied to the quadratic system (2). □

Corollary 1 provides a generalization of Theorem 1 for general quadratic Lyapunov functions. Additionally, it is worthwhile to stress that if $(\mathbf{Q}\mathbf{A})_s$ is stable but not strictly, then $\mathbf{V}(\mathbf{x}(t)) = \mathbf{x}^\top(t)\mathbf{Q}\mathbf{x}(t)$ is a still global Lyapunov function but not in a strict sense. Consequently, the system will still be globally stable but not necessarily asymptotic. In addition to this, when $(\mathbf{Q}\mathbf{A})_s = 0$, then $\mathbf{V}(\mathbf{x}(t))$ is constant for all the trajectories, i.e., the system (2) is Hamiltonian and energy-preserving.

Based on this result, we focus on parametrizing a family of `GAS` quadratic systems having a generalized energy-preserving term. This result is stated in the following lemma.

Lemma 2. *A quadratic system* (2) *represented by the matrices* $\mathbf{A}$ *and* $\mathbf{H}$, *with the matrix* $\mathbf{H}$ *being corresponding to a generalized energy-preserving quadratic term with respect to the SPD matrix* $\mathbf{Q}$, *is* `GAS` *if*

$$\mathbf{A} = (\mathbf{J} - \mathbf{R})\mathbf{Q}, \tag{13}$$

where $\mathbf{J} = -\mathbf{J}^\top$, $\mathbf{R}$ *is a SPD matrix, and*

$$\mathbf{H} = \left[\mathbf{H}_1\mathbf{Q}, \ldots, \mathbf{H}_n\mathbf{Q}\right], \tag{14}$$

with $\mathbf{H}_i \in \mathbb{R}^{n\times n}$ *being* $\mathbf{H}_i = -\mathbf{H}_i^\top$. *Moreover,* $\mathbf{V}(\mathbf{x}(t)) = \frac{1}{2}\mathbf{x}^\top(t)\mathbf{Q}\mathbf{x}(t)$ *is a global Lyapunov function for the underlying system.*

Proof. Firstly, notice that $\mathbf{H} = \left[\mathbf{H}_1\mathbf{Q}, \ldots, \mathbf{H}_n\mathbf{Q}\right]$ is generalized energy-preserving. Thus,

$$\begin{aligned} \mathbf{x}^\top\mathbf{Q}\mathbf{H}(\mathbf{x} \otimes \mathbf{x}) &= \mathbf{x}^\top\left[\mathbf{Q}\mathbf{H}_1\mathbf{Q}, \ldots, \mathbf{Q}\mathbf{H}_n\mathbf{Q}\right](\mathbf{x} \otimes \mathbf{x}) \\ &= \left(\sum_{i=1}^{n} \mathbf{x}_i\left(\mathbf{x}^\top\mathbf{Q}\mathbf{H}_i\mathbf{Q}\mathbf{x}\right)\right) = 0, \end{aligned}$$

since $\mathbf{Q}\mathbf{H}_i\mathbf{Q} = \mathbf{Q}\mathbf{H}_i\mathbf{Q}^\top$ are skew symmetric matrices for $i \in \{1, \dots n\}$. Hence, the parametrization (14) provides a generalized energy-preserving quadratic term. As a consequence, from Corollary 1, $\mathbf{V}(\mathbf{x}(t)) = \mathbf{x}^\top(t)\mathbf{Q}\mathbf{x}(t)$ is a global Lyapunov function if and only if $\frac{1}{2}\left(\mathbf{Q}\mathbf{A} + \mathbf{A}^\top\mathbf{Q}\right)$ is negative definite. Clearly, the parametrization $\mathbf{A} = (\mathbf{J} - \mathbf{R})\mathbf{Q}$ satisfies this condition, because

$$\frac{1}{2}\left(\mathbf{Q}(\mathbf{J} - \mathbf{R})\mathbf{Q} + \mathbf{Q}(\mathbf{J}^\top - \mathbf{R})\mathbf{Q}\right) = -\mathbf{Q}\mathbf{R}\mathbf{Q} = -\mathbf{Q}^\top\mathbf{R}\mathbf{Q} \prec 0,$$

thus concluding the proof. □

Based on the discussions so far, we present a few key observations, which are as follows.

- Notice that Lemma 2 provides a sufficient condition for quadratic systems (2) to be `GAS`. We believe that the converse of Lemma 2 is also true, i.e., if the generalized energy preserving quadratic system represented by the matrices $\mathbf{A}$ and $\mathbf{H}$ is `GAS`, then $\mathbf{A}$ and $\mathbf{H}$ can be factorized as in (13) and (14), respectively. Indeed, every matrix $\mathbf{A}$ for which $\frac{1}{2}\left(\mathbf{Q}\mathbf{A} + \mathbf{A}^\top\mathbf{Q}\right) < 0$ can be written in the form $\mathbf{A} = (\mathbf{J} - \mathbf{R})\mathbf{Q}$, where $\mathbf{J}$ is skew symmetric and $\mathbf{R}$ is SPD. Moreover, (14) is a sufficient condition to $\mathbf{x}^\top\mathbf{Q}\mathbf{H}(\mathbf{x} \otimes \mathbf{x}) = 0$ which, we believe, is necessary as well. Precisely, we are interested to know whether any matrix $\mathbf{H}$, satisfying $\mathbf{x}^\top\mathbf{Q}\mathbf{H}(\mathbf{x} \otimes \mathbf{x}) = 0$, can be re-written as a matrix $\tilde{\mathbf{H}}$ so that $\tilde{\mathbf{H}}$ has form as in (14) and $\mathbf{H}(\mathbf{x} \otimes \mathbf{x}) = \tilde{\mathbf{H}}(\mathbf{x} \otimes \mathbf{x})$. It would require further investigations, which is out of the scope of this work.
- It is worth noticing that the matrix $\mathbf{H}$, satisfying (14) is not symmetric, implying $\mathbf{H}(e_i \otimes e_j) \neq \mathbf{H}(e_j \otimes e_i)$, where e_i and e_j are canonical unit-vectors. However, suppose one is interested in obtaining a symmetric tensor, i.e., $\mathbf{H}(e_i \otimes e_j) = \mathbf{H}(e_j \otimes e_i)$. In that case, we suggest performing a symmetrization trick as, for example, in [23] once an inference or identification of `GAS` systems is done.
- If the quadratic system (2) is globally stable but not asymptotically stable, then $\mathbf{A} = (\mathbf{J} - \mathbf{R})\mathbf{Q}$ with $\mathbf{R} = \mathbf{R}^\top \succeq 0$, instead of being a definite matrix. Hence, the parametrization that allows $\mathbf{R}$ to be semi-definite also includes globally (non-asymptotic) stable systems.

The parametrization from Lemma 2 guides us to cast the inference problem to obtain `GAS` quadratic systems from data, which we discuss next.

4.2. Global-stability informed learning

Benefiting from Lemma 2, we can write down an inference problem to obtain `GAS` quadratic model using the corresponding $\mathbf{X}$ and $\dot{\mathbf{X}}$ data (for the definitions of $\mathbf{X}$ and $\dot{\mathbf{X}}$, see Section 2) as follows:

$$\begin{aligned}(\mathbf{J}, \mathbf{R}, \mathbf{Q}, \mathbf{H}_1, \dots, \mathbf{H}_n) = \underset{\widehat{\mathbf{J}}, \widehat{\mathbf{R}}, \widehat{\mathbf{Q}}, \widehat{\mathbf{H}}_1, \dots, \widehat{\mathbf{H}}_n}{\arg\min} & \left\|\dot{\mathbf{X}} - (\widehat{\mathbf{J}} - \widehat{\mathbf{R}})\widehat{\mathbf{Q}}\mathbf{X} - \left[\widehat{\mathbf{H}}_1\widehat{\mathbf{Q}}, \dots, \widehat{\mathbf{H}}_n\widehat{\mathbf{Q}}\right]\mathbf{X}^{\otimes}\right\|_F, \\ \text{subject to } & \widehat{\mathbf{J}} = -\widehat{\mathbf{J}}^\top, \widehat{\mathbf{R}} = \widehat{\mathbf{R}}^\top \succ 0, \widehat{\mathbf{Q}} = \widehat{\mathbf{Q}}^\top \succ 0, \text{ and} \\ & \widehat{\mathbf{H}}_i = -\widehat{\mathbf{H}}_i,\ i \in \{1, \dots, n\}.\end{aligned} \tag{15}$$

Having the optimal tuple $(\mathbf{J}, \mathbf{R}, \mathbf{Q}, \mathbf{H}_1, \dots, \mathbf{H}_n)$ solving (15), we can construct the matrices $\mathbf{A}$ and $\mathbf{H}$ as follows:

$$\mathbf{A} = (\mathbf{J} - \mathbf{R})\mathbf{Q}, \qquad \mathbf{H} = \left[\mathbf{H}_1\mathbf{Q}, \dots, \mathbf{H}_n\mathbf{Q}\right], \tag{16}$$

thus leading to a quadratic model of the form (2), which is `GAS`. Similar to the case (8), the above problem enforces constraints on the matrices. To remove these constraints, we can parameterize analogously. The parameterize for matrices $\mathbf{J}, \mathbf{R}$, and $\mathbf{Q}$ remain the same as in Subsection 3.2. We can similarly trick for each $\mathbf{H}_i$ as done for $\mathbf{J}$, but it might be computationally inefficient. However, there is a computationally efficient way to parameterize $\mathbf{H}$, satisfying (14). For this, let us consider a tensor $\mathcal{H} \in \mathbb{R}^{n \times n \times n}$, and denotes its mode-1 and mode-2 matricization by $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$, respectively. For the definition of the matricization for tensors, we refer to [48]. Then, we construct the matrix $\mathbf{H}$ as follows:

$$\left(\mathcal{H}^{(1)} - \mathcal{H}^{(2)}\right)(\mathbf{I} \otimes \mathbf{Q}),$$

then it fulfills the condition (14) by construction. Intuitively, matrices obtained using mode-1 and mode-2 matricizations can interpreted as a transpose of each other in a particular way; hence, $\mathcal{H}^{(1)} - \mathcal{H}^{(2)}$ would yield skew-symmetric matrix in the specific blocks. Thus, we can recast the constraint optimization problem (15) as an unconstrained one as follows:

$$(\bar{\mathbf{J}}, \bar{\mathbf{R}}, \bar{\mathbf{Q}}, \bar{\mathcal{H}}) = \underset{\acute{\mathbf{J}}, \acute{\mathbf{Q}}, \acute{\mathbf{R}}, \acute{\mathcal{H}}}{\arg\min} \left\| \dot{\mathbf{X}} - (\acute{\mathbf{J}} - \acute{\mathbf{J}}^\top - \acute{\mathbf{R}}\acute{\mathbf{R}}^\top)\acute{\mathbf{Q}}\acute{\mathbf{Q}}^\top \mathbf{X} - \left(\acute{\mathcal{H}}^{(1)} - \acute{\mathcal{H}}^{(2)}\right)\left(\mathbf{I} \otimes \acute{\mathbf{Q}}\acute{\mathbf{Q}}^\top\right) \mathbf{X}^{\otimes} \right\|_F. \qquad (\texttt{gasMI})$$

This will then allow us to construct the system (2) with matrices $\mathbf{A}$ and $\mathbf{H}$ as

$$\mathbf{A} = \left(\bar{\mathbf{J}} - \bar{\mathbf{J}}^\top - \bar{\mathbf{R}}\bar{\mathbf{R}}^\top\right)\bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top, \quad \text{and} \quad \mathbf{H} \ = \left(\bar{\mathcal{H}}^{(1)} - \bar{\mathcal{H}}^{(2)}\right)\left(\mathbf{I} \otimes \bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top\right).$$

5. Quadratic Models with Attractive Trapping Regions

Until now, we have discussed the stability of quadratic systems from the perspective of asymptotic stability. However, as we mentioned in the introduction, nonlinear dynamical systems might possess attracting trapping regions, even when no fixed point, if it exists, inside this region is a stable equilibrium. One example of such a phenomenon can be given by the widely-known Lorenz's system [47]. Although for such systems, the long-term behavior of the trajectories is bounded, they do not converge asymptotically to a fixed point. In order to study such attracting trapping regions, the authors in [27], inspired by the results in [47] and the reformulation of the Lyapunov direct method in [49], have characterized these regions as monotonic asymptotic attractive trapping regions when the quadratic systems (2) ensure energy-preserving nonlinearities. Building upon them, we first generalize these results by means of the use of a more general quadratic Lyapunov function of the form $\mathbf{V}(\mathbf{x}(t)) = \mathbf{x}^\top(t)\mathbf{Q}\mathbf{x}(t)$ and the generalized energy preserving quadratic terms as in (12). Finally, we propose a parametrization for quadratic systems that exhibit attractive trapping regions (`ATR`). We then exploit it to infer `ATR` quadratic models by construction.

5.1. Boundedness and attractive trapping region characterization

Consider a quadratic system described in (2) with a generalized energy-preserving term with respect to a SPD matrix $\mathbf{Q}$. A trapping region $\mathcal{M} \subset \mathbb{R}^n$ is (globally) asymptotically attracting if there exists a Lyapunov function, which strictly monotonically decreases along all trajectories starting from an arbitrary state outside of $\mathcal{M}$. This implies that outside of the trapping region, all trajectories should cross its border in a finite amount of time. To characterize such behavior, let us consider a close ball $\mathcal{B}(\mathbf{m}, r)$ centered at $\mathbf{m} \in \mathbb{R}^n$ with radius $r > 0$ that contains the trapping region $\mathcal{M}$. Consequently, a strict Lyapunov function exists outside the ball $\mathcal{B}(\mathbf{m}, r)$. Having said that, we consider the state translation $\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \mathbf{m}$ applied to the system (2), which leads to the translated quadratic system as follows:

$$\begin{aligned}\dot{\tilde{\mathbf{x}}}(t) &= \mathbf{A}\left(\tilde{\mathbf{x}}(t) + \mathbf{m}\right) + \mathbf{H}\left((\tilde{\mathbf{x}}(t) + \mathbf{m}) \otimes (\tilde{\mathbf{x}}(t) + \mathbf{m})\right) + \mathbf{B}, \\ &= \tilde{\mathbf{A}}\tilde{\mathbf{x}}(t) + \tilde{\mathbf{H}}\left(\tilde{\mathbf{x}}(t) \otimes \tilde{\mathbf{x}}(t)\right) + \tilde{\mathbf{B}},\end{aligned} \qquad (17)$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{H}(\mathbf{I} \otimes \mathbf{m}) + \mathbf{H}(\mathbf{m} \otimes \mathbf{I})$, $\tilde{\mathbf{H}} = \mathbf{H}$ and $\tilde{\mathbf{B}} = \mathbf{B} + \mathbf{H}(\mathbf{m} \otimes \mathbf{m}) + \mathbf{A}\mathbf{m}$. Notice that the translated system (17) possesses the forcing term $\tilde{\mathbf{B}}$, which means that the origin would not be an equilibrium point. For the translated system, we set a quadratic Lyapunov function as $\mathbf{V}(\tilde{\mathbf{x}}(t)) = \frac{1}{2}\tilde{\mathbf{x}}(t)^\top\mathbf{Q}\tilde{\mathbf{x}}(t)$, which can be used to characterize an `ATR` for quadratic systems. The result below is a generalization of the result provided [27, Theorem 1] for the case of general quadratic Lyapunov functions.

Corollary 2. *Consider a quadratic system* (17) *which has the matrix* $\mathbf{H}$*, corresponding to an energy-preserving quadratic term with respect to an SPD matrix* $\mathbf{Q}$*. Then, the following statements are equivalent:*

(a) There exist a radius $r > 0$ *and a shifting vector* $\mathbf{m} \in \mathbb{R}^n$ *for which* $\mathbf{V}(\tilde{\mathbf{x}}(t)) = \tilde{\mathbf{x}}^\top(t)\mathbf{Q}\tilde{\mathbf{x}}(t)$ *is a strictly global Lyapunov function for the system outside the ball* $\mathcal{B}(\mathbf{m}, r)$*.*

(b) The matrix $(\mathbf{Q}\tilde{\mathbf{A}})_s := \frac{1}{2}\left(\mathbf{Q}\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top\mathbf{Q}\right)$ *is strictly stable, i.e, the eigenvalues* $(\mathbf{Q}\tilde{\mathbf{A}})_s$ *are all strictly negative real numbers, where* $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{H}(\mathbf{I} \otimes \mathbf{m}) + \mathbf{H}(\mathbf{m} \otimes \mathbf{I})$*.*

Moreover, when these conditions hold, the ball $\mathcal{B}(\mathbf{m}, r)$ *contains an asymptotic trapping region of the quadratic system* (2)*, where* $r = \frac{1}{\sigma_{\min}((\mathbf{Q}\mathbf{A})_s)}\|\tilde{\mathbf{B}}\|$.

Proof. This proof follows the steps in [27] for the general case of the Lyapunov function $\mathbf{V}(\tilde{\mathbf{x}}(t)) = \frac{1}{2}\tilde{\mathbf{x}}(t)^\top \mathbf{Q}\tilde{\mathbf{x}}(t)$. The derivative of $\mathbf{V}(\mathbf{x}(t))$ is given as

$$
\begin{aligned}
\frac{d}{dt}\mathbf{V}(\tilde{\mathbf{x}}(t)) &= \frac{1}{2}\left(\frac{d}{dt}\tilde{\mathbf{x}}(t)\right)^\top \mathbf{Q}\mathbf{x}(t) + \frac{1}{2}\tilde{\mathbf{x}}(t)^\top \mathbf{Q}\left(\frac{d}{dt}\tilde{\mathbf{x}}(t)\right) \\
&= \frac{1}{2}\left(\tilde{\mathbf{A}}\tilde{\mathbf{x}}(t) + \tilde{\mathbf{H}}\left(\tilde{\mathbf{x}}(t)\otimes\tilde{\mathbf{x}}(t)\right) + \tilde{\mathbf{B}}\right)^\top \mathbf{Q}\mathbf{x}(t) + \frac{1}{2}\tilde{\mathbf{x}}(t)^\top \mathbf{Q}\left(\tilde{\mathbf{A}}\tilde{\mathbf{x}}(t) + \tilde{\mathbf{H}}\left(\tilde{\mathbf{x}}(t)\otimes\tilde{\mathbf{x}}(t)\right) + \tilde{\mathbf{B}}\right) \\
&= \frac{1}{2}\mathbf{x}(t)^\top\left(\tilde{\mathbf{A}}^\top\mathbf{Q} + \mathbf{Q}\tilde{\mathbf{A}}\right)\mathbf{x}(t) + \cancelto{0}{\tilde{\mathbf{x}}(t)^\top \mathbf{Q}\mathbf{H}\left(\tilde{\mathbf{x}}(t)\otimes\tilde{\mathbf{x}}(t)\right)} + \tilde{\mathbf{x}}(t)^\top\tilde{\mathbf{B}} \\
&= \tilde{\mathbf{x}}(t)^\top(\mathbf{Q}\tilde{\mathbf{A}})_s\tilde{\mathbf{x}}(t) + \tilde{\mathbf{x}}^\top(t)\tilde{\mathbf{B}}.
\end{aligned}
$$

Notice that for large enough $\tilde{\mathbf{x}}(t)$ (e.g., in 2-norm), the quadratic term $\tilde{\mathbf{x}}(t)^\top(\mathbf{Q}\mathbf{A})_s\tilde{\mathbf{x}}(t)$ dominates the linear term. Hence, $\tilde{\mathbf{x}}(t)^\top\mathbf{A}\tilde{\mathbf{x}}(t) < 0$ if and only if there exist a radius $r > 0$ for which $\mathbf{V}(\tilde{\mathbf{x}}(t))$ is a Lyapunov function outside the ball $\mathcal{B}(\mathbf{m}, r)$.

In order to estimate the radius r, notice that

$$
\begin{aligned}
\frac{d}{dt}\mathbf{V}(\tilde{\mathbf{x}}(t)) &= \tilde{\mathbf{x}}(t)^\top(\mathbf{Q}\tilde{\mathbf{A}})_s\tilde{\mathbf{x}}(t) + \tilde{\mathbf{x}}^\top(t)\tilde{\mathbf{B}} \\
&\le -\sigma_{\min}(\mathbf{Q}\tilde{\mathbf{A}})_s\|\mathbf{x}(t)\|^2 + \|\tilde{\mathbf{B}}\|\|\mathbf{x}(t)\|
\end{aligned}
$$

Hence, if $\|\mathbf{x}(t)\| > \frac{1}{\sigma_{\min}(\mathbf{Q}\tilde{\mathbf{A}})_s}\|\tilde{\mathbf{B}}\| := r$, then $\dot{\mathbf{V}}(\tilde{\mathbf{x}}(t)) < 0$. Hence, the ball $\mathcal{B}(\mathbf{m}, r)$ is an attractive trapping region. □

Corollary 2 provides a characterization for `ATR` systems, by means of the shifted system (17). It enables us to parametrize a family of quadratic systems with generalized energy-preserving nonlinearities that has `ATR`.

Lemma 3. *A quadratic system (2) represented by the matrices* $\mathbf{A}$ *and* $\mathbf{H}$*, with* $\mathbf{H}$ *to correspond to a generalized energy-preserving quadratic term with respect to a SPD matrix* $\mathbf{Q}$ *is* `ATR` *if there exists an* $\mathbf{m} \in \mathbb{R}^n$ *such that matrices* $\tilde{\mathbf{A}}$ *and* $\tilde{\mathbf{H}}$ *of the translated system (17) are given by*

$$
\tilde{\mathbf{A}} = (\mathbf{J} - \mathbf{R})\mathbf{Q}, \tag{18}
$$

where $\mathbf{J} = -\mathbf{J}^\top$, $\mathbf{R} = \mathbf{R}^\top \succ 0$*, and* $\mathbf{Q} = \mathbf{Q}^\top \succ 0$*, and*

$$
\tilde{\mathbf{H}} = \left[\mathbf{H}_1\mathbf{Q}, \dots, \mathbf{H}_n\mathbf{Q}\right], \tag{19}
$$

where $\mathbf{H}_i \in \mathbb{R}^{n\times n}$ *with* $\mathbf{H}_i = -\mathbf{H}_i^\top$*. Moreover,* $\mathbf{V}(\tilde{\mathbf{x}}(t)) = \frac{1}{2}\tilde{\mathbf{x}}^\top(t)\mathbf{Q}\tilde{\mathbf{x}}(t)$ *with* $\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \mathbf{m}$ *is a global Lyapunov function for the underlying system outside the ball* $\mathcal{B}(\mathbf{m}, r)$*, with* $r = \frac{1}{\sigma_{\min}((\mathbf{Q}\mathbf{R})_s)}\|\tilde{\mathbf{B}}\|$.

Proof. The proof follows the same lines of thoughts as Lemma 2. We, therefore, skip it for the brevity of the paper. □

Lemma 3 provides a parametrization for `ATR` systems, guiding us to cast the inference problem, which we discuss next. As for globally `ATR` systems, we briefly comment on the case where $\mathbf{R}$ is semi-definite. If $\mathbf{R}$ is allowed to be semi-definite, then $\mathbf{V}(\tilde{\mathbf{x}}(t)) = \tilde{\mathbf{x}}^\top(t)\mathbf{Q}\tilde{\mathbf{x}}(t)$ will be a non-strictly global Lyapunov function if $\mathbf{v}^\top\tilde{\mathbf{B}} = 0$, where $\tilde{\mathbf{B}}$ is defined in (17), for every $\mathbf{v} \in \mathbb{R}^n$ satisfying $\mathbf{R}\mathbf{Q}\mathbf{v} = 0$. Moreover, when $\mathbf{R} = 0$, the system is energy-preserving if and only if $\tilde{\mathbf{B}} = 0$, and $\mathbf{V}(\tilde{\mathbf{x}}(t))$ is constant.

5.2. Attractive trapping regions informed learning

In this section, we propose a methodology to guarantee a learned quadratic system as in (2) to be `ATR`. To this aim, we notice that we can rewrite the dynamics (2) in terms of the matrices of the shifted system in (17). Indeed, since $\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \mathbf{m}$, we have

$$
\dot{\mathbf{x}}(t) = \tilde{\mathbf{A}}(\mathbf{x}(t) - \mathbf{m}) + \tilde{\mathbf{H}}\left((\mathbf{x}(t) - \mathbf{m})\otimes(\mathbf{x}(t) - \mathbf{m})\right) + \tilde{\mathbf{B}}, \tag{20}
$$

where $\tilde{\mathbf{A}}$ is a stable matrix, and $\tilde{\mathbf{H}}$ is energy preserving. Then, we can leverage the parametrization from Lemma 3 to obtain an `ATR` system. Since the vector $\mathbf{m}$ is also unknown, it needs to be a part of the inference problem as well. As a consequence, we can write down the inference problem to obtain ATR systems using the corresponding data as follows:

$$\begin{aligned}
&(\tilde{\mathbf{J}},\tilde{\mathbf{R}},\tilde{\mathbf{Q}},\tilde{\mathbf{H}}_1,\ldots,\tilde{\mathbf{H}}_n,\tilde{\mathbf{B}},\mathbf{m})\\
&\qquad= \underset{\widehat{\mathbf{J}},\widehat{\mathbf{R}},\widehat{\mathbf{Q}},\widehat{\mathbf{H}}_1,\ldots,\widehat{\mathbf{H}}_n,\widehat{\mathbf{B}},\widehat{\mathbf{m}}}{\arg\min} \left\|\dot{\mathbf{X}} - (\widehat{\mathbf{J}}-\widehat{\mathbf{R}})\widehat{\mathbf{Q}}(\mathbf{X}-\widehat{\mathbf{m}}) - \left[\widehat{\mathbf{H}}_1\widehat{\mathbf{Q}},\ldots,\widehat{\mathbf{H}}_n\widehat{\mathbf{Q}}\right](\mathbf{X}-\widehat{\mathbf{m}})^{\otimes} + \widehat{\mathbf{B}}\right\|\\
&\qquad\qquad \text{subject to } \ \widehat{\mathbf{J}} = -\widehat{\mathbf{J}}^\top, \widehat{\mathbf{R}} = \widehat{\mathbf{R}}^\top \succ 0, \widehat{\mathbf{Q}} = \widehat{\mathbf{Q}}^\top \succ 0, \text{ and}\\
&\qquad\qquad\qquad\qquad \widehat{\mathbf{H}}_i = -\widehat{\mathbf{H}}_i,\ i\in\{1,\ldots,n\}, \widehat{\mathbf{B}}, \widehat{\mathbf{m}} \in \mathbb{R}^n.
\end{aligned} \tag{21}$$

Having the optimal tuple $(\tilde{\mathbf{J}},\tilde{\mathbf{R}},\tilde{\mathbf{Q}},\tilde{\mathbf{H}}_1,\ldots,\tilde{\mathbf{H}}_n,\tilde{\mathbf{B}},\mathbf{m})$ for (21), we can construct $\mathbf{A}$, $\mathbf{H}$ and $\mathbf{B}$ as follows:

$$\mathbf{A} = (\tilde{\mathbf{J}}-\tilde{\mathbf{R}})\tilde{\mathbf{Q}} - \mathbf{H}(\mathbf{I}\otimes\tilde{\mathbf{m}}) - \mathbf{H}(\mathbf{m}\otimes\mathbf{I}), \mathbf{B} = \tilde{\mathbf{B}} - \mathbf{H}(\mathbf{m}\otimes\mathbf{m}) - \mathbf{A}\mathbf{m}, \ \text{ and } \ \mathbf{H} = \left[\tilde{\mathbf{H}}_1\tilde{\mathbf{Q}},\ldots,\tilde{\mathbf{H}}_n\tilde{\mathbf{Q}}\right], \tag{22}$$

thus leading to a quadratic model of the form (2) that is ATR. Moreover, the same parametrization tricks, used in Section 4.2 for $\mathbf{J},\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{H}$, can be employed here. Thus, we can recast the constraint optimization problem (21) as an unconstrained one as follows:

$$\begin{aligned}
(\mathbf{m},\bar{\mathbf{J}},\bar{\mathbf{R}},\bar{\mathbf{Q}},\bar{\mathcal{H}}) = \underset{\mathbf{m},\acute{\mathbf{J}},\acute{\mathbf{Q}},\acute{\mathbf{R}},\acute{\mathcal{H}}}{\arg\min} &\left\|\dot{\mathbf{X}} - (\acute{\mathbf{J}} - \acute{\mathbf{J}}^\top - \acute{\mathbf{R}}\acute{\mathbf{R}}^\top)\acute{\mathbf{Q}}\acute{\mathbf{Q}}^\top(\mathbf{X}-\mathbf{m})\right.\\
&\left. - \left(\acute{\mathcal{H}}^{(1)} - \acute{\mathcal{H}}^{(2)}\right)\left(\mathbf{I}\otimes\acute{\mathbf{Q}}\acute{\mathbf{Q}}^\top\right)(\mathbf{X}-\mathbf{m})^{\otimes} - \tilde{\mathbf{B}}\right\|_F.
\end{aligned} \tag{atrMI}$$

This then allows the construction of the system (2) with matrices $\mathbf{A}$ and $\mathbf{H}$ as follows:

$$\begin{aligned}
\mathbf{H} &= \left(\bar{\mathcal{H}}^{(1)} - \bar{\mathcal{H}}^{(2)}\right)\left(\mathbf{I}\otimes\bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top\right),\\
\mathbf{A} &= \left(\bar{\mathbf{J}} - \bar{\mathbf{J}}^\top - \bar{\mathbf{R}}\bar{\mathbf{R}}^\top\right)\bar{\mathbf{Q}}\bar{\mathbf{Q}}^\top - \mathbf{H}(\mathbf{I}\otimes\mathbf{m}) - \mathbf{H}(\mathbf{m}\otimes\mathbf{I}),
\end{aligned} \tag{23}$$

Once again, the inferred is guaranteed to be ATR due to its dedicated parameterization.

6. Optimize Through an Integrator to Infer Operators

To infer operators using lasMI, gasMI, and atrMI, we require the derivative information. However, obtaining this information can be challenging using numerical methods, particularly when the data are noisy, scarce, or irregularly sampled. Therefore, it would be beneficial to infer operators via an integral form of differential equations, which are widely adopted for learning continuous-time models. For this, let us consider the following general form of differential equation:

$$\dot{\mathbf{x}}(t) = \mathbf{g}_{\boldsymbol{\alpha}}(\mathbf{x}(t)),$$

where $\mathbf{g}$ defines a vector field and is parameterized by learnable $\boldsymbol{\alpha}$. In our case, $\mathbf{g}$ has a quadratic form, and $\boldsymbol{\alpha}$ contains all learnable operators such as $\mathbf{J}$ and $\mathbf{R}$.

Given the data $\mathbf{x}(t)$ and $\dot{\mathbf{x}}(t)$ at $t\in\{t_1,\ldots,\mathcal{N}\}$, we can cast the learning $\mathbf{g}$ as follows:

$$\min_{\boldsymbol{\alpha}} \sum_i \|\dot{\mathbf{x}}(t_i) - \mathbf{g}_{\boldsymbol{\alpha}}(\mathbf{x}(t_i))\|.$$

We can also re-cast the above optimization problem in an integral form as follows:

$$\min_{\boldsymbol{\alpha}} \sum_i \left\|\mathbf{x}(t_{i+1}) - \int_{t_i}^{t_{i+1}} \mathbf{g}_{\boldsymbol{\alpha}}(\mathbf{x}(t))\mathtt{dt}\right\|, \tag{24}$$

which then avoids the need for derivative information. However, it comes at the expense that the problem (24) is more involved. The concepts of NeuralODEs [42] are developed, allowing us to solve such problems efficiently by taking a gradient of learnable parameters through any numerical integration method with arbitrary accuracy. Despite this method being memory efficient ($\mathcal{O}(1)$), it might call several function calls (in this case, the function $\mathbf{g}$). This might add to computational costs. Therefore, at the expense of memory and slight accuracy, we approximate the integral in (24) using a fourth-order Runge-Kutta method as follows:

$$\mathbf{x}(t_{i+1}) = \Phi_{\mathtt{dt}}\left(\mathbf{g}_{\boldsymbol{\alpha}}(\mathbf{x}(t)), \mathbf{x}(t_i)\right) := \mathbf{x}(t_i) + \mathtt{dt}\left(\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4\right), \tag{25}$$

where $\mathtt{dt} = t_{i+1} - t_i$ and

$$\begin{aligned} \mathbf{h}_1 &= \tfrac{1}{6}\mathbf{g}_{\boldsymbol{\alpha}}(\mathbf{x}(t_i)), & \mathbf{h}_2 &= \tfrac{1}{3}\mathbf{g}_{\boldsymbol{\alpha}}\left(\mathbf{x}(\mathbf{t}_i) + \frac{\mathtt{dt}}{2}\mathbf{h}_1\right) \\ \mathbf{h}_3 &= \tfrac{1}{3}\mathbf{g}_{\boldsymbol{\alpha}}\left(\mathbf{x}(t_i) + \tfrac{\mathtt{dt}}{2}\mathbf{h}_2\right) & h_3 &= \frac{1}{6}\mathbf{g}_{\boldsymbol{\alpha}}\left(\mathbf{x}(t_i) + \mathtt{dt}\mathbf{h}_3\right). \end{aligned} \tag{26}$$

Subsequently, we can write the objective function (24) as follows:

$$\min_{\bar{\boldsymbol{\alpha}}} \sum_i \|\mathbf{x}(t_{i+1}) - \Phi_{\mathtt{dt}}\left(\mathbf{g}_{\bar{\boldsymbol{\alpha}}}, \mathbf{x}(t_i)\right)\|_F. \tag{27}$$

We can utilize the above formation in our inference problems. For example, we can write the lasMI problem as follows:

$$\bar{\boldsymbol{\alpha}} = \arg\min_{\acute{\boldsymbol{\alpha}}} \sum_i \|\mathbf{x}(t_{i+1}) - \Phi_{\mathtt{dt}}\left(\tilde{\mathbf{g}}_{\acute{\boldsymbol{\alpha}}}, \mathbf{x}(t_i)\right)\|_F, \tag{28}$$

where $\acute{\boldsymbol{\alpha}} = (\acute{\mathbf{J}}, \acute{\mathbf{R}}, \acute{\mathbf{Q}}, \acute{\mathbf{H}})$ and $\tilde{\mathbf{g}}_{\acute{\boldsymbol{\alpha}}}(\mathbf{x}) = (\acute{\mathbf{J}} - \acute{\mathbf{J}}^\top - \acute{\mathbf{R}}\acute{\mathbf{R}}^\top)\acute{\mathbf{Q}}\acute{\mathbf{Q}}^\top\mathbf{x} + \acute{\mathbf{H}}(\mathbf{x}\otimes\mathbf{x})$. This way, we can combine a numerical integration scheme with lasMI to infer the operators, thus avoiding the need of the derivative information at any stage. Analogously can be done for gasMI and atrMI as well.

7. Numerical Experiments

In this section, we demonstrate the effectiveness of the proposed stability-guarantee methodologies, namely lasMI, gasMI, and atrMI. To evaluate the performance of these methodologies, we apply them to two case studies involving high-dimensional data obtained from Burgers' equation and Chafee-Infante equations. Additionally, we use two examples, namely the chaotic Lorenz model and the magneto-hydrodynamics model, to demonstrate the utility of the proposed methodologies for discovering governing equations via sparse regression and learning energy-preserving models. We shall discuss a precise set-up of each example in their respective subsections. However, note that all three proposed methodologies incorporate a numerical integration scheme as discussed in Section 6 so that we are not required to estimate the derivative information from data that could be challenging if they are scarce and noisy. And this also enhances the performance of the classical OpInfas well, as discussed in [44].

Training set-up: Given that the involved optimization problems in inferring (sparse) operators are nonlinear, non-convex, and lack analytical solutions, we adopt a gradient descent method for identification. Specifically, we utilize the Adam optimizer [50] with a triangular cyclic learning rate ranging from 10^{-6} to 10^{-2} over a cycle of 4 000-steps and make 12 000 updates in total. The coefficients of all matrices are initialized with random values drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 0.1. To evaluate the performance of our proposed methodology, we compare it with the operator inference method discussed in [44], which utilizes the integral form of differential equations to learn operators but without imposing any constraints. We refer to it as opInf-benchmark. To ensure a fair comparison, we also use the Runge-Kutta integration scheme as discussed in Section 6 for opInf-benchmark. Furthermore, for sparse regression problems to discover governing equations, we employ a sequential thresholding algorithm [9, 43]. We set the tolerance 0.1 and perform hard-pruning of coefficients below the tolerance after each 12 000 update. We optimize the remaining non-zero coefficients. We repeat this process four times to obtain a sparse parsimonious model which exhibits the desired stability properties. We implement all the methods using PyTorch on 12th Gen Intel® Core™i5-12600K 32GB RAM.

Setting $\mathbf{Q} = \mathbf{I}$ for gasMI and atrMI: The involvement of matrix $\mathbf{Q}$ in defining both matrices $\mathbf{A}$ and $\mathbf{H}$ (see Lemmas 2 and 3) makes the corresponding optimization problem even more challenging to solve. To ease the problem, we set $\mathbf{Q}$ to identity. It results in monotonic Lyapunov functions for GAS and ATR. However, it can indeed limit the flexibility of the learned operators. We had attempted to learn $\mathbf{Q}$ along with other parameters but encountered challenges with convergence and got stuck in bad local minima. In future work, we plan to investigate more efficient ways of solving the optimization problem when $\mathbf{Q} \neq \mathbf{I}$. This will require developing novel strategies to effectively handle the additional complexity introduced by the matrix $\mathbf{Q}$.

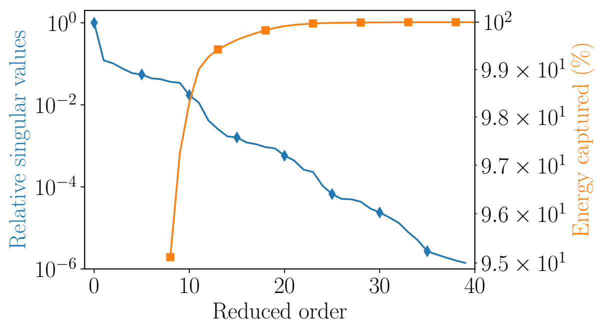


Figure 1: Burgers' equation: A decay of the singular values obtained using training data. It also indicates (in orange color) how much energy is captured by how many dominant modes.

7.1. Learning reduced operators for high-dimensional data via `OpInf`

We begin by considering high-dimensional data and learning corresponding operators in a low-dimension.

7.1.1. Burgers' equations

In our first example, we consider the one-dimensional Burgers' equation, which is governed by the following equations:

$$\begin{aligned} v_t + v v_\zeta &= \mu \cdot v_{\zeta\zeta} && \text{in } (0,1)\times(0,T), \\ v_\zeta(0,\cdot) &= 0, \quad \text{and} \quad u_\zeta(1,\cdot) = 0, && \\ u(\zeta,0) &= v_0(\zeta) && \text{in } (0,1), \end{aligned} \tag{29}$$

where v_t and v_ζ denote the derivative of v with respect to the time t and space ζ, and $v_{\zeta\zeta}$ denotes the double-derivative of v with respect to ζ. We consider the same setup for discretization as in [23]. We consider 1 000 equidistant grid points in the spatial domain, and 500 equidistant data points are collected in the time-interval $[0,1]$. Furthermore, as in [26], we collect data using 17 differential initial conditions, i.e.,

$$v_0(\zeta) = 1 + \sin((2f\zeta + 1)\pi), \qquad f = \{1, 1.25, \dots, 4.75, 5.0\}.$$

Taking data corresponding to three initial conditions ($f = \{1.75, 2.75, 3.75\}$) out for testing, we use the rest of them for learning operators for quadratic systems of the form (2).

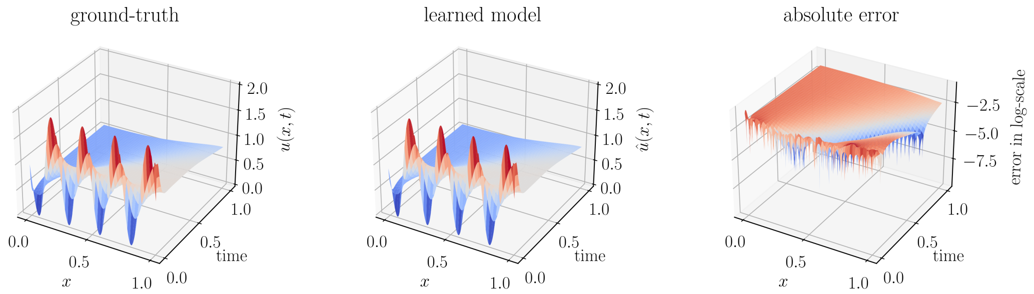
First, we plot the decay of singular values of the training data in Figure 1, which also indicates how many dominant modes capture how much energy presented in the data. We notice a rapid decay of it, thus a possibility of constructing low-order models yet capturing the underlying dynamics accurately. Next, we construct various low-dimensional data by projecting the high-dimensional data onto the r-most dominant vectors, where r ranges from 10 to 28 in increments of 2. We then infer the corresponding operators using `opInf-benchmark`, `lasMI`, and `gasMI`. Since we know that the model is an asymptotic stable, we focus only on `LAS` and `GAS`. We test the quality of these learned models on the left-out testing data, which is discussed in the following.

We first project the left-out testing initial conditions on the corresponding low-dimensional subspace of order 20 determined using the dominant subspace of the training data. We then simulate the reduced-order model in the time interval $[0,1]$ by considering 1 000 points. We then re-project the solutions on the high dimensional using the same projection matrix. We present the solutions obtained using these learned models for one of the test cases in Figure 2, illustrating that the inferred model captures the essential dynamics. For a more qualitative measure, we first compute the relative L_2-norm of the error between the ground truth and the predicted using the learned models as follows:

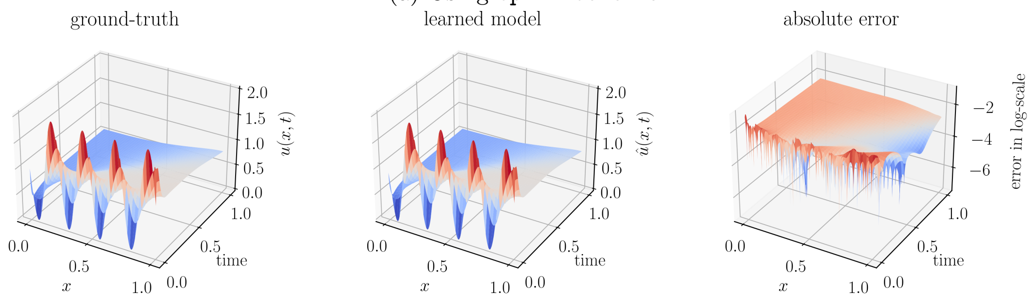
$$\frac{\|\mathbf{X}^{\texttt{ground-truth}} - \mathbf{X}^{\texttt{learned}}\|_2}{\|\mathbf{X}^{\texttt{ground-truth}}\|_2}, \tag{30}$$

where $\mathbf{X}^{\texttt{ground-truth}}$ and $\mathbf{X}^{\texttt{learned}}$, respectively, store the solutions using the high-fidelity models and using the inferred reduced quadratic models for a considered initial condition. It is then followed by taking the mean over all test samples.

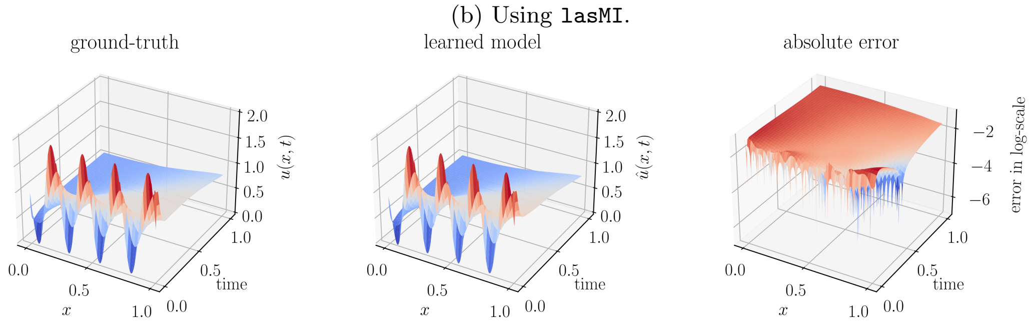
We plot these results in Figure 3, which shows `opInf-benchmark` and `lasMI` produce similar performance on the test case, whereas `gasMI` yields slightly inferior results. However, it should be noted that



(a) Using `opInf-benchmark`.



(b) Using `lasMI`.



(c) Using `gasMI`.

Figure 2: Burgers' equation: A comparison of the time-domain simulations of the inferred models on an initial test condition.

`gasMI` guarantees to be globally stable by construction, even in the cases when the model is not globally stable, which is potentially the case for this example. Another reason may be attributed to the fact that the matrix $\mathbf{Q}$ is set to be the identity for `GAS`, which may not be optimal. Moreover, `lasMI` learns models that are `LAS`, and it is determined by the eigenvalues of the matrix $\mathbf{A}$ in the system (2). None of the stability properties can be guaranteed by `opInf-benchmark`, and to show this, we closely look at the eigenvalues of the learned linear operator $\mathbf{A}$ using all three methods. For this, we plot the eigenvalues of the linear operator $\mathbf{A}$ for reduced-order $r = 14$ by projecting them on the unit circle in Figure 4. We can notice that `opInf-benchmark` is marginally locally unstable, whereas `lasMI` and `gasMI` yield are locally stable models.

7.1.2. Chafee-Infante example

In our second example, we consider the one-dimensional Chafee-Infante equation, which is governed by the following equations:

$$
\begin{aligned}
v_t + v^3 &= v + v_{\zeta\zeta} && \text{in } (0,1)\times(0,T),\\
v_\zeta(0,\cdot) = 0, \quad \text{and} \quad v_\zeta(1,\cdot) &= 0, &&\\
v(\zeta,0) &= v_0(\zeta) && \text{in } (0,1),
\end{aligned}
\tag{31}
$$

where, similar to the Burgers' equation, v_t and $v_{\zeta\zeta}$ denote the derivative of v with respect to the time t, and double derivative of v with respect to the space ζ. It is noteworthy that the governing equation is of cubic nonlinearity. However, these equations can be written as a quadratic system by defining an auxiliary state variable as αv^2, where α is a non-zero scalar, using lifting principles [22, 23, 37–40].

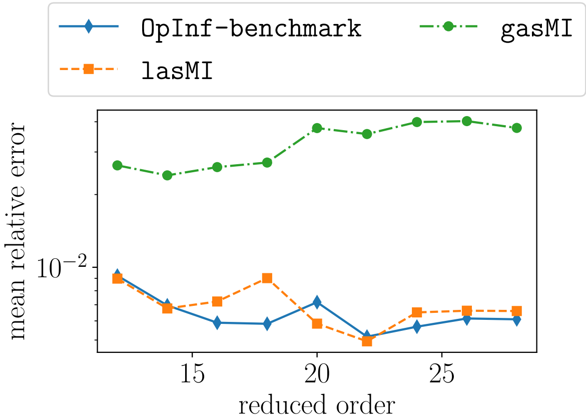


Figure 3: Burgers' equation: A mean performance over all the test data of the inferred models.

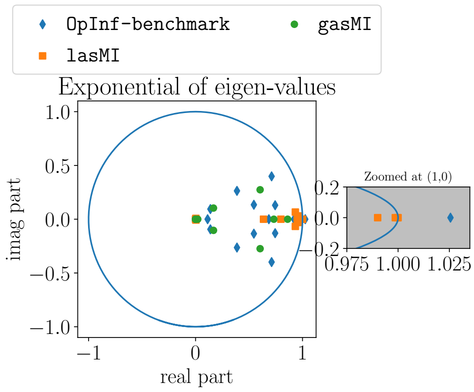


Figure 4: Burgers' equation: The plot shows the eigenvalues of the inferred operators but projected on the unit circle.

Towards creating training and testing datasets, we first discretize the governing PDE using 1 000 equidistant points for a finite-difference scheme. Then, we consider different initial conditions to collect data, as we did for the Burgers' equation in the previous example. We parameterize the initial condition as follows:

$$v(\zeta, 0) = 0.1 + f \cdot (\sin(4\pi\zeta))^2, \tag{32}$$

for a given f. We consider 13 initial conditions by taking 13 equidistant points for f in the interval $[1, 3]$. Assuming the considered values for f are shorted in increasing order, we take 4th, 8th, and 11th indices for testing and the remaining 10 values for training. For each initial condition, we take 500 points in the time interval $[0, 8]$.

To learn models, we augment the state using $w := \alpha v^2$ so that a quadratic model can jointly describe the dynamics of both v and w. Before that, we also shift v by 1 to have zero as an equilibrium point. Moreover, we set α to 0.5 so that the singular values of v and w are in the same order (at least the dominant ones); otherwise, we observed numerical difficulties in inferring operators correctly. We first plot the decay of singular values for v and w in Figure 5. The figure indicates a rapid decay of the singular values, confirming a low-dimensional representation of data. After that, we determine the dominant subspaces for v and w based on how much energy is aimed to capture. We stress that we determine dominant subspaces for v and w separately—denoted by $\mathbf{V}_\mathrm{p}$ and $\mathbf{W}_\mathrm{p}$—so that information of these two quantities are not mixed. In our experiments, this yields a more robust inference. We then compute the projection matrix $\mathbf{V} := \operatorname{diag}(\mathbf{V}_\mathrm{p}, \mathbf{W}_\mathrm{p})$. Using the matrix $\mathbf{V}$, we determine a low-dimensional coordinate system for which we aim to learn reduced operators. By varying the amount of the energy captured by $\mathbf{V}$, we obtain different dimensions of the reduced coordinate system, thus the order of reduced models. We learn reduced-order operators using all three considered methods. The performance of these models is determined using the left-out test data. For reduced-order $r = 7$, we plot the time-domain simulation for one of the test initial conditions in Figure 6.

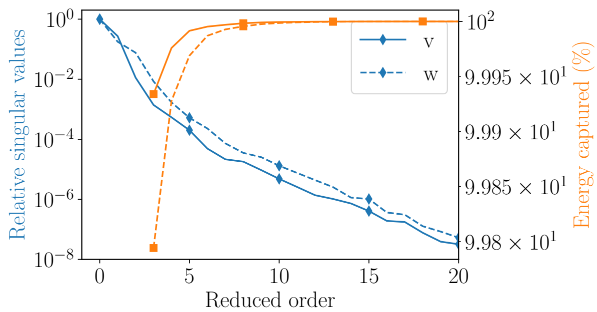
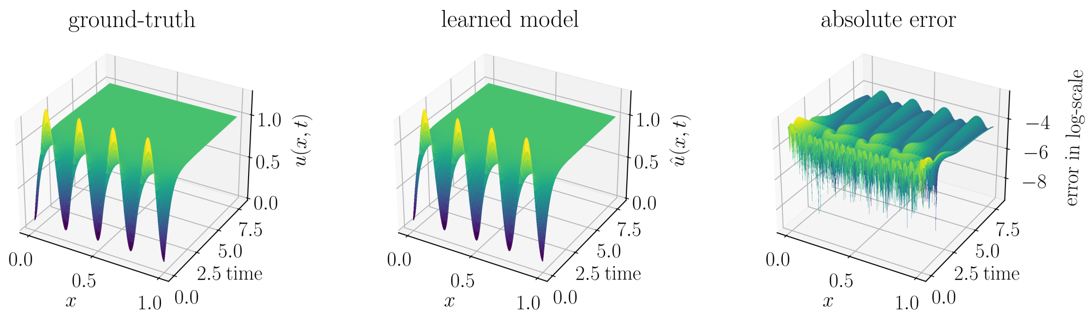
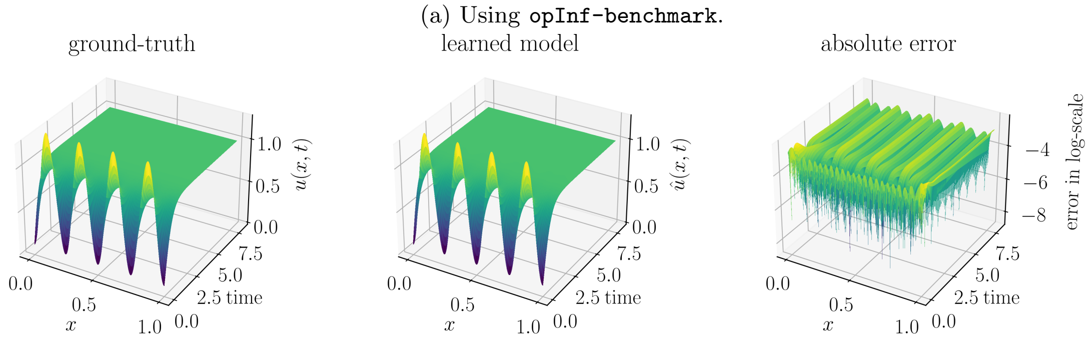


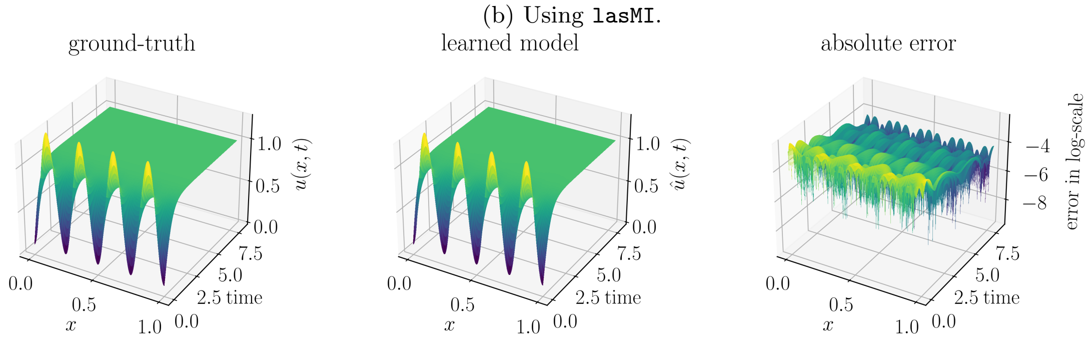
Figure 5: Chafee Infante equation: Decay of the singular values for the v and $w := \frac{1}{2}v^2$ using the training data.



(a) Using `opInf-benchmark`.



(b) Using `lasMI`.



(c) Using `gasMI`.

Figure 6: Chafee-Infante equation: A comparison of the time-domain simulations of the learned models on an initial condition for testing.

Moreover, the mean error over all the initial testing conditions—similar to the previous example—is plotted in Figure 7. For this particular example, we observe that all the inferred models were locally stable, including for `OpInf`, although it does not explicitly impose the stability. However, we notice a better performance of `lasMI` and `gasMI` for all three test cases as compared to `opInf-benchmark`. Moreover, `gasMI`—which guarantees global stability—also produces comparable models to `lasMI` despite `gasMI` having more restrictions on operators.

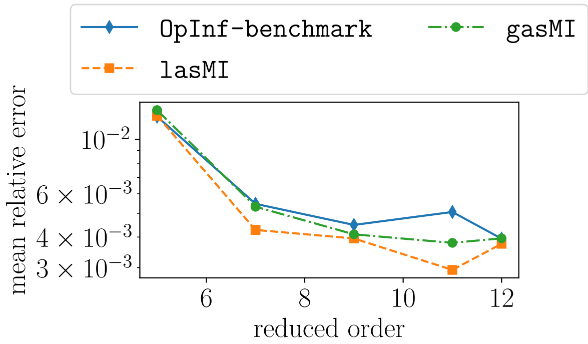


Figure 7: Chafee Infant equation: A mean performance over all the test data using the inferred models.

7.2. Discovery of governing equations via sparse regression

Next, we test the proposed stability-guaranteed learning in the context of sparse regression to discover governing equations.

7.2.1. Chaotic Lorenz models

The Lorenz model, which is one of the simplest models describing chaotic dynamics with roots in fluid dynamics [47]. Its dynamics is given by

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\mathbf{y}}(t) \\ \dot{\mathbf{z}}(t) \end{bmatrix} = \begin{bmatrix} -10\mathbf{x}(t) + 10\mathbf{y}(t) \\ \mathbf{x}(t)(28 - \mathbf{z}(t)) - \mathbf{y}(t) \\ \mathbf{x}(t)\mathbf{y}(t) - \frac{8}{3}\mathbf{z}(t) \end{bmatrix}. \tag{33}$$

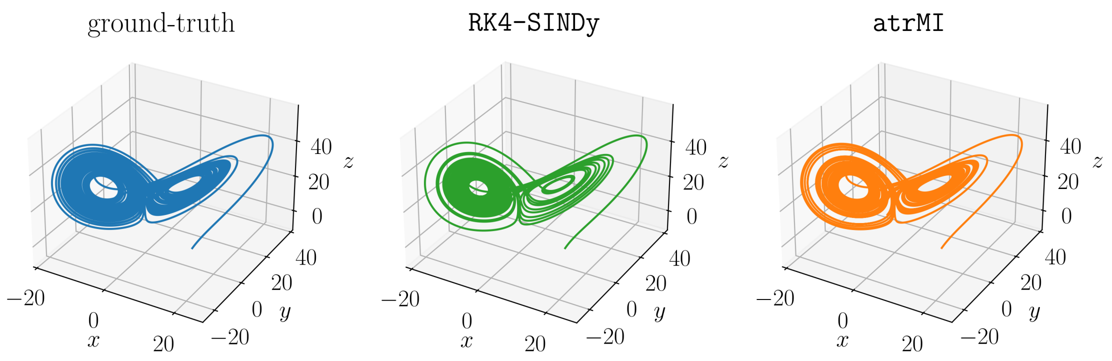
We gather 5 000 data points in the time interval $t = 0$ to $t = 20$ with an initial condition $[-8.0, 7.0, 27.0]$. We corrupt the data by adding the Gaussian noise of mean 0 and standard deviation 0.1. Like done in [43], we shift and re-scale the data as follows:

$$\tilde{\mathbf{x}}(t) = \frac{\mathbf{x}(t)}{8}, \quad \tilde{\mathbf{y}}(t) = \frac{\mathbf{y}(t)}{8}, \quad \text{and} \quad \tilde{\mathbf{z}}(t) = \frac{\mathbf{z}(t) - 25}{8}. \tag{34}$$

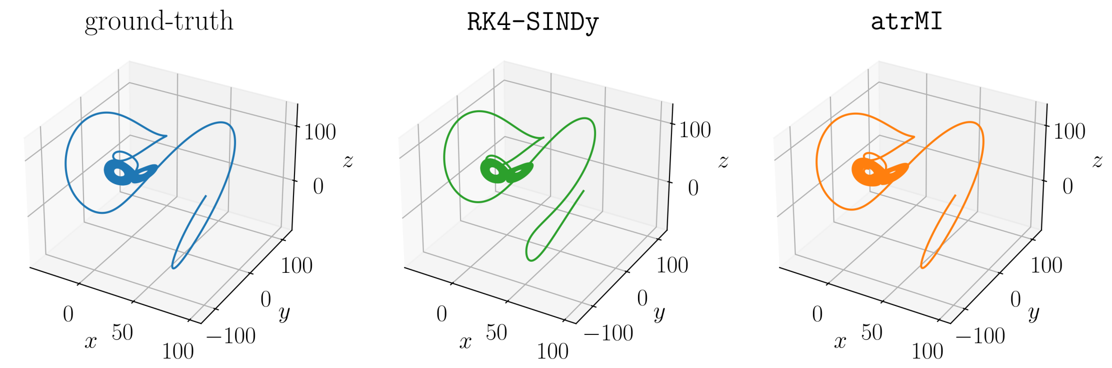
With the shifting, we bring the mean and standard deviation of all the variables close to 0 and 1, respectively, without changing the interaction pattern, see [43].

Note that the Lorenz system does not have any stable equilibrium point, but it is globally bounded and stable. Thus, it allows us to employ the trapping globally stable arguments discussed in Section 5. We employ a `RK4-SINDy` [43], inspired by sparse regression for system identification (SINDy) approach [9] so that we can discover the underlying model without requiring derivative information. Instead of derivative information, it also uses an integral form and approximates it using a fourth-order Runge-Kutta scheme. Note that there exist other methodologies, see, e.g., [51], which also avoid the need for the derivative information to discover governing equations in continuous time.

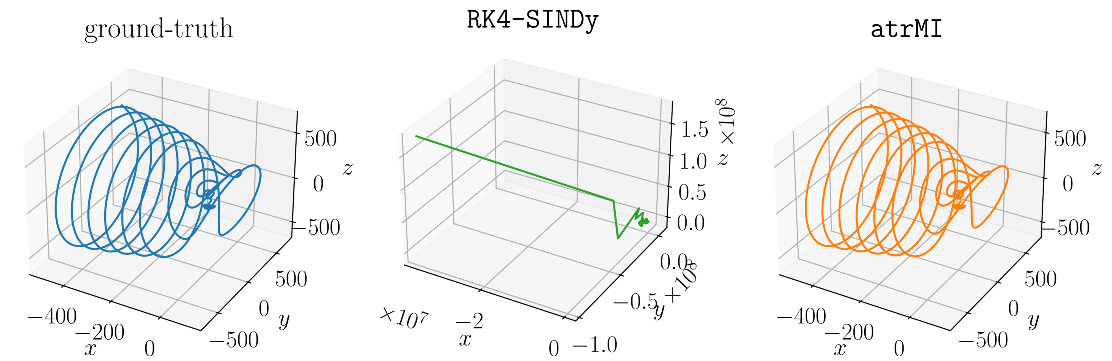
Next, we combine the discussion in Subsection 5 with sparse regression, where we seek to determine sparse matrices $\mathbf{A}$ (or $\mathbf{J}$ and $\mathbf{R}$), $\mathbf{H}$, and $\mathbf{B}$. To that end, we use an iterative thresholding scheme [9, 43], in which small coefficients are hard-pruned. We learn models using both approaches and compare their performance by simulating for different initial conditions. We consider three test initial conditions as $[10, 10, -10]$, $[100, -100, 100]$, and $[-500, 500, 500]$. Note that these test initial conditions are even of a different order than the training one. We plot their responses in Figure 8, where we observe that **`atrMI`** yields the models which are globally stable by construction and can faithfully capture the Lorenz dynamics. The **`atrMI`** model results in dynamics correctly on the attractor. In contrast, `RK4-SINDy` does not guarantee global stability, and for very large initial conditions (for example, in 2-norm), it produces an unstable trajectory.



(a) For initial condition $[10, 10, -10]$.



(b) For initial condition $[100, -100, 100]$.



(c) For initial condition $[-500, 500, 500]$.

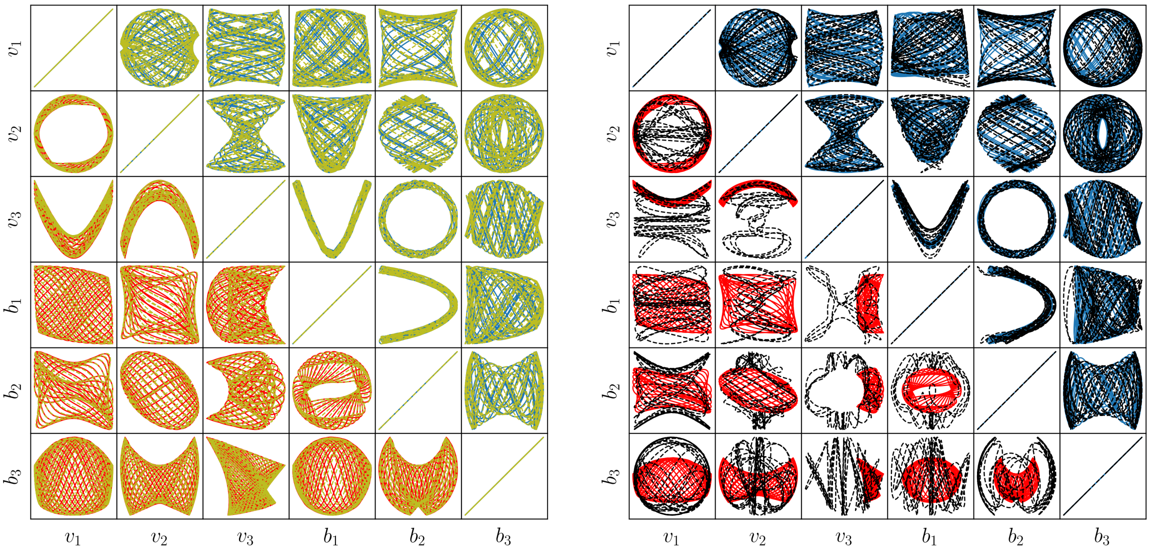
Figure 8: Lorenz model: A comparison of the time-domain simulations of the learned models for test initial conditions.

7.2.2. Magneto-hydrodynamic model

In our last example, we aim to discover a magneto-hydrodynamic (MHD) model with quadratic non-linearities. The model often exhibits energy-preserving properties. In this paper, we consider a simple two-dimensional incomprehensible MHD model from [36, 52], which is given by

$$\begin{bmatrix} \dot{v}_1(t) \\ \dot{v}_2(t) \\ \dot{v}_3(t) \\ \dot{b}_1(t) \\ \dot{b}_2(t) \\ \dot{b}_3(t) \end{bmatrix} = \begin{bmatrix} -2\nu & 0 & 0 & 0 & 0 & 0 \\ 0 & -5\nu & 0 & 0 & 0 & 0 \\ 0 & 0 & -9\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & -2\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & -5\nu & 0 \\ 0 & 0 & 0 & 0 & 0 & -9\nu \end{bmatrix} \begin{bmatrix} v_1(t) \\ v_2(t) \\ v_3(t) \\ b_1(t) \\ b_2(t) \\ b_3(t) \end{bmatrix} + \begin{bmatrix} 4\,(v_2v_3 - b_2b_3) \\ -7\,(v_1v_3 - b_1b_3) \\ 3\,(v_1v_2 - b_1b_2) \\ 2\,(b_3v_2 - v_3b_2) \\ 5\,(b_1v_3 - v_1b_3) \\ 9\,(b_2v_1 - b_1v_2) \end{bmatrix}, \tag{35}$$

where $\mu \geq 0$ and $\mu \geq 0$ are the viscosity and resistivity, respectively. Typically, the system is stable in the absence of any external forces and dissipates to zero. Thus, the model is global asymptotic stable. However, following [36], we consider the in-viscid case by setting $\nu = \mu = 0$. As a result, the system is



(a) Using atrMI. (b) Using RK4-SINDy.

Figure 9: MHD model: A comparison of the time-domain simulations of the learned models for test initial conditions. In both plots, the training data (blue, upper triangle), and testing data (red, lower triangle) are shown, which are indeed very different. Figures (a) and (b) demonstrate the performance of the discovered model using `atrMI` (yellow) and `RK4-SINDy` (black) for both training and testing data.

Hamiltonian and energy-preserving. Before we proceed further, we highlight that when the quadratic term of the model (35) is noted down in the form of $\mathbf{H}(\mathbf{x} \otimes \mathbf{x})$, then it can have the form given in (14) with $\mathbf{Q} = \mathbf{I}$. It fits into our hypothesis on the structure of $\mathbf{H}$ in (2).

Although Section 7 discusses trapping region in a strict sense, it also covers energy-preserving systems, where the energy $\frac{1}{2}(\mathbf{x} - \mathbf{m})^\top \mathbf{Q}(\mathbf{x} - \mathbf{m})$ is preserved when $\mathbf{A}_s = 0$. This implies that $\mathbf{R} = 0$ in Section 5; thus, we assume it to be zero, hence can be removed from the optimization problem (`atrMI`). With this setting, we aim to infer an energy-preserving model with sparse regression and present a comparison with `RK4-SINDy` in Figure 9 for a testing condition, which is different than the initial condition used to generate training data. We notice that the proposed methodology with parameterization (`atrMI`) is able to infer a model that preserves energy. In contrast, without any stability or energy-preserving enforcement, the inferred model using `RK4-SINDy` fails to capture the underlying Hamiltonian dynamics accurately in the test phase, although it has a good fit for the training data (see Figure 9 (b), upper triangle).

8. Conclusions

This paper has discussed data-driven inference of quadratic systems that guarantee stability by design. We have begun by investigating the local and global asymptotic stability of quadratic systems using a generalized Lyapunov quadratic function. We have utilized the concept of energy-preserving non-linearity for global asymptotic stability. Based on these discussions, we have proposed suitable parameterizations, allowing us to learn quadratic systems that ensure local and global asymptotic stability. We have achieved these goals without using any matrix inequality-based constraint or based on eigenvalues of the matrix but by using dedicated parameterizations of operators. Additionally, we have addressed the problem of inferring quadratic systems with no stable equilibrium points. For this, we have discussed a concept of a global attractive trapping region, followed by proposing an appropriate parameterization. Moreover, to avoid using derivative information while inferring the operators, we have blended a numerical integration scheme that is robust to noise and scarce data. Furthermore, since the involved optimization problems for inference do not have a closed-form solution, we have utilized a gradient-based approach, and the stability of the model is preserved at each iteration due to the stability-preserving dedicated parameterization. Several numerical examples have been presented, demonstrating its effectiveness in inferring low-dimensional operators, discovering governing equations, and energy-preserving modeling. In our future work, we plan to extend our frameworks for learning stable parametric operators and a detailed investigation of how the methodologies can be used in the case of highly noisy scenarios. Overall, this paper provides a valuable contribution to the field of quadratic system inference, offering a promising framework for developing stable quadratic models using data by appropriate parameterization.

References

[1] I. Mezić, Analysis of fluid flows via spectral properties of the Koopman operator, Annu. Rev. Fluid Mech. 45 (2013) 357–378.

[2] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28. doi:10.1017/S0022112010001217.

[3] M. O. Williams, I. G. Kevrekidis, C. W. Rowley, A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition, J. Nonlinear Science 25 (6) (2015) 1307–1346. doi:10.1007/s00332-015-9258-5.

[4] J. L. Proctor, S. L. Brunton, J. N. Kutz, Dynamic mode decomposition with control, SIAM J. Appl. Dyn. Syst. 15 (1) (2016) 142–161.

[5] J. L. Proctor, S. L. Brunton, J. N. Kutz, Generalizing Koopman theory to allow for inputs and control, SIAM J. Appl. Dyn. Syst. 17 (1) (2018) 909–930. doi:10.1137/16M1062296.

[6] M. O. Williams, C. W. Rowley, I. G. Kevrekidis, A kernel-based method for data-driven Koopman spectral analysis, J. Comput. Dyn. 2 (2) (2015) 247–265.

[7] S. Le Clainche, J. M. Vega, Higher order dynamic mode decomposition, SIAM J. Appl. Dyn. Syst. 16 (2) (2017) 882–925.

[8] I. V. Gosea, I. Pontes Duff, Toward fitting structured nonlinear systems by means of dynamic mode decomposition, in: Model Reduction of Complex Dynamical Systems, Vol. 171 of International Series of Numerical Mathematics, Birkhäuser, Cham, 2021, pp. 53–74. doi:10.1007/978-3-030-72983-7_3.

[9] S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Nat. Acad. Sci. U.S.A. 113 (15) (2016) 3932–3937.

[10] J.-N. Juang, R. S. Pappa, An eigensystem realization algorithm for modal parameter identification and model reduction, J. Guidance, Control, and Dynamics 8 (5) (1985) 620–627.

[11] P. Van Overschee, B. De Moor, N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems, Automatica 30 (1) (1994) 75–93.

[12] M. Viberg, Subspace-based methods for the identification of linear time-invariant systems, Automatica 31 (12) (1995) 1835–1851.

[13] B. Peherstorfer, K. Willcox, Data-driven operator inference for nonintrusive projection-based model reduction, Comp. Meth. Appl. Mech. Eng. 306 (2016) 196–215. doi:10.1016/j.cma.2016.03.025.

[14] E. Qian, B. Kramer, B. Peherstorfer, K. Willcox, Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems., Physica D: Nonlinear Phenomena 406 (2020) 132401. doi:10.1016/j.physd.2020.132401.

[15] P. Benner, P. Goyal, B. Kramer, B. Peherstorfer, K. Willcox, Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms, Comp. Meth. Appl. Mech. Eng. 372 (2020) 113433. doi:10.1016/j.cma.2020.113433.

[16] H. Sharma, B. Kramer, Preserving lagrangian structure in data-driven reduced-order modeling of large-scale mechanical systems, arXiv preprint arXiv:2203.06361 (2022).

[17] Y. Filanova, I. Pontes Duff, P. Goyal, P. Benner, An operator inference oriented approach for mechanical systems, arXiv preprint arXiv:2210.07710 (2022).

[18] P. Benner, P. Goyal, J. Heiland, I. Pontes Duff, Operator inference and physics-informed learning of low-dimensional models for incompressible flows, Electron. Trans. Numer. Anal. 56 (2022). doi:10.1553/etna_vol56s28.

[19] S. Yıldız, P. Goyal, P. Benner, B. Karasözen, Learning reduced-order dynamics for parametrized shallow water equations from data, Internat. J. Numer. Methods in Fluids 93 (8) (2021) 2803–2821. doi:10.1002/fld.4998.

[20] S. A. McQuarrie, P. Khodabakhshi, K. E. Willcox, Non-intrusive reduced-order models for parametric partial differential equations via data-driven operator inference, arXiv preprint arXiv:2110.07653 (2021).

[21] M. A. Savageau, E. O. Voit, Recasting nonlinear differential equations as S-systems: a canonical nonlinear form, Mathematical Biosciences 87 (1) (1987) 83–115.

[22] C. Gu, QLMOR: A new projection-based approach for nonlinear model order reduction, in: In Proc. International Conference on Computer-Aided Design, 2009, pp. 389–396.

[23] P. Benner, T. Breiten, Two-sided projection methods for nonlinear model order reduction, SIAM J. Sci. Comput. 37 (2) (2015) B239–B260. doi:10.1137/14097255X.

[24] N. Sawant, B. Kramer, B. Peherstorfer, Physics-informed regularization and structure preservation for learning stable reduced models from data with operator inference, Comp. Meth. Appl. Mech. Eng. 404 (2023) 115836.

[25] N. Gillis, P. Sharma, On computing the distance to stability for matrices using linear dissipative Hamiltonian systems, Automatica 85 (2017) 113–121.

[26] P. Goyal, I. Pontes Duff, P. Benner, Stability-gauranteed learning of linear models, arXiv preprint arXiv:2301.10060 (2023).

[27] M. Schlegel, B. R. Noack, On long-term boundedness of Galerkin models, J.Fluid Mechanics 765 (2015) 325–352.

[28] P. Holmes, J. L. Lumley, G. Berkooz, C. W. Rowley, Turbulence, Coherent Structures, Dynamical Systems and Symmetry, Cambridge university press, 2012.

[29] W. D. McComb, The Physics of Fluid Turbulence, Oxford, 1990.

[30] B. Rummler, A. Noske, Direct Galerkin approximation of plane-parallel-Couette and channel flows by stokes eigenfunctions, Notes on Numer. Fluid Mechanics 64 (1998) 3–19.

[31] H. Schlichting, K. Gersten, Boundary-layer Theory, Springer, 2016.

[32] J. P. Freidberg, ideal MHD, Cambridge University Press, 2014.

[33] S. Galtier, Introduction to Modern Magnetohydrodynamics, Cambridge University Press, 2016.

[34] A. A. Kaptanoglu, T. E. Benedett, K. D. Morgan, C. J. Hansen, T. R. Jarboe, Two-temperature effects in Hall-MHD simulations of the HIT-SI experiment, Physics of Plasmas 27 (7) (2020) 072505.

[35] A. A. Kaptanoglu, K. D. Morgan, C. J. Hansen, S. L. Brunton, The structure of global conservation laws in Galerkin plasma models, arXiv preprint arXiv:2101.03436 (2021).

[36] A. A. Kaptanoglu, J. L. Callaham, A. Aravkin, C. J. Hansen, S. L. Brunton, Promoting global stability in data-driven models of quadratic nonlinear dynamics, Physical Review Fluids 6 (9) (2021) 094401.

[37] C. Gu, QLMOR: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems, IEEE Trans. Comput. Aided Des. Integr. Circuits. Syst. 30 (9) (2011) 1307–1320. doi:10.1109/TCAD.2011.2142184.

[38] P. Benner, P. Goyal, Balanced truncation model order reduction for quadratic-bilinear systems, e-prints 1705.00160, arXiv (2017).

[39] P. Benner, P. Goyal, S. Gugercin, $\mathcal{H}_2$-quasi-optimal model order reduction for quadratic-bilinear control systems, SIAM J. Matrix Anal. Appl. 39 (2) (2018) 983–1032. doi:10.1137/16M1098280.

[40] B. Kramer, K. E. Willcox, Nonlinear model order reduction via lifting transformations and proper orthogonal decomposition, AIAA Journal 57 (6) (2019) 2297–2307. doi:10.2514/1.J057791.

[41] S. N. Kumpati, P. Kannan, Identification and control of dynamical systems using neural networks, IEEE Trans. Neural Networks 1 (1) (1990) 4–27.

[42] R. T. Chen, Y. Rubanova, J. Bettencourt, D. K. Duvenaud, Neural ordinary differential equations, Adv. Neural Inform. Processing Sys. 31 (2018).

[43] P. Goyal, P. Benner, Discovery of nonlinear dynamical systems using a Runge-Kutta inspired dictionary-based sparse regression approach, Proc. Royal Society A: Mathematical, Physical and Engineering Sciences 478 (2262) (2022) 20210883. doi:10.1098/rspa.2021.0883.

[44] W. I. T. Uy, D. Hartmann, B. Peherstorfer, Operator inference with roll outs for learning reduced models from scarce and low-quality data, arXiv preprint arXiv:2212.01418 (2022).

[45] S. A. McQuarrie, C. Huang, K. E. Willcox, Data-driven reduced-order models via regularised operator inference for a single-injector combustion process, J. Royal Society of New Zealand 51 (2) (2021) 194–211.

[46] B. Kramer, Stability domains for quadratic-bilinear reduced-order models, SIAM J. Appl. Dyn. Syst. 20 (2) (2021) 981–996.

[47] E. N. Lorenz, Deterministic nonperiodic flow, J. Atmospheric Sciences 20 (2) (1963) 130–141.

[48] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, SIAM Rev. 51 (3) (2009) 455–500.

[49] P. Swinnerton-Dyer, A note on Liapunov's method, Dynamics and Stability of Systems 15 (1) (2000) 3–10. doi:10.1080/026811100281884.

[50] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[51] D. A. Messenger, D. M. Bortz, Weak SINDy: Galerkin-based data-driven model selection, Multiscale Modeling & Simulation 19 (3) (2021) 1474–1497.

[52] V. Carbone, P. Veltri, Relaxation processes in magnetohydrodynamics: a triad – interaction model, Astron. and Astrophy. 259 (1992) 359–372.

A. Appendix

In the following, we show that equations (10) and (11) are equivalent.

Theorem 2. *The following statements are equivalent:*

(a) The $\mathbf{H}$ *matrix, in* (2)*, satisfies*

$$\mathbf{H}_{ijk} + \mathbf{H}_{ikj} + \mathbf{H}_{jik} + \mathbf{H}_{jki} + \mathbf{H}_{kij} + \mathbf{H}_{kji} = 0, \tag{36}$$

where $\{i, j, k\} \in \{1, \dots n\}$ *and* $\mathbf{H}_{ijk} := \mathbf{e}_i^\top \mathbf{H}(\mathbf{e}_j \otimes \mathbf{e}_k)$.

(b) The $\mathbf{H}$ *matrix, in* (2)*, satisfies*

$$\mathbf{x}^\top \mathbf{H}(\mathbf{x} \otimes \mathbf{x}) = 0, \quad \text{for every } \mathbf{x} \in \mathbb{R}^n. \tag{37}$$

Proof. $(a) \Rightarrow (b)$: For a given $\mathbf{x} \in \mathbb{R}^n$, let us write it as $\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e}_i$, where $x_i \in \mathbb{R}$ are the coordinates of $\mathbf{x}$ in the basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Using this notation, the left-hand side of equation (37) becomes

$$\begin{aligned}\mathbf{x}^\top \mathbf{H}(\mathbf{x} \otimes \mathbf{x}) &= \left(\sum_{i=1}^{n} x_i \mathbf{e}_i^\top\right) \mathbf{H} \left(\left(\sum_{j=1}^{n} x_j \mathbf{e}_j\right) \otimes \left(\sum_{k=1}^{n} x_k \mathbf{e}_k\right)\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} x_i x_j x_k \mathbf{e}_i^\top \mathbf{H}\left(\mathbf{e}_j \otimes \mathbf{e}_k\right)\end{aligned}$$

Note that the product $x_i x_j x_k$ gives the same value if we permute the indexes in (i, j, k). Hence, the above sum can be rearranged as follows:

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} x_i x_j x_k \mathbf{e}_i^\top \mathbf{H}\left(\mathbf{e}_j \otimes \mathbf{e}_k\right) = \\ &\qquad = \sum_{i=1}^{n} x_i^3 \left(\mathbf{H}_{iii}\right) + \frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} x_i x_j^2 \left(\mathbf{H}_{ijj} + \mathbf{H}_{jij} + \mathbf{H}_{jji} + \mathbf{H}_{ijj} + \mathbf{H}_{jij} + \mathbf{H}_{jji}\right) \\ &\qquad\quad + \frac{1}{6}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n} x_i x_j x_k \left(\mathbf{H}_{ijk} + \mathbf{H}_{ikj} + \mathbf{H}_{jik} + \mathbf{H}_{jki} + \mathbf{H}_{kij} + \mathbf{H}_{kji}\right),\end{aligned} \tag{38}$$

with $\mathbf{H}_{ijk} := \mathbf{e}_i^\top \mathbf{H}(\mathbf{e}_j \otimes \mathbf{e}_k)$. Using (36), one can note the following:

$$\begin{aligned}\mathbf{H}_{iii} &= 0 \\ \mathbf{H}_{ijj} + \mathbf{H}_{jij} + \mathbf{H}_{jji} + \mathbf{H}_{ijj} + \mathbf{H}_{jij} + \mathbf{H}_{jji} &= 0 \\ \mathbf{H}_{ijk} + \mathbf{H}_{ikj} + \mathbf{H}_{jik} + \mathbf{H}_{jki} + \mathbf{H}_{kij} + \mathbf{H}_{kji} &= 0,\end{aligned}$$

As a result, $\mathbf{x}^\top \mathbf{H}(\mathbf{x} \otimes \mathbf{x}) = 0$.

$(b) \Rightarrow (a)$: Using (37) and (38), we have

$$\begin{aligned}&\sum_{i=1}^{n} x_i^3 \left(\mathbf{H}_{iii}\right) + \frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} x_i x_j^2 \left(\mathbf{H}_{ijj} + \mathbf{H}_{jij} + \mathbf{H}_{jji} + \mathbf{H}_{ijj} + \mathbf{H}_{jij} + \mathbf{H}_{jji}\right) \\ &\quad + \frac{1}{6}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{k=1\\ k\neq i\\ k\neq j}}^{n} x_i x_j x_k \left(\mathbf{H}_{ijk} + \mathbf{H}_{ikj} + \mathbf{H}_{jik} + \mathbf{H}_{jki} + \mathbf{H}_{kij} + \mathbf{H}_{kji}\right) = 0.\end{aligned}$$

Since the above equation can be viewed as a multi-dimensional polynomial in $(x_1, \dots, x_n)$ and it is zero, each coefficient must be zero. This leads to the condition (36). □