SPECTRAL TEST FUNCTION SELECTION FOR WEAK-FORM SPARSE SYSTEM IDENTIFICATION UNDER MEASUREMENT NOISE

A Thesis

Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Master of Science

by

Zhiheng Chen

May 2026

ABSTRACT

Learning accurate dynamical models from data is a central challenge in control systems, robotics, and aerospace engineering, where governing equations are needed for prediction, analysis, and controller design. Classical system identification approaches perform well when model structure is known a priori, but struggle when this structural knowledge is unavailable and trajectory data are corrupted by substantial measurement noise. In the presented work, we address the problem of robust learning of unknown dynamic equations from noise-corrupted data.

We build on the Sparse Identification of Nonlinear Dynamics (SINDy) method that selects governing equations by assuming sparsity in a large library of candidate functions and solving a sparse regression problem. Classical SINDy formulations rely on numerical estimates of derivatives, which render the method fragile against measurement noise. Weak-form SINDy mitigates the effect of measurement noise by projecting the equations onto a set of test functions before regression, avoiding numerical differentiation; however, the typically used "bump"-shaped test functions lack physical interpretability, have many hyperparameters that require tuning, and are not mutually orthogonal, therefore introducing redundancy into the regression problem.

To address the above robustness and interpretability challenges in sparse system identification, we introduce Fourier Weak SINDy, a minimal noise-robust and interpretable derivative-free equation learning method that combines weak-form sparse equation learning with spectral density estimation for data-driven test function selection. By using orthogonal sinusoidal test functions inspired by their prevalence in Modulating Function-based system identi-

fication, the weak-form sparse regression problem reduces to a regression over Fourier coefficients. Dominant frequencies are then selected via multitaper estimation of the frequency spectrum of the data. This formulation unifies weak-form learning and spectral estimation within a compact and flexible framework. We illustrate the effectiveness of this approach in numerical experiments across multiple chaotic and hyperchaotic ODE benchmarks.

BIOGRAPHICAL SKETCH

Zhiheng Chen received his Bachelor of Science in Mechanical Engineering from University of Wisconsin-Madison in 2024. He then joined the graduate program at Cornell University to pursue a Master of Science in Mechanical Engineering. During his time at Cornell, his research and academic interest focused on nonlinear dynamical systems and sparse system identification. This focus culminated in his thesis work developing Fourier Weak SINDy, an approach to robust and interpretable learning of dynamical equations from noisy measurement data.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor Professor Anastasia Sergeyevna Bizyaeva for her unwavering support and kindness throughout the two years. Her guidance is the crucial part for my Master's research, which has led to a published paper and this thesis. She consistently provided me with helpful new insights for the problems I got stuck on, and gave me a lot of encouragement and useful advice to move on when I was in a hard time trying to balance research, coursework, and Ph.D. applications.

Throughout my Master's research, Professor Bizyaeva introduced me to a wealth of mathematical tools and guided me toward useful courses to enhance my technical background. Thanks to her generous mentorship, I have significantly improved my foundational mathematical knowledge over the past two years, laying a solid groundwork for my future Ph.D. studies.

I also want to thank my other committee members, Professor Mark Campbell and Professor Samuel Epstein Otto, for their support and valuable feedback on my research. Professor Campbell also taught me important concepts regarding parameter estimation and Bayesian filters in his Model-Based Estimation course, which provided me with new perspectives on the system identification problems in my research.

Last but certainly not least, I want to thank my parents, Xiu Wang and Guanjie Chen. I would not be where I am today without their selfless support and love across continents. They have always been a constant source of strength, and are always there to lift my spirits whenever I feel confused, overwhelmed, or exhausted.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Equation learning for dynamical systems is the process of identifying governing differential equations of a system given measurements of its trajectory data. In recent years, equation learning has been extensively applied in scientific machine learning for physics [20, 22], biology [24, 23], and various other fields [21, 44]. Specific applications of equation learning include time-series predictions for disease models [17], data-driven model predictive control [11], and distilling structures and key relations in biological systems [25]. Among these equation learning methods, sparse system identification, which selects key terms from a library, presents a more interpretable framework compared to black-box methods such as methods based on standard neural networks [29].

In this thesis, we present an interpretable framework for noise-robust derivative-free equation learning that combines recent advances in sparse system identification with classical signal processing tools to address the challenge of data-driven test function selection in the algorithm design. In recent years, sparse regression based methods for data-driven modeling have received significant attention. Among these, the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm is a popular data-driven framework for learning the governing equations of Ordinary Differential Equation (ODE) models, that uses sparsity-promoting optimization to identify unknown model coefficients while avoiding overfitting [6]. SINDy has been successfully applied to equation learning problems in different domains from fluid dynamics to disease modeling [13, 10, 17], and has found many extensions including learning partial differential equations [36], simultaneously learning coordinates and differen-

tial equations [8, 4], and quantifying uncertainty in the learned model structure and coefficients [47, 30, 28, 14].

Measurement noise and limited data are key challenges for learning models from data, motivating extensions of the SINDy framework such as ensembling [12], denoising signals [18], and weak-form reformulations [38, 35, 26, 27, 41]. The weak-form or integral-form family of methods are particularly promising noise-robust variants of SINDy, as they bypass the need for estimation of derivative information from noise corrupted data, which makes the standard SINDy framework fragile. Weak-form methods insteaed rely on integrating data against user-defined analytic smooth test functions and their derivatives. Success of these methods therefore relies on selecting and tuning a good test function basis given characteristics of the collected data, to which significant attention has been devoted in recent work. For example, [26] propose adaptive location and radius selections for compactly-supported "bump"-shaped test functions based on approximate error covariance, the width-at-half-max parameter, and dominant wave modes; [41] propose an optimization framework for test function tuning for weak-form PDE learning; [5] start with compactly supported test functions and orthogonalize them using SVD, with the orthogonalized test functions resembling local Fourier basis or wavelets; [16] provide a detailed heuristic discussion of test function selection based on the frequency content of the data.

In modern weak-form sparse equation learning formulations, "bump"-shaped test functions are a popular design choice due to their compact support and interpretable parameters. However, this choice introduces significant degrees of freedom into the test function selection problem, as the support re-

gion boundaries as well as the shape of the functions become optimization parameters. Furthermore, such test functions are not orthogonal, which means redundant information in the signal is reused in different components of the regression problem. In the work presented in this thesis, we will illustrate that a simple test function basis of orthogonal sinusoids is an interpretable and robust choice for weak-form learning. In this formulation, we take inspiration from classical Modulating Function Method system identification literature [39], in which orthogonal test functions (or modulating functions) have been widely used for system identification [32, 34, 33]. The utility of orthogonal test functions has also recently been noted in the weak SINDy literature with data-driven constructions [5] and derivation of general error bounds [37]. Furthermore, we will show that with sinusoidal test functions, classical spectral density estimation (SDE) presents a natural strategy for data-driven selections of test function frequencies. Common approaches to SDE include nonparametric methods such as Bartlett method and Welch method [42]; a more advanced approach with higher resolutions and lower variances is the multitaper method [43]. Due to its strong frequency specificity, the multitaper method in particular is popular across disciplines for the spectral analysis of complex data, from seismic data analysis to brain wave analysis [31, 46, 3]. We will utilize multitaper to inform the selection of a sinusoidal test function basis.

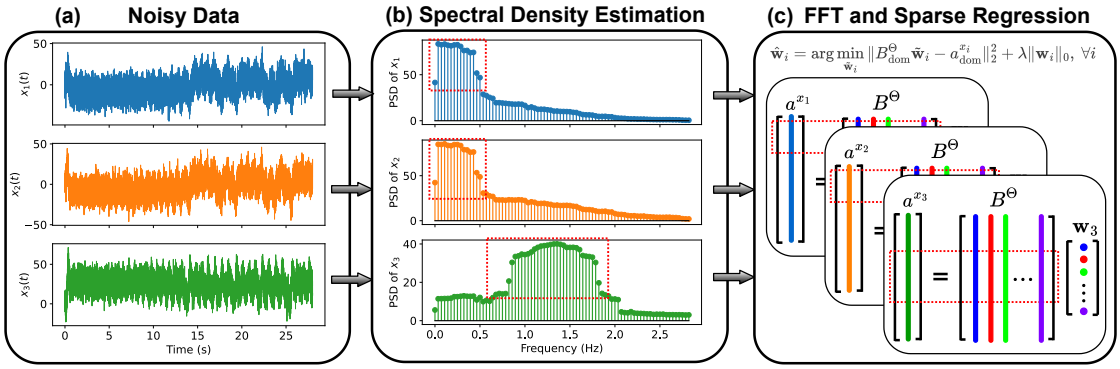

Figure 1.1: Graphical summary of proposed Fourier Weak SINDy method for sparse equation learning described in detail in chapter 3. First, multitaper spectral density estimation is used to identify dominant temporal frequencies in the observed noisy signals. Second, spectral sparse regression problem is solved to learn model coefficients in a selected function dictionary, leveraging the identified dominant Fourier frequencies as test functions in a weak-form system identification problem.

This thesis is adapted from our published work on Fourier Weak SINDy [9]. In this thesis, we present the following contributions. First, we introduce Fourier Weak SINDy, a minimal method that combines weak-form sparse regression with spectral density estimation to learn parsimonious sparse ODE models from noisy measurement data. A high-level graphical summary of the method can be found in Figure 1.1. We show that under the choice of sinusoidal test functions, the Weak SINDy sparse regression problem simplifies to a regression over Fourier series coefficients, which are efficiently computable with the Fast Fourier Transform (FFT). The choice of coefficients to regress over is made in a data-driven way by selecting dominant frequency modes in the data using multitaper estimation or Fourier phase index. Second, we illustrate effectiveness of this minimal approach in simulation studies. In particular, we show numerical evidence for improvements of Fourier Weak SINDy

over baseline SINDy and Weak SINDy methods across different chaotic and hyperchaotic benchmark systems. It is important to note that the Weak SINDy we compare against are from the PySINDy package [19]; we leave principled comparisons against Weak SINDy with adaptive hyperparameter tuning (for example the framework presented in [26]) for future work. Finally, we investigate the robustness of Fourier Weak SINDy against hyperparameter selections such as the number of test function frequencies and the bandwidth of taper functions. Through these numerical experiments, we show the improved robustness of Fourier Weak SINDy against measurement noise and hyperparameter selections for learning the governing equations of low-dimensional ODE benchmarks with reasonably small dictionary function library.

This thesis is structured as follows. Chapter 2 reviews least squares and sparse regression, signal processing techniques used in this study, and the fundamental formulations and limitations of current SINDy and weak SINDy frameworks; Chapter 3 presents our Fourier weak SINDy method, with mathematical derivations for the frequency-domain representation and the frequency selections based on spectral density estimations; Chapter 4 shows numerical experiment results, where Fourier weak SINDy is compared with standard SINDy and weak SINDy on equation learning tasks of four benchmark systems; Chapter 5 discusses limitations and future work.

CHAPTER 2

BACKGROUND

This section discusses the prerequisite knowledge for the core Fourier weak SINDy method. The background knowledge includes a) least squares problems and sparse regression which are the core of all sparse system identification algorithms presented in this thesis, b) the classic Sparse Identification of Nonlinear Dynamics (SINDy) and its limitations, c) bump-function weak-form SINDy which attempts to resolve some of SINDy's limitations, and d) Fourier analysis and signal processing techniques used in the Fourier weak SINDy method.

2.1 Notations

In this thesis, we use $\mathbb{R}$ to denote real numbers, $\mathbb{C}$ to denote complex numbers, $\mathbb{Z}$ to denote integers, and $\mathbb{F}$ to denote a general field. We use $\mathbf{i}$ to denote the imaginary unit, and thus a complex number can be written as $x \in \mathbb{C} = a + \mathbf{i}b = re^{\mathbf{i}\phi}$. We represent matrices by capital letters (e.g., $A \in \mathbb{R}^{m\times n}$ is an m by n matrix, and $A^T \in \mathbb{R}^{n\times m}$ is its transpose) and vectors by boldfaced symbols (e.g., $\mathbf{v} \in \mathbb{R}^n$). We represent the space of all continuous-time signals as $\mathbb{F}^{\mathbb{R}}$.

We denote the 2-norm of a vector $\mathbf{v} \in \mathbb{R}^n$ as $\|\mathbf{v}\|_2 = (\sum_{i=1}^n v_i^2)^{1/2}$, where v_i is the i-th component of $\mathbf{v}$. For a matrix $A \in \mathbb{R}^{m\times n}$, we denote its 2-norm (i.e., largest singlar value) as $\|A\|_2$ and its Frobenius norm as $\|A\|_F = (\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2)^{1/2}$, where a_{ij} is the entry on the i-th row and j-th column of A.

2.2 Least Squares and Sparse Regression

The core of sparse system identification is the formulation of least squares problems and sparsity-promoting algorithms used to solve them. Suppose $A \in \mathbb{R}^{m \times n}$ is a known matrix, and $\mathbf{b} \in \mathbb{R}^m$ is a known vector. Then the least squares approximate of an unknown $\mathbf{x} \in \mathbb{R}^n$ is

$$\hat{\mathbf{x}}^{\text{LS}} = \arg\min_{\mathbf{x}} ||A\mathbf{x} - \mathbf{b}||_2^2. \tag{2.1}$$

The $\mathbf{x}$ that minimizes $||A\mathbf{x} - \mathbf{b}||_2$ corresponds to the projection of $\mathbf{b}$ onto range(A), and thus $\mathbf{b} - A\mathbf{x}$ need to be orthogonal to range(A). That is, $A^T(\mathbf{b} - A\mathbf{x}) = 0$. This then gives the normal equation

$$A^T A\mathbf{x} = A^T \mathbf{b}, \tag{2.2}$$

and $\hat{\mathbf{x}}^{\text{LS}}$ can be obtained by solving (2.2) for $\mathbf{x}$.

Adding an ℓ^2-regularizatin term to the cost function prevents the entries of the learned $\hat{\mathbf{x}}$ from getting excessively large. This has two major benifits: a) the regulariztion improves numerical stability when solving the optimization problem and b) adding a small bias reduces overfitting since the model is less likely to fit the noise. Therefore, the ridge regression is extensively used in this thesis:

$$\hat{\mathbf{x}}^{\text{ridge}} = \arg\min_{\mathbf{x}} ||A\mathbf{x} - \mathbf{b}||_2^2 + \beta^2 ||\mathbf{x}||_2^2 \tag{2.3}$$

where $\beta \in \mathbb{R}$. Since

$$||A\mathbf{x} - \mathbf{b}||_2^2 + \beta^2 ||\mathbf{x}||_2^2 = \left\| \begin{bmatrix} A \\ \beta \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix} \right\|_2^2, \tag{2.4}$$

the ridge regression problem can be reduced to a least squares problem, and $\hat{\mathbf{x}}^{\text{ridge}}$ can be obtained by solving

$$(A^T A + \beta^2 \mathbf{I})\mathbf{x} = A^T \mathbf{b} \tag{2.5}$$

for $\mathbf{x}$.

If $\mathbf{x}$ is inherently sparse, then sparse regression algorithms are useful to identify the sparse structures. In this thesis, the sequentially thresholded least squares (STLS) and the sequentially thresholded ridge regressions (STRR) are extensively used to perform sparse regressions. The STRR algorithm that performs ridge regressions with sequential hard-thresholding is shown in Alg. 1 [36]. The STLS algorithm is presented in [6].

Algorithm 1: Sequentially Thresholded Ridge Regression

Require: $A \in \mathbb{R}^{m \times n},\ \mathbf{b} \in \mathbb{R}^n,\ \beta \in \mathbb{R},\ \lambda \in \mathbb{R}^+$

$\hat{\mathbf{x}} \leftarrow \arg\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \beta^2 \|\mathbf{x}\|_2^2$ ▷ Full ridge regression

while not converged **do**

smallInds $\leftarrow \{i :\ |\hat{x}[i]| < \lambda\}$ ▷ Find indices of small entries in $\hat{\mathbf{x}}$

$\hat{\mathbf{x}}$[smallInds] $\leftarrow 0$ ▷ Hard threshold

$A \leftarrow A[:, \sim \text{smallInds}]$ ▷ Extract columns of A for large entries

$\hat{\mathbf{x}}[\sim \text{smallInds}] \leftarrow \arg\min_{\mathbf{x}} \|A\mathbf{x} - b\|_2^2 + \beta^2 \|\mathbf{x}\|_2^2$ ▷ Ridge regression on remaining columns

end while

2.3 Sparse Identification of Nonlinear Dynamics (SINDy)

Consider the problem of learning the structure of the governing equations for a finite-dimensional autonomous dynamical system given noisy observations of its state,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \qquad \mathbf{y}(t) = \mathbf{x}(t) + \boldsymbol{\xi}(t) \tag{2.6}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$, $\dot{\mathbf{x}}$ is the time derivative, $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$, and $\boldsymbol{\xi}(t) \in \mathbb{R}^n$ is a vector of additive measurement noise with components $\xi_i(t) \sim \mathcal{N}(0, \sigma^2)$ independent across state dimensions and time. We assume the output is sampled at regular time intervals $t \in \{t_1, \ldots, t_k\}$, $t_{i+1} - t_i = \Delta t$, with the collected data snapshots summarized in the matrix

$$Y = \begin{bmatrix} \mathbf{y}(t_1) & \ldots & \mathbf{y}(t_k) \end{bmatrix}^T \in \mathbb{R}^{k \times n}. \tag{2.7}$$

The Sparse Identification of Nonlinear Dynamics (SINDy) method for equation learning has received significant interdisciplinary attention [6, 7]. SINDy combines classical least-squares system identification [2, 45] with a sparsity-promoting sequentially thresholded least squares solver to identify the coefficients of nonlinear terms in the function $\mathbf{f}(\mathbf{x})$ in (2.6) from a large dictionary of candidate functions, under the assumption that $\mathbf{f}(\mathbf{x})$ is linear-in-coefficients and sparse in the selected dictionary.

To set up the sparse identification problem, define m dictionary functions $\{\theta_1, \ldots, \theta_m\}$ where each $\theta_i : \mathbb{R}^n \to \mathbb{R}$. Common dictionary choices include polynomials up to degree d, i.e. $\{\theta_j(\mathbf{x})\}_{j=1}^m = \{1, x_1, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_n^d\}$, or other function bases such as trigonometric functions. It is assumed that each component

$f_i(\mathbf{x})$ of the vector field $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_n(\mathbf{x})]^T$ admits a representation

$$f_i(\mathbf{x}(t)) = \sum_{j=1}^{m} w_{ij}\theta_j(\mathbf{x}(t)) = \Theta(\mathbf{x}(t))\mathbf{w}_i \tag{2.8}$$

where $\mathbf{w}_i \in \mathbb{R}^m$ is a coefficient vector and $\Theta(\mathbf{x}(t)) = \begin{bmatrix} \theta_1(\mathbf{x}(t)) & \ldots & \theta_m(\mathbf{x}(t)) \end{bmatrix} \in \mathbb{R}^{1\times m}$. The true coefficient matrix $W = \begin{bmatrix} \mathbf{w}_1 & \ldots & \mathbf{w}_n \end{bmatrix} \in \mathbb{R}^{m\times n}$ parametrizes the right hand side of (2.6) with respect to the function dictionary as $\mathbf{f}(\mathbf{x}(t)) = \Theta(\mathbf{x}(t))W$. The objective is to recover the unknown coefficients W from the observed snapshots (2.7) under the assumption that each $\mathbf{w}_i$ is sparse, i.e. each function component $f_i(\mathbf{x}(t))$ comprises only a small number of dictionary terms.

Given time series Y of state measurements defined in (2.7), an estimate of the corresponding time derivatives of the state $\dot{X} = \begin{bmatrix} \dot{\mathbf{x}}(t_1) & \ldots & \dot{\mathbf{x}}(t_k) \end{bmatrix}^T \in \mathbb{R}^{k\times n}$ is constructed, for example using finite-difference approximations or polynomial smoothing. By stacking evaluations of the dictionary functions at each snapshot, we can construct the dictionary matrix

$$\Theta(Y) = \begin{bmatrix} \Theta(\mathbf{y}(t_1) \\ \vdots \\ \Theta(\mathbf{y}(t_k)) \end{bmatrix} = \begin{bmatrix} \theta_1(\mathbf{y}(t_1)) & \ldots & \theta_m(\mathbf{y}(t_1) \\ \vdots & \ddots & \vdots \\ \theta_1(\mathbf{y}(t_k)) & \ldots & \theta_m(\mathbf{y}(t_k)) \end{bmatrix} \in \mathbb{R}^{k\times m}. \tag{2.9}$$

Approximation of the unknown coefficient matrix W can be formulated as the least squares problem

$$\widehat{W} = \arg\min_{\tilde{W}} \left\| \dot{X} - \Theta(Y)\tilde{W} \right\|_2^2. \tag{2.10}$$

The regression problem (2.10) is then solved using a sparsity promoting method such as sequentially thresholded least squares or sequentially thresholded ridge regression as presented in the previous section.

2.4 Weak SINDy

While SINDy is an efficient and interpretable framework for model discovery, its reliance on derivative estimates $\dot{X}$ from noisy measurements limits its robustness. Specifically, the accuracy of the finite difference methods implemented to approximate derivatives are easily corrupted by measurement noises, as the noise can be signifiantly amplified when we divide a difference by a very small time step. In practice, SINDy coefficient estimates become highly inaccurate even with reasonably large signal-to-noise ratios in the data matrix (2.7).

To mitigate this noise sensitivity, a number of recent research efforts proposed integral-form or weak-form reformulations of SINDy, which bypass the need for estimating derivatives [38, 16, 35, 26, 5, 41]. In the Weak SINDy formulation, each component $f_i(\mathbf{x})$ of (2.6) is projected onto a compactly-supported test function $\phi_j \in C_c^1([t_a, t_b])$ and integrated over its support region, $\int_{t_a}^{t_b} \dot{x}_i(t)\phi_j(t)dt = \int_{t_a}^{t_b} f_i(\mathbf{x}(t))\phi_j(t)dt$. Imposing that test functions vanish at the end points of the time interval, i.e. $\phi_j(t_a) = \phi_j(t_b) = 0$, and applying integration-by-parts yields

$$-\int_{t_a}^{t_b} x_i(t)\phi_j'(t)dt = \int_{t_a}^{t_b} f_i(\mathbf{x}(t))\phi_j(t)dt. \tag{2.11}$$

In discrete time, the integrals in (2.11) are approximated using numerical quadrature. The matrices of evaluations of $\phi_j(t)$ and its derivative $\phi_j'(t)$ at the sampled time points are defined as $\Phi_j = \begin{bmatrix}\phi_j(t_1) & \dots \phi_j(t_k)\end{bmatrix} \in \mathbb{R}^{1\times k}$ and $\Phi_j' = \begin{bmatrix}\phi_j'(t_1) & \dots \phi_j'(t_k)\end{bmatrix} \in \mathbb{R}^{1\times k}$. Defining $X \in \mathbb{R}^{k\times n}$ as the data snapshot matrix and leveraging the library matrix expansion (2.8) along with the dictionary matrix definition (2.9), the integral expression (2.11) becomes $-\Delta t\Phi_j' X = \Delta t\Phi_j\Theta(X)W$. Collecting evaluations of p distinct test functions into the matrices $\Phi \in \mathbb{R}^{p\times k}$ and $\Phi' \in \mathbb{R}^{p\times k}$, and incorporating noisy data observations (2.7), coefficients W of the

ODE are then estimated by solving the least squares problem

$$\widehat{W} = \arg\min_{\tilde{W}} \left\| \Delta t \Phi \Theta(Y) \tilde{W} + \Delta t \Phi' Y \right\|_2^2 . \tag{2.12}$$

As in the standard SINDy formulation, coefficient sparsity is enforced by solving (2.12) using least squares or ridge regression with sequential thresholding.

A key design choice that significantly affects the performance of weak-form methods is the form of test function $\phi_j(t)$. Grounded in the definition of a weak derivative, a common constraint on the choice of ϕ_j in recent literature is that it is compactly supported over the observation interval. Thus, typical test functions have the form of a localized "bump", for example [35] utilize functions of the form

$$\phi_j(t) = (1 - \underline{t}^2)^q \tag{2.13}$$

where $q \in \mathbb{Z}^+$ is a hyperparameter (larger q gives narrower bumps around the subdomain center), and $\underline{t}$ is the dimensionless time obtained via $\underline{t} = (t - t_j)/H_t$ with t_j being the center of the j-th subdomain and H_t being the duration of each subdomain. Optimizing the performance of weak SINDy then requires strategic selections of the sharpness and support interval boundaries of the test functions. For example, [26, 5] present an adaptive-grid strategy that places more test functions at locations where large changes in the dynamics occur, where the gradient is approximated by formulating and solving a nonlinear root finding problem; then they select hyperparameters and support intervals based on the approximate error covariance, the width-at-half-max parameter, and optionally the dominant wave modes.

2.5 Fourier Analysis and Power Spectral Density

As discussed in the previous section, selecting the optimal hyperparameters for the bump test functions is generally a subtle, not interpretable, and computationally expensive process that involves formulating and solving sophisticated numerical optimization problems. The resulting test functions are also not orthogonal and contain repeated information, usually requiring orthogonalizations via the Gram-Schmidt process or the Singular Value Decomposition (SVD). In contrast to bump functions, sinusoids are naturally orthogonal to one another; more importantly, we can implement well-developed signal processing tools, such as Fourier analysis and spectral density estimation, to find the components of a given trajectory signal along a sinusoid, which implies that selecting sinusoids with frequencies that closely match the dominant frequencies of the system can make the weak-form projections more effective. Thus, utilizing sinusoidal test functions can enable us to perform hyperparamter selections in a more interpretable manner.

For a deterministic continuous-time signal $x \in \mathbb{F}^{\mathbb{R}}$, its Fourier transform, if exists, is

$$\hat{X}(\Omega) = \int_{-\infty}^{\infty} x(t) e^{-\mathbf{i}\Omega t} dt \tag{2.14}$$

where $\mathbf{i}$ is the imaginary unit and $\Omega \in \mathbb{R}$. The Fourier transform creates a frequency-domain representation of the, with its value at Ω representing the component of the signal along the complex sinusoid $e^{-\mathbf{i}\Omega t}$. The inverse Fourier transform is

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{X}(\Omega) e^{\mathbf{i}\Omega t} d\Omega. \tag{2.15}$$

Moreover, denote the Fourier transform pair as $x \overset{\mathcal{F}}{\leftrightarrow} \hat{X}$.

This thesis primarily uses two properties of the Fourier transform. The first property is the frequency-domain convolution rule: if $x_1 \overset{\mathcal{F}}{\leftrightarrow} \hat{X}_1$, $x_2 \overset{\mathcal{F}}{\leftrightarrow} \hat{X}_2$, and $\hat{X} = \hat{X}_1 * \hat{X}_2$ exists, then $x_1 x_2 \overset{\mathcal{F}}{\leftrightarrow} 2\pi\hat{X}$. The second property is the definition of power spectral density (PSD). The PSD of a deterministic signal $x \in \mathbb{F}^{\mathbb{R}}$ at a given frequency is defined as

$$S_x(\Omega) = \lim_{T\to\infty} \frac{1}{T} \left| \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) e^{-\mathbf{i}\Omega t} dt \right|^2 . \tag{2.16}$$

That is, the PSD of a signal represents its power at a given frequency. Moreover, the PSD and the autocorrelation function $R_x(\tau)$ is related by the Wiener-Khinchin theorem:

$$S_x(\Omega) = \int_{-\infty}^{\infty} R_x(\tau) e^{-\mathbf{i}\Omega\tau} d\tau. \tag{2.17}$$

Therefore, the PSD of a signals captures the correlations (i.e., linear dependence) within the signal.

2.6 Multitaper Method for Spectral Density Estimation

Although the PSD definition (2.16) is for continuous-time deterministic signals, it can be readily extended to discrete-time random signals that are more common in equation learning problems. In this section, we review classical ideas from spectral analysis of signals [42], which are used in the spectral density estimation component of our method. Given a right-sided random signal sequence $x(n)$ with length N, its PSD $S_x(\omega)$ is defined as

$$S(\omega) = \lim_{N\to\infty} E\left[\frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-\mathbf{i}\omega n} \right|^2 \right] \tag{2.18}$$

where E denotes expectation. Note that $\sum_{n=1}^{N} x(n)e^{-\mathbf{i}\omega n}$ is the discrete-time Fourier transform of the right-sided $x(n)$; therefore, $S(\omega)$ is a measurement of the amount of energy in $x(n)$ at frequency ω, and the function S represents the distribution of $x(n)$'s energy spanning a continuous spectrum of frequencies, which is a direct result from (2.17) by letting $\tau = 0$.

Direct computation of the PSD is not possible since it requires an infinite number of ensembles for $y(n)$, and thus spectral density estimations (SDEs) are needed. A common approach to SDE is via the periodogram, defined as

$$\hat{S}_P(\omega) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-\mathbf{i}\omega n}\right|^2. \tag{2.19}$$

The periodogram by itself cannot provide decent spectral density estimations due to its high variance. To obtain a consistent PSD estimate, the multitaper method averages modified periodograms obtained using a set of orthogonal window functions (tapers); a typical set of tapers is the Discrete Prolate Spheroidal (Slepian) Sequences [40, 43]. Denote the j-th Slepian sequence as g_j, and denote the bandwidth normalized by the sampling rate as W, then g_j can be defined by its convolution with the sinc function:

$$\sum_{m=0}^{N-1} \frac{\sin(2\pi W(n-m))}{\pi(n-m)} g_j(m) = \lambda_k g_j(n), \quad n, j = 0, 1, 2, \ldots, N-1, \tag{2.20}$$

and g_j's can be obtained by solving the eigenvalue problem for the sinc function matrix. Then the j-th modified periodogram is constructed by first tapering the signal with g_j, then computing the squared magnitude of its discrete-time Fourier transform:

$$\hat{S}_j(\omega) = \frac{1}{N}\left|\sum_{n=0}^{N-1} g_j(n)x(n)e^{-\mathbf{i}\omega n}\right|^2 \tag{2.21}$$

There are approximately $2NW$ sequences with eigenvalues close to unity, and thus we choose $M = 2NW$ tapers. Then the PSD estimate from the multitaper method is the average of the M modified periodograms:

$$\hat{S}_{MT}(\omega) = \frac{1}{M} \sum_{j=0}^{M-1} \hat{S}_j(\omega) \tag{2.22}$$

Note that the sinc function is the impulse response of an ideal lowpass filter with passband $[-W, W]$, and thus choosing a larger bandwidth will result in the tapers having a wider frequency content, causing the tapers to have a stronger frequency-averaging effect that can be visualized if we convert the Fourier transform of the time-domain multiplication in (2.21) into frequency-domain convolution. Therefore, if we choose a smaller and more concentrated bandwidth, then the PSD estimate will have more spectral leakage but higher resolution; conversely, if we choose a larger and more distributed bandwidth, then the PSD estimate will have less leakage but lower resolution.

2.7 Fourier Phase Index

As shown in (2.17), the PSD defined using the Fourier amplitude only captures the autocorrelation within a signal, which is the linear part. To evaluate the strength of nonlinearity within a trajectory signal, we also need to investigate its Fourier phases. In this thesis, we implement the Fourier phase index method proposed in [1].

Consider a high-dimensional nonlinear dynamical system with two of its state variables denoted as x_1 and x_2. In our equation learning setup, we can obtain two N-point discrete-time signals for the two state variables. Denote

discrete Fourier transform (DFT) pairs $x_1 \overset{\mathcal{F}}{\leftrightarrow} \hat{X}_1$ and $x_2 \overset{\mathcal{F}}{\leftrightarrow} \hat{X}_2$. Also let $\hat{X}_1(f) = A_1(f)e^{\mathbf{i}\Phi_1(f)}$ and $\hat{X}_2(f) = A_2(f)e^{\mathbf{i}\Phi_2(f)}$ for $f = 0, \frac{1}{N}, \ldots, \frac{N-1}{N}$. Since the $\Phi(f)$ is defined on a unit circle and thus is always within the interval $[0, 2\pi]$, the sequence of the pairs $(\Phi_1(f), \Phi_2(f))$ defines a three-dimensional torus, as shown in Fig. 2.1.

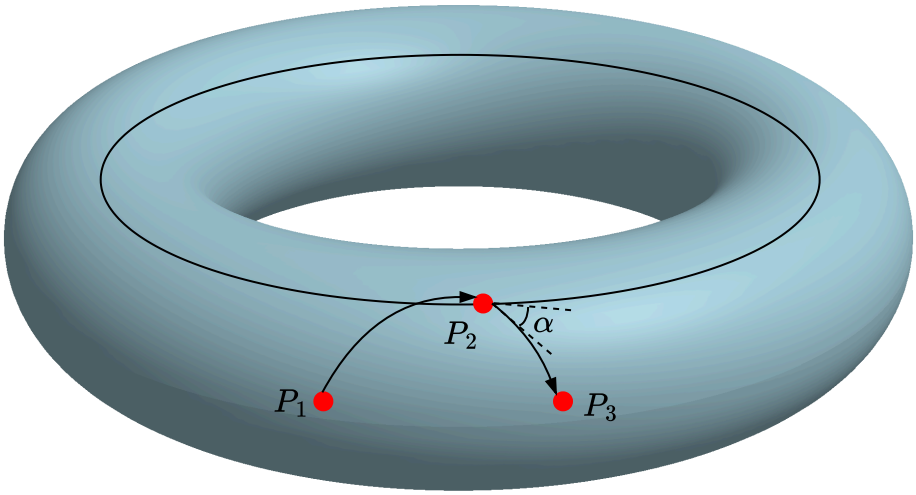

Figure 2.1: The torus defined by the sequence of the Fourier phase pairs $(\Phi_1(f), \Phi_2(f))$, where each point (e.g., P_1, P_2, and P_3) corresponds to a Fourier phase pair at a specific frequency. This schematic is adapted from an image in the published paper on Fourier Phase Index by Aguilar-Hernández et al. [1].

The Fourier phase method focus on the deviation angles as we move across the frequencies along the torus. Although there are multiple alternatives to move from one frequency to another along the torus, we choose the shortest path to maintian consistency on the deviation angle metric. The J-index of the two trajectory signals is then defined as

$$J[\Phi_1, \Phi_2] = 1 - \left| \frac{1}{N-1} \sum_{\ell=1}^{N-1} e^{i\alpha(f_\ell)} \right| \tag{2.23}$$

where $\alpha(f_\ell)$ is the shortest-path deviation angle from frequency f_ℓ to $f_{\ell+1}$ on the torus formed by Φ_1 and Φ_2.

The J-index is computed solely based on the Fourier phase, excluding linear dependence (i.e., autocorrelation) contained in the Fourier amplitude. Therefore, the J-index captures the level of irregularity within the nonlinear part of the deterministic dynamics. A larger J (closer to 1) implies a stronger irregularity, while a smaller J (closer to 0) implies weaker irregularity.

CHAPTER 3

FOURIER WEAK SINDY

The use of orthogonal test functions has been explored in several prior works; for example [5] obtain test functions by orthogonalizing smooth bump functions via SVD. Moreover, in classical systems and control literature, such as [32, 34, 33], Fourier-mode test functions (also known as modulating functions) are used for identification of linear systems, without requiring them to be compactly supported, and only requiring that they should vanish at domain boundaries.

Motivated by these precedents, we present in this section the sinusoidal test functions for the weak SINDy framework and show that the problem can be reformulated into a frequency-domain regression over the Fourier coefficients of trajectory measurements and dictionary functions. This reformulation then converts test function hyperparameter selections into the interpretable process of selecting the most informative Fourier frequencies, which can be addressed by classical signal processing methods. This frequency-domain reformulation offers a robust and interpretable approach to sparse equation learning. This frequency-domain reformulation is also presented in our published paper [9].

3.1 Sinusoidal Test Functions

Suppose the data we collect has duration T (i.e., $T = t_k - t_1$), then define the ℓ-th sinusoidal test function and its derivative, where $\ell \in \mathbb{Z}^+$, as

$$\phi_\ell(t) = \sin\left(\frac{2\pi\ell}{T}t\right), \qquad \phi'_\ell(t) = \frac{2\pi\ell}{T}\cos\left(\frac{2\pi\ell}{T}t\right). \tag{3.1}$$

Since $\phi_\ell(0) = \phi_\ell(T) = 0$ for all $\ell \in \mathbb{Z}^+$, the weak form equation (2.11) becomes

$$-\frac{2\pi\ell}{T}\int_0^T x_i(t)\cos\left(\frac{2\pi\ell}{T}t\right)dt = \int_0^T f_i(\mathbf{x}(t))\sin\left(\frac{2\pi\ell}{T}t\right)dt. \tag{3.2}$$

Let $x_i(t)$ and $f_i(t)$ defined over $[0, T]$ be expanded as Fourier series

$$x_i(t) = a_0^{x_i} + \sum_{\ell=1}^{\infty}\left(a_\ell^{x_i}\cos\left(\frac{2\pi\ell}{T}t\right) + b_\ell^{x_i}\sin\left(\frac{2\pi\ell}{T}t\right)\right), \tag{3.3a}$$

$$f_i(x(t)) = a_0^{f_i} + \sum_{\ell=1}^{\infty}\left(a_\ell^{f_i}\cos\left(\frac{2\pi\ell}{T}t\right) + b_\ell^{f_i}\sin\left(\frac{2\pi\ell}{T}t\right)\right). \tag{3.3b}$$

Since the Fourier basis are mutually orthogonal, substituting (3.3) into (3.2) lets most of the inner products vanish, giving

$$-\frac{2\pi\ell}{T}a_\ell^{x_i} = b_\ell^{f_i}, \quad \text{for all } \ell \in \mathbb{Z}^+. \tag{3.4}$$

Note that the $a_\ell^{x_i}$ and $b_\ell^{f_i}$ coefficients can be efficiently computed using FFT of the data matrix Y and the dictionary matrix $\Theta(Y)$.

3.2 Dictionary Function Expansion

In the equation learning problem, we do not know the exact form of f_i and instead expand it in a dictionary of functions $\Theta(\mathbf{x}) = \begin{bmatrix}\theta_1(\mathbf{x}) & \dots & \theta_m(\mathbf{x})\end{bmatrix}$, with the relation described in (2.8). Let the Fourier series expansion of $\theta_j(\mathbf{x}(t))$ defined over $[0, T]$ be

$$\theta_j(\mathbf{x}(t)) = a_0^{\theta_j} + \sum_{\ell=1}^{\infty}\left(a_\ell^{\theta_j}\cos\left(\frac{2\pi\ell}{T}t\right) + b_\ell^{\theta_j}\sin\left(\frac{2\pi\ell}{T}t\right)\right). \tag{3.5}$$

By linearity of the Fourier operator, the right-hand side of (3.4) can then be written as

$$b_\ell^{f_i} = \sum_{j=1}^{m} w_{ij} b_\ell^{\theta_j}. \tag{3.6}$$

Define $B_\ell^\Theta = \begin{bmatrix} b_\ell^{\theta_1} & \dots & b_\ell^{\theta_m} \end{bmatrix}$; then $b_\ell^{f_i} = B_\ell^\Theta \mathbf{w}_i$, and (3.4) becomes

$$-\frac{2\pi\ell}{T} a_\ell^{x_i} = B_\ell^\Theta \mathbf{w}_i. \tag{3.7}$$

Since for each j, $b_\ell^{\theta_j}$ for all ℓ up to the index for Nyquist frequency can be efficiently computed using one FFT, this formulation can be significantly faster to implement than quadrature-based approaches, where an evaluation of quadrature is needed for each test function.

3.3 Frequency Selection Based on Spectral Density Estimation

An approach to selecting the frequencies of test functions is to choose frequencies that matches the most energetic frequencies of the trajectory signals. To this end, we need to estimate the PSD of the trajectory signals. The multitaper method provides a low-variance, low-leakage, and high-resolution approach to spectral density estimation (SDE) compared to simple single-taper approaches [43]. Therefore, we use the multitaper method to perform SDEs for trajectory data.

For each f_i, we pick frequencies at which x_i has the highest energies (i.e., dominant frequencies) by implementing multitaper SDE to approximate the power spectral density from noisy measurements. This provides the most reliable and effective frequencies of sinusoids for the governing equations to project onto for weak-form system identification. Suppose we pick K dominant frequencies in total, and define

$$a_{\text{dom}}^{x_i} = \begin{bmatrix} -\frac{-2\pi}{T} a_1^{x_i} & \dots & -\frac{-2\pi K}{T} a_K^{x_i} \end{bmatrix}^T, \quad B_{\text{dom}}^\Theta = \begin{bmatrix} (B_1^\Theta)^T & \dots & (B_K^\Theta)^T \end{bmatrix}^T. \tag{3.8}$$

Then we can obtain, and solve using sequentially thresholded methods, the least squares problem

$$\widehat{\mathbf{w}}_i = \arg\min_{\tilde{\mathbf{w}}_i} \left\| B^{\Theta}_{\text{dom}} \tilde{\mathbf{w}}_i - a^{x_i}_{\text{dom}} \right\|_2^2 . \tag{3.9}$$

This approach to test function selection is interpretable in a physically meaningful sense, since the selected frequencies correspond directly to the dominant modes of the observed signal. The choice of test functions is directly grounded in the spectral structure of the data, rather than indirectly through abstract optimization criteria, and utilizes classical signal processing methods.

3.4 Frequency Selection Based on Fourier Phase Index

Although the multitaper PSD estimation is effective in extracting the most energetic frequencies of a trajectory signal, it only captures the linear part of the trajectory of a nonlinear system (as mentioned in Chapter 2). For nonlinear systems, more meaningful information could be contained in features beyond the trajectory signal's most energetic modes.

To determine the frequencies showing the most nonlinear behavior from a trajectory signal, we invoke the Fourier phase index method introduced in 2. Given N-point trajectory time series $x_1, x_2 \in \mathbb{R}^N$, we find their DFT phases, extract frequencies up to the index $N-1$, exclude DC, and thus form $\Phi_1, \Phi_2 \in [0, 2\pi]^{N-1}$. Then we obtain moving slices (windows) of Φ_1 and Φ_2, with each paired window denoted as $(\Phi_1^{(k)}, \Phi_2^{(k)})$, and compute a J_k for each k. This J_k then represents the strength of deterministic chaotic behavior at frequency f_k, where f_k is the frequency at the center of window k. We then pick frequencies whose

corresponding windows have the largest J-indices. A schematic of selecting test function frequencies based on Fourier phase indices is shown in Fig. 3.1.

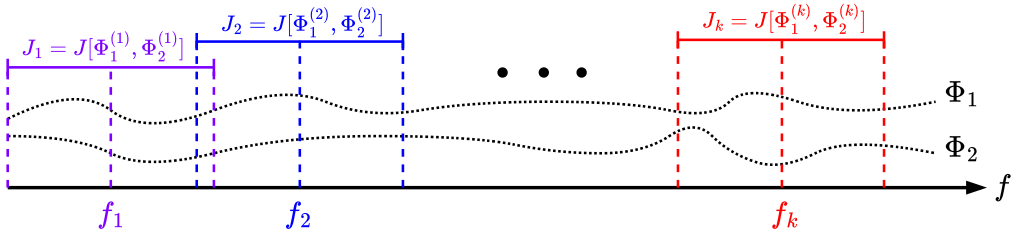

Figure 3.1: Schematic of using Fourier phase index for selecting test function frequencies. The whole Φ_1 and Φ_2 are divided into multiple small windows, and each window has a J-index and a frequency value.

CHAPTER 4

SIMULATION RESULTS ON BENCHMARK SYSTEMS

We implement our Fourier Weak SINDy framework in Python and conducted further numerical experiments based on the results presented in our published work [9], testing it on simulated trajectories of different nonlinear dynamical systems with varying measurement noise levels. The trajectory data have durations of 10 s and are evenly sampled with 1000 samples per second. We use benchmark systems chosen from [15], and follow the noise level definition and error metrics in [26]. Specifically, the noise ratio σ_{NR} is defined using

$$\sigma = \sigma_{NR} \frac{\|X\|_F}{\sqrt{kn}} \tag{4.1}$$

where $\|X\|_F$ is the Frobenius norm of the trajectory data matrix $X \in \mathbb{R}^{k \times n}$, and the noise level is then $\sigma_{NR} \times 100\%$.

We use three metrics to evaluate the error of equation learning results; the first metric is the relative coefficient error $E_2(\hat{W})$, defined as

$$E_2(\hat{W}) = \frac{\|\hat{W} - W\|_F}{\|W\|_F}. \tag{4.2}$$

The second metric is the true positive ratio (TPR)

$$\mathrm{TPR}(\hat{W}) = \frac{\mathrm{TP}(\hat{W})}{\mathrm{TP}(\hat{W}) + \mathrm{FP}(\hat{W}) + \mathrm{FN}(\hat{W})} \tag{4.3}$$

where $\mathrm{TP}(\hat{W})$ is the number of correctly identified nonzero terms, $\mathrm{FP}(\hat{W})$ is the number of falsely identifd nonzero terms, and $\mathrm{FN}(\hat{W})$ is the number of terms that are falsely identified as having a coefficient of zero. While the first two metrics focuses on equation error, the third metric compares the simulated trajectory error, defined as

$$\varepsilon_2 = \frac{\|\hat{X} - X\|_F}{\|X\|_F} \tag{4.4}$$

where $\hat{X}$ is the simulated trajectory matrix of the identified system.

In the equation learning of all benchmark systems using Fourier weak SINDy, we use the sequentially thresholded ridge regression with a sparsity threshold of 0.5 and a ridge regularization value of 0.001. We implement two frequency selection methods for Fourier weak SINDy – the first method is to use the first 500 frequencies of the FFT result (i.e., using a sweep of frequencies up to 50 Hz, $\frac{1}{10}$ of the Nyquist frequency), and the second method is to use the top 100 dominant frequencies from the SDE. We compare the results of Fourier weak SINDy with those of standard SINDy and weak SINDy methods in the PySINDy package [19], where we choose between sparse relaxed regularized regression and sequentially thresholded ridge regression for the one with better performance, and use 1000 subdomains for the weak SINDy in PySINDy (i.e., 1000 test functions of the form (2.13)).

4.1 Second-Order Polynomial Dictionary with PSD Frequency Selections

We perform our first numerical experiment on the Lorenz system with sturcture

$$\dot{x} = \sigma(y - x) \tag{4.5a}$$

$$\dot{y} = x(\rho - z) - y \tag{4.5b}$$

$$\dot{z} = xy - \beta z \tag{4.5c}$$

where we use $\sigma = 10$, $\rho = 28$ $\beta = 8/3$, and initial condition $\begin{bmatrix} 20 & 12 & -30 \end{bmatrix}^T$ for benchmarking. We use library terms up to second-order polynomials. We set

the initial conditions of the simulation to $\mathbf{x}_0 = \begin{bmatrix} 20 & 12 & -30 \end{bmatrix}^T$. As shown in Fig. 4.1, we first compare the performances of SINDy (PySINDy), weak SINDy (PySINDy), and Fourier weak SINDy based on both the frequency sweep and the spectral density estimation methods, with noise ratios ranging from 0.0001% to 100% and 20 noise instances per noise level. Moreover, we investigate the effect of choosing different numbers of dominant frequencies for the Fourier weak SINDy at different noise levels (the dominant frequencies are extracted using the FFT on the clean data, to eliminate the effect of SDE error and focus on the effect of frequency numbers), as well as how the bandwidth of the multitaper SDE influences equation learning results.

According to Fig. 4.1(a), at high noise regions, SINDy has a median coefficient error at the order of 10 and a median true positive ratio near 0.2. This implies that the classical SINDy approach is not providing meaningful equation learning results at high measurement noise. In contrast, the Fourier weak SINDy approach with SDE has a median coefficient error at the order of 0.1 and a median true positive ratio near 0.7. This implies that even with significant measurement noise, Fourier weak SINDy can still reveal the core structure of the model with reasonably small coefficient error.

Fig. 4.1(b) reveals the robustness of Fourier weak SINDy against the number of test functions being selected. Once the number of test functions is more than 30, the coefficient error and true positive ratio does not vary significantly as the number of test functions increases, though there is a slight decline in performance as test functions of higher frequencies are included. The slight decline in performance is likely caused by the fact that the frequency content of the system is mostly in the top 50 to 100 dominant frequencies, and thus the components

projected onto the test functions of other frequencies are more likely to be measurement noise.

Fig. 4.1(c) reveals the robustness of Fourier weak SINDy against the selection of the taper function bandwidths of the multitaper SDE. The performance improves with the size of the bandwidth at first, but once the bandwidth becomes wide enough, the performance does not vary much. This is a reasonable trend as the taper functions are essentially frequency averagers, and averaging more frequencies (i.e., increasing bandwidth) will provide saturating gains. Thus, it is usually a safe choice to set a reasonably wider bandwidth; however, a bandwidth that is too wide will introduce extra computational burden, so it is still important to make judicious trade-offs.

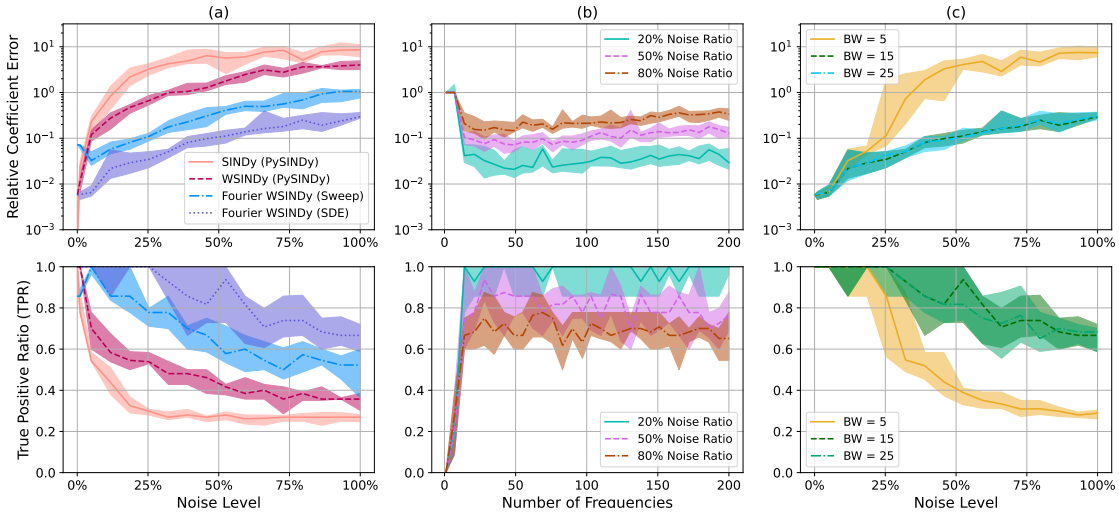

Figure 4.1: Lorenz system equation learning results with second-order polynomial library. Median and quartiles of the relative coefficient error and TPR of (a) four different SINDy and weak SINDy methods at varying noise levels, (b) Fourier weak SINDy with different numbers of dominant frequencies chosen, and (c) Fourier weak SINDy with different bandwidth settings (in Hz) for the multitaper SDE.

The simulated trajectory error for the Lorenz system, with second-order polynomial library, is shown in Fig. 4.2. For numerical simulations, the identified models can be unstable or largely stiff as noise level increases, which causes the adaptive solvers to halt even if implicit methods are used; thus, we perform this numerical experiment for noise ratios up to 30%.

Fig. 4.2(a) shows that at 30% noise ratio, both SINDy and bump-function weak-SINDy has a median trajectory error at the order of 1, while Fourier weak SINDy with multitaper SDE has a median trajectory error at the order of 0.1. Moreover, Fig. 4.2(b) shows that for noise ratios up to 30%, SINDy and bump-function weak SINDy both identified 12 out of 200 dynamic models that are unstable, while the models identified by Fourier weak SINDy are all stable. This is a direct showcase that Fourier weak SINDy can be more accurate in doing predictions for nonlinear dynamical systems.

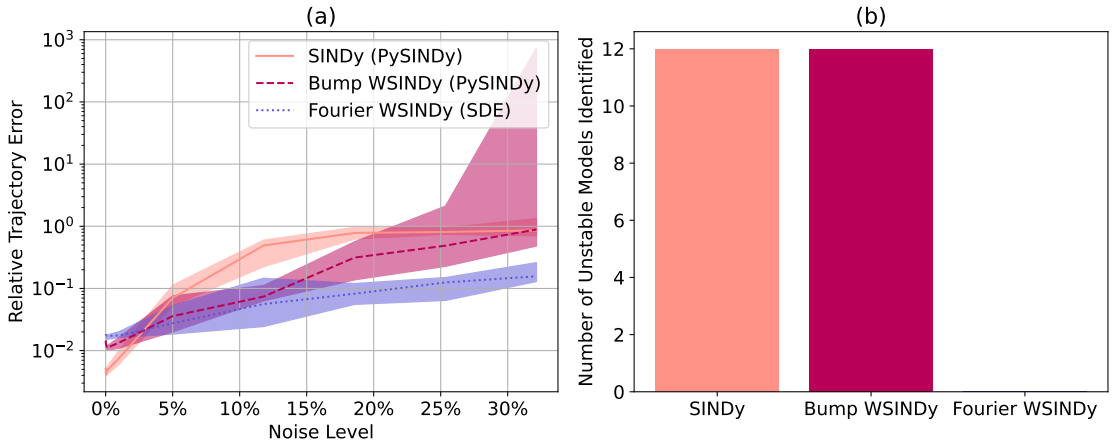

Figure 4.2: Lorenz system equation learning results with second-order polynomial library. Median and quartiles of the simulated trajectory error of (a) regular SINDy (PySINDy), bump weak SINDy (PySINDy), and Fourier weak SINDy with spectral density estimation and (b) number of unstable models identied by the three methods (out of 200 numerical experiments for each method).

We also perform numerical experiments on the Lotka-Volterra equations

$$\dot{x} = 3x - \beta xy \tag{4.6a}$$

$$\dot{y} = -6y + \beta xy \tag{4.6b}$$

where we use $\beta = 1$ and initial conditions $\begin{bmatrix} 1 & 1 \end{bmatrix}^T$, the hyperchaotic Lorenz system

$$\dot{x} = a(y - x) + w \tag{4.7a}$$

$$\dot{y} = -xz + cx - y \tag{4.7b}$$

$$\dot{z} = -bz + xy \tag{4.7c}$$

$$\dot{w} = dw - xz \tag{4.7d}$$

where we use $a = 10$, $b = 2.667$, $c = 28$, $d = 1.1$, and initial conditions $\begin{bmatrix} 5 & 8 & 12 & 21 \end{bmatrix}^T$, and the hyperchaotic Jha system

$$\dot{x} = a(y - x) + w \tag{4.8a}$$

$$\dot{y} = -xz + bx - y \tag{4.8b}$$

$$\dot{z} = xy - cz \tag{4.8c}$$

$$\dot{w} = -xz + dw \tag{4.8d}$$

where we use $a = 10$, $b = 28$, $c = 8/3$, $d = 1.3$, and initial conditions $\begin{bmatrix} 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix}^T$, all with second-order polynomial libraries; the equation learning results are shown in Fig. 4.3.

Based on the simulation results, Fourier Weak SINDy with PSD frequency selections shows either improved or comparable performance in equation learning tasks for the four benchmark systems, compared to the weak SINDy method in the PySINDy package. Specifically, Fourier weak SINDy shows improved performance on the Lotka-Volterra equations (Fig. 4.3(a)) and the hyperchaotic

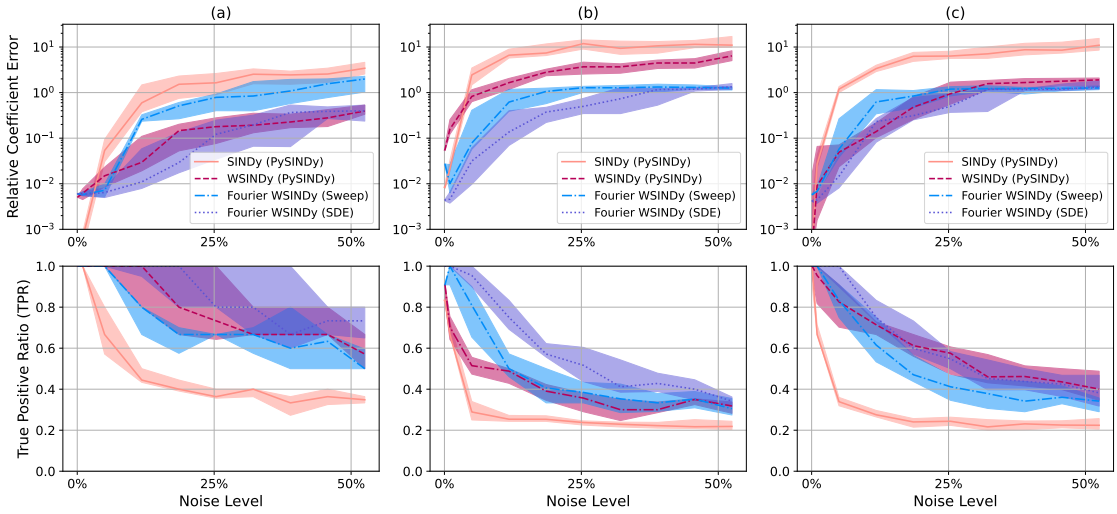

Figure 4.3: Equation learning results of four different SINDy and weak SINDy methods for (a) Lotka-Volterra equations, (b) hyperchaotic Lorenz system, and (c) hyperchaotic Jha system.

Lorenz system (Fig. 4.3(b)); it shows comparable performance to bump-function weak SINDy on the hyperchaotic Jha system, but provides higher true positive ratios on the low-noise region.

In addition to numerical experiments carried out with second-order polynomial libraries, we conducted numerical experiments with higher-order polynomial dictionaries, as shown in Fig. 4.4.

Fig. 4.4(a) shows that for the equation learning task on the Lorenz system with fifth-order polynomial library, at 100% noise ratio, Fourier weak SINDy with SDE can have a 5-times smaller coefficient error compared to bump-function weak SINDy, and it provides a higher true positive ratio (0.5 compared to 0.3). Although for the two hyperchaotic systems with third-order polynomial library (Fig. 4.4(b) and Fig. 4.4(c)) the advantage of Fourier weak SINDy is less salient, the accuracy of Fourier weak SINDy is still more consistent than the

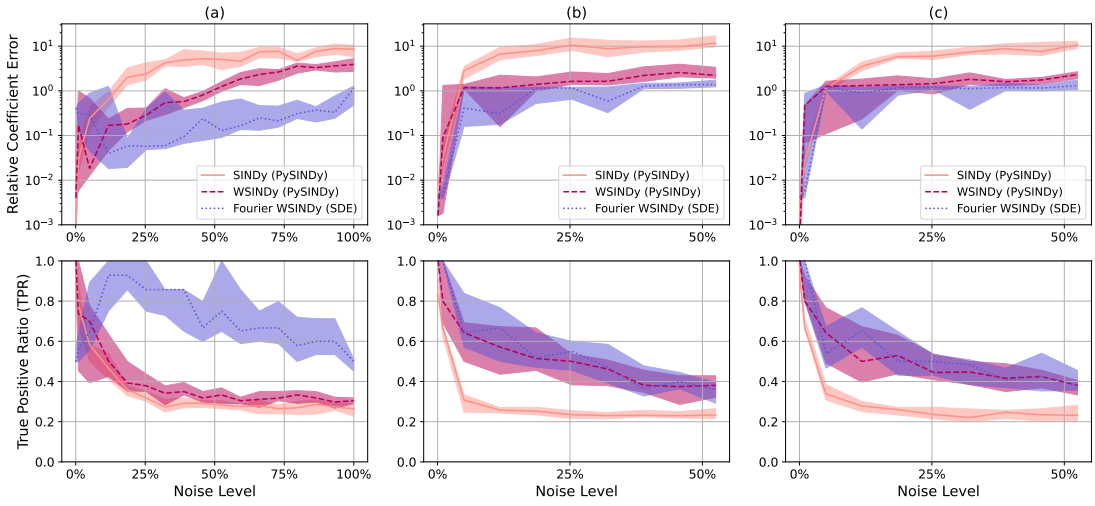

Figure 4.4: Equation learning results comparisons of (a) Lorenz system with fifth-order polynomial library, (b) hyperchaotic Lorenz system with third-order polynomial library, and (c) hyperchaotic Jha system with third-order polynomial library.

bump-function method across the different benchmark systems and different dictionary function libraries.

4.2 Second-Order Polynomial Dictionary with Fourier Phase Index

We tested the model identification result where we select the test function frequencies based on the Fourier phases indices of the x and y trajectories of the Lorenz system. We use a simulation trajectory of 800 seconds with a sampling interval of 0.01 seconds. Additionally, we use second-order polynomials as the library. The result is shown in Fig. 4.5.

According to Fig. 4.5, the Fourier weak SINDy based on Fourier phase index did not outperform the Fourier weak SINDy that uses the first 500 FFT frequen-

cies. In fact, the Fourier phase index approach did worse in terms of the true positive ratio. A possible reason is that a more judicious choice of coordinates is needed, rather than simply selecting x and y of the Lorenz system. Moreover, the lengths, overlaps, and numbers of the frequency windows may also require more careful design and computations.

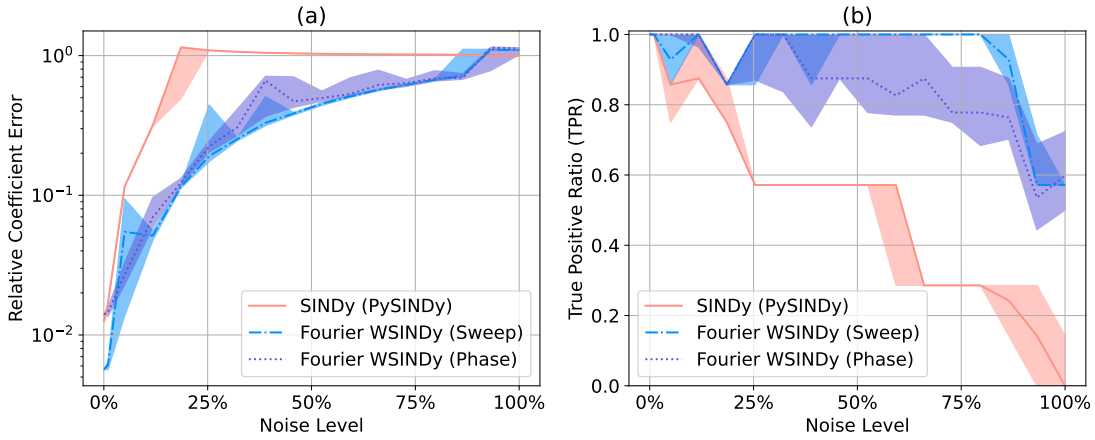

Figure 4.5: Equation learning result comparisons of Lorenz system with second-order polynomial library. (a) Relative coefficient error and (b) true positive ratio of equation learning results from SINDy, Fourier weak SINDy with first 500 FFT frequencies, and Fourier weak SINDy with frequency selections based on Fourier phase index.

CHAPTER 5
DISCUSSION

In this thesis, we showed that combining weak-form equation learning over orthogonal sinusoidal test functions with classical signal processing techniques for the selection of test function frequencies to match maximally energetic modes in the data, presents an interpretable, robust, and effective approach to equation learning that requires minimal tuning. Specifically, Fourier weak SINDy shows lower equation coefficient and trajectory error, and higher true positive ratios on the benchmark systems, particularly at the high-noise region.

While Fourier weak SINDy shows consistent performance and interpretability over different benchmarks, there are many promising directions of future work. For example, the current FFT implementations of Fourier weak SINDy are limited to uniformly sampled trajectory data. Therefore, a future direction is to experiment with Nonuniform Fast Fourier Transform to perform equation learning using data with irregular time samples. Moreover, an assumption made in this thesis is that the measurement noise is additive white Gaussian noise. Although this assumption holds for many equation learning tasks, noise profiles beyond white noise are common in practice. In these cases, Fourier weak SINDy can be further augmented by more advanced spectral density estimation methods, such as the adaptive multitaper method and the Capon estimator. Additionally, existing techniques for improving the accuracy and noise-robustness of SINDy, such as ensembling and Bayesian coefficient selection strategies, can be incorporated into the current framework to further improve accuracy and robustness. Furthermore, as discussed in the introduction, the bump-function weak SINDy methods we are comparing against are the baseline

methods from the PySINDy package; principled comparisons to bump-function weak SINDy with adaptive hyperparameter tuning can provide more insights.

Moreover, the frequency selection based on the Fourier phase index still has room for improvement. The selection of coordinates, and the number and size of the frequency windows, can have significant effect on the effectiveness of J-index computations. Also, further hypothesis tests on the J-index results need to be carried out to verify the validity of resulting Fourier phase indices.

BIBLIOGRAPHY

[1] Alberto Isaac Aguilar-Hernández, David Michel Serrano-Solis, Wady A Ríos-Herrera, José Fernando Zapata-Berruecos, Gloria Vilaclara, Gustavo Martínez-Mekler, and Markus F Müller. Fourier phase index for extracting signatures of determinism and nonlinear features in time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(1), 2024.

[2] Karl Johan ström and Peter Eykhoff. System identification—a survey. *Automatica*, 7(2):123–162, 1971.

[3] Behtash Babadi and Emery N Brown. A review of multitaper spectral analysis. *IEEE Transactions on Biomedical Engineering*, 61(5):1555–1564, 2014.

[4] Joseph Bakarji, Kathleen Champion, J Nathan Kutz, and Steven L Brunton. Discovering governing equations from partial measurements with deep delay autoencoders. *Proceedings of the Royal Society A*, 479(2276):20230422, 2023.

[5] David M Bortz, Daniel A Messenger, and Vanja Dukic. Direct estimation of parameters in ode models using wendy: Weak-form estimation of nonlinear dynamics. *Bulletin of Mathematical Biology*, 85(11):110, 2023.

[6] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

[7] Steven L Brunton, Nicholas Zolman, J Nathan Kutz, and Urban Fasel. Machine learning for sparse nonlinear modeling and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 8, 2025.

[8] Kathleen Champion, Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019.

[9] Zhiheng Chen, Urban Fasel, and Anastasia Bizyaeva. Fourier weak sindy: Spectral test function selection for robust model identification. *arXiv preprint arXiv:2604.20141*, 2026.

[10] Robin Delabays, Giulia De Pasquale, Florian Dörfler, and Yuanzhao Zhang. Hypergraph reconstruction from dynamics. *Nature Communications*, 2025.

[11] Urban Fasel, Eurika Kaiser, J Nathan Kutz, Bingni W Brunton, and Steven L Brunton. Sindy with control: A tutorial. In *2021 60th IEEE conference on decision and control (CDC)*, pages 16–21. IEEE, 2021.

[12] Urban Fasel, J Nathan Kutz, Bingni W Brunton, and Steven L Brunton. Ensemble-sindy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proceedings of the Royal Society A*, 478(2260):20210904, 2022.

[13] Kai Fukami, Takaaki Murata, Kai Zhang, and Koji Fukagata. Sparse identification of nonlinear dynamics with low-dimensionalized flow representations. *Journal of Fluid Mechanics*, 2021.

[14] Lloyd Fung, Urban Fasel, and Matthew Juniper. Rapid bayesian identification of sparse nonlinear dynamics from scarce and noisy data. In *Proceedings A*. The Royal Society, 2025.

[15] William Gilpin. Chaos as an interpretable benchmark for forecasting and data-driven modelling, 2023.

[16] Daniel R Gurevich, Patrick AK Reinbold, and Roman O Grigoriev. Robust and optimal sparse regression for nonlinear pde models. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(10), 2019.

[17] Jonathan Horrocks and Chris T Bauch. Algorithmic discovery of dynamic models from infectious disease data. *Scientific reports*, 2020.

[18] Kadierdan Kaheman, Steven L Brunton, and J Nathan Kutz. Automatic differentiation to simultaneously identify nonlinear dynamics and extract noise probability distributions from data. *Machine Learning: Science and Technology*, 3(1):015031, 2022.

[19] Alan A. Kaptanoglu, Brian M. de Silva, Urban Fasel, Kadierdan Kaheman, Andy J. Goldschmidt, Jared Callaham, Charles B. Delahunt, Zachary G. Nicolaou, Kathleen Champion, Jean-Christophe Loiseau, J. Nathan Kutz, and Steven L. Brunton. Pysindy: A comprehensive python package for robust sparse system identification. *Journal of Open Source Software*, 7(69):3994, 2022.

[20] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[21] Christian Legaard, Thomas Schranz, Gerald Schweiger, Ján Drgoňa, Basak Falay, Cláudio Gomes, Alexandros Iosifidis, Mahdi Abkar, and Peter Larsen. Constructing neural network based models for simulating dynamical systems. *ACM Computing Surveys*, 55(11):1–34, 2023.

[22] Dongzhuo Li, Kailai Xu, Jerry M Harris, and Eric Darve. Coupled time-lapse full-waveform inversion for subsurface flow problems using intrusive automatic differentiation. *Water Resources Research*, 56(8):e2019WR027032, 2020.

[23] Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12):1053–1058, 2018.

[24] Bernardo Maestrini, Gordan Mimić, Pepijn AJ van Oort, Keiji Jindo, Sanja Brdar, Ioannis N Athanasiadis, and Frits K van Evert. Mixing process-based and data-driven approaches in yield prediction. *European Journal of Agronomy*, 139:126569, 2022.

[25] Gemma Massonis, Alejandro F Villaverde, and Julio R Banga. Distilling identifiable and interpretable dynamic models from biological data. *PLoS computational biology*, 19(10):e1011014, 2023.

[26] Daniel A Messenger and David M Bortz. Weak sindy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3):1474–1497, 2021.

[27] Daniel A Messenger, April Tran, Vanja Dukic, and David M Bortz. The weak form is stronger than you think. *arXiv preprint arXiv:2409.06751*, 2024.

[28] Robert K Niven, Laurent Cordier, Ali Mohammad-Djafari, Markus Abel, and Markus Quade. Dynamical system identification, model selection, and model uncertainty quantification by bayesian inference. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2024.

[29] Ben Noordijk, Monica L Garcia Gomez, Kirsten HWJ Ten Tusscher, Dick de Ridder, Aalt DJ van Dijk, and Robert W Smith. The rise of scientific machine learning: a perspective on combining mechanistic modelling with machine learning for systems biology. *Frontiers in Systems Biology*, 4:1407994, 2024.

[30] Joshua S North, Christopher K Wikle, and Erin M Schliep. A bayesian ap-

proach for data-driven dynamic equation discovery. *Journal of Agricultural, Biological and Environmental Statistics*, 2022.

[31] Jeffrey Park, Craig R Lindberg, and Frank L Vernon III. Multitaper spectral analysis of high-frequency seismograms. *Journal of Geophysical Research: Solid Earth*, 92(B12):12675–12684, 1987.

[32] A Pearson and F Lee. On the identification of polynomial input-output differential systems. *IEEE Transactions on Automatic Control*, 30(8):778–782, 1985.

[33] AE Pearson. Aerodynamic parameter estimation via fourier modulating function techniques. Technical report, NASA, 1995.

[34] AE Pearson and JQ Pan. Frequency analysis via the method of moment functionals. Technical report, NASA, 1990.

[35] Patrick AK Reinbold, Daniel R Gurevich, and Roman O Grigoriev. Using noisy or incomplete data to discover models of spatiotemporal dynamics. *Physical Review E*, 101(1):010203, 2020.

[36] Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science advances*, 3(4):e1602614, 2017.

[37] Benjamin P Russo and M Paul Laiu. Convergence of weak-sindy surrogate models. *SIAM Journal on Applied Dynamical Systems*, 23(2):1017–1051, 2024.

[38] Hayden Schaeffer and Scott G McCalla. Sparse model selection via integral terms. *Physical Review E*, 96(2):023302, 2017.

[39] Marvin Shinbrot. On the analysis of linear and nonlinear systems. *Transactions of the American Society of Mechanical Engineers*, 79(3):547–551, 1957.

[40] David Slepian. Prolate spheroidal wave functions, fourier analysis, and uncertainty—v: The discrete case. *Bell System Technical Journal*, 57(5):1371–1430, 1978.

[41] Robert Stephany and Christopher Earls. Weak-pde-learn: A weak form based approach to discovering pdes from noisy, limited data. *Journal of Computational Physics*, 506:112950, 2024.

[42] Petre Stoica, Randolph L Moses, et al. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[43] David J Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 2005.

[44] Alexander Tong, David van Dijk, Jay S Stanley III, Matthew Amodio, Kristina Yim, Rebecca Muhle, James Noonan, Guy Wolf, and Smita Krishnaswamy. Interpretable neuron structuring with graph spectral regularization. In *International Symposium on Intelligent Data Analysis*, pages 509–521. Springer, 2020.

[45] Heinz Unbehauen and GP Rao. Continuous-time approaches to system identification—a survey. *Automatica*, 26(1):23–35, 1990.

[46] Marieke K van Vugt, Per B Sederberg, and Michael J Kahana. Comparison of spectral analysis methods for characterizing brain oscillations. *Journal of neuroscience methods*, 162(1-2):49–63, 2007.

[47] Sheng Zhang and Guang Lin. Robust data-driven discovery of governing physical laws with error bars. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2018.