Robust spectral representations and model inference for biological dynamics

by

Alasdair D. Hastewell

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MATHEMATICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2024

Authored by: Alasdair D. Hastewell
Department of Mathematics
April 18, 2024

Certified by: Jörn Dunkel
MathWorks Professor of Mathematics, Thesis Supervisor

Accepted by: Jonathan Kelner
Professor of Mathematics
Chairman, Department Committee on Graduate Theses

Robust spectral representations and model inference for biological dynamics

by

Alasdair D. Hastewell

Submitted to the Department of Mathematics
on April 18, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MATHEMATICS

ABSTRACT

Current developments in automated experimental imaging allow for high-resolution tracking across various scales, from whole animal behavior to single-cell dynamics to spatiotemporal gene expression. Transforming these high-dimensional data into effective low-dimensional models is an essential theoretical challenge to enable the characterization, comparison, and prediction of dynamics both within and across biological systems. Spectral mode representations have been used successfully across physics, from quantum mechanics to fluid dynamics, to compress and model dynamical data. However, their use in analyzing biological systems has yet to become prevalent. Here, we present a set of noise-robust, geometry-aware mathematical tools that enable spectral representations to extract quantitative measurements directly from experimental data. We demonstrate the practical utility of these methods by applying them to the extraction of defects in signaling fields on membranes, the inference of partial differential equations directly from videos of active particle dynamics, and the categorization of emergent patterns in spatiotemporal gene expression during bacterial swarming.

An additional challenge occurs for complex biophysical processes where the underlying dynamics are only partially determined. We wish to use experimental data directly to infer effective dynamical models that elucidate the system's underlying biological and physical mechanisms. Building on spectral mode representations, we construct a generic computational framework for inferring the dynamics of living systems through the evolution of their mode representations. The framework can incorporate prior knowledge about biological and physical constraints. We apply this framework first to single-cell imaging data during zebrafish embryogenesis, demonstrating how our framework can compactly characterize developmental symmetry-breaking and reveal similarities between pan-embryo cell migration and Brownian particles on curved surfaces. Next, we apply the framework to the undulatory locomotion of worms, centipedes, robots, and snakes to distinguish between locomotion behaviors. Finally, we present an extension of the framework to the case of nonlinear dynamics when all relevant degrees of freedom are only partially observed, with applications to neuronal and chemical dynamics.

Thesis supervisor: Jörn Dunkel
Title: MathWorks Professor of Mathematics

Contents

Acknowledgments

Firstly, I would like to thank my advisor, Jörn Dunkel, for his tireless support and mentorship over the last eight years. Since our first meeting during my undergraduate studies, I have been inspired by and learned so much from your approach to science and mathematics. Additionally, I would like to thank the other faculty members in Physical Applied Math, John Bush and Ruben Rosales; their comments and discussions over the years have been invaluable. I would also like to thank my committee members, John Bush and Alexander Mietke, for their helpful comments and mentorship.

I have been fortunate to collaborate with many exceptional experimental groups throughout my Ph.D, which has heavily motivated and enabled the work presented here. I want to thank all my collaborators, especially the groups of Nikta Fakhri, Knut Drescher, Adam Martin, Steven Flavell, and Kirsty Wan. I am grateful to my coauthors Jinghui, Yu-Chen, Lisa, Shreyas, Hyun, Hannah, Niklas, Kazuki, Sreeparna, and Alexander for their patience and openness to teaching me about biology and for sharing their experimental data.

I have also been fortunate to work alongside many wonderful people within the Dunkel group. I am especially thankful to my collaborators Jan, Pearson, Rohit, Boya, Alex Mietke, Nico, Alex Cohen, George, and Dominic; I have enjoyed working and learning alongside you all. Additionally, I appreciate my many conversations with Henrik Ronellenfitsch, Keaton Burns, and the entire Physical Applied Math community at MIT.

I had many excellent teachers, both before and at MIT. Notably, Homer Reid and Luiz Faria opened my eyes to the world of numerical and physical applied mathematics. I will forever be grateful to Clifford Dammers; without his support, I would never have made it to MIT in the first place.

Outside of MIT, I have been fortunate to have many incredible friends who have provided, at times, much-needed distractions and breaks from work over the years. Finally, I would like to thank my parents. None of this would have been possible without your unconditional support and love over the years.

List of Figures

Chapter 1

Introduction

Experimental advances in live imaging and high-throughput sequencing make it possible to study biological processes in unprecedented detail across scales. From whole animal behavioral states [1–3], to high resolution microscopy of signaling [4, 5] and single-cell [6–9] dynamics, to the spatiotemporal physical [10, 11] and genetic [12] dynamics of multicellular bacterial colonies. Combined with algorithmic progress in automated tracking and segmentation [8–10], experiments can yield thousands of tracked long-lived single-particle trajectories, often with simultaneous measurements of gene expression [5] or neuronal dynamics [13]. These novel data enable us to ask questions about the nature of interactions within a system and the interplay between biological signaling and physical behavior. However, making quantitative statements and theoretical predictions directly from experimental data requires developing new mathematical tools that are robust to noise, scattered sampling, and partial observations. We require methods to translate high-dimensional input data into low-dimensional representations that provide insight into a system's dynamics.

Modal decompositions have been widely applied in physics [14–16], applied mathematics [17–20], and numerical analysis [21, 22] to decompose a systems' dynamics into its dominant components. Modal representations, therefore, are a powerful tool for providing low-dimensional representations and differentiating structure from noise. Different types of modes may be used to analyze a system: data-driven modes from techniques such as dynamic mode decomposition [19] and proper orthogonal decomposition [20]; prescribed modes like Fourier bases or orthogonal polynomials chosen based on the geometry of the system [22, 23]; or eigen-modes from theoretical models [14, 15]. Here, we aim to build generic methods readily applicable across experimental realizations and systems. Therefore, we focus here on spectral representations where the basis is chosen based on the system's geometry in a data and model-free way. One benefit of spectral bases is that fast methods exist for working with them [24], but many of these algorithms require data sampled on fixed transform

grids. While recent works have applied data-driven mode decomposition to analyze multicellular communities [25] and developmental dynamics [26], the use of spectral methods to analyze experimental data is not yet common due to the challenges posed by working with data not sampled on the fast transform grid and the presence of noise. In the first part of this thesis, we develop tools to apply spectral representations to experimental data and use these representations to robustly extract theoretical quantities directly from experimental measurements, enabling direct comparison between theory and experiment.

A full understanding of the underlying mechanisms of complex biophysical phenomena may not always be available. Therefore, we would like to infer predictive models directly from experimental data and use these models to interpret the dynamic structure. Dynamical system inference has gained popularity lately due to the increased availability of experimental data and computational power [27, 28]. Commonly used methods include library-based regression [27, 29–33], symbolic regression [34, 35] and neural networks [28, 36, 37]. Many of these previous works have focused on synthetic data, whereas applications of model discovery to biological data are less frequent. In the second part of this thesis, we develop dynamical system inference techniques that can be applied directly to experimental biological data. Specifically, we focus on developing and inferring models directly in spectral mode space, leading to methods robust to noise, scattered data samples, and partial observations.

The organization of this thesis is as follows. In Chapter 2, we introduce spectral representations and develop techniques to apply them to experimental data. In particular, we focus on accounting for noise during spectral projection and show that fast algorithms exist in the presence of noise, even for data not sampled on the fast transform grid. Chapter 3 contains three applications of spectral methods for analyzing experimental data:

1. In collaboration with Nikta Fakhri's lab at MIT, we apply spectral methods to live imaging data of Rho-GTP signaling protein on the surface of starfish egg cells. We use the representations to extract smooth, differentiable, denoised signaling fields. We then use this denoised representation for further analysis, enabling the robust tracking of long-lived defects, revealing a correlation between cellular activity and braiding exponents, and suggesting phenomenological similarities between living matter and certain quantum systems. The results have previously been published in J. Liu, J. F. Totz, P. W. Miller, A. D. Hastewell, Y. C. Chao, J. Dunkel, and N. Fakhri, "Topological braiding and virtual particles on the cell membrane," *Proceedings of the National Academy of Sciences, 118(34), e2104191118*, 2021.

2. We apply spectral methods to coarse-grained active particle data. We use the resulting differentiable fields to compute high-accuracy derivatives of the data, improving the

robustness of library-based model inference regression. The results have previously been published in R. Supekar, B. Song, A. D. Hastewell, G. P. Choi, A. Mietke, and J. Dunkel "Learning hydrodynamic equations for active matter from particle simulations and experiments," *Proceedings of the National Academy of Sciences, 120(7), e2206994120*, 2023.

3. Finally, we extend the spectral representation framework to a non-rectangular domain. In collaboration with Knut Drescher's lab, we apply the framework to spatiotemporally measured gene expression patterns during bacterial swarming. We use the low-dimensional representation provided by spectral projection to identify and categorize genes with strong spatiotemporal patterns. The resulting analysis reveals a strong relationship between the spatiotemporal expression patterns of genes associated with metabolism and the spatiotemporal dynamics of phenotypic properties, leading to the discovery of spatiotemporal cross-feeding interactions within the swarm. The results have previously been published in H. Jeckel*, K. Nosho*, K. Neuhaus, A. D. Hastewell, D. J. Skinner, D. Saha, N. Netter, N. Paczia, J. Dunkel and K. Drescher, "Simultaneous spatiotemporal transcriptomics and microscopy of Bacillus subtilis swarm development reveal cooperation across generations," *Nature Microbiology, 8(12), 2378-2391,* 2023.

We study dynamic system inference for mode space dynamics in the second part. In Chapter 4, we consider the theory behind mode space dynamics and develop a sparse linear dynamical systems inference framework that is robust to noise. In Chapter 5, we apply the inference framework to the developmental dynamics of zebrafish during the early stages of embryogenesis. The results in Chapters 4 and 5 have previously been published in N. Romeo*, A. D. Hastewell*, A. Mietke, and J. Dunkel, "Learning developmental mode dynamics from single-cell trajectories". *Elife, 10, e68679,* 2021.

Chapter 6 shows how to incorporate physical and biological constraints into mode space models using planar curve dynamics as an example. Additionally, we develop an efficient inference algorithm for the constrained dynamics. In Chapter 7, we apply the results in Chapter 6 to undulatory locomotion and show how the resulting inferred dynamical models can be used to classify different animals' motion and behavioral states. The results in Chapters 6 and 7 have previously been published in A. E. Cohen*, A. D. Hastewell*, S. Pradhan, S. W. Flavell, and J. Dunkel, "Schrödinger Dynamics and Berry Phase of Undulatory Locomotion," *Physical Review Letters, 130(25), 258402,* 2023.

Finally, in Chapter 7, we discuss extensions of the inference framework to incorporate nonlinearities and account for unobserved variables. We show the utility of this extension by applying it to neural, chemical, and chaotic dynamics. The results in Chapter 7 have

previously been published in a preprint G. Stepaniants*, A. D. Hastewell*, D. J. Skinner, J. F. Totz, and J. Dunkel, "Discovering dynamics and parameters of nonlinear oscillatory and chaotic systems from partial observations," *arXiv preprint arXiv:2304.04818,* 2023.

Chapter 8 concludes with some final remarks about the general framework presented and possible future extensions.

Chapter 2

Spectral representations

The key idea underlying spectral representations, formalized in the next section, is that a function is well approximated by a series expansion in a fixed basis $\{\phi_n(\mathbf{x})\}_{n=1}^N$,

$$f(\mathbf{x}) \approx \sum_{n=0}^{N} c_n \phi_n(\mathbf{x}). \tag{2.1}$$

We call expansions of the form (2.1) a spectral representation of $f(\mathbf{x})$ in the basis $\phi_n(\mathbf{x})$. Common examples of spectral representations are Fourier series where the basis is trigonometric functions, for example, $\phi_n(x) = \cos nx$, or orthogonal polynomial expansions, such as Chebyshev series where $\phi_n(x) = T_n(x) = \cos(n \arccos x)$. When performing expansions of the form (2.1), we are shifting information about the function $f(\mathbf{x})$ into a finite number of scalar coefficients c_n. At the same time, operations on $f(\mathbf{x})$, such as differentiation and integration, are shifted to corresponding manipulations on the known basis $\phi_n(\mathbf{x})$. This split has several advantages. From a theoretical perspective, we can compare functions by comparing their coefficients and properties of $f(\mathbf{x})$, such as continuity and differentiability, follow directly from the basis. From a practical perspective, we can perform accurate computations on spectral representations since operations on the functions become operations on known bases.

Spectral representations have been used with great success in numerical analysis [21] to develop efficient and accurate algorithms for function approximation [24, 38], solving differential equations [22, 39], and root solving [40]. The success of these methods relies on the observation that when the function $f(\mathbf{x})$ is continuous and has continuous derivatives, then the coefficients c_n in the expansion (2.1) decay rapidly, enabling accurate approximations with small N, for a large class of functions with a single fixed basis. Additionally, the existence of fast algorithms [41] for calculating coefficients c_n directly from samples of f

enables the development of practical and efficient algorithms.

In biophysics, many experimentally measured quantities, from chemical signaling fields and gene expression profiles to physical densities and velocity fields, are assumed to vary smoothly in space and time. Such data, therefore, lend themselves naturally to spectral representations. Furthermore, the existence of fast algorithms means that these approaches can be readily scaled to high-resolution imaging data.

In this chapter, we summarize some definitions and theoretical results for spectral representations. We then discuss how spectral representations behave when applied to noisy inputs and the consequences when working with experimental data.

2.1 Spectral representations in orthogonal bases

We consider spectral representations of real or complex, scalar differentiable functions defined over some possibly infinite subset D of $\mathbb{R}^n$, $f(\mathbf{x}) : D \subseteq \mathbb{R}^n \to \mathbb{R}$ or $\mathbb{C}$. We define an inner product over the domain D, using $*$ to denote complex conjugation,

$$\langle f(\mathbf{x}), g(\mathbf{x}) \rangle_w = \int_D \mathrm{d}\mathbf{x}\, w(\mathbf{x}) f(\mathbf{x}) g^*(\mathbf{x}) \tag{2.2}$$

with $w(\mathbf{x})$ a non-negative weight function that is integrable over the domain D. We will work with bases that are orthogonal under the inner product

$$\langle \phi_n(\mathbf{x}), \phi_m(\mathbf{x}) \rangle_w = k_n^\phi \delta_{n,m} \tag{2.3}$$

where $\delta_{n,m}$ is the Kronecker delta defined by,

$$\delta_{n,m} = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{otherwise} \end{cases}. \tag{2.4}$$

When the choice of basis is clear, we will drop the ϕ and w labels on the inner product and k_n. The coefficients in (2.1) are given by,

$$c_n = \frac{\langle f, \phi_n \rangle_w}{\langle \phi_n, \phi_n \rangle_w} = \frac{1}{k_n^\phi} \int_D \mathrm{d}\mathbf{x}\, w(\mathbf{x}) f(\mathbf{x}) \phi_n^*(\mathbf{x}). \tag{2.5}$$

We define the weighted square approximation error of the spectral representation by,

$$\mathcal{E}_N = \int_D \mathrm{d}\mathbf{x}\, w(\mathbf{x}) \left| f - \sum_{n=1}^{N} c_n \phi_n(\mathbf{x}) \right|^2. \tag{2.6}$$

By differentiating (2.6) with respect to c_n, we see that the values of c_n that minimize the square error are given by (2.5).

We will frequently work with two sets of basis functions: the Fourier basis for periodic data and the Chebyshev basis for non-periodic data. For convenience, we summarize some details about these bases below.

2.1.1 Fourier series

In one dimension, the Fourier basis is defined by the representation,

$$f(x) = a_0 + \sum_{n=1}^{N} a_n \cos(2\pi n x) + b_n \sin(2\pi n x) = \sum_{n=-N}^{N} c_n e^{2\pi i n} \tag{2.7}$$

which are orthogonal on $[0, 1]$ under the weight function $w(x) = 1$. We will use the series' real and complex exponential forms interchangeably based on convenience. Using integration by parts, we can bound the coefficients a_n and b_n. If $f(x)$ is periodic and is k times differentiable with $k-1$ periodic derivatives and the kth derivative of f has bounded variation, then the coefficients decay as $a_n, b_n \sim n^{-(k+1)}$. If the function is analytic, then the coefficients decay geometrically as $a_n, b_n \sim \rho^{-n}$ for some $\rho > 1$ [42]. These decay rates have immediate consequences for the accuracy of the spectral representation since the error will generally be on the order of the last retained coefficient [23].

Given M samples of f on the uniform grid $x_m = m/M$ for $0 \leq m < M$ we can discretize the integral in (2.5) using the trapezoidal rule, which is known to be exponentially convergent for periodic functions [43],

$$\tilde{c}_k = \frac{1}{M} \sum_{m=0}^{M-1} f(x_m) e^{-\frac{2\pi i m k}{M}}. \tag{2.8}$$

The coefficients $\tilde{c}_k$ are accurate approximations of c_k and we will treat them interchangably and drop the tilde. The summation (2.8) can be evaluated efficiently using the Fast Fourier Transform (FFT) in $\mathcal{O}(n \log n)$ operations, enabling the accurate and efficient evaluation of coefficients in the spectral representation using uniform function samples. We discuss dealing with noisy or non-uniform samples in Section 2.2.

We can evaluate a Fourier representation at an arbitrary point using Horner's method for polynomial evaluation. We can calculate the representation's derivative as,

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = 2\pi \sum_{n=1}^{N} -n a_n \sin(2\pi n x) + n b_n \cos(2\pi n x) = 2\pi i \sum_{n=-N}^{N} n c_n e^{2\pi i n}, \tag{2.9}$$

which corresponds to another series of the form (2.7) but with coefficients $a'_n = 2\pi n b_n$ and $b'_n = -2\pi n a_n$ or $c'_n = 2\pi i n c_n$.

2.1.2 Chebyshev polynomials of the first kind

The Chebyshev polynomials of the first kind defined by

$$T_n(x) = \cos(n \arccos x) \tag{2.10}$$

are degree n polynomials and are orthogonal on $[-1, 1]$ under the weight function $w(x) = (1 - x^2)^{-1/2}$ [23, 44]. The Chebyshev polynomials form a complete orthogonal basis for continuous functions on the domain $[-1, 1]$. A Chebyshev polynomial representation of a function is given by,

$$f(x) = \sum_{n=0}^{N} c_n T_n(x) \tag{2.11}$$

The Chebyshev coefficients for a function f are given by,

$$c_n = \frac{2 - \delta_{n0}}{\pi} \int_{-1}^{1} \mathrm{d}x \, \frac{1}{\sqrt{1 - x^2}} T_n(x) f(x) = \frac{2 - \delta_{n0}}{\pi} \int_0^{\pi} \mathrm{d}\theta \cos(n\theta) f(\cos\theta) \tag{2.12}$$

Analogous theorems for the decay of coefficients exist for the Chebyshev series. Suppose that $f(x)$ and its first $k - 1$ derivatives are absolutely continuous on $[-1, 1]$ and the kth derivative has bounded variation then the Chebyshev coefficients decay as $c_n \sim n^{-(k+1)}$. If $f(x)$ is analytic on $[-1, 1]$ then its Chebyshev coefficients decay geometrically $c_n \sim \rho^{-n}$ for some $\rho > 1$ [45].

The $N + 1$ point Chebyshev grid $\{x_n^c\}_{n=0}^N$ is defined by the extremal points of $T_N(x)$

$$x_n^c = -\cos(n\pi/N) \text{ for } 0 \leq n \leq N. \tag{2.13}$$

Given a function $f(x)$ sampled on the Chebyshev grid, the first $N + 1$ coefficients can be approximated to high accuracy by the sum,

$$c_n \approx \frac{2 - \delta_{0,n} - \delta_{N,n}}{N} \left[\frac{f(-1)(-1)^n + f(1)}{2} + (-1)^n \sum_{m=1}^{N-1} f(x_m^c) \cos\left(\frac{\pi n m}{N}\right) \right]. \tag{2.14}$$

The sum can be considered the trapezoidal rule applied to the periodic integral on the right-hand side of (2.12). The sum in the coefficient transform can be calculated efficiently using a fast DCTII in $\mathcal{O}(N \log N)$ operations [41].

As a general rule of thumb, the magnitude of the error (2.6) will be governed by the magnitude of the last coefficient c_N [23]. An important observation for smooth differentiable functions is that the magnitude of c_n decays rapidly, implying that spectral representations can provide accurate representations of data with significant compression – the number of terms retained, N, for an accurate representation of the data is much less than the number of data samples .

Given a representation of the form (2.11), we can evaluate the series and its derivatives at arbitrary points using Clenshaw-Smith recurrence [46]. The representation of the derivative is given by,

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x) = \sum_{n=0}^{N} c_n \frac{\mathrm{d}}{\mathrm{d}x}T_n(x) = \sum_{n=0}^{N-1} c'_n T_n(x) \tag{2.15}$$

which is a polynomial of degree $N-1$. The coefficients in the new expansion are given by [23, 44],

$$c'_n = (2-\delta_{n,0}) \sum_{m=0}^{\lfloor \frac{N-1-n}{2} \rfloor} (n+1+2m) c_{n+1+2m}. \tag{2.16}$$

2.1.3 Extension to multiple dimensions

A natural extension of spectral representation bases into multiple dimensions is given by the tensor product basis,

$$\phi_{n_1,n_2,\cdots,n_N}(\mathbf{x}) = \prod_{i=1}^{N} \phi_{n_i}(x_i). \tag{2.17}$$

Moreover, we can extend this concept further by introducing mixed tensor product bases, allowing for different bases in each dimension. Frequently, we will use these mixed bases to analyze video data where the time dimension is treated separately from the spatial dimensions.

2.2 Calculation of spectral coefficients from data

A unique challenge posed by experimental data is fixed grid sampling. In many applications in numerical computing, we assume that the function can be sampled on an arbitrary grid, enabling fast algorithms. Here, we present several approaches for calculating coefficients in spectral representations when data samples are on a fixed grid and contain noise. We will illustrate the approaches using a sampled one-dimensional function, although higher-dimensional generalizations are readily possible. Consider a function $f(x)$ sampled on an ordered grid $\{x_n^s\}_{n=0}^N$, $f_n = f(x_n^s)$. The function samples and the coefficients in the spectral

representation can be stacked into vectors $\mathbf{f} = [f_0, \ldots, f_N]$ and $\mathbf{c} = [c_0, \ldots, c_M]$. We will frequently not have access to $\mathbf{f}$ but a noise corrupted version $\tilde{\mathbf{f}} = \mathbf{f} + \boldsymbol{\epsilon}$, where we assume that the noise is additive and $\|\mathbf{f}\| \gg \|\boldsymbol{\epsilon}\|$

2.2.1 Least squares fit coefficients

We define a discrete square approximation error by,

$$E_N = \sum_{m=1}^{M} \left(f_m - \sum_{n=1}^{N} c_n \phi_n(\mathbf{x}_m) \right)^2 = \|\mathbf{f} - \Phi\mathbf{c}\|^2 \tag{2.18}$$

where the Vandermonde matrix Φ has entries $\Phi_{mn} = \phi_n(\mathbf{x}_m)$, and $\|\cdot\|$ is the standard Euclidean vector norm. When the number of observations is small enough, the coefficient vector $\mathbf{c}$ may be determined from the least squares solution of the problem $\Phi\mathbf{c} = \mathbf{f}$ implied by (2.18). When the samples are not well distributed or when N approaches M, the least squares solution may become ill-conditioned, and it becomes necessary to add regularization to the least squares problem [47].

2.2.2 Interpolated DCT coefficients

The solution of the least squares problem scales as $\mathcal{O}(M^2N)$, which becomes computationally too expensive for high-resolution data. Instead, we would like to use fast transforms that scale as $\mathcal{O}(M\,logM)$. However, we need the function to be sampled on a fixed grid to use fast transforms. We consider Chebyshev transforms but the ideas presented can be generalized to other bases if needed. We transform between sample points and the fixed grid using interpolation. Consider a function $f(x)$ sampled on a uniform grid $f_m = f(m\Delta x)$ for $0 \leq m \leq M$ and $\Delta x = 1/M$. We must obtain samples of the function on the Chebyshev grid f^c_m. The Chebyshev coefficients are given on this grid by performing the DCTII. Various approaches to resampling points between the two grids have been proposed, including polynomial interpolation [48], rational interpolation [21], and kernel methods [49].

For efficiency, we choose to use linear interpolation. For a given Chebyshev point x^c_k define the interval $[x^s_k, x^s_{k+1}]$ to be the shortest interval with endpoints in the sample grid such that x^c_k is contained in the interval. Then linear interpolation gives

$$f(x^c_k) = f(x^s_k)\left(\frac{x^s_{k+1} - x^c_k}{x^s_{k+1} - x^s_k}\right) + f(x^s_{k+1})\left(\frac{x^c_k - x^s_k}{x^s_{k+1} - x^s_k}\right) + r_k \tag{2.19}$$

where the error term by Rolle's Theorem is bounded by [48],

$$|r_k| \leq \frac{(x_{k+1}^s - x_k^s)^2}{8} \max_{x \in [x_{k-1}^s, x_k^s]} |f''(x)| \tag{2.20}$$

and scales as the square in the sample spacing and with the smoothness of the function through the second derivative term. The operations required for the interpolation step are linear in the number of Chebyshev samples, meaning that the dominant computational cost is still the $\mathcal{O}(N, \log N)$ computation.

2.2.3 Noise robustness of DCT interpolated coefficients

If instead of $\mathbf{f}$ we use samples $\tilde{\mathbf{f}}$ then the interpolated sample vector will be given by $\tilde{\mathbf{f}}^c = \mathbf{f}^c + \tilde{\boldsymbol{\epsilon}} + \mathbf{r}$, with two sources of error; the interpolated noise $\tilde{\boldsymbol{\epsilon}}$ and the interpolation error $\mathbf{r}$. The interpolated noise $\tilde{\boldsymbol{\epsilon}}$ is given by,

$$\tilde{\epsilon}_k = \epsilon_k \left(\frac{x_{k+1}^s - x_k^c}{x_{k+1}^s - x_k^s} \right) + \epsilon_{k+1} \left(\frac{x_k^c - x_k^s}{x_{k+1}^s - x_k^s} \right). \tag{2.21}$$

Under the assumption that the ϵ_k are independent and identically distributed Normal random variables with mean 0 and variance σ^2, the interpolated noise will also be a random variable with modified variance

$$\tilde{\sigma}^2 = \sigma^2 \left(\frac{\alpha^2 + (\Delta - \alpha)^2}{\Delta^2} \right) \leq \sigma^2 \tag{2.22}$$

where $\Delta = x_{k+1}^s - x_k^s$, $\alpha = x_{k+1}^s - x_k^c$. The bound comes from maximizing the quadratic coefficient over $0 \leq \alpha \leq \Delta$.

When we use $\tilde{\mathbf{f}}^c$ to calculate Chebysehv coefficients using (2.14), by the linearity of the transform, the coefficient vector will be given by $\tilde{\mathbf{c}} = \mathbf{c} + \hat{\boldsymbol{\epsilon}} + \hat{\mathbf{r}}$. The first term is the true coefficients accurate to machine precision. The second term comes from the Chebyshev transform of the interpolated noise. Again assuming that the noise on each term is an independent Normal variable with mean 0 and variance $\tilde{\sigma}^2$, the transformed noise will also be a Normal random variable with zeros mean and new variance, $1 \leq k \leq N-1$,

$$\hat{\sigma}_k^2 = \frac{4}{N^2} \left[\frac{1}{4} + \frac{1}{4} + \sum_{n=1}^{N-1} \cos^2 \left(\frac{\pi n k}{N} \right) \right] \tilde{\sigma}^2 = \frac{4}{N^2} \frac{N-1}{2} \tilde{\sigma}^2 \approx \frac{2}{N} \tilde{\sigma}^2. \tag{2.23}$$

The Chebyshev transform decreases the variance in the noise; the mean magnitude of the noise is $\mathbb{E}[|\hat{\epsilon}_k|] = \sqrt{4/\pi N}\tilde{\sigma}$. The third term comes from the Chebyshev transform of the

interpolation error. The triangle inequality can bound the interpolation error sum,

$$\hat{r}_k \leq \frac{2}{N}\left[\sum_{n=1}^{N-1}\left|\cos\left(\frac{\pi nm}{N}\right)\right||r_n| + \frac{|r_0|+|r_N|}{2}\right] \leq \frac{2}{N}\left[\sum_{n=1}^{N-1}|r_n| + \frac{|r_0|+|r_N|}{2}\right] \leq \frac{2}{N}\|\mathbf{r}\|_1. \tag{2.24}$$

In practice, however, the interpolation error will not saturate the bound, and numerically, we see similar behavior to the noise term with a factor of $\sqrt{N}$ decrease in the error, which can be justified by thinking of the interpolation error as a random variable with variance proportional to the upper bound.

Similar results for Fourier representations and, more generally, any quadrature-based coefficient calculation are possible. The results above show that we can efficiently and robustly use spectral representations for many types of noisy data where function samples are provided on a dense enough grid and the underlying function is sufficiently smooth such that $\hat{r}_k \ll \hat{\epsilon}_k \ll c_k$ for small k. The interpolation step does not introduce significant errors while enabling efficient computations.

2.2.4 Noise reduction of spectral representations

A key property of spectral representations that makes them useful for data analysis is their ability to separate information about a function's smooth structure – contained in the low-degree coefficients – from noise in the measurement – which dominates higher-degree coefficients where the contribution from the smooth structure has become negligible. This enables us to remove noise from data by thresholding the coefficients in the spectral representation.

Given full information about the underlying function and the noise statistics, the results in the previous section may be used to determine the transition point between coefficients that capture the structure and those that are noise-polluted and should, therefore, be thresholded. In practice, however, we do not have either piece of information. We, therefore, have to come up with heuristic arguments to determine where to truncate the spectral expansion. Heuristics have to be used even in the noise-free case; see, for example, [50] for a detailed discussion. In general, we wish to balance a few key properties,

- Accuracy in the representation with compression in the number of coefficients retained;
- Structure in the underlying function with noise reduction.

If we plot the square error (2.18) as a function of the number of the number of coefficients used in the expansion N, we would expect to see two distinct regimes in the curve. Initially, the error will be dominated by the large coefficients that are not retained, and the error

should decay geometrically ρ^{-N} if the underlying function is sufficiently smooth. Eventually, the noise will dominate $\rho^{-N} \ll \epsilon$, and the error will begin to decay linearly in N to some fixed level. In general, the square error will not decrease to machine precision due to the interpolation errors. This suggests a method to extract thresholds. We choose a cutoff in the elbow of the error curve where we do not retain coefficients associated with the noise. We can identify the start of this noise regime by looking for regions where the error is approximately linear on a log-log plot. We show applications of this approach to experimental data in Chapter 3.

Chapter 3

Applications of spectral representations to experimental data

3.1 Application to Rho-GTP imaging data

The contents of this section have been adapted from the publication, J. Liu, J. F. Totz, P. W. Miller, A. D. Hastewell, Y. C. Chao, J. Dunkel, and N. Fakhri, "Topological braiding and virtual particles on the cell membrane" Proceedings of the National Academy of Sciences, 118(34), e2104191118, 2021.

As a first application we apply spectral representations to experimental imaging data from Nikta Fakhri's lab performed by J. Liu and Y. C. Chao. The resulting spectral representations were then used in further analyses by J. Liu, J. F. Totz and P. W. Miller.

3.1.1 Motivation

Braiding confers remarkable robustness to static and dynamic structures, from plaited hair and fabrics [51] to the entangled worldlines of classical [52] and quantum particles [53]. Stabilized by an inherent topological protection, braided threads, ropes, and wires have long been used to transmit forces and shield signals [54]. Over the last decade, dynamic braiding processes [55–57] have attracted major interest in soft matter [58, 59] and quantum physics [53] as promising candidates for robust information storage and processing [60, 61]. A widely studied application is topological quantum algorithms that perform computations by braiding the worldlines of 2D quasiparticle excitations [53, 60, 61]. Of similar importance to information processing in living systems – albeit much less well understood – is the braiding dynamics of chemical spiral wave signals on cell membranes, which control a wide range of developmental and physiological functions, including cell division [62], cardiac rhythm [63–66]

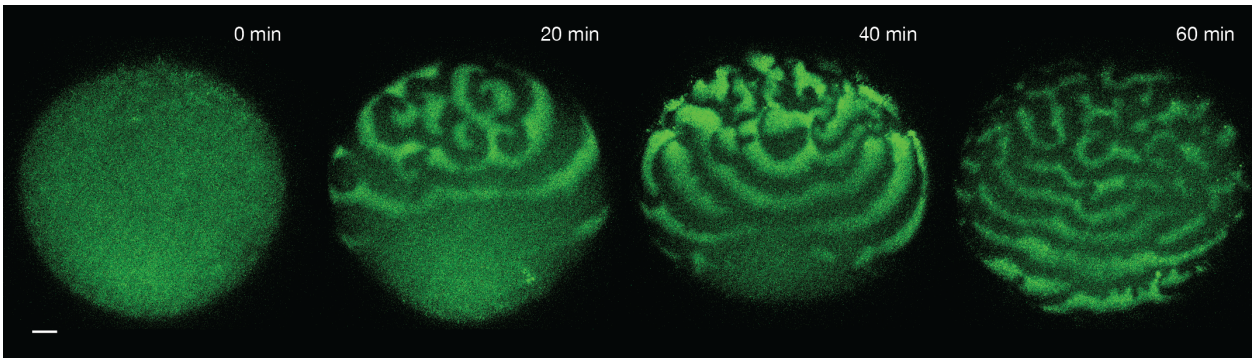

Figure 3.1: Time evolution of chemical Rho signaling wave patterns on the starfish oocyte from a homogeneous initial state to a quasi-steady state exhibiting turbulent spiral patterns. Snapshots show maximal intensity projections of three near-membrane Z-stack confocal slices spanning 5 µm. Scale bar: 40 µm. Experimental images provided by the Fakhri lab [70]. Note the visible presence of noise in the images on top of relatively smooth intensity profiles, lending these types of microscopic images well to spectral representation.

and brain activity [67]. These spiral waves belong to a rapidly expanding class of recently discovered biological phenomena [68, 69] in which topological structures serve as robust organizers of essential life processes.

To investigate the braiding dynamics of biochemical spiral waves in living cells, we consider experimental observations of Rho-GTP activation waves, a highly conserved signaling protein [71, 72] across a wide variety of eukaryotic species [73], on starfish oocyte membranes performed by J. Liu and Y. C. Chao in the Fakhri lab [4, 70](Figure 3.1). Since the biological functions of Rho-GTP have been widely investigated previously [74], we focused here on the topological characterization of the biochemical signaling dynamics. Different steady-state patterns of Rho-GTP are induced by a systematic increase of the GEF (guanine exchange factor) responsible for activating Rho-GTP (experimental states i–v) [75]. Rho-GTP waves maintained constant oscillatory periods within all observed non-equilibrium steady states with different cellular activity [4], enabling the reconstruction of spatiotemporal phase fields. We compare the experimental phase fields and the resulting defect statistics to predictions from the complex Ginzburg-Landau (CGL) equation [76]

$$\partial_t \psi = \psi - (1 + \mathrm{i}c)|\psi|^2\psi + (1 + \mathrm{i}b)\nabla^2\psi \ ,$$

a generic model describing the spatio-temporal evolution of oscillatory continuum systems $\psi(\mathbf{x}, t)$ near a Hopf bifurcation [70].

Overcoming previous observational and algorithmic limitations, we achieved the spatiotemporal resolution required for dynamical analysis by combining *in vivo* imaging with

spectral signal representation, quantitative mathematical modeling, and large-scale computational parameter estimations. Full details about the experimental setup and computational modeling can be found in [70]. Here, we detail the spectral representation step that enabled the extraction of phase fields and defect tracking from the experimental movies for comparison with theoretical predictions.

3.1.2 Spectral representations of microscopic imaging data.

We constructed the phase field from microscope images using the following procedure: Raw time-lapse Z-stack images (covering ∼5 µm near the membrane) were first combined into a single intensity field video through a maximal intensity projection. The rectangular intensity field video, $I(\mathbf{x}, t)$, was extracted from the experimental intensity video by least square fitting an ellipse to the boundary of the oocyte. The data was then rotated to axis align the major and minor axes of the ellipse. We extract the rectangular region by choosing the largest rectangle inscribed in the ellipse. Following the approach outlined in Chapter 2, we use the tensor product basis to construct a multidimensional basis (Section 2.1.3); the (2+1)D space-time data cube with N_x, N_y and N_t pixels in each dimension respectively is represented as a sum over basis functions, Chebyshev polynomials of the first kind in space (Section 2.1.2) and Fourier in time (Section 2.1.1), assuming N_t is even,

$$I(\mathbf{x}, t) = \sum_{n=0}^{N_x-1} \sum_{m=0}^{N_y-1} \sum_{k=-N_t/2}^{N_t/2-1} c_{n,m,k} T_n(x) T_m(y) e^{2\pi i k t/N_t}. \tag{3.1}$$

The coefficients $c_{n,m,k}$ are found following the linear interpolation approach outlined in Section 2.2 generalized to multiple dimensions. The coefficients can be calculated efficiently using fast algorithms for the Discrete Cosine Transform and the Discrete Fourier Transform [41, 44] applied along their respective dimensions. These transforms assume that the function is sampled on the Chebyshev grid in both spatial dimensions and uniformly spaced in time. The data is converted from the uniformly spaced experimental grid to the Chebyshev grid using linear interpolations. Given the noise in the images, we expect that the errors introduced in the coefficients from the linear interpolations will be of lower order than the errors already introduced from the noise in the data (Section 2.2).

From Figure 3.1, we expect that the image $I(\mathbf{x}, t)$ will be well approximated by a smooth differentiable function $f(\mathbf{x}, t)$ with additive noise. We, therefore, expect that we can remove noise and get a smooth representation of the data by cutting off the summation in (3.1) at

the thresholds M_c and M_f to get the new denoised representation of the data (Section 2.2.4),

$$f(\mathbf{x},t) \approx \sum_{n,m=0}^{(Ny/Nx)n+m\leq M_c} \sum_{k=-M_f}^{M_f} c_{n,m,k} T_n(x) T_m(y) e^{2\pi i k t/N_t}. \tag{3.2}$$

We define the threshold M_c to include all combined spatial polynomial basis functions below a given total degree. The N_y/N_x weighting is incorporated to account for the different sizes of the two spatial dimensions, leading to different resolutions of polynomials with the same degree along the two different dimensions. The thresholds M_c and M_f are chosen by sweeping over all possible thresholds $0 \leq M_c \leq N_y$ and $1 \leq M_f \leq N_t/2$ and calculating the relative reconstruction error,

$$\mathcal{E}(M_c, M_f) = \frac{\sqrt{\sum_{\mathrm{x},t}(I(\mathbf{x},t) - f(\mathbf{x},t))^2}}{\sqrt{\sum_{\mathbf{x},t} I(\mathbf{x},t)^2}} \tag{3.3}$$

and the effective compression at each point. The effective compression is defined by the mean of the spatial and temporal compression $\mathcal{C} = 0.5(\mathcal{C}_s + \mathcal{C}_t)$ where $\mathcal{C}_t = (2M_f + 1)/(2N_t + 1)$ and $\mathcal{C}_s = (2 + 2M_c - N_y)(1 + N_x)/(2N_xN_y)$. A scatter plot of compression versus reconstruction error shows a front corresponding to the points of best reconstruction error for a given compression (Figure 3.2). These correspond to points where the reconstruction error and compression contours are tangent. Following the discussion in Section 2.2.4, we expect the error to show two different regimes: a rapid decay as the compression is initially increased, followed by a slower decrease when the noise starts to dominate. Therefore, we choose the threshold at the start of the elbow of this error compression front, which is determined by fitting a line to the linear part of the curve on a log-log plot.

Validation on mock CGL data

The method was first tested on mock CGL data with added Poisson noise of varying strengths to approximate experimental noise. Figure 3.2(b, c) illustrates the process for choosing the thresholds, and Figure 3.2(a) shows a comparison of the resulting reconstruction to the noisy input image and the true image for one noise strength.

The phase was calculated from the spectral data representation. Let τ be approximately a quarter of the oscillation period of the oscillating time signal of each pixel. First, the moving center, $\bar{I}(t)$, calculated by convolving the signal with a Gaussian window of width τ, was subtracted from each pixel. This centers the pixel time signal around 0. Next we computed the analytic extension of each pixel trajectory $(I - \bar{I})(t)$ into the complex plane, $z(t) = (I - \bar{I})(t) + \mathrm{HT}[(I - \bar{I})(t)]$, where $\mathrm{HT}[(I - \bar{I})(t)]$ is the Hilbert transform of the signal

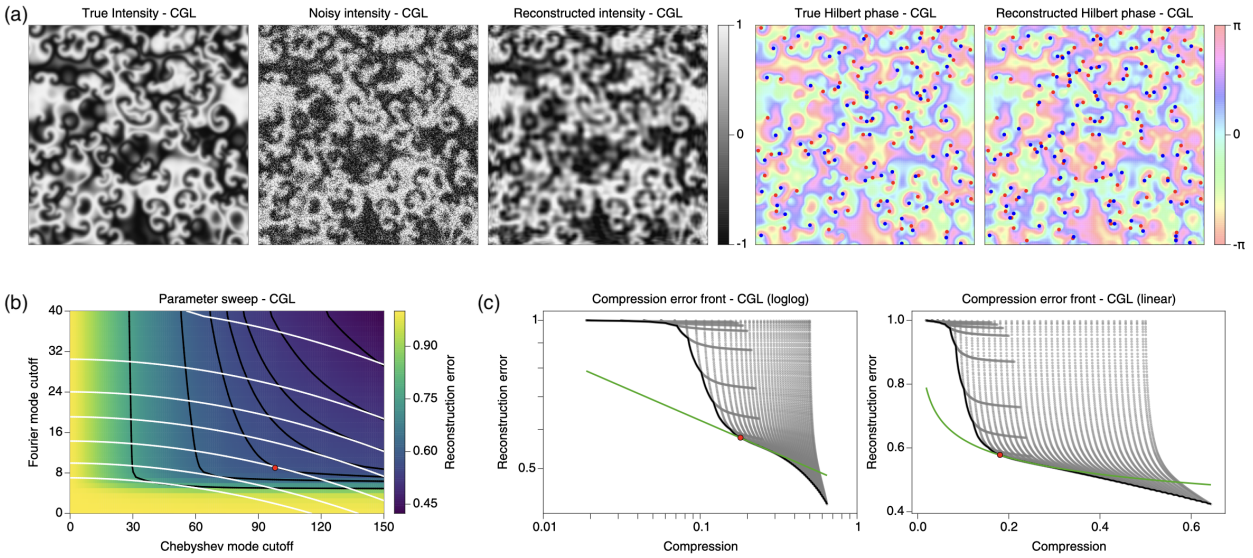

Figure 3.2: (a) Truncated spectral representation (middle) shows significant noise reduction compared to noise-polluted input data (second left) while retaining the structure of the true underlying field (left) for simulated CGL test data with added Poisson noise. A comparison of the extracted phase field from the true CGL data (second right) and the denoised reconstructed field (right) shows they contain the same structure and very similar defect statistics. (b) Parameter sweep over possible Fourier and Chebyshev mode cutoffs for the CGL test data. Reconstruction error contours (black lines) and compression contours (white lines) are shown. Points where the two contours are tangent correspond to points on the error compression front. (c) Compression error front extracted from the parameter sweep in (b) along with the linear fit to the log-log curve (green line) and the chosen cutoff point (red dot)

[77]. In the frequency domain, this corresponds to setting all coefficients corresponding to negative frequencies to 0 and doubling those corresponding to positive frequencies. For a discrete signal of length N with discrete Fourier coefficients $\hat{I}_k$ this is given by

$$\hat{z}_k = \begin{cases} 2\hat{I}_k & \text{for } 1 \leq k \leq \frac{N}{2} - 1 \\ 0 & \text{for } \frac{N}{2} + 1 \leq k \leq N - 1 \\ \hat{I}_k & \text{for } k = 0 \text{ or } k = \frac{N}{2} \end{cases}. \tag{3.4}$$

We then define the phase of the signal $\phi(t)$ as the angle of the analytic signal in the complex plane $\phi(t) = \tan^{-1}(\mathrm{HT}[(I - \bar{I})(t)], (I - \bar{I})(t))$. The phase field is extracted for each pixel in the image to obtain the time-lapse phase field, $\phi(\mathbf{x}, t)$. Figure 3.2(a, right) compares the phase extracted using the Hilbert phase between the true CGL simulation data and the denoised spectral representation. We see that the denoised phase closely matches the true phase, highlighting the ability of spectral representations to extract quantitative quantities

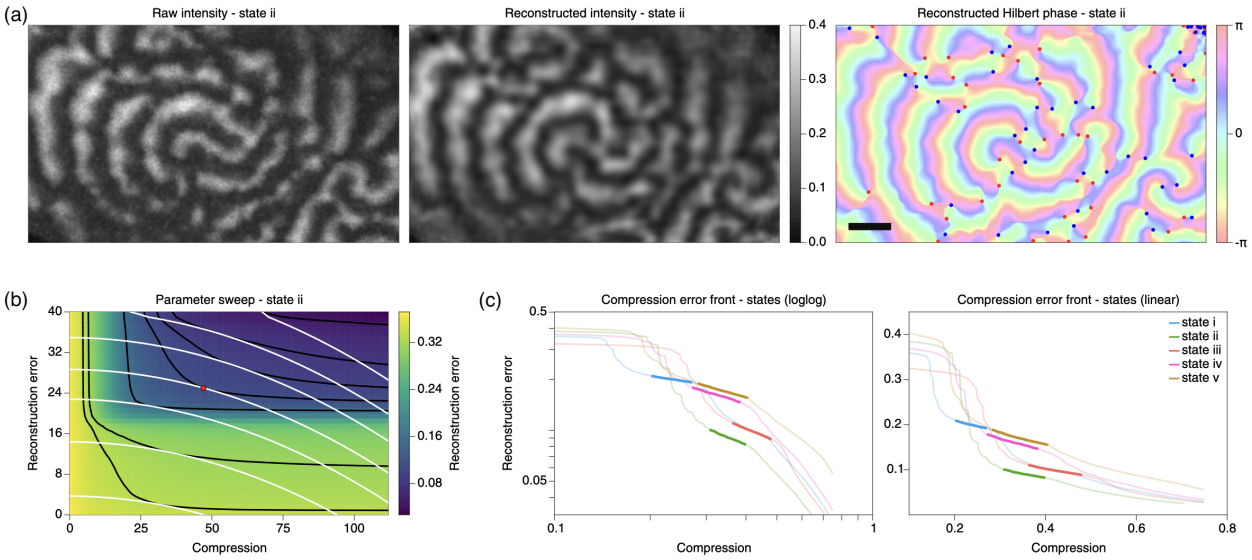

Figure 3.3: (a) Comparison of an experimental snapshot with the corresponding data and phase field representation (3.2) for state ii shows significant noise reduction while maintaining the main underlying structure. The extracted phase using the Hilbert transform is shown on the right. Scale bar 20 μm. (b) Parameter sweep over possible Fourier and Chebyshev mode cutoffs for state ii. Reconstruction error contours (black lines) and compression contours (white lines) are shown. (c) Compression error fronts for the experimental states. We see that all the curves have a similar shape across all states. The linear fit to each of the experimental fronts on the log-log graph is shown with thick lines.

from noisy input data robustly.

Application to experimental data

For the experimental data, the same parameter sweeps were performed for each state, and a cutoff was chosen for each state [Figure 3.3(a–c)]. Once we have the basis representation for the experimental data, we can then evaluate the expansion in (3.2) at any points (x, y, t) inside the domain interpolating the data at higher resolution. Evaluating the representations at higher resolutions provides increased robustness for defect extraction and tracking. The denoised expansions were, therefore, evaluated at 2 times higher spatial resolution and at 1 frame per second temporal resolution before calculating the phase field and extracting defects. Raw videos have frame rates of 10 s to 12 s and pixel resolution of 0.625 µm. Figure 3.4(a) shows the reconstructed intensity and phase field snapshots for all experimental states. An example of the phase extraction for an individual pixel is shown on the right.

The phase defects were subsequently tracked based on the reconstructed phase field. To locate the positions of phase defects harbored in the phase field of an instantaneous time frame, $\phi(\mathbf{x})$, we performed a line integral $\oint_C \nabla\phi(\mathbf{x}) \cdot d\mathbf{s}$ for every two-pixel-by-two-pixel local

window. Plus defects have a value of 2π, and minus defects have a value of -2π. We used particle-tracking software to retrieve the trajectories of phase defects through continuous time frames (see [70] for details). An example of the defect trajectories that were extracted from the spectral representation is shown in Figure 3.4(b, c).

3.1.3 Conclusions

The defect statistics extracted from the spectral representations of the experimental images enable the study of topological braiding of the defects and the quantitative matching of the experiments to parameters in the CGL equation; see [70] for further details. In summary, our quantitative analysis showed that these defects exhibit complex braiding, pair creation, and annihilation dynamics. Experimentally measured worldline braiding exponents and topological entropy correlate with cellular activity and agree with CGL theory predictions. More broadly, our analysis reveals the creation and annihilation of phase defects during scattering events, suggesting phenomenological parallels between information transport in quantum and living matter. Spectral representations played a critical role in providing robust measurements of the quantities needed to compare with the theory and observe these phenomena, demonstrating the power of spectral representations for analyzing high-resolution microscopic imaging data.

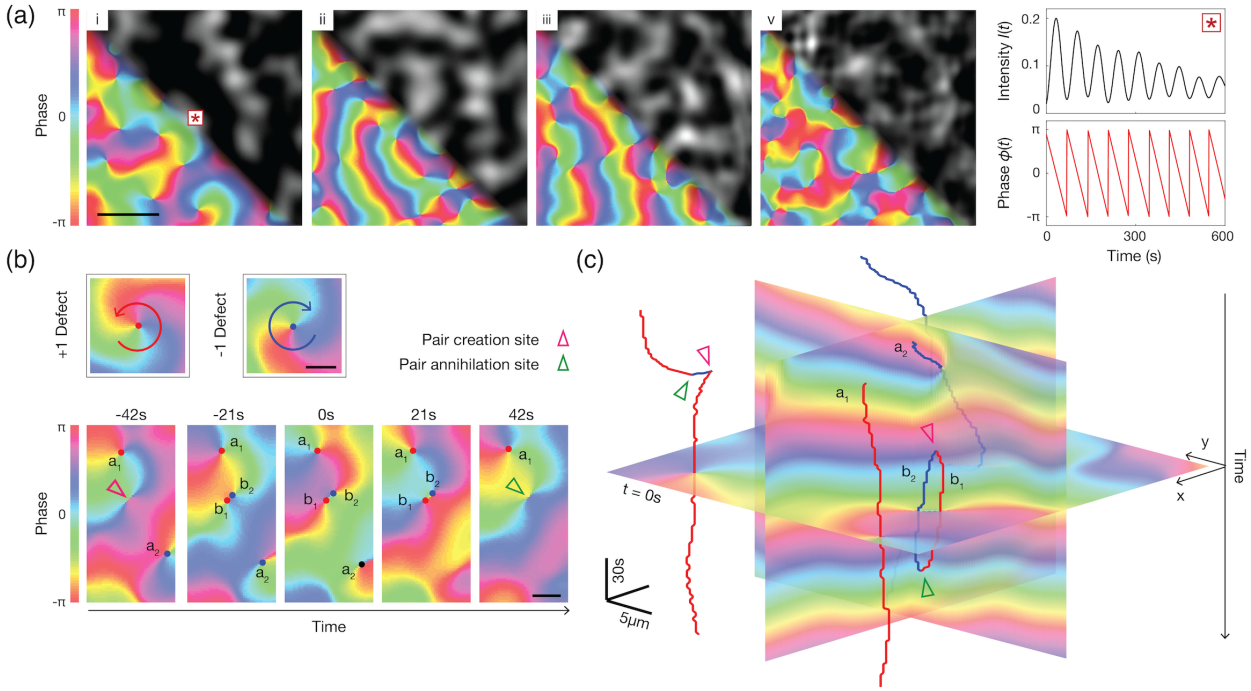

Figure 3.4: (a) Quasi-steady wave patterns ($t > 60$ min) of Rho-GTP intensity field from four starfish egg cells, aligned with phase fields reconstructed from oscillations in pixel fluorescence intensity signal (inset). Scale bar: 20 µm. (b) The reconstructed phase fields harbor topological defects of winding number +1 (red, counter-clockwise rotating spiral core) and -1 (blue, clockwise rotating spiral core). Lower: Time-lapse snapshots of localized creation (annihilation) events that produce (destroy) oppositely charged defects in pairs. Scale bar: 5 µm. (c) Worldline representation of topological defects embedded in 2+1-dimensional phase field. The time-lapse snapshots in (b) correspond to the formation of the simple space-time loop in (c). Scale bars: 5 µm; 30 s (vertical).

3.2 Application of spectral methods to coarse grained active particle dynamics

The contents of this section have been adapted from R. Supekar, B. Song, A. D. Hastewell, G. P. Choi, A. Mietke, and J. Dunkel "Learning hydrodynamic equations for active matter from particle simulations and experiments," Proceedings of the National Academy of Sciences, 120(7), e2206994120, 2023.

For a second application of spectral representations, we consider coarse-grained active particle data. We use spectral representations to evaluate high-quality derivatives of input data. Additionally, we use the spectral representations to justify the selection of coarse-graining length scale. The results are applied to particle simulations by B. Song and experimental data from Denis Bartolo's and Iain Couzin's lab. The spectral representations are incorporated into the larger model inference framework developed by R. Supekar.

3.2.1 Introduction

Active systems, from self-propelled colloids to animal swarms, can form complex dynamical patterns as their microscopic constituents exchange energy and momentum with the environment. Reflecting this complexity, continuum models for active matter typically possess many more parameters than those of classical fluids like water. While much progress has been made in the qualitative understanding of active pattern formation, measuring the various hydrodynamic parameters of an active system still poses significant challenges.

Recent advances in high-resolution imaging techniques [6–8, 10] and agent-based computational modeling [78], mean active matter systems can now be observed and analyzed at unprecedented spatiotemporal [11, 79, 80] resolution. To infer interpretable predictive theories, the high-dimensional data recorded in experiments or simulations must be compressed and translated into low-dimensional models. Such models must faithfully capture the macroscale dynamics of the relevant collective properties. Macroscale properties can be efficiently encoded through hydrodynamic variables, continuous fields that are linked to the symmetries, and conservation laws of the underlying microscopic system [81, 82]. Although much theoretical progress has been made in the field of dynamical systems learning over the last two decades [27, 29–31, 34, 35, 83], the inference of hydrodynamic models and their parameters from particle data has remained largely unsuccessful in practice, not least due to severe complications arising from measurement noise, inherent fluctuations and self-organized scale-selection in active systems. We present a framework that leverages spectral basis representations and sparse regression algorithms to discover PDE models from microscopic

simulation and experimental data while incorporating the relevant physical symmetries.

3.2.2 Summary of the learning framework

The full learning framework is summarized in Figure 3.5 and consists of three main steps:

Coarse-graining. Spatial coarse-graining of time-series data for particle positions and orientations measured from experiments or simulations.

Spectral representation. Coarse-grained data is projected onto spectral basis functions, enabling accurate computation of spatiotemporal derivatives.

Sparse regression. Sparse regression on a library of candidate terms to determine hydrodynamic PDEs. The candidate terms contain derivatives of the coarse-grained fields, which are calculated from the data using the spectral representation. The resulting sparse PDE models are then validated by simulation to ensure stability and accuracy.

Full details about the particle simulations and regression framework can be found in [84]. Here, we provide details on the coarse-graining, specifically how to choose the correct coarse-graining length scale and the spectral representation used for library construction.

3.2.3 Selecting coarse-graining length scales

Given the particle positions $\mathbf{x}_i(t)$ and orientations $\mathbf{p}_i(t)$, an associated particle number density field $\rho(t, \mathbf{x})$ and polarization density field $\mathbf{p}(t, \mathbf{x})$ can be defined by convolution with a kernel $K(\mathbf{x})$,

$$\rho(t, \mathbf{x}) = \sum_i K[\mathbf{x} - \mathbf{x}_i(t)], \tag{3.5a}$$

$$\mathbf{p}(t, \mathbf{x}) = \sum_i K[\mathbf{x} - \mathbf{x}_i(t)]\, \mathbf{p}_i(t). \tag{3.5b}$$

The symmetric kernel $K(\mathbf{x})$ is centered at $\mathbf{x} = 0$ and normalized, $\int d^2\mathbf{x}\ K(\mathbf{x}) = 1$, so that the total number of particles is recovered from $\int d^2\mathbf{x}\, \rho(t, \mathbf{x}) = N$. We found that, in the context of hydrodynamic model learning, the coarse-graining equations (3.5) with a Gaussian kernel $K(\mathbf{x}) \propto \exp[-|\mathbf{x}|^2/(2\sigma^2)]$ are suitable and provide a useful preprocessing step that simplifies the use of fast transforms at later stages. The coarse-graining scale σ determines the spatial resolution of the hydrodynamic theory. In practice, σ must be chosen larger than the particles' mean-free path length or interaction scale to ensure the smoothness of the hydrodynamic fields but also smaller than the emergent collective structures. To understand

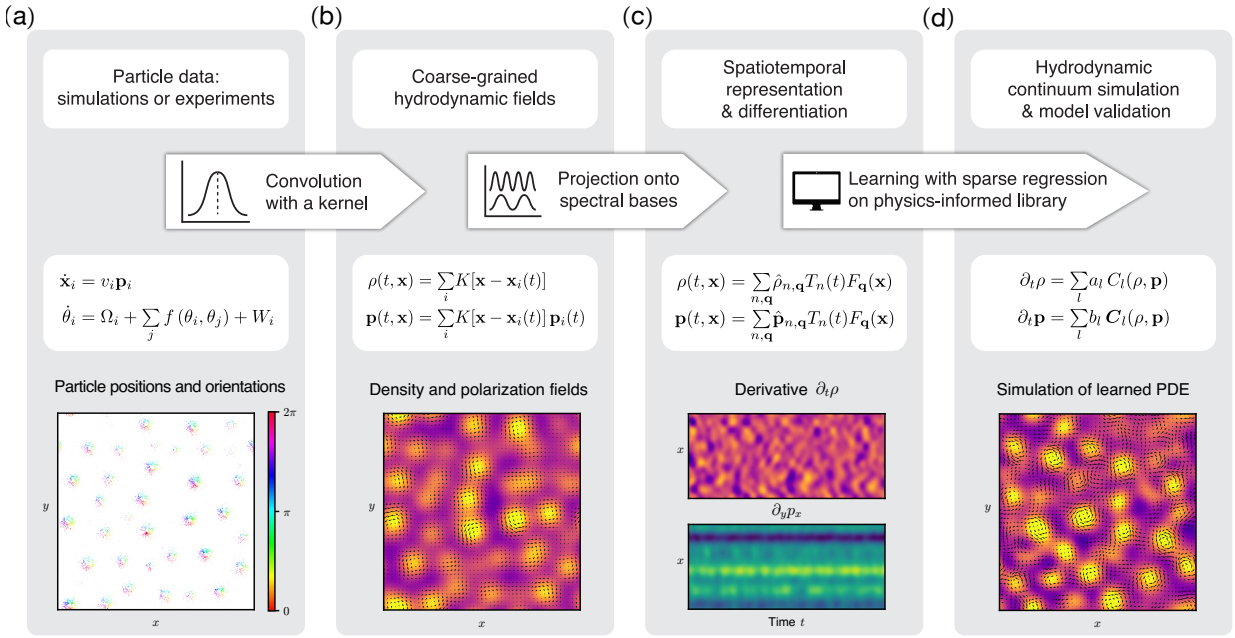

Figure 3.5: Learning hydrodynamic models from particle simulations and experiments. (a) Inputs are time-series data for particle positions $\mathbf{x}_i(t)$, and particle orientations $\mathbf{p}_i(t) = (\cos\theta_i, \sin\theta_i)^\top$, measured in simulations or experiments with microscale resolution. (b) Spatial kernel coarse-graining of the discrete microscopic variables provides continuous hydrodynamic fields, such as the density $\rho(t,\mathbf{x})$ or the polarization density $\mathbf{p}(t,\mathbf{x})$. (c) Coarse-grained fields are sampled on a spatiotemporal grid and projected onto suitable spectral basis functions. Systematic spectral filtering (compression) ensures smoothly interpolated hydrodynamic fields, enabling efficient and accurate computation of spatiotemporal derivatives. (d) Using these derivatives, a library of candidate terms $C_l(\rho,\mathbf{p})$ and $\boldsymbol{C}_l(\rho,\mathbf{p})$ consistent with prior knowledge about conservation laws and broken symmetries is constructed. A sparse regression algorithm determines subsets of relevant phenomenological coefficients a_l and b_l. The resulting hydrodynamic models are sparse and interpretable, and their predictions can be directly validated against analytic coarse-graining results or experiments. Bottom: Snapshots illustrating the workflow for microscopic data generated from simulations of chiral active Brownian particles [84].

suitable choices of σ, we can observe how the information loss due to coarse-graining changes as a function of σ.

To quantify the information loss due to coarse-graining as a function of the coarse-graining length scale σ, we use spectral entropy [85, 86] as a measure of the information content that remains in the coarse-grained fields. Specifically, we define the spectral entropy as

$$H(\sigma) = -\sum_{\mathbf{q}} \hat{S}^{(\sigma)}_{\rho;\mathbf{x}}(t,\mathbf{q}) \log_2 \hat{S}^{(\sigma)}_{\rho;\mathbf{x}}(t,\mathbf{q}), \tag{3.6}$$

where the normalized spatial power spectral density $\hat{S}^{(\sigma)}_{\rho;\mathbf{x}}$ is defined as

$$\hat{S}^{(\sigma)}_{\rho;\mathbf{x}}(t,\mathbf{q}) = S^{(\sigma)}_{\rho;\mathbf{x}}(t,\mathbf{q}) \left(\int d^2\mathbf{q}\, S^{(\sigma)}_{\rho;\mathbf{x}}(t,\mathbf{q}) \right)^{-1}, \tag{3.7}$$

with $S^{(\sigma)}_{\rho;\mathbf{x}}(t,\mathbf{q})$ the spatial power spectral density defined by

$$S^{(\sigma)}_{\rho;\mathbf{x}}(t,\mathbf{q}) = \mathcal{A}^{-1} \left| \int d^2\mathbf{x}\, \rho^{(\sigma)}(t,\mathbf{x}) \exp(2\pi i \mathbf{q}\cdot\mathbf{x}) \right|^2. \tag{3.8}$$

Here, $\mathcal{A}$ is the domain area. The index σ indicates the Gaussian kernel smoothing width ('coarse-graining length scale') with which the underlying density field $\rho^{(\sigma)}(t,\mathbf{x})$ was computed from the raw particle data. For our analysis, we rescale the spectral entropy H given in (3.6) by the spectral entropy of the raw particle data, yielding a normalized spectral entropy between 0 and 1. Figures 3.6 and 3.7 show example applications to simulation [84] and experimental Quinke roller [87] particle data. The shape of the curve $H(\sigma)$ reveals important information about the effect of the coarse-graining length scale. Plateaus in the curve suggest regions where changing σ does not affect the information content in the fields. Rapid changes in the curve indicate regions where changing σ changes the information stored in the fields. Characteristically, we see that there is an initial plateau where the coarse-graining is not obscuring the particle nature of the data (i in Figures 3.6 and 3.7). This plateau is followed by an initial decrease in the entropy when the particle nature of the data starts to become obscured, but sharp boundaries between high and low-density regions exist (ii in Figures 3.6 and 3.7). There is then a large decrease in the entropy when the boundaries start to become smoothed out, followed by a steady decrease in the entropy as the field becomes progressively smoother (iii – v Figures 3.6 and 3.7). Coarse-graining length scales should be chosen in this region (iv in the figures) to balance the smoothness of the fields with the collective structures.

Interestingly, measuring the spectral entropy at the chosen coarse-graining length scale

for both simulated and experimental data showed that coarse-grained hydrodynamic fields typically maintain only about 1% of the spectral information contained in the fine-grained particle data (Figure. 3.6 and 3.7).

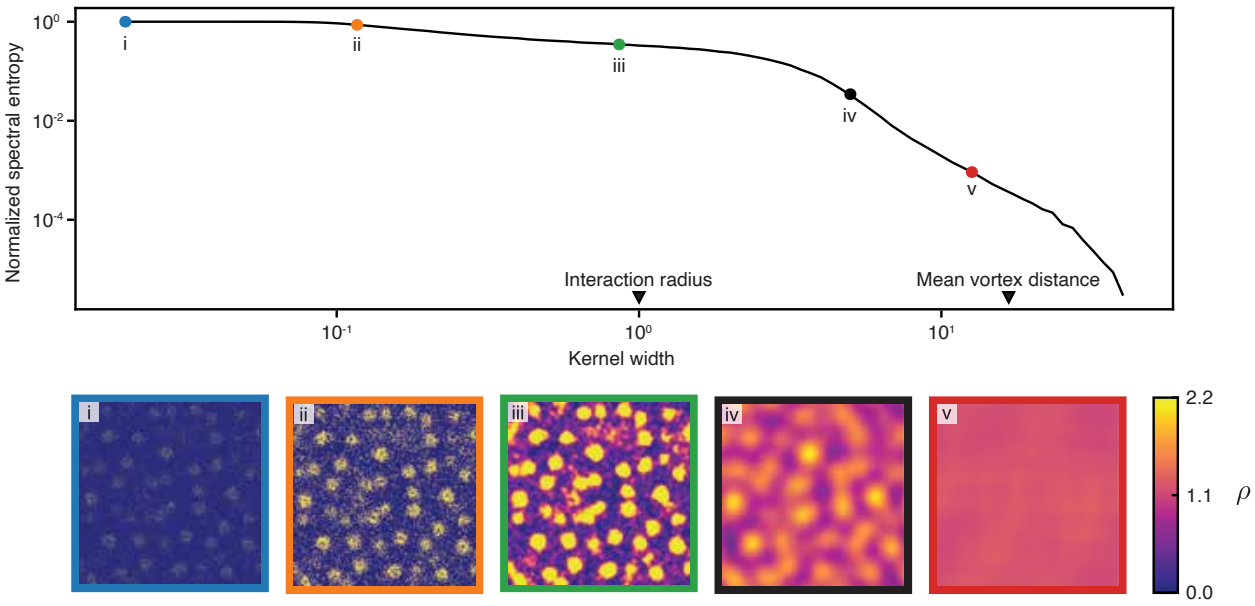

Figure 3.6: Normalized spectral entropy, Eq. (3.6), as a function of the Gaussian kernel width σ for the chiral particle model data (top) quantifies the fraction of information that remains in Fourier space after coarse-graining. Representative snapshots of coarse-grained fields are shown in the bottom panels. Characteristic scales in units of particle-particle interaction distance: Median vortex distance ~ 17 (obtained from a Delaunay triangulation of density peaks), box size 100. (i, $\sigma = 0.02$): Raw image before coarse-graining. (i, $\sigma = 0.02$)–(ii, $\sigma = 0.12$): The particle data's discrete nature remains present, leading to little information loss. (ii, $\sigma = 0.12$)–(iii, $\sigma = 0.86$): As the coarse-graining scale approaches the interaction length scale, $\sigma \to 1$, coarse-grained data starts losing single-particle information and vortices become more prominent than individual particles. (iii, $\sigma = 0.86$)–(iv, $\sigma = 5$): Vortices start to be smoothed out as σ exceeds the particle interaction distance and vortex size. Data from (iv, $\sigma = 5$) was used for inferring a continuum model from the chiral-particle simulation date; this choice of the coarse-graining scale ensures that the hydrodynamic fields are sufficiently smooth while still containing sufficient information about density fluctuations and vortex patterns. (v, $\sigma = 12.6$): When the kernel width σ approaches the typical vortex-vortex distance, coarse-graining results in a constant homogeneous density, and all spatially heterogeneous information is lost.

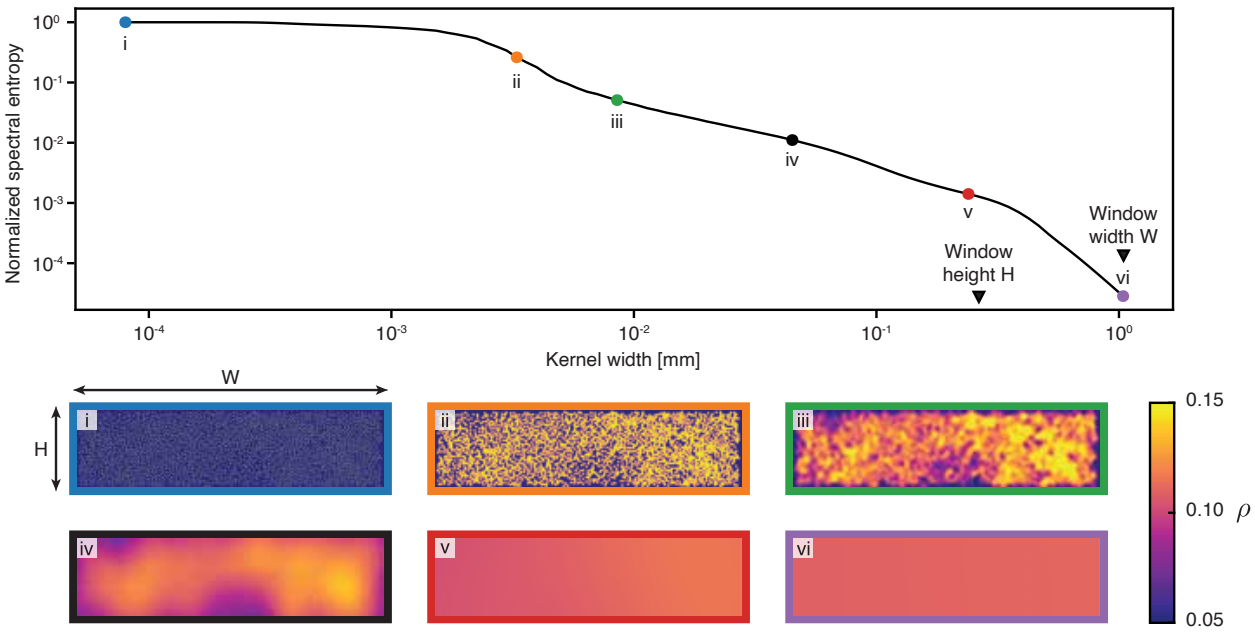

Figure 3.7: Normalized spectral entropy, Eq. (3.6), as a function of the Gaussian kernel width σ for the Quincke roller system (top) quantifies the fraction of information that remains in Fourier space after coarse-graining. Representative snapshots of coarse-grained fields are shown in the bottom panels. Characteristic length scales: Roller diameter $4.8\,\mu$m, mean roller-roller centroid distance $\sim 11\mu$m, window height $H = 0.286\,$mm, window width $W = 1.146\,$mm. (i, $\sigma = 10^{-5}\,$mm): Raw image before coarse-graining. (ii, $\sigma = 0.0033\,$mm)–(iii, $\sigma = 0.0085$mm): Single rollers are increasingly smoothed out, leading to an initial decrease in information. (iii, $\sigma = 0.0085\,$mm)–(iv, $\sigma = 0.045\,$mm): Large-scale density fluctuations become increasingly smoothed out by the coarse-graining. Data from (iv, $\sigma = 0.045\,$mm) was used for model learning from experimental Quincke roller data, providing a compromise between sufficiently smooth data and well-resolved details of density fluctuation in both spatial directions. (v, $\sigma = 0.24\,$mm): As the coarse-graining scale σ becomes comparable to the window height H, density fluctuations in the vertical direction have been smoothed out, leading to an effective one-dimensional density pattern that varies only along the horizontal direction. (vi, $\sigma = 1.04\,$mm): As σ becomes comparable to the window width W, all density variations disappear, and the coarse-graining yields a constant homogeneous density.

3.2.4 Spectral representations for library construction

A central challenge in PDE learning is the computation of spatial and temporal derivatives of the coarse-grained fields. Our framework exploits that hydrodynamic models aim to capture the long-wavelength dynamics of the slow collective modes [81]. As discussed in Chapter 2, this allows us to project the coarse-grained fields on suitable basis functions that enable sparse representations (high compression), fast transforms, and efficient differentiation. Here, we work with representations of the form

$$\rho(t, \mathbf{x}) = \sum_{n,\mathbf{q}} \hat{\rho}_{n,\mathbf{q}} T_n(t) F_{\mathbf{q}}(\mathbf{x}), \tag{3.9a}$$

$$\mathbf{p}(t, \mathbf{x}) = \sum_{n,\mathbf{q}} \hat{\mathbf{p}}_{n,\mathbf{q}} T_n(t) F_{\mathbf{q}}(\mathbf{x}), \tag{3.9b}$$

where $T_n(t)$ denotes a degree-n Chebyshev polynomial of the first kind [23, 44] (Section 2.1.2), and $F_{\mathbf{q}}(\mathbf{x})$ is a tensor product spatial basis (Section 2.1.3). Generally, the choice of the spatial basis functions should be adapted to the spatiotemporal boundary conditions of the microscopic data. For periodic simulation data we use the Fourier basis $F_{\mathbf{q}}(\mathbf{x}) = \exp(2\pi i \mathbf{q}\cdot\mathbf{x})$ with wave vector $\mathbf{q} = (q_x, q_y)^\top$ (Section 2.1.1), where $\hat{\rho}_{n,\mathbf{q}}$ and $\hat{\mathbf{p}}_{n,\mathbf{q}}$ are complex mode coefficients. For non-periodic experimental data we use the Chebyshev basis $F_{\mathbf{q}}(\mathbf{x}) = T_{q_x}(x)T_{q_y}(y)$, where $\hat{\rho}_{n,\mathbf{q}}$ and $\hat{\mathbf{p}}_{n,\mathbf{q}}$ are real mode coefficients. Since the coarse-graining can be evaluated at arbitrary points in space, the fast transforms can be applied directly to the spatial dimensions. In time, the coefficients are found using the interpolation method discussed in Section 2.2.

The spectral representation (3.9) enables the efficient and accurate computation of space and time derivatives [88] as discussed in Chapter 2. Preprocessing via spatial coarse-graining ensures that the mode coefficients $\hat{\rho}_{n,\mathbf{q}}$ and $\hat{\mathbf{p}}_{n,\mathbf{q}}$ decay quickly for $|\mathbf{q}| \gg 1/(2\pi\sigma)$ (Figure 3.8(a), left). If the asymptotic decay of the mode amplitudes with the temporal mode number n is at least exponential, then deterministic PDE descriptions are sufficient, whereas algebraically decaying temporal spectra indicate that stochastic PDEs may be required to capture essential aspects of the coarse-grained dynamics. For the simulated and experimental systems considered in this work, temporal spectra were found to decay exponentially [Figure 3.8(b)] or super-exponentially (Figure 3.9), suggesting the existence of deterministic PDE-based hydrodynamic models. To infer such models from data, we focus on the slow hydrodynamic modes and filter out the fast modes with $n > n_0$ by keeping only the dominant Chebyshev terms in (3.9). The cut-off value n_0 can usually be inferred from a characteristic steep drop-off in the power spectrum of the data, which signals the transition to hydrodynamically irrelevant fast fluctuations [50] [Figure 3.8(a, right)]. Choosing n_0 according to

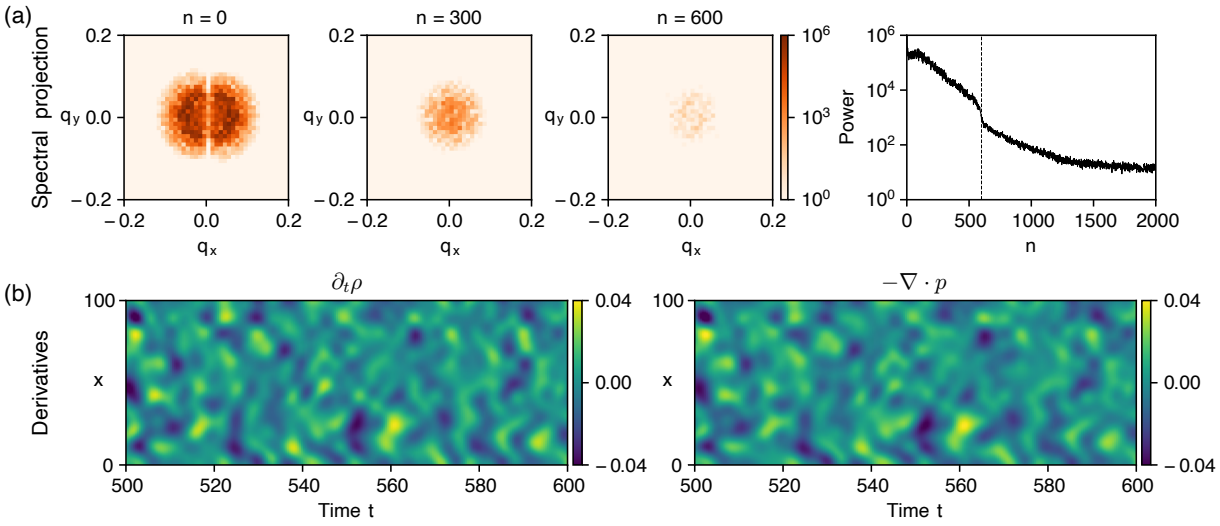

Figure 3.8: (a) Slices through the spatio-temporal power spectrum $S_{x;n,\mathbf{q}} = |\mathbf{e}_x \cdot \hat{\mathbf{p}}_{n,\mathbf{q}}|^2$ for different values of the Chebyshev polynomial order $n \in \{0, 300, 600\}$, corresponding to modes with increasing temporal frequencies. The rightmost panel depicts the total spatial spectral power $\sum_{\mathbf{q}} S_{x;n,\mathbf{q}}$ [see Eq. (3.9b)] of each Chebyshev mode n. The slowly decaying long tail of fast modes indicates a regime in which fluctuations dominate over a smooth signal. The cut-off $n_0 = 600$ removes these modes, which is in line with the goal of learning a hydrodynamic model for the slow, long-wavelength modes. (b) Kymographs of the spectral derivatives $\partial_t \rho$ and $-\nabla \cdot \mathbf{p}$ at $y = 50$, obtained from the spectral representation of the data are spatiotemporally consistent.

this criterion yields accurate, spatiotemporally consistent derivatives as illustrated for the kymographs of the derivative fields $\partial_t \rho$ and $-\nabla \cdot \mathbf{p}$ as seen in Figure 3.8(b). More generally, combining kernel-based coarse-graining and spectral representations mitigates measurement noise, enabling a direct application to experimental data.

3.2.5 Conclusions

Leveraging spectral representations of field observables and recent advances in the sparse PDE-inference [27, 31, 32, 83], we presented a PDE learning framework that robustly identifies hydrodynamic models for the self-organized dynamics of active matter systems [84]. The quality of the derivatives was essential for the inference framework to discover consistent PDEs [84] robustly. Spectral representations contributed to the robustness of the overall framework, allowing for insights into coarse-graining length scales from the spatial coefficients and the form of the temporal dynamics from the decay of temporal coefficients. Additionally, spectral representations enabled the robust and efficient extraction of derivatives from experimental data, providing direct comparisons between experimental data and

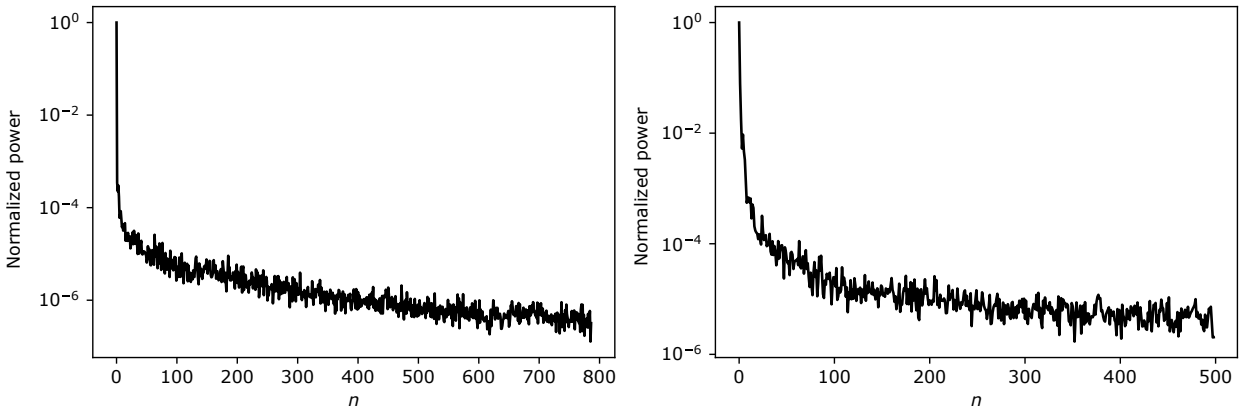

Figure 3.9: Power spectra of coarse-grained data for experimental applications to tracked particle data: the Quincke roller system from Denis Bartolo's lab [87] (left) and for sunbleak fish from Iain Couzin's lab [89] (right), where n denotes temporal Chebyshev mode numbers. The normalized power shown in the two panels is given by $S_{x;n}/S_{x;n=0}$. For both the Quincke roller and the sunbleak fish data spectral powers decay exponentially with increasing temporal mode number n.

inferred theoretical models.

3.3 Applications to gene expression profiles during bacterial swarming

The results in this section are adapted from H. Jeckel, K. Nosho*, K. Neuhaus, A. D. Hastewell, D. J. Skinner, D. Saha, N. Netter, N. Paczia, J. Dunkel and K. Drescher, "Simultaneous spatiotemporal transcriptomics and microscopy of Bacillus subtilis swarm development reveal cooperation across generations," Nature Microbiology, 8(12), 2378-2391, 2023*

As a final application of spectral representations we consider spatiotemporal gene expression data during bacterial swarming from by Knut Drescher's lab [12]. The results of the spectral representation were incorporated into further analyses by H. Jeckel and K. Nosho.

3.3.1 Motivation

Dynamically evolving microbial communities with spatial structure are ubiquitous in nature, from the intestinal microbiota in humans to soil-based biofilms and bacterial swarms expanding across moist surfaces [90–94]. Spatiotemporal patterns in microbial communities can emerge for the arrangement of genotypes, for phenotypic subpopulations of the same genotype, and for emergent community properties, such as resource gradients, biophysical properties, and stress tolerance [95–99]. Pattern formation in microbial communities and other multicellular systems is a complex multi-scale process influenced by cellular growth, division, differentiation, and motility, as well as many types of chemical and physical cell-cell interactions. All these factors can change in space and time due to the varying resource availabilities within developing communities [100, 101]. Even for the simplest microbial communities, such as single-species bacterial swarms, the number and spatiotemporal variability of the parameters influencing community development lead to a degree of complexity that makes it difficult to disentangle which intracellular processes and cellular interactions determine the emergent spatial structure of the community. To understand how spatial structure arises during bacterial community development, multi-scale spatiotemporal measurements of intracellular states, cellular phenotypes, and multicellular structures are required.

Using *Bacillus subtilis* swarm development as a model system for the emergence of spatial structure in multicellular communities, our collaborators in Knut Drescher's lab developed an experimental platform for the measurement of spatiotemporal transcriptomes from live communities with high genome coverage and the simultaneous acquisition of microscopy-based measurements of cellular phenotypes, multicellular phenotypes, and the whole swarm development [12].

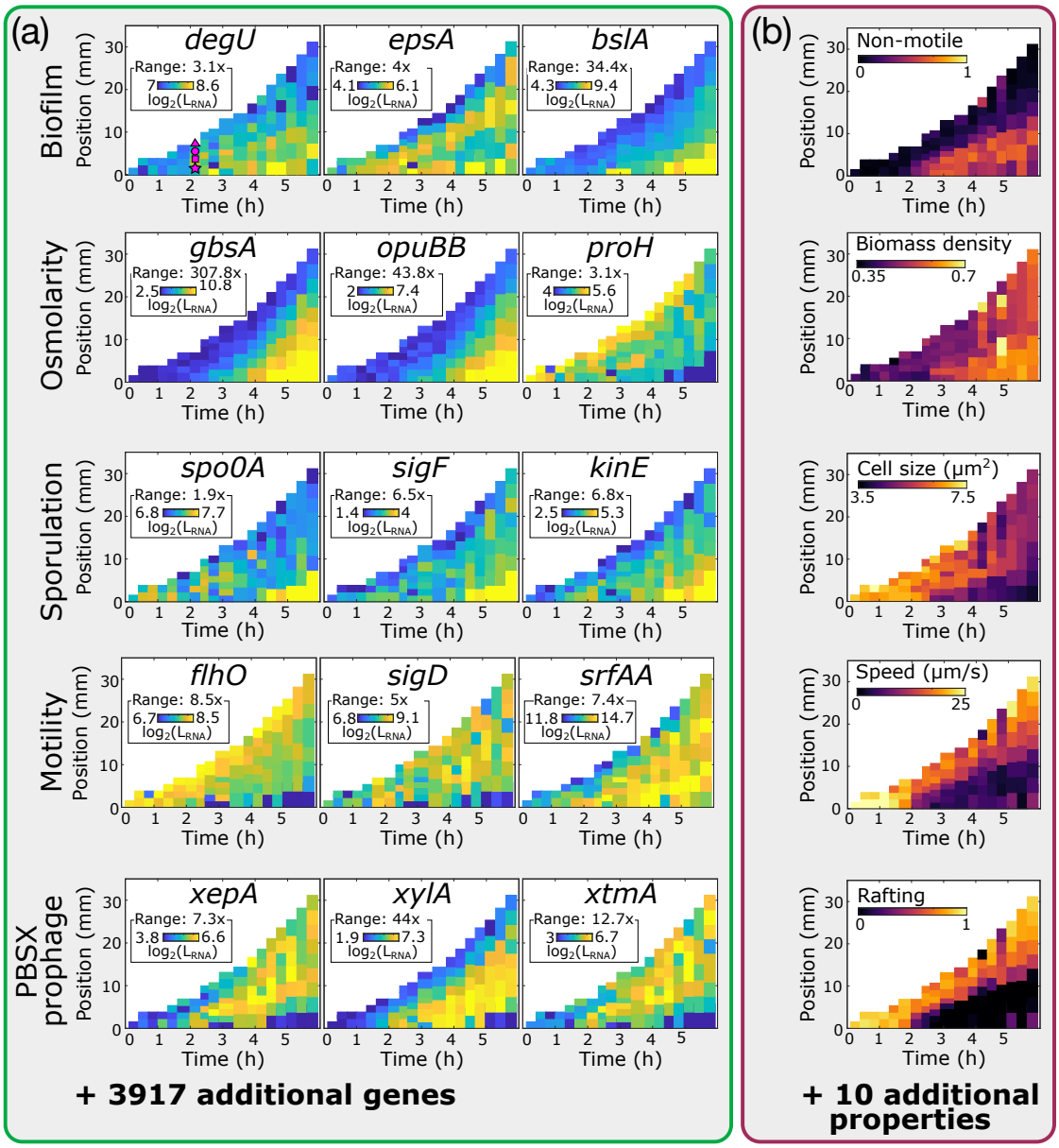

Figure 3.10: Spatiotemporal transcriptome measurements and microscopy-based phenotyping during Bacillus subtilis swarm development provided by the Drescher lab. (a) Spatiotemporal transcriptome results are summarized in kymograph heatmaps with each colored tile corresponding to one sample - the color of each tile in the heatmap indicates the expression level of a particular gene, L_{RNA}. The "Range" value corresponds to the dynamic range of gene expression, defined as the ratio between the highest and lowest color bar values, which are the 5th and 95th percentile of the gene expression values, taking all three replicates into account (additional replicates can be found in [12]). Three spatiotemporal heatmaps depicting the expression pattern of genes related to various processes are shown. Thousands of additional spatiotemporal gene expression heatmaps are available [12]. (b) Spatiotemporal phenotype heatmaps, analogous to the gene expression heatmaps in panel (a). Similar spatiotemporal phenotype maps are available for ten additional properties [12].

By integrating these different levels of biological and biophysical information, we systematically uncovered spatiotemporally varying processes and properties of the swarm. Full experimental details can be found in [12]. Here, we will focus on the spatiotemporal pattern analysis framework based on spectral representations.

In all previous examples of spectral representations, we have considered rectangular domains that lend themselves naturally to tensor product extensions into multiple dimensions (Section 2.1.3). Bacterial swarming, however, occurs on an exponentially growing domain (see example experimental images from Knut Drescher's lab, Figure 3.10), requiring the development of new basis functions for the spectral representation of spatiotemporal patterns from swarming experiments.

3.3.2 Non-dimensionalizing and aligning spatiotemporal domains

The transcriptome measurements and the microscopy-based measurement of phenotypic properties are sampled at a set of radial space-time points $\{r_l = (t_l, p_l)\}_{l=1}^{L}$, where t_l is the time and p_l is the radial position from the center of the swarm at which the sample is acquired. For each gene, we have a sample vector $\mathbf{g}_n$ with length L, where the lth entry is the gene expression at the point (t_l, p_l). Similarly, we have a sample vector $\boldsymbol{\phi}_n$ for each phenotypic property with length L. To enable spectral representation across experiments, we fit a common domain between the three experimental replicates. Experimentally, the radial position of the boundary b_l of the swarm at each time t_l was determined automatically by detecting the presence and location of bacteria in the microscopy field of view and moving the microscope stage until the field of view was split between colonized agar containing bacteria and uncolonized agar in approximately equal proportions. We simultaneously fitted a boundary of the form $b(t) = b_0 \exp(t/\tau)$ to all replicates by first minimizing the loss function,

$$L(\tau, b_0^{(1)}, b_0^{(2)}, b_0^{(3)}) = \sum_{k=1}^{3} \sum_{l=1}^{L^{(k)}} \left(b_i^{(k)} - b_0^{(k)} \exp(t_l^{(k)}/\tau) \right)^2 \tag{3.10}$$

where superscript (n) denotes the index of the three replicates. This exponential fit approximates the experimental data very well [Figure 3.11(a)].

To obtain non-dimensional data, we rescaled data as follows. Let r be the index corresponding to the largest $b_0^{(k)}$, we defined the time shift $t_s^{(r)} = t_0^{(r)}$ and scaled the initial value $b_0 = b_0^{(r)} \exp(t_s^{(r)}/\tau)$. The other time shifts are given by

$$t_s^{(k)} = \tau \left[\log(b_0^{(r)}) - \log(b_0^{(k)}) \right] + t_s^{(r)}.$$

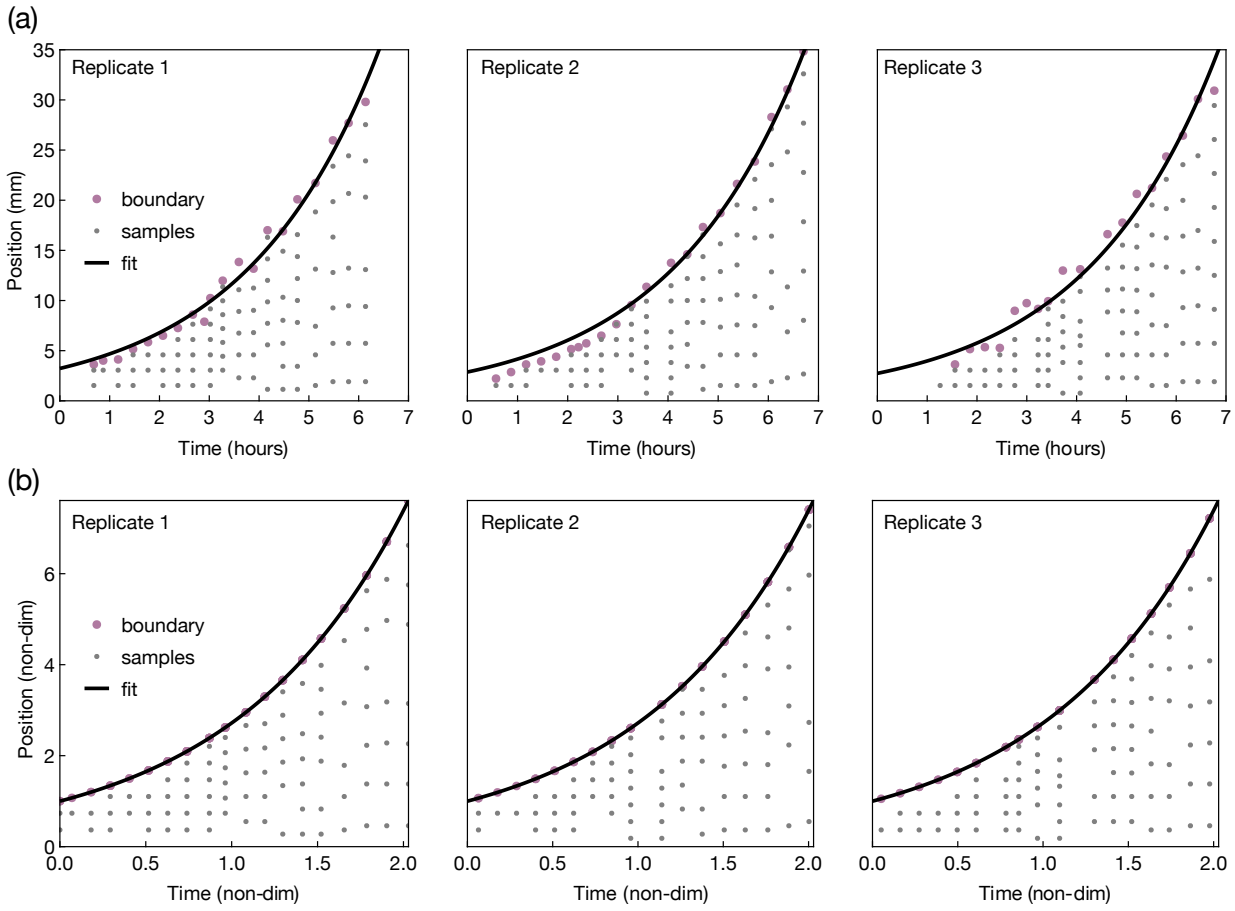

Figure 3.11: (a) An exponential function is fitted to the space-time location of boundary points (indicated in purple color) for each data set. Space-time coordinates are then rescaled. (b) The rescaled space-time coordinates are shown in their common non-dimensionalized domain.

We then non-dimensionalized the data ($\tilde{\cdot}$ variables) using $\tilde{t}_l = (t_l - t_s^{(k)})/\tau$ and $\tilde{p}_l = p_l/b_0$. In non-dimensional variables, the domain boundary is given by $0 \leq \tilde{t} \leq T$ and $0 \leq \tilde{r} \leq \exp(\tilde{t})$ where T is maximum non-dimensional time present in all three replicates. Data points outside the domain after rescaling are not used in the spectral representation. The non-dimensionalized domain and how the individual sampling data points are distributed within this domain are shown in Figure 3.11(b).

3.3.3 Construction of spectral representations on the swarming domain

We built a domain-specific orthogonal polynomial basis $\{P_m(\tilde{t}, \tilde{p})\}_{m=0}^{M}$ by applying Gram-Schmidt orthogonalization [102] over the non-dimensional domain $0 \leq \tilde{t} \leq T$, $0 < \tilde{p} \leq \exp(\tilde{t})$

to the monomial set,

$$\{1, \tilde{t}, \tilde{p}, \tilde{t}^2, \tilde{t}\tilde{p}, \tilde{p}^2, \ldots\},$$

under the inner product

$$\langle f, g \rangle = \int_0^T \mathrm{d}\tilde{t} \int_0^{\exp(\tilde{t})} \mathrm{d}\tilde{p} e^{-\tilde{t}} f g. \tag{3.11}$$

The space-time dependence of each gene is compressed by expanding each sample vector,

$$\mathbf{g}_n = \sum_{m=0}^{M} c_{m,n} \mathbf{P}_m \tag{3.12}$$

where $\mathbf{P}_m$ is the length L vector formed by evaluating $P_m(\tilde{t}, \tilde{p})$ at each non-dimensional space-time point $(\tilde{t}_l, \tilde{p}_l)$. The coefficients are fit using least squares on the matrix equation (Section 2.2),

$$\mathbf{g}_n = [\mathbf{P}_0 \, \mathbf{P}_1 \, \cdots \, \mathbf{P}_M] \begin{bmatrix} c_{0,n} \\ c_{1,n} \\ \vdots \\ c_{M,n} \end{bmatrix} \tag{3.13}$$

for each gene and property.

To combine information from all replicates, we fitted a single coefficient vector for the replicates by stacking the least square problems on top of each other to form a single linear regression problem,

$$\begin{bmatrix} \mathbf{g}_n^{(1)} \\ \mathbf{g}_n^{(2)} \\ \mathbf{g}_n^{(3)} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_0^{(1)} & \mathbf{P}_1^{(1)} & \cdots & \mathbf{P}_M^{(1)} \\ \mathbf{P}_0^{(2)} & \mathbf{P}_1^{(2)} & \cdots & \mathbf{P}_M^{(2)} \\ \mathbf{P}_0^{(3)} & \mathbf{P}_1^{(3)} & \cdots & \mathbf{P}_M^{(3)} \end{bmatrix} \begin{bmatrix} \bar{c}_{0,n} \\ \bar{c}_{1,n} \\ \vdots \\ \bar{c}_{M,n} \end{bmatrix}, \tag{3.14}$$

from which a smooth average spatiotemporal gene expression was formed,

$$\bar{\mathbf{g}}_n(\tilde{t}, \tilde{p}) = \sum_{m=0}^{M} \bar{c}_{m,n} P_m(\tilde{t}, \tilde{p}). \tag{3.15}$$

The same procedure was also used to produce average properties $\bar{\mathbf{\Phi}}_n$.

3.3.4 Analysis of spatiotemporal gene expression

To systematically characterize the different types of spatiotemporal patterns in gene expression and phenotypes during swarm development, we performed an unbiased analysis of the

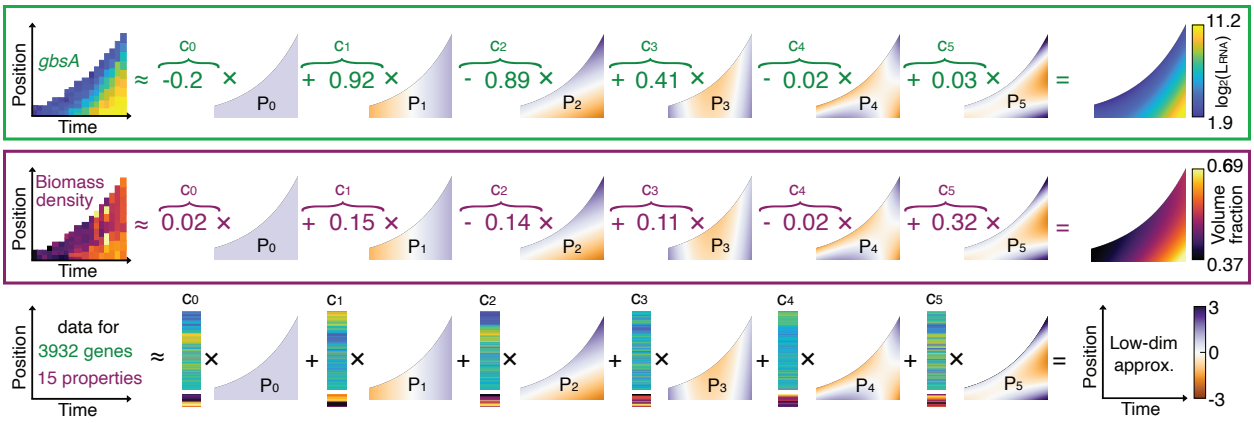

Figure 3.12: Spectral representation of spatiotemporal expression pattern data from the Drescher lab. The spatiotemporal expression heatmap for each gene and each phenotypic property was approximated by a linear combination of six orthogonal basis functions P_i, $(i = 0, ..., 5)$ that are optimized for the spatiotemporal swarm domain. The coefficients c_i of the basis functions were determined using all three biological replicates.

spatiotemporal data. For this, each spatiotemporal swarming dataset was represented by the coefficients $c_{m,n}$ in the domain-adapted orthogonal basis functions expansion (Figure 3.12).

Using the coefficients $c_{m,n}$ for each spatiotemporal dataset, we assesed how strongly each gene varies in space and time during swarming. With the analysis, we would like to identify the underlying patterns independent of global shifts and scaling. Therefore, we defined coefficients with the mean subtracted and scaled by the standard deviation,

$$k_{m,n} = \frac{c_{m,n} - \delta_{m,0}\mu_n/P_0}{\sigma_n} \tag{3.16}$$

where σ_n is the standard deviation of the expression of gene n, μ_n is the mean of the expression of gene n, $\delta_{n,k}$ is the Kronecker delta, and P_0 is constant since it is a degree 0 polynomial. Note that $k_{0,n}$ is no longer an independent parameter under this rescaling since it is fully determined using the higher-order polynomials and their respective coefficients. Using these rescaled coefficients $k_{m,n}$, we defined a score for how strongly patterned a gene expression profile is. A spatiotemporally patterned expression profile should have two components:

- The pattern should vary smoothly, so the pattern should be well approximated by the spectral representation, which means that the scaled representation error

$$\mathcal{E}_n = \|(\mathbf{g}_n - \mu_n)/\sigma_n - \sum_{m=0}^{M} k_{m,n}\mathbf{P}_m\|^2$$

  should be small.

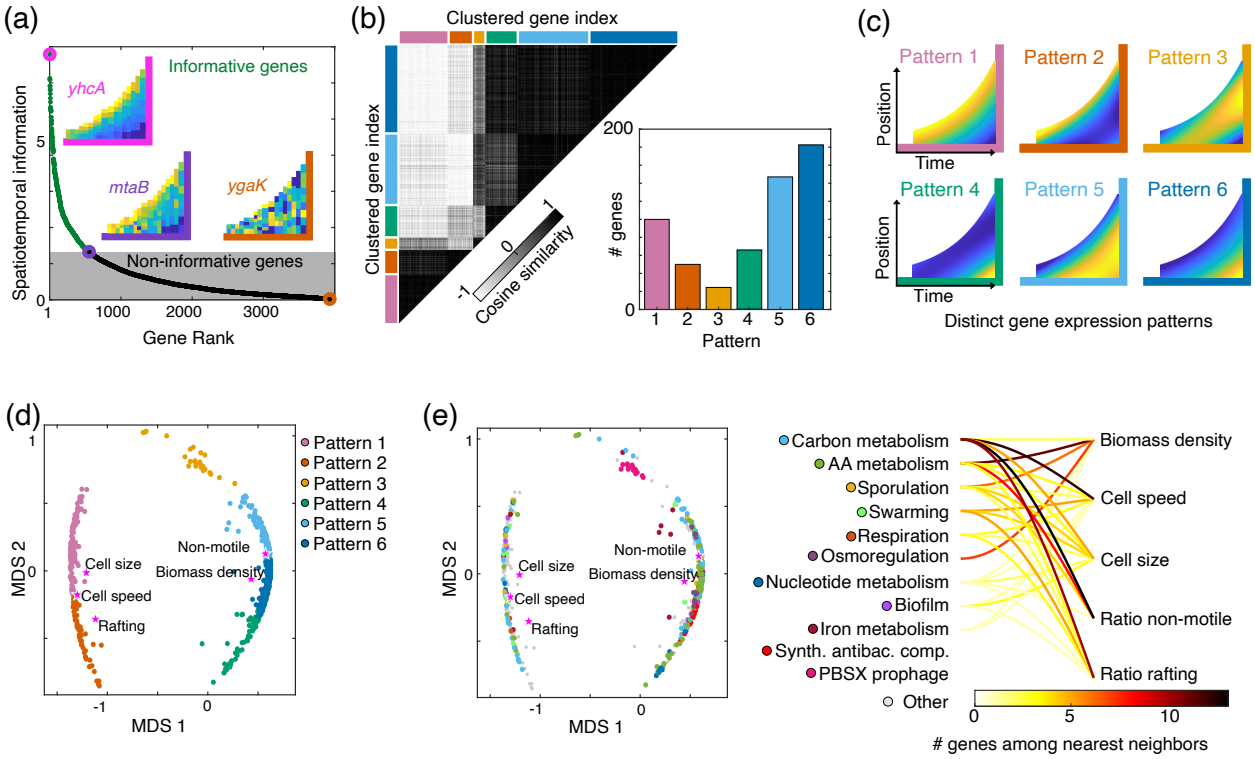

Figure 3.13: Identification of genes with spatiotemporal regulation and different spatiotemporal expression patterns. (a) A spatiotemporal information score was defined, quantifying the spatiotemporal information observed for a given gene. By ranking genes according to this spatiotemporal information score and defining a cutoff at the weighted median of the spatiotemporal information, we identified all genes with spatiotemporal gene expression patterns. Insets show gene expression heatmaps with low, intermediate, and high spatiotemporal information scores. (b) For the 572 genes with spatiotemporal gene expression patterns identified in panel (a), we used cosine similarity based on the coefficients c_i to identify clusters of highly correlated spatiotemporal patterns, revealing six major distinct spatiotemporal gene expression patterns (indicated by colored lines on the side of the similarity matrix). (c) To visualize these, the pattern corresponding to the mean of all coefficients c_i within each pattern cluster is shown. (d-e) Multidimensional scaling (MDS) was applied to genes and swarm properties based on their cosine similarity, as shown in panel (b). The expression pattern of each gene is represented as a point with color indicating their expression pattern category in panel (d) or gene function in panel (e). Gene functions are based on subtiWiki [103, 104]. Gene function categories with fewer than ten genes assigned to them are grouped into the category "other". Five phenotypic properties of the swarm (see Figure 3.10 for heatmaps) are shown as stars, revealing that their location in the MDS space is near some gene patterns and functions. (e) The 50 nearest neighbor genes in the MDS space (corresponding to similar gene expression patterns) were identified and grouped into gene function categories for each of these five phenotypic properties. The number of genes in each gene function category for each phenotypic property is visualized in the connection plot (right).

- The pattern should not just be constant, meaning that the higher coefficients should be important, which means that the pattern score

$$\mathcal{P}_n = \sum_{m=1}^{M} k_{m,n}^2$$

should be large.

We, therefore, define a space-time ranking,

$$\mathcal{R}_n = \frac{\mathcal{P}_n}{\mathcal{E}_n} \tag{3.17}$$

which is large when a gene expression has a strong spatiotemporal pattern and small when there is no spatiotemporal pattern [Figure 3.13(a)]. We define the cutoff for which genes are designated as displaying a spatiotemporal pattern by ordering the genes by their ranking smallest to largest and then finding the largest integer N_c such that,

$$\frac{\sum_{n=1}^{N_c} \mathcal{R}_n}{\sum_{n=1}^{N} \mathcal{R}_n} \leq 0.5. \tag{3.18}$$

This results in 572 genes with a spatiotemporal pattern.

The coefficients $k_{m,n}$ for the basis functions P_m encode information about the spatiotemporal expression pattern of the genes and phenotypic properties, which allows us to cluster genes based on their spatiotemporal expression pattern using the spectral coefficients $k_{m,n}$ directly. For the 572 genes that displayed a high degree of spatiotemporal variation during swarming, we computed the similarity in their spatiotemporal expression pattern, using cosine similarity between the coefficients,

$$d_{n,p} = \frac{\sum_{m=1}^{M} k_{m,n} k_{m,p}}{\sqrt{\sum_{m=1}^{M} k_{m,n}^2} \sqrt{\sum_{m=1}^{M} k_{m,p}^2}}. \tag{3.19}$$

We use cosine distances since we need a metric independent of a global scaling to the expression level. The distance matrix $D = (d_{n,p})$ is then separated into clusters using the k-medoids algorithm [105] to produce k distinct patterns. The number of clusters is chosen based on a plot of the total cost versus number of clusters and choosing a value in the elbow of the curve (Figure 3.14). This analysis revealed 6 clusters of genes with similar spatiotemporal expression patterns [Figure 3.13(b); inset shows the number of genes with a particular pattern]. To illustrate the six different patterns, the pattern corresponding to the mean of all coefficients is shown in Figure 3.13(c). Interestingly, all six patterns vary in space and

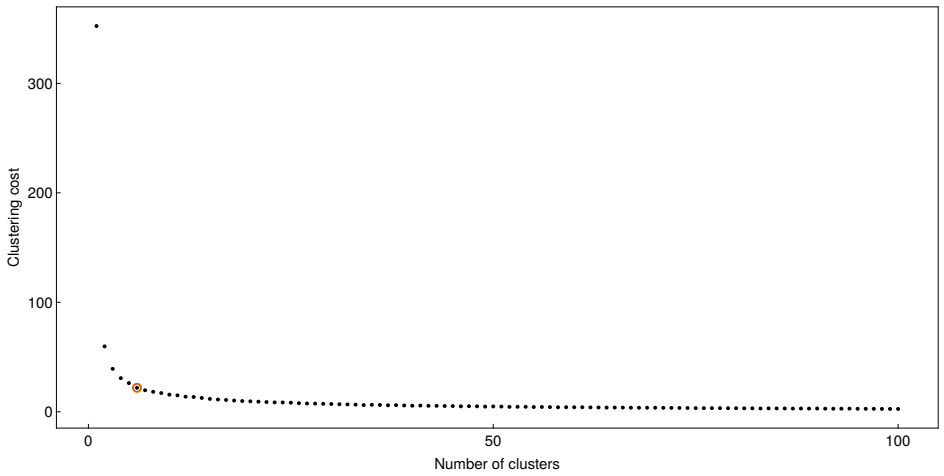

Figure 3.14: With increasing clusters, the minimum clustering cost decreases, with an initially sharp decline that levels off for high cluster numbers. Highlighted in red is the location of 6 clusters, the number which was chosen for further analyses in Figure 3.13

time, not only in space or time.

Spatiotemporal phenotype patterns correlate with metabolism gene expression patterns. To reveal connections between our measurements of gene expression and microscopy-based phenotypic properties, we performed a 2D embedding using multidimensional scaling [106] for the spatiotemporally expressed genes and the phenotypic properties [Figure 3.13(d, e)]. This analysis shows that the biomass density and the abundance of non-motile cells in the swarm are closely related and are close to patterns 5 and 6. In contrast, motility-related phenotypic properties, such as the cell speed and the abundance of rafting cells, as well as the cell size, are close to patterns 1 and 2 [Figure 3.13(d, e)]. By looking at the gene functions of the 50 closest genes to the phenotypic properties in Figure 3.13(e), we found that metabolic genes are most closely associated with the spatiotemporal dynamics of the phenotypic properties [Figure 3.13(e) right]. This insight led us to investigate the spatiotemporal changes in metabolism during swarm development, with the aim of understanding how and why metabolic changes occur in space and time and how they influence swarm development. Spatiotemporal measurements of metabolites secreted and consumed by the different subpopulations led us to discover spatiotemporal cross-feeding interactions within the swarm.

3.3.5 Conclusions

Spectral representations of the spatiotemporal gene expression and phenotypic properties enabled an unbiased and quantitative analysis of the emergent patterns. This pattern analysis led to the discovery of a strong relationship between metabolic genes and spatiotemporal dynamics of the phenotypic properties and, on further investigation, a spatiotemporally

organized cross-feeding mechanism. These results further emphasize the strength and extensibility of spectral representations for the robust analysis of biological data, even on more complicated experimental domains.

Chapter 4

Dynamical systems inference in spectral mode space

As highlighted in the previous chapter, mode space representations of data have many benefits that make them useful for analyzing experimental data. We used spectral representations to extract quantitative information from experimental data and make comparisons with theoretical models. However, the underlying dynamical system of complex biophysical phenomena is often unknown. We wish, therefore, to infer the dynamics directly from experimental data when this is the case. Inferring models directly in mode space has several theoretical advantages. As shown in the following two sections, a large class of dynamics in real space can be reduced to coupled ordinary differential equations (ODEs) in mode space, allowing us to focus on dynamical systems inference techniques specifically for ODEs.

4.1 Projection of real space dynamics into mode space

Until now, we have used mode space representation as an intermediary step where operations are easy to compute before projecting back into real space. However, the projection into mode space is orthogonal, and the mode space representation contains the same information as the real space representation. Therefore, understanding a system's dynamics in real space is equivalent to understanding its dynamics in mode space. Concretely, consider a scalar field $\rho(t, \mathbf{x})$ that evolves according to a possibly nonlinear and non-local integrodifferential equation,

$$\frac{\partial}{\partial t}\rho(t, \mathbf{x}) = F(t, \mathbf{x}, \rho, \nabla\rho, \nabla^2\rho, \ldots). \tag{4.1}$$

A spatial mode representation of ρ is given by,

$$\rho(t, \mathbf{x}) \approx \sum_{n=1}^{N} \hat{\rho}_n(t)\phi_n(\mathbf{x}) \tag{4.2}$$

where the dynamical information about the field is now shifted into the finite number of time-varying mode amplitudes $\hat{\rho}_n(t)$. We substitute (4.2) into (4.1) and write the right hand side as a mode representation

$$\sum_{n=1}^{N} \frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_n(t)\phi_n(\mathbf{x}) = F\left(t, \mathbf{x}, \sum_{n=1}^{N} \hat{\rho}_n(t)\phi_n(\mathbf{x})\right) = \sum_{n=1}^{N} f_n(t, \hat{\rho}_1, \ldots, \hat{\rho}_N)\phi_n(\mathbf{x}) \tag{4.3a}$$

where the functions f_n are given by the integrals,

$$f_n(t, \hat{\rho}_1, \ldots, \hat{\rho}_N) = \int_D \mathrm{d}\mathbf{x}\, w(\mathbf{x})\phi_n(\mathbf{x}) F\left(t, \mathbf{x}, \sum_{m=1}^{N} \hat{\rho}_m(t)\phi_n(\mathbf{x})\right). \tag{4.3b}$$

The function F approximated by a mode representation will have coefficients that depend, possibly nonlinearly, on the original coefficients. Projecting both sides of (4.3a) onto the basis $\phi_n(\mathbf{x})$ gives a set of coupled nonlinear ODEs for the mode amplitudes,

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}_n(t) = f_n(t, \hat{\rho}_1, \ldots, \hat{\rho}_N) \tag{4.4}$$

The approach is intricately linked with spectral methods for solving differential equations [22], where a truncated basis function expansion approximates solutions to the differential equation. Many of the same technical challenges for spectral methods, such as aliasing and closure, are also present here.

The procedure outlined above is also readily generalized to sets of coupled differential equations; see Chapter 5 for an example.

Linear partial differential equations

An important sub-class of problems is linear partial differential equations with constant coefficients, for example

$$\frac{\partial}{\partial t}\rho(t, x) = \sum_{k=0}^{K} c_k \frac{\partial^k}{\partial x^k}\rho(t, x). \tag{4.5}$$

under the projection into mode space, the dynamics become,

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\boldsymbol{\rho}} = \left[\sum_{k=0}^{K} c_k (D_\phi)^k\right] \hat{\boldsymbol{\rho}} \tag{4.6}$$

where D_ϕ is the differentiation matrix in mode space for the basis given by $\phi_n(x)$ and $\hat{\boldsymbol{\rho}} = [\hat{\rho}_0, \ldots, \hat{\rho}_N]$ is a vector of the representation coefficients. Equation (4.6) is a linear set of coupled ODEs. Note that, in general, for polynomial bases, the differentiation matrix will be strictly upper triangular. This can have severe consequences for the truncated dynamics. Consider, as a simple example, the heat equation on $[-1, 1]$ with initial condition $\rho(0, x) = \rho_0$ and boundary conditions $\rho(t, -1) = \rho(t, 0) = 0$.

$$\frac{\partial}{\partial t}\rho(t, x) = \frac{\partial^2}{\partial x^2}\rho(t, x) \implies \frac{\mathrm{d}}{\mathrm{d}t}\hat{\boldsymbol{\rho}} = D_T^2 \hat{\boldsymbol{\rho}} \tag{4.7}$$

in the Chebyshev basis, where D_T is strictly upper triangular. The last two rows of the D_T^2 contain all 0 entries. This decoupling of the last coefficients can have catastrophic consequences when the ODE is reintegrated since it is unstable. To see this we look at the highest modes' dynamics, $\hat{\rho}_N(t) = \hat{\rho}_N(0)$, since the row is all 0, and $\mathrm{d}/\mathrm{d}t\, \hat{\rho}_{N-2}(t) = 280\hat{\rho}_N(0)$, which shows that $\rho_{N-2}(t)$ will grow unbounded in time. Some form of closure is required for the dynamics to be stable. In Chebyshev spectral methods, the zero rows are typically replaced with boundary conditions to close the system. When we apply model inference in mode space, we will discover these closures alongside the dynamics by enforcing that our inferred models are stable.

Generically, we find that linear partial differential equations in real space will lead to linear ODEs in mode space. Modeling dynamics directly in mode space has several theoretical and practical advantages. The theory of ODEs, with robust uniqueness and existence theorems, is more straightforward than of partial differential equations and robust numerical methods exist to solve them efficiently.

4.2 Projection of mode space dynamics into real space

We now show that linear models in mode space generally lead to linear integrodifferential equations in real space. Consider a linear dynamics in mode space,

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\boldsymbol{\rho}} = M\hat{\boldsymbol{\rho}}. \tag{4.8}$$

Define the vector function $\boldsymbol{\phi}(x) = [\phi_1(x), \dots \phi_N(x)]$, which gives a compact formula for the real space dynamics $\rho(t,x) = \boldsymbol{\phi}^\top(x)\hat{\boldsymbol{\rho}}(t)$. Multiplying both sides of the dynamics by $\boldsymbol{\phi}(x)^\top$ and using the definition of the coefficients (2.5) the dynamics becomes,

$$\boldsymbol{\phi}^\top(x)\frac{\mathrm{d}}{\mathrm{d}t}\hat{\boldsymbol{\rho}} = \boldsymbol{\phi}^\top(x)MK^{-1}\int_D \mathrm{d}y\, w(y)\rho(t,y)\boldsymbol{\phi}(y). \tag{4.9}$$

where K is a diagonal matrix with entries $K_{nn} = \langle\phi_n, \phi_n\rangle$. Defining the two-dimensional kernel,

$$G(x,y) = w(y)\boldsymbol{\phi}^\top(x)MK^{-1}\boldsymbol{\phi}(y) \tag{4.10}$$

the real space dynamics become,

$$\frac{\partial}{\partial t}\rho(t,x) = \int_D \mathrm{d}y\, G(x,y)\rho(t,y) \tag{4.11}$$

a linear integro-differential equation for the time evolution of $\rho(t,x)$.

The non-local behavior of generic linear mode space models makes them particularly well suited to study biological phenomena where non-local effects appear frequently, for example, in Lévy processes [107], chemotaxis [108, 109], or cell migration [110]. By studying the resulting real space kernels, we can infer information about the locality of a given mode of space dynamics.

4.3 Previous approaches to dynamical systems inference

This section has adapted from the preprint G. Stepaniants, A. D. Hastewell*, D. J. Skinner, J. F. Totz, and J. Dunkel, "Discovering dynamics and parameters of nonlinear oscillatory and chaotic systems from partial observations," arXiv preprint arXiv:2304.04818, 2023.*

In canonical form, data-driven dynamics discovery consists of fitting the parameters $\mathbf{p}$ of an ODE $\dot{x} = f(x, \mathbf{p})$ to a given dataset. This problem has been considered since early computers were used to solve differential equations numerically [111]. Renewed interest in the problem came with increasing computational power in the 1980s and 1990s [112]. With the increasing prevalence of machine learning techniques and the explosion of high-resolution data acquisition, once again, attention has turned to this problem in recent years [37, 113–116].

4.3.1 Parametric models

The simplest case is when the form of the ODE right-hand side $f(x, \mathbf{p})$ is fixed ahead of time by some knowledge of the system, and the parameter vector $\mathbf{p}$ is the only unknown. Frequently, $\mathbf{p}$ will be low dimensional, consisting of only a few parameters. Fitting these parameters can be formulated as an optimization problem minimizing a data loss $\sum_i |\dot{x}_i - f(x_i, \mathbf{p})|^2$. Approximations of the time derivative can be calculated by finite differences, smooth derivative approximations [112, 117, 118] or weak formulations [119]. Minimizing the data loss to find $\mathbf{p}$ can be done using either linear [27, 120] or nonlinear [118, 121] least squares based on the structure of the dynamical equations. For instance, the parameters of the canonical Lorenz system enter linearly while the parameters of the nonlinear pendulum $\ddot{x} = -A\sin(\omega x)$ do not. Such methods are not robust and can overfit noisy data, as they impose no causal constraints that the observed data at time $t + \Delta t$ is related to the data at time t through the evolution of an underlying differential equation.

4.3.2 Optimization methods

In the case of partial observations or noise-corrupted data, we need a causal formulation of the problem, which encodes prior knowledge that observed data come from the evolution of a differential equation. When the integrated solution to the ODE model has a closed form, such as linear time-invariant systems, we can use least squares fitting to match the integrated solution to the data [122]. For more complicated models, we need to use numerical approximations to the integrated ODE solution and fit the parameters by solving suitably chosen, often nonlinear, and non-convex optimization problems. The causal nature of the problem can be encoded by adding additional penalty terms to the data loss enforcing an ODE structure, leading to a large class of approaches known as data assimilation methods [114, 123]. These approaches typically enforce that the fit dynamics is consistent with a single time step of an ODE model, although generalizations to n steps have shown better stability and noise robustness [124–126]. While better suited for data with partial observations and noise, there is still no guarantee that the integrated model is stable over long times or fits the data when these finite steps are performed sequentially. A stronger constraint on the ODE structure is that the solution $x(t, \mathbf{p})$ comes from numerically integrating the model, which is then fitted to the data. We can achieve this hard constraint by directly optimizing through a numerical ODE solver. Such optimization techniques roughly fall into two categories: (1) gradient-free methods like root finding [127], multiple shooting [128], Nelder-Mead, particle swarm optimization [129], or Kalman filters [130–132], and (2) gradient-based methods like Newton's method, ADAM [133], and BFGS [134] where derivatives are approximated either using fi-

nite differencing on the ODE solutions, sensitivity methods or automatic differentiation [37, 135]. These techniques have been applied successfully to learn periodic, chaotic, and transient dynamics [37]. The question of parameter identifiability, which studies whether model coefficients can be determined uniquely from the data, has also been extensively explored for parametric models [136, 137]. However, parametric models are only applicable when we have some prior knowledge about the form of the model governing the dynamics in our data, which is frequently not the case for experimental systems.

4.3.3 Nonparametric models

A more expressive and less constrained right-hand side function f is needed if we cannot assume a particular parametric form of the equations. Taking f as a general function with a large set of model parameters, such as a neural network, results in a nonparametric model. It is called nonparametric as the large set of parameters has no inherent interpretation. Examples of such functions f that have been applied to dynamical inference include locally linear models [138, 139], bases expansions [140] (sometimes referred to as atlases or libraries including polynomial [27], rational polynomial [141, 142], wavelet [143], and radial basis functions [144]), discrete-time perceptrons [145, 146], and neural networks (referred to as neural ODEs) [37]. The techniques for fitting parametric models outlined above can also be applied to these general functions. These techniques often discretize the ODE during fitting; directly learning a discretized form of the dynamics $x_{t+\Delta t} = f(x_t, \mathbf{p})$, is called a discrete nonparametric model [140, 141, 144–146].

4.3.4 Model selection

Without prior knowledge about the form of the dynamics, nonparametric models can fit an ODE to data. However, the resulting model lacks physical interpretability. This drawback prompted the development of automated methods for learning physically interpretable ODE models without assuming prior knowledge of the dynamics. A popular approach is first to learn a nonparametric model given as a linear combination of basis terms and to subsequently perform model selection keeping only a few basis terms in the expansion [27, 140]. Various techniques for basis term selection include sparse identification [27], bootstrapping [147], information criteria (MLE, AIC, BIC, MDL) ranking [140, 148, 149], and hypothesis testing [150, 151]. When partial knowledge of the ODE model is available, a parametric interpretable ODE model can be added to a nonparametric model and jointly fit to the data, an approach called hybrid modeling [131].

4.4 Linear dynamical systems inference

Even for complex nonlinear dynamics, effective linear models can capture the relevant dynamics over a characteristic region of phase space. Studying linearizations of nonlinear dynamics to reveal information about the underlying dynamical system is frequently used in dynamical systems theory [152], for example, in the analysis of fixed point stability. More recently, data-driven approaches have shown the power of effective linear dynamical models in characterizing complex nonlinear phenomena. In fluid dynamics, dynamic mode decomposition extracts dominant spatio-temporal motifs from data [19]. In neuronal dynamics, switching linear dynamical systems are used to accurately approximate nonlinear dynamics using different linear models for different regions of phase space [153]. In the following two chapters, we will restrict ourselves to inferring linear dynamics in mode space and assume that all relevant degrees of freedom have been observed. We discuss extensions to the case of nonlinear dynamics and unobserved degrees of freedom in Chapter 8.

Dynamic Mode Decomposition (DMD) [19] discovers the eigen-decomposition of a linear operator assuming deterministic discrete or continuous linear dynamics. The N mode continuous time DMD assumes that the dynamics of the system follow the linear matrix ODE $\dot{\mathbf{d}}(t) = M\mathbf{d}(t)$. The solution of the linear dynamics gives a series approximation to the data $\mathbf{d}(t)$ of the form

$$\mathbf{d}(t) \approx \sum_{n=1}^{N} c_n \mathbf{v}_n e^{\lambda_n t}, \tag{4.12}$$

where $\mathbf{v}_n$ are the eigenvectors of the matrix M with corresponding eigenvalues λ_n. Multiple approaches have been developed for fitting approximations of the form (4.12) [19] using linear algebra [17] and variable projection [154]. In general, the expansion (4.12) can be sensitive to noise, leading to eigenvalues with spurious positive real parts and unstable dynamics. Additionally, the expansion (4.12) will generally correspond to a dense unstructured matrix with no prior knowledge about the structure imposed by physical or biological constraints used. The recent development of physics-informed DMD (piDMD) [155] shows that incorporating prior knowledge about the dynamics adds robustness to the inference step. A key challenge with biological data is noise in the data, making differentiating data challenging and ill-posed. It may also be unclear ahead of time what the correct model structure is, or only a portion of the structure may be known. As we saw in Section 4.1, mode space dynamics arising from linear differential equations in real space often have a sparse structure in their dynamical matrix. We, therefore, seek to develop a method that can infer sparse linear dynamical systems that match the data well and are stable over the time region of interest.

In the rest of this chapter, we describe in detail an inference procedure to learn parsimonious linear ODE models. We consider an approach where known sparse structures can be imposed ahead of time, and additional sparsity can be inferred directly from the data. We gain robustness by directly matching integrated predictions of the linear dynamical system to experimental data. By comparing integrated quantities, we avoid numerically differentiating noisy data, which is known to be ill-posed. We also ensure that long-time integrated solutions of the linear system match the data well and are stable.

4.5 Inferring sparse linear models

In this section we outline an optimization procedure for robustly fitting sparse linear dynamical systems to experimental mode trajectories. In the next chapter so show how this is integrated into a full analysis of zebrafish embrogenesis in collaoration with N. Romeo and A. Mietke. *The results in this section have previously been published in N. Romeo*, A. D. Hastewell*, A. Mietke, and J. Dunkel, "Learning developmental mode dynamics from single-cell trajectories". Elife, 10, e68679, 2021.*

Given a dynamical mode vector $\mathbf{a}(t)$, the goal is to learn a linear minimal model

$$\frac{\mathrm{d}\mathbf{a}(t)}{\mathrm{d}t} = M \cdot \mathbf{a}(t) \tag{4.13}$$

of the mode dynamics. Here, M is an unknown $n \times n$ mode coupling matrix. Our inference algorithm combines adjoint techniques [37] and a multi-step sequential thresholding approach inspired by the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm introduced by Brunton et al. [27].

To describe the algorithm used to infer the mode coupling matrix M, we parameterize M by a vector $\mathbf{p}$ that contains all non-zero entries and introduce a function $\mathcal{M}$ that represents the underlying matrix structure. Together, they generate the explicit form $M = \mathcal{M}(\mathbf{p})$ of the mode coupling matrix. Imposing structure on the matrix, such as rank constraints or sparsity, leads to a shorter vector $\mathbf{p}$ and modifies the definition of $\mathcal{M}$ accordingly. Denoting $\mathbf{A}(t; \mathcal{M}, \mathbf{p}, \mathbf{a}_0)$ as the result of numerically integrating the system of ODEs (4.13) up to time t from initial condition $\mathbf{a}_0$ with $M = \mathcal{M}(\mathbf{p})$, we define the loss function

$$L(\mathbf{p};\, \mathcal{M},\, t_I,\, t_N) = \frac{1}{N-I} \sum_{i=I}^{N} \|\mathbf{a}(t_i) - \mathbf{A}(t_i; \mathcal{M}, \mathbf{p}, \mathbf{a}(t_I)\|_2^2, \tag{4.14}$$

where the t_i are time points in an interval $[t_I, t_N]$ at which the data and the ODE solution are sampled. Using the ODE solvers and optimization functions provided by

`DifferentialEquations.jl` and `DiffEqFlux.jl` [37], we can differentiate through the ODE solver to calculate derivatives of the loss function (4.14) with respect to parameters $\mathbf{p}$ and subsequently apply gradient-based optimization algorithms. The loss function is minimized using the ADAM algorithm [133], followed by the Broyden-Fletcher-Goldfarb-Shannon (BFGS) algorithm [156]. To increase the robustness of the optimization and promote sparsity, we use a sequentially thresholded algorithm [27, 33, 84]. A complete overview of this procedure is shown in Figure 4.1, and the details of the specific design decisions made in the algorithm are discussed in the following:

1. To account for the variation in scale between the different modes in the data $\mathbf{a}(t)$, each mode is normalized by its median absolute deviation (MAD) across the full period in which the data are available. Specifically, we scale each mode by

$$\text{mad}(a_i) = \text{median}_k \left(|a_i(t_k) - \bar{a}_i| \right), \tag{4.15}$$

   where $\bar{a}_i = \text{median}_k[a_i(t_k)]$ and the median is taken over all time-points, giving rise to a scaled mode vector $\tilde{\mathbf{a}}(t)$. Losses analogous to equation (4.14) that are computed using scaled data are denoted in the following by $\tilde{L}$.

2. To prevent over-fitting, we divide the data into two regions, a learning region from t_I to t_N and a validation region from t_N to t_F. Only data from the learning region is used to optimize the loss function (4.14). However, the model is integrated into the validation region, and a corresponding validation loss using only the data in the validation region is calculated. We choose the model with the lowest loss in the validation region during each optimization run, lowering the likelihood of over-fitting to the specific data in the learning region.

3. To prevent the optimization from getting stuck in local minima, we incrementally increase the time span of the data included in the optimization objective (blue box in Figure 4.1). We increase the time window backward from a fixed endpoint $t_1 = t_F$, choosing an earlier initial condition at each iteration at time $t_i < t_{i-1}$. The advantage of stepping backward rather than forward from a fixed initial condition is twofold. First, the validation region stays unchanged throughout the optimization, making comparisons of the validation loss easy. Second, because the initial condition changes with each run, the learned matrix tends to be more robust to fluctuations in the initial condition.

4. After the optimization step, sparsity is promoted by thresholding the elements in the matrix [27], removing small magnitude elements that do not noticably contribute to

the mode dynamics (purple box in Figure 4.1). The optimization procedure is then repeated until the thresholding converges. The threshold is chosen to generate a sparse matrix that faithfully reproduces the dynamics.

5. Once the sparsity pattern is obtained from the sequential thresholding and optimization procedure, a final run of the optimization is performed on the unscaled mode data to find the final dynamical matrix M, which removes any potential slight bias the MAD scaling might have introduced in the parameter values $\mathbf{p}$.

Finally, the numerical stability of the model can be checked by examining the eigenvalues of the learned matrix.

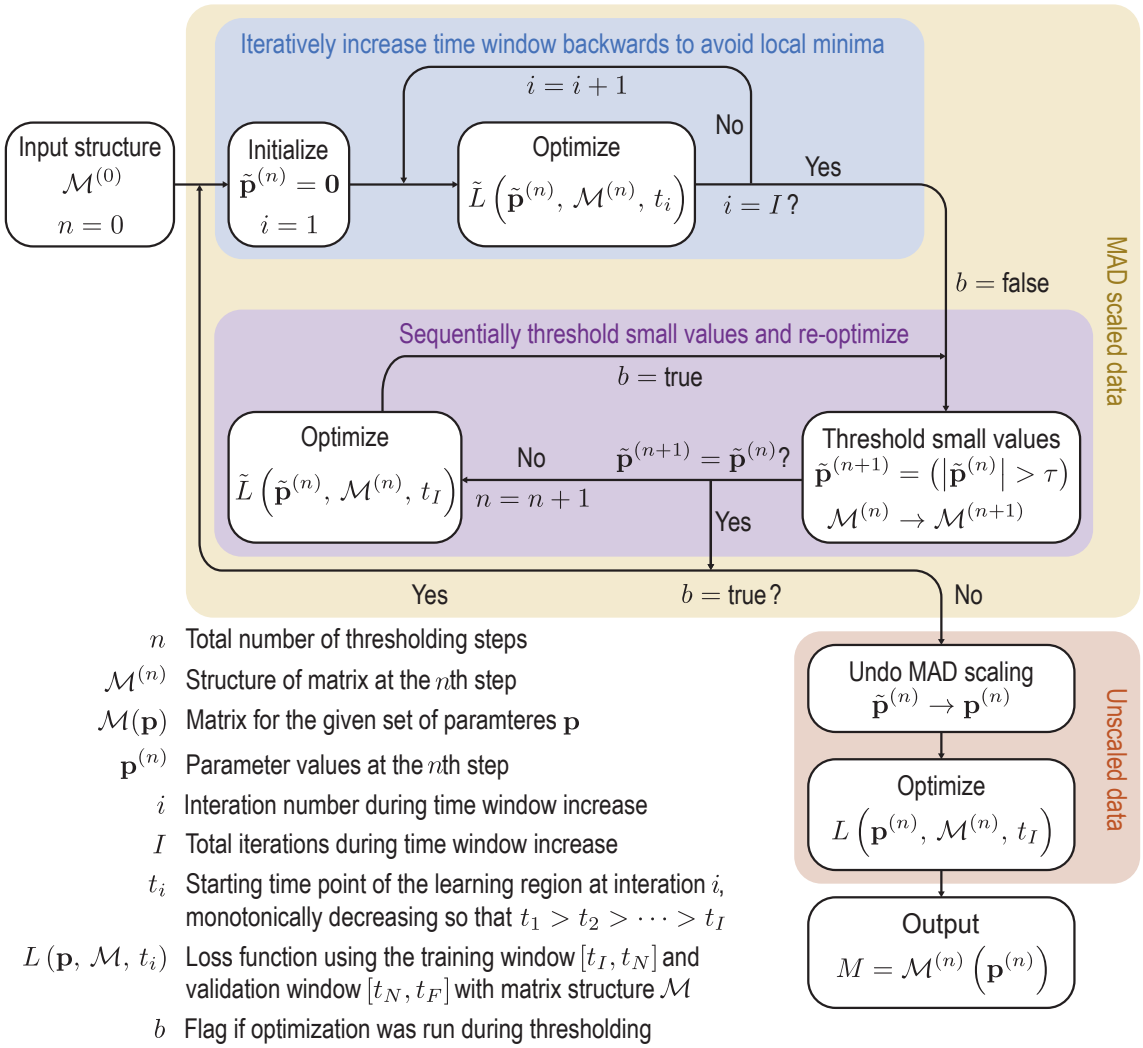

Figure 4.1: Schematic of the learning procedure. Initially, the data is rescaled using the median absolute deviation (MAD) (4.15) to account for variation in scales across the modes. Tildes denote scaled variables. To avoid local minima of the optimization function, we iteratively feed more data into the cost function. Next, we sequentially threshold the small terms in the matrix until convergence is reached. These procedures are repeated until the sparsity pattern converges. Finally, the scaling is undone, and the parameters are optimized on the unscaled data to produce the final matrix.

Chapter 5

Learning developmental mode dynamics from single-cell trajectories

The work in this chapter has been adapted from N. Romeo, A. D. Hastewell*, A. Mietke, and J. Dunkel, "Learning developmental mode dynamics from single-cell trajectories," Elife, 10, e68679, 2021.*

In this chapter, we apply the model inference framework outlined in the previous chapter. We summarize a coarse-graining and spectral representation procedure on the sphere developed by N. Romeo and A. Mietke. We also show an additional application of spectral entropy to determine coarse-graining length scales. We then apply the dynamical systems inference procedure to a test data set from N. Romeo and the results from the coarse-graining procedure. Finally, we discuss the implications of the resulting models.

5.1 Motivation

Embryogenesis, the development of a multicellular organism from a single fertilized egg cell, requires coordinated collective motions of thousands of cells across a wide range of length and time scales [157, 158]. Understanding how a highly reproducible and robust tissue organization arises from the dynamics and interactions of individual cells presents a major interdisciplinary challenge [159]. Recent advances in high-resolution live imaging make it possible to track the internal biological states and physical movements of many individual cells on pan-embryonic scales throughout various stages of development [6–8, 10]. This unprecedented wealth of data poses two intertwined compression problems of equal practical and conceptual importance: the first concerns the efficient reduction of high-dimensional tracking data without losing relevant information; the second concerns inferring predictive low-dimensional models for developmental dynamics. Mathematical solutions to the first

problem are aided by considering the geometry and symmetries of the developing embryo, which suggest suitable basis functions for a coarse-grained and sparse mode representation of raw data [160]. Efficient algorithmic approaches tackling the second problem are within reach thanks to recent advances in the direct inference of dynamical systems equations from data [27, 37] as discussed in the previous chapter. Building on these ideas, we construct and demonstrate a computational framework that translates developmental single-cell trajectory data on curved surfaces into quantitative models for the dominant hydrodynamic modes.

Widely applied in physics [14, 161–163], engineering [164, 165] and spectral computing [21, 22, 166], mode representations [17, 18] provide a powerful tool to decompose and study system dynamics at and across different energetic, spatial and temporal scales. In quantum systems, for example, mode representations in carefully constructed eigenstates are used to characterize essential energetic system properties [167, 168]. Similarly, turbulence theory has seen significant progress by studying the coupling between Fourier modes representing dynamics at different length scales. This approach enabled a better understanding of energy cascades [169, 170] and provided insights into the nature of turbulence in non-living [171, 172] and in living [173–176] systems. Additionally, the multi-scale nature of many biological processes makes them particularly amenable to a representation in terms of spatial and temporal modes [81]. Despite this, mode representations are not yet widely used to characterize and compress cell tracking data or infer dynamic models from such data.

To demonstrate the practical potential of mode representations for the description of multicellular developmental processes, we develop here a computational framework that takes cell tracking data as inputs, translates these data into a sparse mode representation by exploiting symmetries of the biological system, and utilizes recently developed ODE inference techniques [37] to infer a predictive dynamical model. The model will be specified in terms of a learned Green's function – linear dynamics in mode space – that propagates initial cell density and flux data forward in time. To validate the approach, we demonstrate that it correctly recovers the hydrodynamic equations for active Brownian particle (ABP) dynamics on curved surfaces. Subsequently, as a first example application to experimental single-cell tracking data, we consider the pan-embryonic cell migration during early gastrulation in zebrafish [8], an important vertebrate model system for studying various morphogenetic events [158, 177, 178]. During gastrulation, complex migratory cell movements organize several thousand initially undifferentiated cells into different germ layers that lay out the primary body plan [179]. The underlying high-dimensional single-cell data make this process a prototypical test problem for illustrating how spatiotemporal information can be efficiently compressed to analyze and model biological structure formation.

Broadly, our goal is to translate experimentally measured single-cell trajectories on a

curved surface into a quantitative model of collective cell migration dynamics. As a specific example, we consider recently published light sheet microscopy data that captures the individual movements of thousands of cells during early zebrafish development from epiboly onset at 4 hours post-fertilization (hpf) to about 18 hpf from the Huisken lab [8]. This developmental period is characterized by a collective symmetry-breaking event during which cells collectively migrate over the yolk cell surface [179]. Namely, they rearrange from an initial localization around the animal pole (AP) [Figure 5.1(a), left] into a more elongated configuration that already indicates the basic geometry of the fully developed zebrafish larva [Figure 5.1(a), right]. Working with a two-dimensional (2D) sphere projection of the experimental data, we first coarse-grain the trajectories on a curved surface. Next, we construct a sparse mode representation of the resulting hydrodynamic fields in terms of scalar and vector spherical harmonic basis functions and then use the resulting spectral representation to infer linear dynamics that capture the initial symmetry-breaking event. We validate this mode representation framework and the subsequent model inference using synthetic data of ABPs on a sphere, for which coarse-grained fields and learned models can be directly compared against analytical predictions. Finally, we infer a linear model for the mode dynamics of the experimental zebrafish data, which enables us to study the characteristics of cell interactions through kernels that couple cell density and flux and compare their features with the hydrodynamic mean-field signatures of ABPs on a sphere.

5.2 Coarse-graining of cellular dynamics on a spherical surface

The experimentally observed cell motions are approximately two-dimensional (2D): The radius of the yolk cell surface on which the dynamics take place is much larger than the average height changes of the evolving cell mass [8]. Therefore, we adopt a thin film approximation, in which the cellular motion is represented on an effective spherical mid-surface [gray surface in Figure 5.1(b)]; refined future models should aim to account for the full 3D dynamics. Focusing on the in-plane dynamics, we project all cell positions and velocities onto a spherical mid-surface $\mathcal{S}$ of radius $R_s = 300\,\mu$m. On this spherical surface, each cell $\alpha = 1, 2, ..., N$ has a position $\mathbf{r}_\alpha(t)$ and in-plane velocity $\mathbf{v}_\alpha(t) = \mathrm{d}\mathbf{r}_\alpha/\mathrm{d}t$.

As a second processing step, a coarse-grained representation of the single-cell dynamics on a spherical surface is determined. To facilitate the applicability of our framework to a wide range of experimental inputs, we propose a coarse-graining approach that can flexibly integrate cell number variations stemming from cell divisions, as well as those from experi-

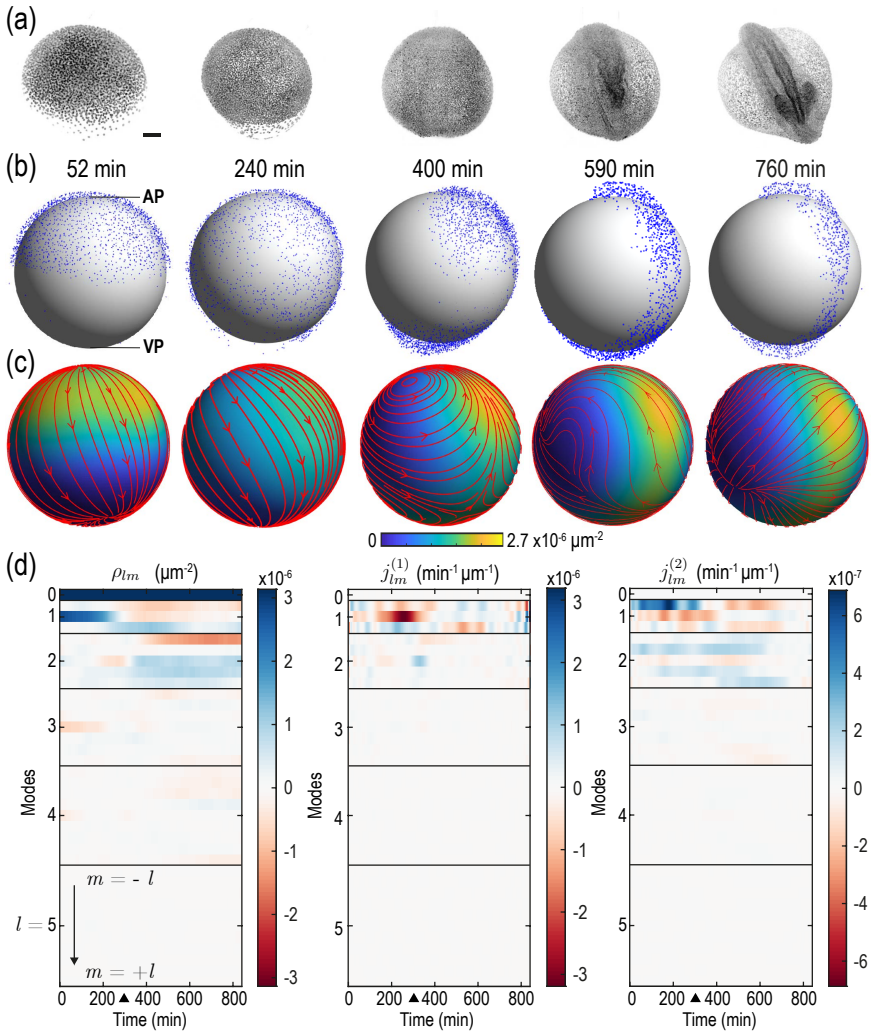

Figure 5.1: From single-cell tracking data to sparse mode amplitude representations. (a) Microscopic imaging data of early zebrafish development (adapted from the Huisken lab [180]) shows cell migration from an initially homogeneous pole of cells (left) towards an elongated structure that indicates the head-tail axis of the fully developed organism. Scale bar, $100\,\mu m$. (b) Experimental single-cell tracking data from [8] (blue dots) during similar developmental time points (± 20 min) as in (a). $t = 0$ min for the indicated time points in $\mathbf{B}$ corresponds to a developmental time of 4 hours post fertilization. The z-axis points from the ventral pole (VP) to the animal pole (AP). (c) Coarse-grained relative cell density $\rho(\mathbf{r}, t)$ (color) and associated coarse-grained flux $\mathbf{J}(\mathbf{r}, t)$ (streamlines) determined from single cell positions and velocities from data in (b) via equations (5.1). The thickness of streamlines is proportional to the logarithm of the spatial average of $|\mathbf{J}|$. (d) Dynamic harmonic mode representation of the relative density $\rho(\mathbf{r}, t)$ [equation (5.5), left panel] and of the flux $\mathbf{J}(\mathbf{r}, t)$ [equation (5.6), middle and right panel] for fields shown in (c). The modes $j_{lm}^{(1)}$ correspond to compressible, divergent cell motion, the modes $j_{lm}^{(2)}$ describe incompressible, rotational cell motion. Mode amplitudes become negligible for $l \geq 5$. For all panels, horizontal black lines delineate blocks of constant harmonic mode number l, and black triangles denote the end of the epiboly phase.

mental uncertainties in cell imaging and tracking. Details of the coarse-graining procedure can be found in [181]. Briefly, we translate microscopic cell positions $\mathbf{r}_\alpha(t)$ and velocities $\mathbf{v}_\alpha(t)$ into a continuous cell surface density $\rho(\mathbf{r}, t)$ and an associated flux $\mathbf{J}(\mathbf{r}, t)$ at any point $\mathbf{r}$ of the spherical mid-surface through a kernel coarse-graining of the form

$$\rho(\mathbf{r}, t) = \frac{1}{N} \sum_{\alpha=1}^{N} K\left[\mathbf{r}, \mathbf{r}_\alpha(t)\right] \tag{5.1a}$$

$$\mathbf{J}(\mathbf{r}, t) = \frac{1}{N} \sum_{\alpha=1}^{N} \mathcal{K}\left[\mathbf{r}, \mathbf{r}_\alpha(t)\right] \cdot \bar{\mathbf{v}}_\alpha, \tag{5.1b}$$

where N is the total number of cells and $\bar{\mathbf{v}}_\alpha = \mathbf{v}_\alpha / |\mathbf{r}_\alpha|$ is the angular velocity of a given cell on a reference unit sphere. The kernels K and $\mathcal{K}$ are chosen so that the field ρ and $\mathbf{J}$ are related by mass conservation,

$$\frac{\partial \rho}{\partial t} + \nabla_{\mathcal{S}} \cdot \mathbf{J} = 0. \tag{5.2}$$

However, because cell divisions are essential to most developmental processes, total cell numbers will, in many cases – including early zebrafish gastrulation [180] – vary over time. In the following analysis, we thus focus on normalized fields, $\tilde{\rho} = N(t)\rho$ and $\tilde{\mathbf{J}} = N(t)\mathbf{J}$ that satisfy the modified mass conservation equation,

$$\frac{\partial \tilde{\rho}}{\partial t} + \nabla_{\mathcal{S}} \cdot \tilde{\mathbf{J}} = k(t)\tilde{\rho}. \tag{5.3}$$

where $k(t) = \dot{N}(t)/N(t)$ denotes a time-dependent effective growth rate.

Information loss due to coarse graining

Coarse-graining microscopic data into smooth fields is an irreversible operation, during which some of the original particle information is irretrievably lost. As seen in the second example of Chapter 3, the choice of coarse-graining scale is thus dictated by a trade-off between smoothness and information content - choosing larger coarse-graining scales leads to smoother fields but blurs finer scale structures which may be of interest. To inform our choice of coarse-graining scale, we quantify the loss of information incurred by the coarse-graining operation.

The measure we introduce to quantify information loss is based on the the well-known relationship between the smoothness of functions in real space and Fourier space [182]: A smooth function in real space should have a peaked, quickly decaying spectrum in Fourier space while a collection of point-like objects such as delta functions should have a uniform non-decaying spectrum. Specifically, we describe a uniformly sampled field as a $M \times N$ matrix with components being the field values $X_{i,j} = X(\theta_i, \phi_j)$. In our case, $X_{i,j}$ represents

either the density field ρ or any of the Cartesian components of the flux vector field $\mathbf{J}$ at a given time point. We find the complex discrete Fourier spectrum $\hat{X}_{i,j}$ of this matrix using the two-dimensional fast Fourier transform. We then calculate the power spectral density (PSD) of the Fourier spectrum as $R_{i,j} = |\hat{X}_{i,j}|^2$ and interpret the normalized PSD

$$P_{i,j} = \frac{R_{i,j}}{\sum_{a,b} R_{a,b}}$$

as a discrete probability distribution. The spectral entropy S characterizing the information content of the field X is then defined by

$$S(X) = -\frac{1}{\log_2 NM} \sum_{i,j} P_{i,j} \log_2 P_{i,j}. \tag{5.4}$$

Smooth fields are sharply peaked in Fourier space and have a low spectral entropy, whereas fields that resolve discrete single particle information are rather flat in Fourier space and have a large spectral entropy. The difference in entropy between particle data and smoothed fields then measures the information eliminated by the coarse-graining procedure. If we additionally normalize by the entropy of the spectral entropy $S_0(X)$ of the raw particle data, we finally obtain a relative measure of the information that is lost in the coarse-graining process. In general, a measure as given in equation (5.4) can be defined for any transform with the property that smoothness in real space leads to a fast decaying spectrum in transform space.

We compute the spectral entropy of density and flux component fields at a representative time point and for varying coarse-graining length scales (Figure 5.2). Specifically, we coarse-grain density and flux for different values of the kernel parameter k that determines the coarse graining length scale. Large values of k correspond to small coarse-graining length scales, with the effective half-width at half-maximum of the kernels scaling as $\mathrm{HWHM} = \arccos(2^{-1/k})$. Normalized spectral entropies $S(X)/S_0(X)$ with $X \in \{\rho, \mathbf{J}\}$ are then computed using (5.4). For the flux field, we define $S(\mathbf{J}) := S(J_x) + S(J_y) + S(J_z)$ ("Flux sum" in Figure 5.2) and interpret the sum of these three contributions ("Flux x", "Flux y", "Flux z" in Figure 5.2) as the total information contained in the flux field. We find that the spectral entropies of all fields show similar features. In particular, an increasing coarse-graining width first results in a sharp loss of information as individual particle positions are blurred, followed by less steep information loss as continuous fields progressively lose details of finer structures. In this work, we use an intermediate value of the coarse-graining parameter $k = 6$ (yellow data in Figure 5.2).

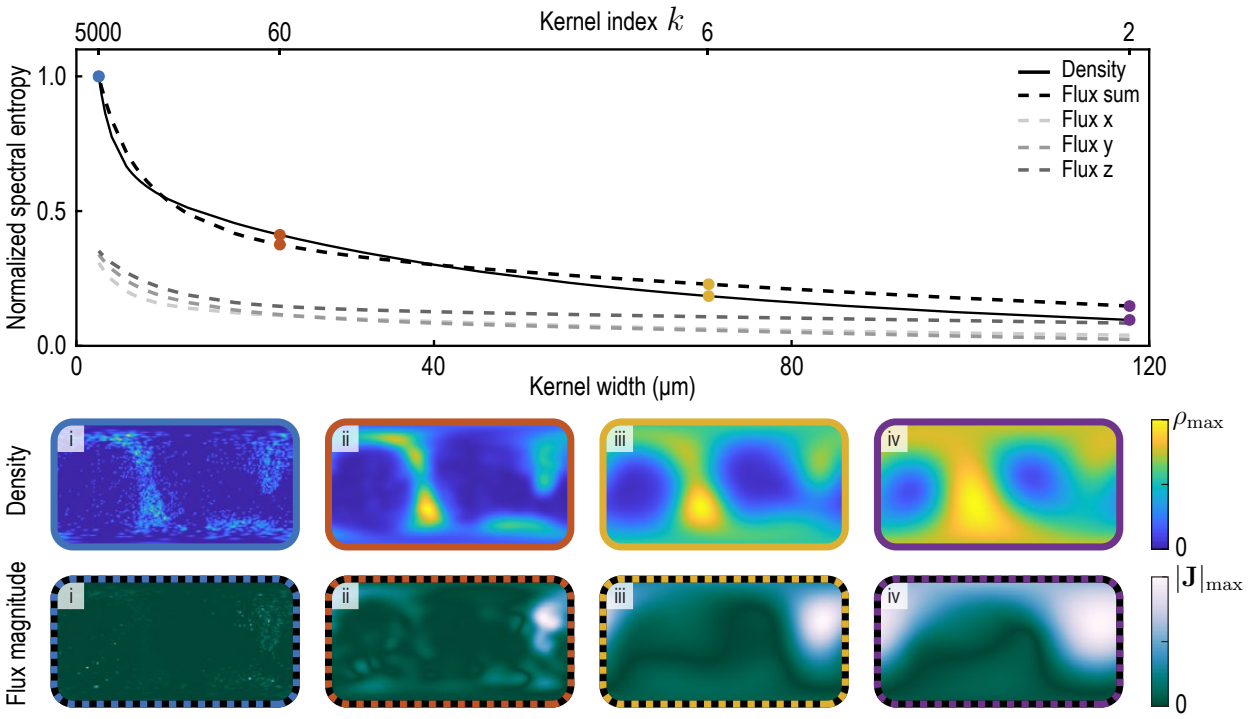

Figure 5.2: Normalized spectral entropy as a function of the coarse-graining kernel width (top) computed for density ρ and flux field $\mathbf{J}$ using equation (5.4). To evaluate the spectral entropy for the vector-valued flux, we define $S(\mathbf{J}) := S(J_x)+S(J_y)+S(J_z)$ ("Flux sum"). The coarse-graining width – the half-width at half-maximum $\arccos(2^{-1/k})$ is varied by varying the kernel index k. The fields ρ and $|\mathbf{J}|$ are shown in the two bottom rows for different values of k. i. $k = 5000$ (blue, data used to compute the reference spectral entropies $S_0(\rho)$ and $S_0(\mathbf{J})$) ii. $k = 60$ (brown) iii. $k = 6$ (yellow, used in main text) iv. $k = 2$ (purple).

5.3 Mode representation of spatiotemporal dynamics on a spherical surface

5.3.1 Spatial mode representation

To obtain a sparse mode representation of the hydrodynamic fields $\rho(\mathbf{r},t)$ and $\mathbf{J}(\mathbf{r},t)$ on the spherical surface, we expand them in terms of scalar and vector spherical harmonics (SHs) [183, 184] [181]. Spherical harmonics are defined on points $\hat{\mathbf{r}} = \mathbf{r}/R_s$ of the unit sphere, where $R_s = 300\,\mu$m is the mid-surface radius. In this basis, the scalar density field is represented as

$$\rho(\mathbf{r},t) = \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} \rho_{lm}(t) Y_{lm}(\hat{\mathbf{r}}), \tag{5.5}$$

which conveniently separates the time- and space-dependence of $\rho(\mathbf{r},t)$ into mode amplitudes $\rho_{lm}(t)$ and scalar harmonic functions $Y_{lm}(\hat{\mathbf{r}})$, respectively. The maximal mode number $l_{\max}$ is a proxy for the maximal spatial resolution at which $\rho(\mathbf{r},t)$ is faithfully represented. Similarly, the vector-valued flux $\mathbf{J}(\mathbf{r},t)$ can be decomposed into time-varying mode amplitudes $j^{(1)}_{lm}(t)$ and $j^{(2)}_{lm}(t)$, while its spatial dependence is described by vector spherical harmonics $\mathbf{\Psi}_{lm}(\hat{\mathbf{r}})$ and $\mathbf{\Phi}_{lm}(\hat{\mathbf{r}})$ [184],

$$\mathbf{J}(\mathbf{r},t) = \sum_{l=1}^{l_{\max}} \sum_{m=-l}^{l} \left(j^{(1)}_{lm}(t) \mathbf{\Psi}_{lm}(\hat{\mathbf{r}}) + j^{(2)}_{lm}(t) \mathbf{\Phi}_{lm}(\hat{\mathbf{r}}) \right). \tag{5.6}$$

Besides the in-plane divergence $\nabla_{\mathcal{S}} \cdot \mathbf{J}$ that leads to local density changes, the cell number flux $\mathbf{J}(\mathbf{r},t)$ also contains an in-plane curl component $\nabla_{\mathcal{S}} \times \mathbf{J}$ that is associated with locally rotational cell flux. The two sets of vector spherical harmonics $\{\mathbf{\Psi}_{lm}\}$ and $\{\mathbf{\Phi}_{lm}\}$ conveniently decompose the flux into these contributions: $j^{(1)}_{lm}(t)$ corresponds to modes that drive density changes and $j^{(2)}_{lm}(t)$ represents modes of local rotational cell motion that change relative cell positions but do not change local density.

5.3.2 Mass conservation constraint in mode space

Using harmonic mode representations of the cell number density (5.5) and the cell number flux (5.6) directly in the continuity equation (5.2) we find a set of ordinary differential equation in mode space

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_{lm}(t) = \frac{l(l+1)}{R_s} j^{(1)}_{lm}(t), \tag{5.7}$$

where $l = 0, 1, ..., l_{\max}$ and for each value of l, $m = -l, -l+1, ..., l-1, l$. Equation (5.2) offers an alternative way of determining the modes $j^{(1)}_{lm}(t)$ directly from the modes $\rho_{lm}(t)$ of the coarse-grained cell number density, while ensuring that the resulting fields obey mass conservation exactly. In practice, the modes $j^{(1)}_{lm}(t)$ found from a vector harmonic representation of the coarse-grained cell number flux (5.1b) will often deviate from modes $j^{(1)}_{lm}(t)$ determined from (5.2), even if cell numbers are expected to be conserved. This can be, for example, due to limited accuracy in determining velocities $\mathbf{v}_\alpha(t)$ from noisy single-cell trajectories $\mathbf{r}_\alpha(t)$, or due to spatially inhomogeneous appearances and disappearances of cells in tracking data. Consistent with our simplifying assumption that a globally homogeneous growth rate can sufficiently well approximate cell number changes in the data, the subsequent analysis uses the modes $j^{(1)}_{lm}(t)$ as determined from the density modes $\rho_{lm}(t)$ via. (5.7), together with modes $j^{(2)}_{lm}(t)$ from the explicit velocity coarse-graining Eq. (5.1b).

The representation of $\rho(\mathbf{r}, t)$ and $\mathbf{J}(\mathbf{r}, t)$ in terms of spherical harmonic modes with $l \leq l_{\max}$ leads in total to $3(l_{\max}+1)^2$ mode amplitude trajectories, displaying only a few dominant contributions (Figure 5.1(d)) with almost no signal remaining for $l \geq 5$. This demonstrates that the underlying coarse-grained experimental data is sufficiently smooth and implies that a spectral representation is indeed meaningful.

5.3.3 Temporal mode representation

We further compress the dynamical information by representing the time series of the modes in terms of Chebyshev polynomial basis functions $T_n(t)$ [21, 44] (Section 2.1.2). To simplify notation, we define a dynamic mode vector $\mathbf{a}(t) = [\rho_{lm}(t), j^{(1)}_{lm}(t), j^{(2)}_{lm}(t)]^\top$ that collects all the modes up to $l = l_{\max}$ determined in the previous section and consider an expansion

$$\mathbf{a}(t) = \sum_{n=0}^{n_{\max}} T_n(t)\, \hat{\mathbf{a}}_n \tag{5.8}$$

in terms of the spatiotemporal mode coefficients $\hat{\mathbf{a}}_n$ with temporal mode number n. This compression allows us to accurately evaluate time derivatives of the mode amplitudes [84], an important step when using (5.7) to determine flux modes $j^{(1)}_{lm}(t)$ directly from density modes ρ_{lm}. Fixing $l_{\max} = 4$ and $n_{\max} = 30$ in the remainder, the initial single-cell data set of about 1.4 million recorded cell position entries, or 4.2 million degrees of freedom, has thus been reduced to 2250 mode coefficients, corresponding to a compression ratio $\gtrsim 1800$.

5.3.4 Characterization of the developmental mode dynamics

The associated harmonic basis functions and vectors have an intuitive physical meaning, convenient algebraic properties, and encode information about the length scales and symmetries of the collective dynamics. By studying the evolution and relative importance of different mode trajectories [181], we find that mode signatures of the symmetry breaking and progression through developmental stages are robust, illustrating that mode-based analysis can provide a systematic and meaningful characterization of the developmental symmetry-breaking event.

5.4 Learning a linear hydrodynamic model of the developmental mode dynamics

A low-dimensional mode representation can capture essential characteristics of developmental symmetry-breaking processes [181]. The mode representation, therefore, provides a natural starting point for the inference of hydrodynamic models from coarse-grained cell-tracking data. For a given time-dependent mode vector $\mathbf{a}(t) = [\rho_{lm}(t), j^{(1)}_{lm}(t), j^{(2)}_{lm}(t)]^\top$ that contains all modes up to $l = l_{\max}$, the simplest hydrodynamic model corresponds to the linear dynamical equation

$$\frac{\mathrm{d}\mathbf{a}(t)}{\mathrm{d}t} = M \cdot \mathbf{a}(t), \tag{5.9}$$

where the constant coefficient matrix M encodes the couplings between different modes. Intuitively, (5.9) aims to describe an experimentally observed density and flux dynamics in terms of a relaxation process, starting from inhomogeneous initial conditions represented by $\mathbf{a}(0)$. The mathematical learning problem is then to find a coefficient matrix M such that the linear model (5.9) holds for the mode vector time series $\mathbf{a}(t)$ that was determined from the coarse-graining procedure described in the previous sections. The method used to infer the linear matrix was detailed in Chapter 4

5.4.1 Validation of the learning framework using active Brownian particle dynamics

Before applying the combined coarse-graining and inference framework to experimental data, we illustrate and validate the learning approach on synthetic data for which coarse-graining results and hydrodynamic mean-field equations are analytically tractable. To this end, we consider the stochastic dynamics of non-interacting active Brownian particles (ABPs) on

the unit sphere of radius $R_0 = 1$ [181, 185–187]. Similar to a migrating cell, an ABP at position $\mathbf{x}(t)$ moves across the unit sphere at constant speed v_0 in the direction of its fluctuating orientation unit vector $\mathbf{u}(t)$. The strength of the orientational Gaussian white noise is characterized by a rotational diffusion constant D_r. Details of the full model can be found in [181]. We use simulated non-interacting ABPs on a unit sphere as a test data set, with approximately 1.2×10^5 particles in total. The resulting trajectory data were coarse-grained following the procedure outlined in the previous sections, yielding dynamic density fields $\rho(\mathbf{r}, t)$ and fluxes $\mathbf{J}(\mathbf{r}, t)$, together with their mode representations $\rho_{lm}(t), j^{(1)}_{lm}(t)$ and $j^{(2)}_{lm}(t)$.

In the 'learning' step, we infer a sparse mode coupling matrix M that approximates the dynamics (5.9) for the dynamical mode vectors $\mathbf{a}(t) = [\rho_{lm}, j^{(1)}_{lm}, j^{(2)}_{lm}]^\top$ obtained from the coarse-grained simulated ABP data. The full algorithm is detailed in Chapter 4 and illustrated in the summary flowchart Figure 4.1. In the inference, we enforce the sparsity structure given by mass conservation but do not enforce the values of the entries. Checking that these values agree with mass conservation provides a good way to check that the inference has converged to a good minimum. Importantly, we perform the sparse regression using dynamical mode vectors $\mathbf{a}(t)$ rescaled by their median absolute deviation (MAD) to compensate for substantial scale variations between different modes. The final output matrix M of this learning algorithm is shown in the right panel of Figure 5.3(c) and can be compared against the analytically coarse-grained dynamics of ABPs on curved surfaces [186, 187]. Under suitable closure assumptions [181], the mean-field dynamics of ABPs on a unit sphere is given in harmonic mode space by

$$\frac{\mathrm{d}\rho_{lm}}{\mathrm{d}t} = \frac{l(l+1)}{R_0} j^{(1)}_{lm} \tag{5.10a}$$

$$\frac{\mathrm{d}j^{(1)}_{lm}}{\mathrm{d}t} = -\frac{v_0^2}{2R_0}\rho_{lm} - D_r j^{(1)}_{lm} \tag{5.10b}$$

$$\frac{\mathrm{d}j^{(2)}_{lm}}{\mathrm{d}t} = -D_r j^{(2)}_{lm}, \tag{5.10c}$$

from which we can read off the mode coupling matrix M shown in the left panel of Figure 5.3(c). A direct comparison between the theoretical and the inferred matrices shows that our framework recovers both the structure and the quantitative values of M with good accuracy. Due to the finite number of ABPs used to determine the coarse-grained fields, we do not expect that the theoretically predicted coupling matrix will be recovered perfectly from the data. Instead, some mode couplings suggested by (5.10) may not be present or modified in the particular realization of the ABP dynamics that was coarse-grained. Indeed, direct

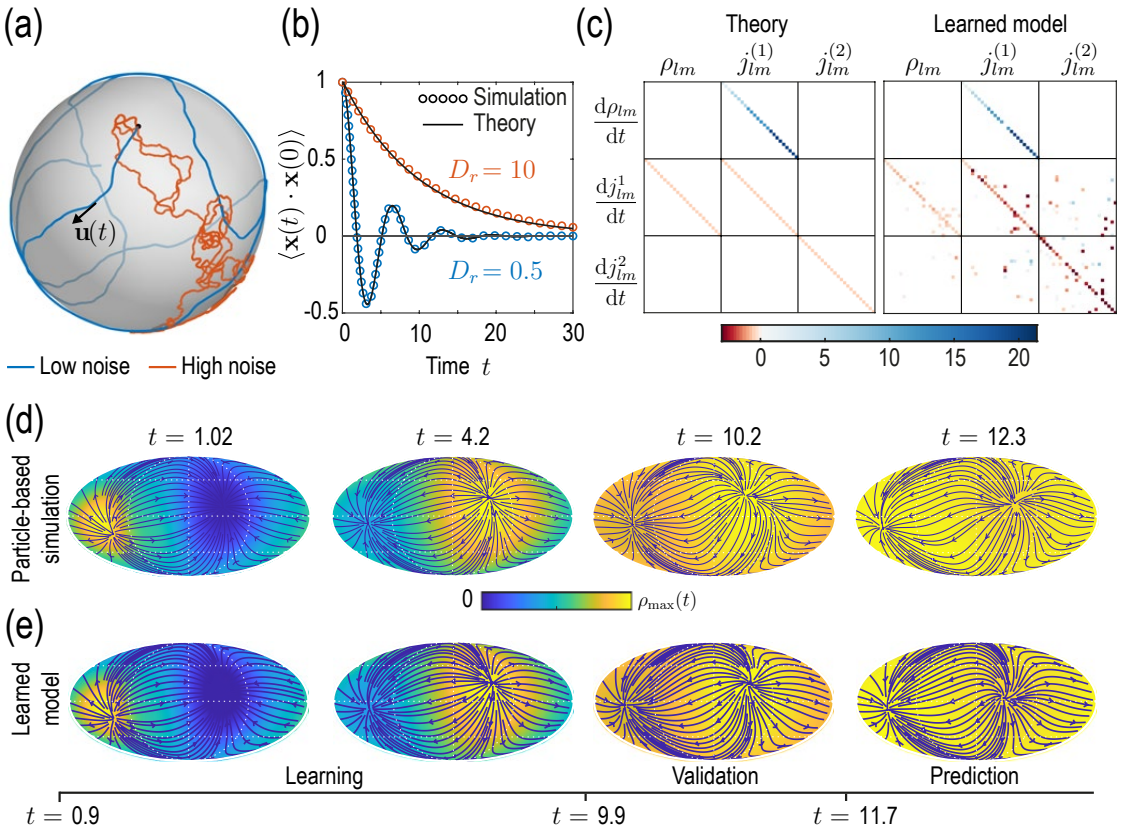

Figure 5.3: Learning active Brownian particle (ABP) dynamics on a sphere. (a) ABPs move on a unit sphere (radius $R_0 = 1$) with angular speed $v_0 = 1$ along a tangential unit vector $\mathbf{u}(t)$ that is subject to stochastic in-plane fluctuations. Example single-particle trajectories are shown in the high-noise (orange, $D_r = 10$ in units of $R_0 v_0$) and in the low-noise regime (blue, $D_r = 0.5$). Time t is measured in units of R_0/v_0 in all panels. (b) Position correlation function $\langle \mathbf{x}(t) \cdot \mathbf{x}(0) \rangle$ averaged over 3×10^4 independent ABP trajectories show distinct oscillations of period $\approx 2\pi$ in the low-noise regime, as ABPs orbit the spherical surface more persistently. The standard error of the mean is smaller than the symbol size. (c) Analytically predicted (left) and inferred (right) dynamical matrices M [see equation (5.9)] describing the mean-field dynamics of a large collection of non-interacting ABPs show good quantitative agreement. (d) Mollweide projections of coarse-grained ABP simulations with $v_0 = 1$ and $D_r = 0.5$ using cell positions from the first time point in the zebrafish data (Figure 5.1) as the initial condition: At each position, 60 particles with random orientation were generated and their ABP dynamics simulated, amounting to approximately 1.2×10^5 particles in total. The density fields homogenize over time, where the maximum density at $t = 12.3$ has decayed to about 5 % of the maximum density at $t = 1.02$. Blue lines and arrows indicate streamlines of the cell flux $\mathbf{J}(\mathbf{r}, t)$. (e) Simulation of the learned linear model (5.9) with M shown in Figure 5.3(c) (right), for the same initial condition as in (d). Marked time points indicate intervals of learning, validation, and prediction phases of the model inference.

simulation of the learned model projected in real space [Figure 5.3(e)] reveals density and flux dynamics that agree very well with the dynamics of the coarse-grained input data [Figure 5.3(d)]. Altogether, these results demonstrate that the proposed inference framework enables us to faithfully recover expected mean-field dynamics from coarse-grained fields of noisy particle-based data.

5.4.2 Learning developmental mode dynamics from experimental data

The same inference framework can now be directly applied to the coarse-grained experimental zebrafish embryo data shown in Figure 5.1(c, d), yielding a sparse coefficient matrix M [Figure 7.4(a, b)] that encodes the dynamics of the developmental mode vector $\mathbf{a}(t) = [\rho_{lm}(t),\, j^{(1)}_{lm}(t),\, j^{(2)}_{lm}(t)]^\top$ according to (5.9). The inferred coupling between the time derivative of density modes ρ_{lm} and flux modes $j^{(1)}_{lm}$ faithfully recovers mass conservation [Figure 7.4(c) and (5.7)]. Overall, the learned matrix M has 395 non-zero elements, effectively providing further compression of the experimental data, which required 2250 spatiotemporal mode coefficients collected in $\hat{\mathbf{a}}_n$ [see (5.8)] for its representation. Using the mode vector $\mathbf{a}(t = 0)$ of the first experimental time point as the initial condition, the inferred minimal model (5.9) with M shown in Figure 7.4(a, b) faithfully recovers both the mode and real-space dynamics seen in the coarse-grained fields of the experimental input data [Figure 7.4(e–g)].

It is instructive to analyze the inferred matrix M and the linear model it encodes in more detail. Comparing the MAD-rescaled matrix learned for the experimental zebrafish data [Figure 7.4(b)] with the non-dimensionalized matrix learned for the active Brownian particle dynamics [Figure 5.3(c)], we find similar patterns of prominent diagonal and block-diagonal couplings. Consistent with the analysis of single-cell trajectories [8], this suggests that a random but persistent movement of cells akin to ABPs moving on a sphere partially contributes to the early gastrulation process in zebrafish. This is complemented in the minimal model of the experimental dynamics by significant off-diagonal contributions [Figure 7.4(b)], which are absent in the non-interacting ABP model. Such off-diagonal contributions represent effective linear approximations of cell-cell interactions, environmental influences, or other external stimuli reflected in the experimental time-series data. Ultimately, such contributions to the mode coupling matrix M help realize the symmetry-breaking process observed in the underlying experimental data.

The inferred mode coupling matrix M shown in Figure 7.4(b) together with (5.9) provides a highly robust minimal model. Specifically, despite being linear, it is numerically stable over a period of approximately four times as long as the input data from which the matrix M

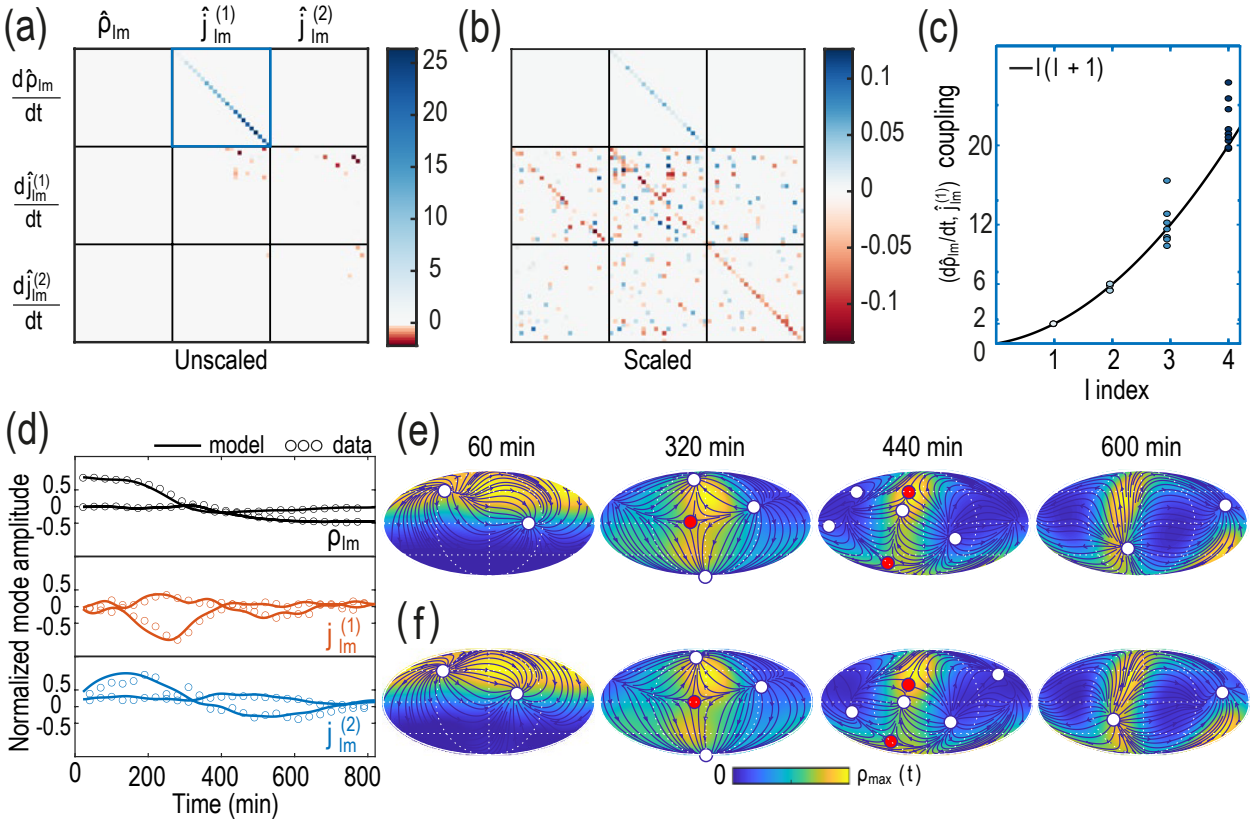

Figure 5.4: Model learning for experimental data of collective cell motion during early zebrafish development. (a) Visualization of the constant mode coupling matrix M that was learned from experimental data and describes the dynamics of the mode vector $\mathbf{a} = [\rho_{lm}(t),\, j^{(1)}_{lm}(t),\, j^{(2)}_{lm}(t)]^T$ via (5.9). Dimensionless fields are defined by $\hat{\rho}_{lm} = R_s^2 \rho_{lm}$ and $\hat{j}^{(i)}_{lm} = R_s \Delta t\, j^{(i)}_{lm}$ $(i = 1, 2)$ with $R_s = 300\,\mu$m and $\Delta t = 2\,$min. (b) Scaling the learned matrix M by the Mean Absolute Deviation (MAD) of the modes reveals structures reminiscent of the mode coupling matrix learned for ABPs (Figure 5.3(c)). (c) The learned model recovers mass conservation (5.2) in mode space. (d) Comparison of experimental mode dynamics (circles) with numerical solution (solid line) of the minimal model (5.9) for learned matrix M visualized in Figure 5.4(a). For clarity, the comparison is shown for the two dominant modes of each set of harmonic modes ρ_{lm}, $j^{(1)}_{lm}$ and $j^{(2)}_{lm}$. (e, f) Mollweide projections of the experimental data (e) and of the numerical solution of the learned model (f) show very good agreement. Blue lines and arrows illustrate streamlines defined by the cell flux $\mathbf{J}(\mathbf{r}, t)$, circles depict defects with topological charge $+1$ (white) and -1 (red).

was learned. Furthermore, simulations with modified initial conditions (see Figure 5.5) still exhibit a characteristic symmetry breaking and lead to the emergence of density and flux patterns similar to those seen in Figure 7.4(f, g). For example, simulating (5.9) using the initial condition of a different experimental data set leads to final patterns with the same symmetry as in the original training data, further corroborating that the observed symmetry breaking is directly encoded in the interactions represented by the matrix M. A similar robustness is observed under moderate perturbations of the initial condition, such as a rotation of initial cell density patterns relative to the coordinate system in which M was inferred or a local depletion of the initial density, emulating a partial removal of cells as experimentally realized in [178]. Taken together, these numerical experiments demonstrate that the inferred mode coupling matrix M meaningfully captures the dynamics and interactions of cells that facilitate the symmetry breaking observed during early zebrafish development.

Additional real-space analysis and comparison of inferred interaction kernels further highlight potential ABP-like contributions to the collective cellular organization during early zebrafish development and reveal an effectively non-local coupling between density and flux dynamics [181]. The latter could result, for example, from unresolved fast-evolving morphogens [188], through mechanical interactions with the surrounding material [189] or due to other relevant degrees of freedom that are not explicitly captured in this linear hydrodynamic model. More generally, a real-space representation of kernels provides an alternative interpretable way to study the interactions and symmetry-breaking mechanisms encoded by models directly learned in mode space.

5.5 Discussion

Leveraging a sparse mode representation of collective cellular dynamics on a curved surface, we have presented a learning framework that translates single-cell trajectories into quantitative hydrodynamic models. This work complements traditional approaches to find quantitative continuum models of complex multicellular processes [178, 189–192] that match problem-specific constitutive relations of active materials in real-space with experimental observations. The successful applications to synthetic ABP simulation data and experimental zebrafish embryo data show that model learning in mode space provides a promising and computationally feasible approach to infer quantitative interpretable models in complex geometries.

The learned linear minimal model for cell migration during early zebrafish morphogenesis quantitatively recapitulates the spatiotemporal dynamics of a complex developmental process [Figure 7.4(f, g)], and highlights similarities between collective cell migration and

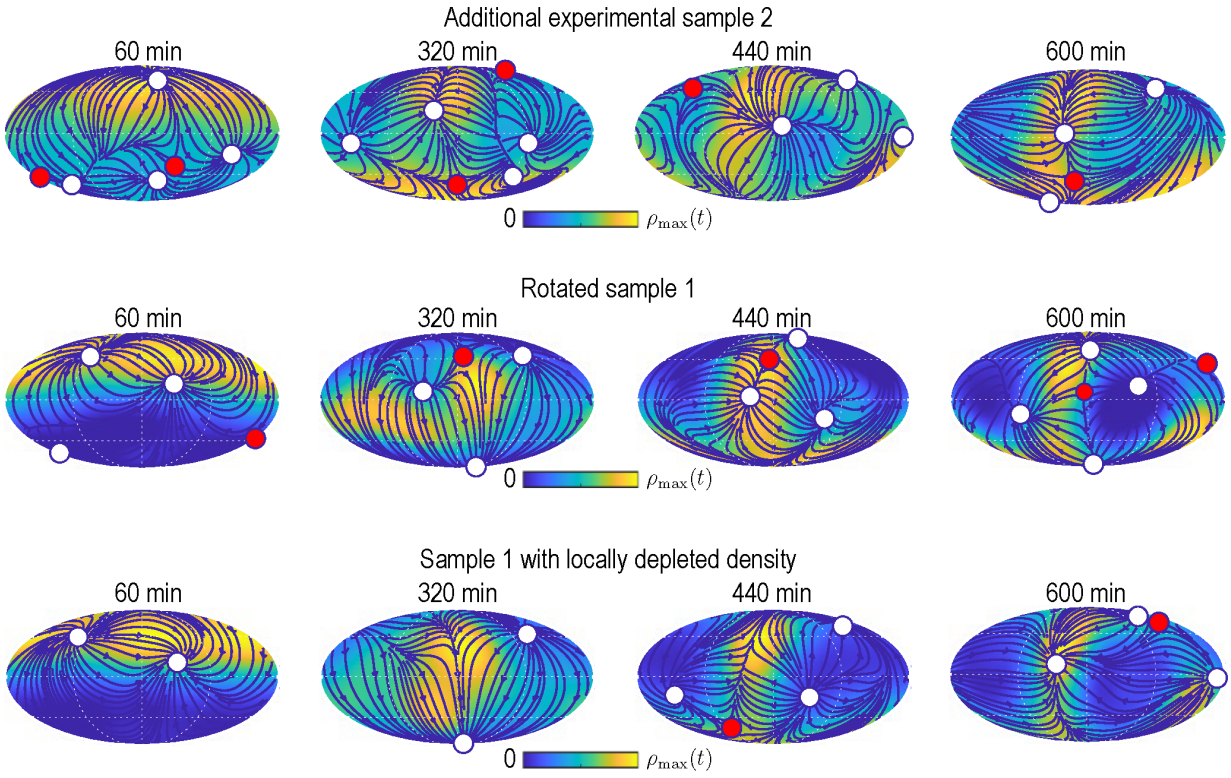

Figure 5.5: Simulating the learned model with different initial conditions. Mollweide projections from simulations of the model Eq. (5.9) with M depicted in Figure 7.4B that was learned for experimental data from sample 1, but using different initial conditions (from top to bottom): initial condition from experimental data set sample 2; initial condition from sample 1 rotated by 10° away from the animal pole; initial condition from sample 1 with $\epsilon = 10\%$ of the density at the animal pole removed. Blue lines and arrows illustrate streamlines defined by the cell flux $\mathbf{J}(\mathbf{r}, t)$, circles depict defects with topological charge $+1$ (white) and -1 (red).

analytically tractable ABP dynamics on a curved surface. An extension to nonlinear mode-coupling models or integration of additional, experimentally measured degrees of freedom, such as concentration fields of morphogens involved in mechanochemical feedbacks [188], is in principle straightforward by including nonlinear terms in the model (5.9). Furthermore, the above framework could be generalized to describe the dynamics within a spherical shell of finite height by complementing the surface vector spherical harmonics used in this work by their radial counterpart [193].

To provide a concrete example, we focused here on applying the model learning framework to single-cell tracking data of early zebrafish morphogenesis. However, the essentially spherical organization of cells during gastrulation observed in zebrafish is shared by many species whose early development occurs through a similar discoidal cleavage [157], and the framework introduced here is directly applicable once tracking data becomes available for these systems. More generally, as novel imaging technologies are being developed [8, 194, 195], we expect that even more extensive and more detailed imaging data will further facilitate the exploration of finer scales and length-scale bridging processes [196] through learning approaches that directly build on mode-based data representations.

Chapter 6

Structured linear dynamical systems inference: planar curve dynamics

The results in this chapter have previously been published in A. E. Cohen, A. D. Hastewell*, S. Pradhan, S. W. Flavell, and J. Dunkel, "Schrödinger Dynamics and Berry Phase of Undulatory Locomotion," Physical Review Letters, 130(25), 258402, 2023.*

In this chapter, we develop the theoretical framework for constrained linear dynamics of planar curve dynamics in mode space. At the end of the chapter, we outline adaptations to the optimization procedure introduced in Chapter 4 for inferring the resulting structured linear dynamical system, developed in collaboration with A. E. Cohen.

6.1 Motivation

Many physical and biological systems, from quantum [197] and many-body systems [198] to animal locomotion [199] and front propagation [200] can be characterized through the evolution of planar curves. In many cases, however, a complete characterization of the system may be intractable, and we, therefore, wish to construct low-dimensional effective curve dynamics in mode-space and infer these dynamics directly from data. In this chapter, we consider linear mode-space models of open planar curves evolving with approximately constant length and how we can efficiently infer such dynamics directly from experimental data. As we have seen previously, compared with traditional continuum descriptions of planar curve dynamics in position space, formulating models in mode space [17, 18, 181, 201] offers several theoretical and practical advantages:

1. high-dimensional experimental data can be efficiently compressed to obtain an interpretable low-dimensional representation;

2. the mode dynamics reduces to a system of linear ordinary differential equations (ODEs);

3. physical symmetries and biological constraints can be efficiently encoded through the structure of the dynamical matrix;

4. all model parameters can be directly inferred from experimental data using ODE sensitivity methods [37, 202] that exploit the imposed matrix structure [203].

In particular, for undulatory curve locomotion, we find that translational invariance, rotational invariance, and length constraints generically lead to a Schrödinger equation [204] in mode space.

6.2 Linear mode space models of planar curve dynamics

6.2.1 Spectral representation of planar curves

Starting from a vector field representation of the curve, $\mathbf{r} = (x(t,s), y(t,s))$, where $s \in [-1, 1]$ is the parameter along the curve and $\mathbf{r}$ is the Cartesian coordinate of the curve at s, we construct a single complex field $z(t,s) = x(t,s) + iy(t,s)$. We can represent this field with an expansion in Chebyshev polynomials

$$z(t,s) = \sum_{k=0}^{n} T_k(s)\hat{z}_k(t) = \sum_{k=0}^{n} T_k(s)\left[\hat{x}_k(t) + i\hat{y}_k(t)\right] \tag{6.1}$$

which have advantageous analytical and computational properties; in principle other basis functions could also be chosen.

Interpretation of coefficients

The coefficients in the Chebyshev expansion have nice interpretations as physical properties of the curve. Starting from the definition of the Chebyshev coefficients (2.5) and using the fact that $T_0 = 1$, the $n = 0$ coefficient is given by,

$$\hat{z}_0(t) = \hat{x}_0(t) + 1i\hat{y}_0(t) = \frac{1}{\pi}\int_{-1}^{1} \mathrm{d}s \; \frac{z(t,s)}{\sqrt{1-s^2}}, \tag{6.2}$$

which is the w-weighted center of mass of the curve.

We can express the $n = 1$ coefficients as,

$$\hat{z}_1(t) = \hat{x}_1(t) + 1i\hat{y}_1(t) = \frac{2}{\pi}\int_{-1}^{1} \mathrm{d}s \; \frac{sz(t,s)}{\sqrt{1-s^2}} = \frac{2}{\pi}\int_{-1}^{1} \mathrm{d}s \; \sqrt{1-s^2}\frac{\partial}{\partial s}z(t,s) \tag{6.3}$$

where we use the fact that $T_1(s) = s$, $\int \mathrm{d}s\; sw(s) = -(1-s^2)^{1/2} = -1/w(s)$ and integration by parts. The resulting expression is the $1/w$-weighted Chebyshev orientation of the curve. Higher-order coefficients will pick up increasingly large curvatures.

6.2.2 Imposing physical symmetries in planar curve dynamics

From the spectral representation of the curve we define the complex coefficient state vector $\boldsymbol{\Psi}_z = [\hat{z}_0,\, \hat{z}_1,\, \ldots,\, \hat{z}_n]$ from (6.1) and consider a general linear dynamics $\dot{\boldsymbol{\Psi}}_z = M\boldsymbol{\Psi}_z$ for some complex matrix M. We seek a dynamics that is invariant under rotation and translation. We also incorporate one additional symmetry typical for curve dynamics – the length of the curve is constant. Each of these symmetries will impose structure on the matrix M.

Rotational invariance

Rotation of the coordinate system corresponds to multiplying z by $e^{i\theta}$, $z' = e^{i\theta}z$, which implies that the coefficients transform as $\hat{z}'_k(t) = e^{i\theta}\hat{z}_k(t)$. Hence the complex coefficient state vector $\boldsymbol{\Psi}_z$ transforms as $\boldsymbol{\Psi}'_z = e^{i\theta}\boldsymbol{\Psi}_z$. Considering a general linear dynamics

$$\begin{aligned} \dot{\boldsymbol{\Psi}}'_z &= M\boldsymbol{\Psi}'_z \\ \implies e^{i\theta}\dot{\boldsymbol{\Psi}}_z &= Me^{i\theta}\boldsymbol{\Psi}_z \\ \implies \dot{\boldsymbol{\Psi}}_z &= M\boldsymbol{\Psi}_z \end{aligned}$$

we see that using this construction, the dynamics are automatically rotationally invariant.

Translational invariance

Invariance under translation requires that the dynamics are invariant under the shift $z' = z + c_z$. Under this shift the coefficients transform as $\boldsymbol{\Psi}'_z = \boldsymbol{\Psi}_z + c_z\boldsymbol{e}_1$ and the transformed dynamics

$$\begin{aligned} \dot{\boldsymbol{\Psi}}'_z &= M\left(\boldsymbol{\Psi}_z + c_z\boldsymbol{e}_1\right) \\ \implies \dot{\boldsymbol{\Psi}}_z &= M\boldsymbol{\Psi}_z + c_zM\boldsymbol{e}_1 \end{aligned}$$

show that we need $M\boldsymbol{e}_1 = \boldsymbol{0}$. The first column of M is zero decoupling the $\hat{z}_0$ dynamics from the higher mode dynamics of $\hat{\boldsymbol{\psi}}_z = [\hat{z}_1,\, \hat{z}_2,\, \ldots,\, \hat{z}_n]$,

$$\dot{\hat{z}}_0 = \boldsymbol{m}^\dagger\hat{\boldsymbol{\psi}}_z \tag{6.6a}$$

$$\dot{\hat{\boldsymbol{\psi}}}_z = \hat{M}\hat{\boldsymbol{\psi}}_z. \tag{6.6b}$$

6.2.3 An approximate length constraint on planar curve dynamics

The length of the curve in terms of the real fields $x(s,t)$ and $y(s,t)$ is,

$$\ell(t) = \int_{-1}^{1} \mathrm{d}s \sqrt{\left(\frac{\partial}{\partial s}x(t,s)\right)^2 + \left(\frac{\partial}{\partial s}y(t,s)\right)^2} = \int_{-1}^{1} \mathrm{d}s \left|\frac{\partial}{\partial s}z(t,s)\right|. \tag{6.7}$$

To allow for a convenient representation of an approximate length constraint in mode space, we consider ℓ^2. Using the Cauchy-Schwarz inequality, $\langle f,g\rangle^2 \leq \langle f,f\rangle\langle g,g\rangle$, with metric $\langle f,g\rangle = \int_{-1}^{1} \mathrm{d}s\, f\cdot g$, we can derive a convex upper-bound on the square length using $f=1$ and $g = \sqrt{x_s(s,t)^2 + y_s(s,t)^2}$,

$$\begin{aligned}
\ell^2 &= \left(\int_{-1}^{1} \mathrm{d}s\, 1 \cdot \sqrt{\left(\frac{\partial}{\partial s}x(t,s)\right)^2 + \left(\frac{\partial}{\partial s}y(t,s)\right)^2}\right)^2 \\
&\leq \left(\int_{-1}^{1} \mathrm{d}s\, 1^2\right)\left(\int_{-1}^{1} \mathrm{d}s \left[\left(\frac{\partial}{\partial s}x(t,s)\right)^2 + \left(\frac{\partial}{\partial s}y(t,s)\right)^2\right]\right) \\
\ell^2 &\leq 2\tilde{\ell}^2
\end{aligned}$$

where we define a convex approximate square length $\tilde{\ell}^2$,

$$\tilde{\ell}^2 = \int_{-1}^{1} \mathrm{d}s \left[\left(\frac{\partial}{\partial s}x(t,s)\right)^2 + \left(\frac{\partial}{\partial s}y(t,s)\right)^2\right].$$

We can also bound ℓ^2 from below. Assuming that the curve is regular (has non-zero and bounded differential arc length),

$$0 < m \leq \left(\frac{\partial}{\partial s}x(t,s)\right)^2 + \left(\frac{\partial}{\partial s}y(t,s)\right)^2 \leq M = m + \Delta m, \tag{6.8}$$

then the following lower bound is true by Cassel's inequality [205],

$$\frac{2(m^2 + m\Delta m)}{(m^2 + m\Delta m) + (\Delta m)^2}\tilde{\ell}^2 = \frac{8mM}{(m+M)^2}\tilde{\ell}^2 \leq \ell^2. \tag{6.9}$$

Enforcing that $\tilde{\ell}$ is constant, bounds ℓ giving an effective length constraint. The tightness of the bound is controlled by how much the differential arc length varies along the curve. When $\Delta m = 0$ then we get equality $\sqrt{2}\tilde{\ell} = \ell$. To illustrate this, we consider a couple of common curves,

1. A straight curve translating in space

$$(x(t,s), y(t,s)) = (x_0(t) + s, y_0(t) + ms) \implies \frac{\partial}{\partial s}(x,y) = (1,m).$$

The differential arc length is constant $\mathrm{d}\ell = \sqrt{1+m^2}$. The length of the curve is, therefore, $\ell = 2\sqrt{1+m^2}$ and the approximate length is $\tilde{\ell} = \sqrt{2}\sqrt{1+m^2}$.

2. A sinusoidal wave,

$$(x(t,s), y(t,s)) = (s, \sin(ks)) \implies \frac{\partial}{\partial s}(x,y) = (1, k\cos(ks - \omega t)).$$

The curve speed is $x_s^2 + y_s^2 = 1 + k^2\cos^2(ks)$. The bound is therefore given by,

$$\sqrt{2}\sqrt{\frac{4(1+k^2)}{(2+k^2)^2}}\tilde{\ell} \leq \ell \leq \sqrt{2}\tilde{\ell}.$$

The bound is exact, $\ell = \sqrt{2}\tilde{\ell}$, when $k = 0$, which corresponds to a straight line along the x-axis. The approximate length is $\tilde{\ell}^2 = 2 + k^2 + k\cos(k)\sin(k)$. When $k = \pi$ the bound is given by $0.81\tilde{\ell} \leq \ell \leq 1.42\tilde{\ell}$. Numerically, we find in this case $\ell = 4.61$ and $\tilde{\ell} = 3.44$, which satisfy the bounds. $\tilde{\ell}$, however, is not a good approximation of ℓ in this case.

Error in approximate length constraint

We can find a better approximation for ℓ in terms of $\tilde{\ell}$ by considering the Taylor expansion of $f(a,b) = \sqrt{a+b}$ around a_0 and b_0. Any $k = n + m$th order derivative of $f(a,b)$, is given by, using $(2n-1)!! = (2n-1)(2n-3)\ldots 1$ as the double factorial,

$$\frac{\partial^n}{\partial a^n}\frac{\partial^m}{\partial b^m} f(a,b) = -\frac{(-1)^{n+m}}{2^{n+m}}(2(n+m)-3)!!\frac{1}{(a+b)^{n+m-1/2}}$$

which only depends on $n + m$. The Taylor series then becomes, summing over all terms of the same polynomial degree,

$$\begin{aligned}\sqrt{a+b} &= \sqrt{a_0+b_0} + \frac{(a+b)-(a_0+b_0)}{2\sqrt{a_0+b_0}} - \sqrt{a_0+b_0}\sum_{k=2}^{\infty}\frac{(-1)^k(2k-3)!!}{2^k k!}\left(\frac{a+b}{a_0+b_0}-1\right)^k \\ &= \frac{1}{2}\sqrt{a_0+b_0} + \frac{a+b}{2\sqrt{a_0+b_0}} - \sqrt{a_0+b_0}\sum_{k=2}^{\infty}\frac{(-1)^k(2k-3)!!}{2^k k!}\left(\frac{a+b}{a_0+b_0}-1\right)^k. \quad (6.10)\end{aligned}$$

Setting $a = x_s^2$ and $b = y_s^2$ we can expand around $a_0 = \langle x_s^2 \rangle$ and $b_0 = \langle y_s^2 \rangle$ in (6.10), where we use $\langle \cdot \rangle$ to represent the average value over s and t

$$\langle f \rangle = \frac{1}{2T} \int_0^T \mathrm{d}t \int_{-1}^1 \mathrm{d}s \, f(s,t),$$

to get an expansion of the square-root term in (6.7),

$$\sqrt{x_s^2 + y_s^2} = \frac{1}{2}\sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle} + \frac{x_s^2 + y_s^2}{2\sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle}} - \sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle} \sum_{k=2}^{\infty} \frac{(-1)^k (2k-3)!!}{2^k k!} \left(\frac{x_s^2 + y_s^2}{\langle x_s^2 \rangle + \langle y_s^2 \rangle} - 1 \right)^k. \quad (6.11)$$

Integrating (6.11) over s, we get the following relationship between ℓ and $\tilde{\ell}$,

$$\ell = \int_{-1}^1 \mathrm{d}s \, \sqrt{x_s^2 + y_s^2} = \sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle} + \frac{1}{2\sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle}} \int_{-1}^1 \mathrm{d}s \, \left(x_s^2 + y_s^2 \right) + R(\Delta)$$
$$\ell \approx \ell_a = \sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle} + \frac{1}{2\sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle}} \tilde{\ell}^2 \quad (6.12)$$

where we define

$$\Delta = \frac{x_s^2 + y_s^2}{\langle x_s^2 \rangle + \langle y_s^2 \rangle} - 1$$

a measure of how much the deviations vary from their average. We can get a bound for the magnitude of $R(\Delta)$ by evaluating the remaining summation and utilizing the triangle inequality,

$$\begin{aligned}
|R(\Delta)| &= \sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle} \left| \int_{-1}^1 \mathrm{d}s \sum_{k=2}^{\infty} \frac{(-1)^k (2k-3)!!}{2^k k!} \Delta^k \right| \\
&\leq \sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle} \int_{-1}^1 \mathrm{d}s \sum_{k=2}^{\infty} \frac{(2k-3)!!}{2^k k!} \Delta_M^k \\
&\leq 2\sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle} \sum_{k=2}^{\infty} \frac{(2k-3)!!}{2^k k!} \Delta_M^k \\
&= \sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle} \left(2 - \Delta_M - 2\sqrt{1 - \Delta_M} \right) \quad (6.13a) \\
&\approx \sqrt{\langle x_s^2 \rangle + \langle y_s^2 \rangle} \frac{\Delta_M^2}{4} \quad (6.13b)
\end{aligned}$$

provided that $\Delta_M \leq 1$, where $\Delta_M = \max_{s,t} |\Delta|$. The error bound given in (6.13) allows us to say when the approximate length constraint will be suitable. The error will be small for

curves that do not have rapid or large oscillations in their shape. The approximate length constraint should work well for relatively smooth curves, as highlighted by our two examples,

1. For a translating straight line $\Delta = 0$ and $\langle x_s^2 + y_s^2 \rangle = 1 + m^2$. The approximate length is
$$\ell_a = \sqrt{1+m^2} + \frac{2(1+m^2)}{2\sqrt{1+m^2}} = 2\sqrt{1+m^2} = \sqrt{2}\tilde{\ell} = \ell.$$
The approximate length becomes exact for a straight line and agrees with the above-mentioned bounds.

2. For a sinusoidal wave the average gives $\langle x_s^2 + y_s^2 \rangle = 1 + k^2/2$. The approximate length is,
$$\ell_a = 2\sqrt{1+k^2/2} + \frac{k\cos(k)\sin(k)}{2\sqrt{1+k^2/2}}.$$
The maximum fluctuation is $\Delta_M = k^2/(2+k^2)$ and the error bound is given by,
$$|R| \leq \frac{4+k^2}{\sqrt{4+2k^2}} - 2.$$
We see that the error is monotonically increasing in $k > 0$ consistent with the expectation that the approximate length will get worse for more oscillatory curves.
When $k = \pi$ the approximate length is $\ell_a = 2\sqrt{1+\pi^2/2} \approx 4.87$ which is within 10% of the true value 4.61 and the error bound is given by $|R| \leq 0.85$. Note that in practice, the error is much less than the bound.

Since ℓ_a is a function solely of $\tilde{\ell}^2$, keeping ℓ_a constant is the same as keeping $\tilde{\ell}^2$ constant. We, therefore, continue working under the assumption that $\tilde{\ell}^2$ is a constant of the motion.

6.2.4 Imposing an approximate length constraint in mode-space

One of the benefits of working with $\tilde{\ell}^2$ rather than ℓ is that x and y appear quadratically, making it convenient to represent $\tilde{\ell}^2$ in mode space. We can express, $z_s(t,s)$ in mode space by differentiating (2.12),
$$\frac{\partial}{\partial s} z(t,s) = \sum_{k=0}^{n} \frac{\mathrm{d}}{\mathrm{d}s} T_k(s)\hat{z}_k(t). \tag{6.14}$$
Substituting (6.14) into the expression for $\tilde{\ell}^2$ (6.2.3) gives
$$\tilde{\ell}^2 = \int_{-1}^{1} \mathrm{d}s \sum_{k,m=1}^{n} \hat{z}_k^\dagger(t)\hat{z}_m(t) \frac{\mathrm{d}T_k}{\mathrm{d}s}(s) \frac{\mathrm{d}T_m}{\mathrm{d}s}(s)$$

$$
\begin{aligned}
&= \sum_{k,m=1}^{n} \hat{z}_k^\dagger(t)\hat{z}_m(t) \int_{-1}^{1} \mathrm{d}s \, \frac{\mathrm{d}T_k}{\mathrm{d}s}(s)\frac{\mathrm{d}T_m}{\mathrm{d}s}(s) \\
&= \sum_{k,m=1}^{n} \hat{z}_k^\dagger(t)\hat{z}_m(t) W_{k,m} \\
&= \hat{\boldsymbol{\psi}}^\dagger W \hat{\boldsymbol{\psi}}
\end{aligned} \tag{6.15a}
$$

where we define the symmetric matrix W with elements given by,

$$
W_{k,m} = \int_{-1}^{1} \mathrm{d}s \, \frac{\mathrm{d}T_k}{\mathrm{d}s}(s)\frac{\mathrm{d}T_m}{\mathrm{d}s}(s). \tag{6.15b}
$$

Note that the choice of basis fixes the value of W. Therefore, W is a basis-dependent constant. The matrix W is symmetric positive definite. From the definition of $\tilde{\ell}^2$ we have that $\mathbf{x}^\dagger W \mathbf{x} \geq 0$ for all $\mathbf{x}$. For $\tilde{\ell} = 0$ we need $|z_s| = 0$ which immediately implies that $\mathbf{x} =$. We can derive the values of $W_{m,n}$ (6.15b) explicitly for the Chebyshev basis,

$$
W_{m,n} = \begin{cases} 0 & \text{if } m-n \text{ odd} \\ 2nm \sum_{k=|n-m|+1}^{n+m-1} \frac{1}{k} & \text{if } m-n \text{ even} \end{cases}. \tag{6.16}
$$

Details of the calculation are provided in Appendix A.

The approximate length constraint requires that $\hat{\boldsymbol{\psi}}$ lives on a hyperellipsoid defined by the matrix W. We can use the Cholesky decomposition of $W = LL^\dagger$ to define the rescaled coefficients $\boldsymbol{\psi}_z = L^\dagger \hat{\boldsymbol{\psi}}_z / \ell$ under which the relaxed length constraint becomes

$$
1 = \boldsymbol{\psi}_z^\dagger \boldsymbol{\psi}_z.
$$

Under the rescaling the dynamics (6.6) becomes

$$
\begin{aligned}
\dot{\boldsymbol{\psi}}_z = L^\dagger \dot{\hat{\boldsymbol{\psi}}}_z / \ell &= L^\dagger \hat{M} \hat{\boldsymbol{\psi}}_z / \ell \\
&= L^\dagger \hat{M} (L^\dagger)^{-1} \boldsymbol{\psi}_z \\
&= M \boldsymbol{\psi}_z.
\end{aligned}
$$

In order to satisfy the unit norm constraint, we require that,

$$
\dot{1} = 0 = \dot{\boldsymbol{\psi}}_z^\dagger \boldsymbol{\psi}_z + \boldsymbol{\psi}_z^\dagger \dot{\boldsymbol{\psi}}_z = \boldsymbol{\psi}_z^\dagger (M^\dagger + M) \boldsymbol{\psi}_z \tag{6.17}
$$

giving the condition that $M = -M^\dagger$ is skew-Hermitian. Since M is skew-Hermitian, we

can write it as iH where H is Hermitian, yielding the following set of equations for the constrained complex dynamics of the center line,

$$\boldsymbol{\psi}_z^\dagger \boldsymbol{\psi}_z = 1 \tag{6.18a}$$

$$\boldsymbol{h}^\dagger \boldsymbol{\psi}_z = \dot{\hat{z}}_0 \tag{6.18b}$$

$$iH\boldsymbol{\psi}_z = \dot{\boldsymbol{\psi}}_z \tag{6.18c}$$

We can also write an equivalent real dynamics for the center line motion. Writing $M = A - iS$ where A and S are real skew-symmetric and symmetric matrices respectively, the dynamics become $M\boldsymbol{\psi}_z = (A - iS)(\boldsymbol{\psi}_x + i\boldsymbol{\psi}_y) = A\boldsymbol{\psi}_x + S\boldsymbol{\psi}_y + i(A\boldsymbol{\psi}_y - S\boldsymbol{\psi}_x)$. Taking real and imaginary components we get, $\dot{\boldsymbol{\psi}}_x = A\boldsymbol{\psi}_x + S\boldsymbol{\psi}_y$ and $\dot{\boldsymbol{\psi}}_y = -S\boldsymbol{\psi}_x + A\boldsymbol{\psi}_x$. These two equations can be summarized in a matrix equation,

$$\begin{bmatrix} \dot{\boldsymbol{\psi}}_x \\ \dot{\boldsymbol{\psi}}_y \end{bmatrix} = \begin{bmatrix} A & S \\ -S & A \end{bmatrix} \begin{bmatrix} \boldsymbol{\psi}_x \\ \boldsymbol{\psi}_y \end{bmatrix}.$$

Sometimes, it may be easier to work numerically with the real formulation.

6.2.5 Projection of mode-space model into real-space

Projecting the dynamics back into real space reveals that the linear mode space model gives an integro-differential equation for the curve dynamics. We define the Chebyshev vector functions $\mathbf{T}(s) = [T_1(s), T_2(s), \ldots, T_N(s)]$. The shape dynamics can be rewritten as,

$$\begin{aligned}
\dot{z}(t,s) - \dot{\hat{z}}_0 &= \mathbf{T}^\dagger(s)\dot{\hat{\boldsymbol{\psi}}}_z = i\mathbf{T}^\dagger(s)(L^\dagger)^{-1}HL^\dagger\hat{\boldsymbol{\psi}}_z \\
&= i\mathbf{T}^\dagger(s)(L^\dagger)^{-1}HL^\dagger \int_{-1}^{1} \mathrm{d}s'\, w(s')\mathbf{T}(s')z(t,s') \\
&= \int_{-1}^{1} \mathrm{d}s' \left[iw(s')\mathbf{T}^\dagger(s)(L^\dagger)^{-1}HL^\dagger\mathbf{T}(s')\right] z(t,s') \\
&= \int_{-1}^{1} \mathrm{d}s'\, G(s,s')z(t,s')
\end{aligned}$$

where we define the shape integral kernel,

$$G(s,s') = iw(s')\mathbf{T}^\dagger(s)(L^\dagger)^{-1}HL^\dagger\mathbf{T}(s'). \tag{6.19a}$$

A similar calculation yields,

$$\begin{aligned}\dot{\hat{z}}_0 = \frac{1}{\ell}\mathbf{h}^\dagger L^\dagger \hat{\boldsymbol{\psi}}_z &= \int_{-1}^{1} \mathrm{d}s' \, \left[w(s')\mathbf{h}^\dagger L^\dagger \mathbf{T}(s')\right] z(t,s') \\ &= \int_{-1}^{1} \mathrm{d}s' \, g(s') z(t,s')\end{aligned}$$

where we define the center of mass integral kernel,

$$g(s') = w(s')\mathbf{h}^\dagger L^\dagger \mathbf{T}(s'). \tag{6.19b}$$

The full real space dynamics is, therefore, given by

$$\dot{z}(t,s) = \int_{-1}^{1} \mathrm{d}s' \, \left[g(s') + G(s,s')\right] z(t,s'). \tag{6.19c}$$

The shape of the kernels allows us to infer properties of the motion, such as locality and heterogeneity along s.

6.2.6 Analytic solution of constrained linear model

Since the matrix H in (6.18) is Hermitian we know that it has real eigenvalues, λ_n and orthogonal eigenvectors $\mathbf{v}_n$, $H = V\Lambda V^\dagger$ with $V^\dagger V = \mathbb{I}$. The general solution of a linear ODE of the form, $\dot{\mathbf{x}} = M\mathbf{x}$, is $\mathbf{x}(t) = \exp(Mt)\mathbf{x}(t_0)$. From the orthogonality of the eigenvectors, the matrix exponential is given by $\exp(itH) = V\exp(it\Lambda)V^\dagger$. The solution is given by,

$$\boldsymbol{\psi}_z(t) = V\exp(it\Lambda)V^\dagger \boldsymbol{\psi}_z(t_0) \tag{6.20a}$$

$$= \sum_{n=1}^{N} (\mathbf{v}_n^\dagger \boldsymbol{\psi}_z(t_0))\mathbf{v}_n e^{it\lambda_n}$$

$$= \sum_{n=1}^{N} c_n \mathbf{v}_n e^{it\lambda_n} \tag{6.20b}$$

We can project the solution back onto the hyper-ellipsoid,

$$\hat{\boldsymbol{\psi}}_z(t) = \ell \sum_{n=1}^{N} c_n (L^\dagger)^{-1}\mathbf{v}_n e^{it\lambda_n}$$

and then back into real space,

$$z(t,s) = \hat{z}_0(t) + \ell \sum_{n=1}^{N} c_n \mathbf{T}(s)^\dagger (L^\dagger)^{-1} \mathbf{v}_n e^{it\lambda_n} = \hat{z}_0(t) + \ell \sum_{n=1}^{N} c_n v_n(s) e^{it\lambda_n} \tag{6.21}$$

with $v_n(s) = \mathbf{T}(s)^\dagger (L^\dagger)^{-1} \mathbf{v}_n$ being the real space eigenfunctions of the motion.

6.2.7 Inference of constrained linear models

The problem of learning an equation of the form (6.18) from data for $\boldsymbol{\psi}_z$ and $\boldsymbol{\psi}_0$ at discrete time points $\{t_n\}_{n=0}^M$ can be formulated as a physics-informed dynamic mode decomposition (PI-DMD) optimization problem. In continuous time, the inference problem for the shape dynamics becomes,

$$\min_H \sum_{n=0}^{T} \|\dot{\boldsymbol{\psi}}_z(t_n) - iH\boldsymbol{\psi}_z(t_n)\|_2^2 = \min_H \|\dot{P} - AP\|_F^2 \tag{6.22}$$

where $P = [\boldsymbol{\psi}_z(t_0)\, \boldsymbol{\psi}_z(t_1)\, \boldsymbol{\psi}_z(t_2) \cdots \boldsymbol{\psi}_z(t_T)]$ is the matrix whose columns consists of the discrete-time samples of $\boldsymbol{\psi}$. The minimization problem in (6.22) has an analytical minimum in terms of the singular value decomposition (SVD) of P but requires numerically differentiating noisy data to calculate $\dot{P}$ an ill-posed and challenging problem. We, therefore, formulate the problem in discrete time. If the data are separated by a constant time step Δt we can reformulate (6.22) in the form,

$$\min_A \sum_{n=0}^{T-1} \left\|\boldsymbol{\psi}_z(t_{n+1}) - e^{iH\Delta t}\boldsymbol{\psi}_z(t_n)\right\|_2^2 = \min_A \left\|P_{2:T} - e^{iH\Delta t} P_{1:T-1}\right\|_F^2 \tag{6.23}$$

where $P_{1:T-1}$ consists of the first $T-1$ columns of P and $P_{2:T}$ consists of the last $T-1$ columns. The skew-Hermitian structure of the continuous time problem does not transfer to the discrete-time problem; instead, $\exp(iH\Delta t)$ is a unitary matrix with a fixed form of its eigenvalues. Additionally, there is no guarantee that a matrix iH found through the single time-step minimization problem in (6.23) will produce reintegrated trajectories close to the original input data when integrated over a long time. Using the matrix exponential solution of linear ODEs, we can modify (6.23)

$$\min_A \sum_{m=1}^{T} \sum_{n=0}^{N} w_n \left|\boldsymbol{\psi}_z(t_m) - e^{iHt_m}\boldsymbol{\psi}_z(t_0)\right|_n^2 \tag{6.24}$$

where we introduce the possibility of a weighting function w_n on the nth mode to account for the magnitude variations across the modes. Using the solution of the dynamical system, the optimization loss function (6.24) can then be written as,

$$\min_{Q,\{\lambda_i\}_{i=1}^r} \sum_{m=1}^{T} \sum_{n=0}^{N} w_n \left[\boldsymbol{\psi}_z(t_m) - V \exp(i\Lambda t_m) V^\dagger \boldsymbol{\psi}(t_0) \right]_n^2 . \tag{6.25}$$

Writing the formula in this way enables us to optimize or constrain the eigenvectors and eigenvalues of H separately. To optimize V, we follow the procedure in [203], parameterizing the orthogonal matrix as the product of Householder matrices, $V = H_1 H_2 \cdots H_N$, where each Householder matrix has the form

$$H_n = \begin{bmatrix} \mathbb{I}_{n-1} & 0 \\ 0 & \mathbb{I}_{N-n+1} - \frac{\mathbf{p}_n \mathbf{p}_n^\dagger}{|\mathbf{p}_n|^2} \end{bmatrix} \tag{6.26}$$

and can be further parameterized by a vector $\mathbf{p}_n$ of length $N - n + 1$. Fast in-place matrix-vector multiplication algorithms exist for both Householder matrices and the diagonal $\exp(i\Lambda t)$, enabling us to efficiently compute the loss function from the combined parameter vector $\mathbf{p} = [\mathbf{p}_1, \cdots \mathbf{p}_N, \lambda_1, \cdots, \lambda_r]$. To minimize the loss function, we calculate gradients using automatic differentiation of the real and complex components of the vector $\mathbf{p}$ and perform gradient descent using the AdaBelief algorithm followed by BFGS.

To learn the time-dependent dynamics, each entry of $\mathbf{p}_n(t)$ is parameterized by a basis function expansion in time. The expansion coefficients are then optimized to find time-varying models by differentiating through the numerical solution of the time-varying linear model. See [37, 135] for more details on the differentiation.

Chapter 7

Schrödinger dynamics and Berry phase of undulatory locomotion

The results in this chapter have previously been published in A. E. Cohen, A. D. Hastewell*, S. Pradhan, S. W. Flavell, and J. Dunkel, "Schrödinger Dynamics and Berry Phase of Undulatory Locomotion," Physical Review Letters, 130(25), 258402, 2023.*

In this chapter, in collaboration with A. E. Cohen, we apply the planar curve dynamics developed in the previous chapter to experimental data of undulatory locomotion provided by the labs of Steven Flavell and Daniel Goldman. We also summarize clustering analyses from the planar curve modeling and extensions to time-varying matrices developed with and primarily by A. E. Cohen.

7.1 Motivation

Undulatory propulsion is the natural locomotion strategy [206, 207] of many aquatic and terrestrial animals, from worms [208–212] and fish [213, 214] to lizards [215, 216] and snakes [199, 217]. The mechanical wave patterns that drive undulatory motion reflect an animal's behavioral state [218], providing a macroscopic physical readout of the underlying biochemical and neuronal excitations. Recent advances in automated live-imaging [219, 220] enable simultaneous observations of macroscopic locomotion dynamics and microscopic cellular activity [3, 221–225], producing rapidly growing multi-scale data sets [226] that have to be tracked [1, 227, 228] and translated into predictive and interpretable models. Despite recent major progress in the experimental characterization [3, 221–224] and biophysical description of specific organisms [139, 209, 215, 229–235], a quantitative model inference framework for comparing experimentally observed undulatory dynamics within and across species has yet to be developed. In addition to providing unifying biophysical insights spanning different an-

imal kingdoms, such a framework would also allow for a direct comparison of living systems with computational models [236, 237] and biomimetic robotic devices [238, 239].

Here, we use spectral mode representations to identify symmetry-constrained dynamical models that can capture and distinguish the undulatory locomotion of worms (*Caenorhabditis elegans*) [218], neuro-mechanical worm models [236], Mojave shovel-nosed snakes (*Chionactis occipitalis*) [240], mechanical snakes, and centipedes (*Lithobius forficatus*). Using the theoretical and practical advantages of models for planar curve dynamics formulated in mode space as outlined in the previous chapter, we model the undulatory dynamics using a Schrödinger equation [204] in mode space. Similar to the characterization of quantum systems in terms of their spectra and eigenstates [15], the eigenspaces of the effective Hamiltonians enable an efficient classification of the locomotion dynamics of worms, snakes, robots, and computational models. Furthermore, transitions between animal behavioral states are encoded in the time evolution of the Hamiltonian and thus can be detected using Berry phases [241]. While our discussion focuses on an important subclass of biophysical dynamics, the underlying approach generalizes to other physical or living systems that permit a mode representation while being subject to exact or approximate geometric constraints.

7.2 Mode space representation of undulatory locomotion

The planar undulatory locomotion of an elongated worm-like object can be described by its centerline position in the complex plane $z(s,t) = x(s,t) + iy(s,t)$, where $s \in [-1,1]$ is the arc length and t denotes time [Figure 7.1(a) and (b)]. While tens to hundreds of points are typically required for an accurate depiction of an organism's shape in position space [3], interpretable lower-dimensional representations can often be obtained by projecting on suitable polynomial, trigonometric, or other basis functions [14, 229]. Although system-dependent representations, such as PCA-based eigenworms [229, 233], yield near-optimal compression for a specific organism under fixed experimental conditions, system-independent orthogonal basis expansions enable direct comparisons across different systems and experimental conditions. Moreover, system-dependent representations are often non-differentiable, making physically constrained modeling analytically intractable. Here, following (6.1), we use Chebyshev polynomials [23] of the first kind, $T_k(s)$; in principle other basis functions could be chosen as well. The dynamics of the complex scalar field $z(s,t) = x(s,t) + iy(s,t)$ can then be represented in terms of its leading Chebyshev coefficients $\hat{z}_k(t) = \hat{x}_k(t) + i\hat{y}_k(t)$ up to degree n, calculated using (2.12). For the experimental imaging data analyzed below, $n+1 = 10$ modes suffice for achieving reconstruction errors less than 1% [Figure 7.1(c)]. We illustrate the physical meaning of the Chebyshev modes using recent tracking microscopy

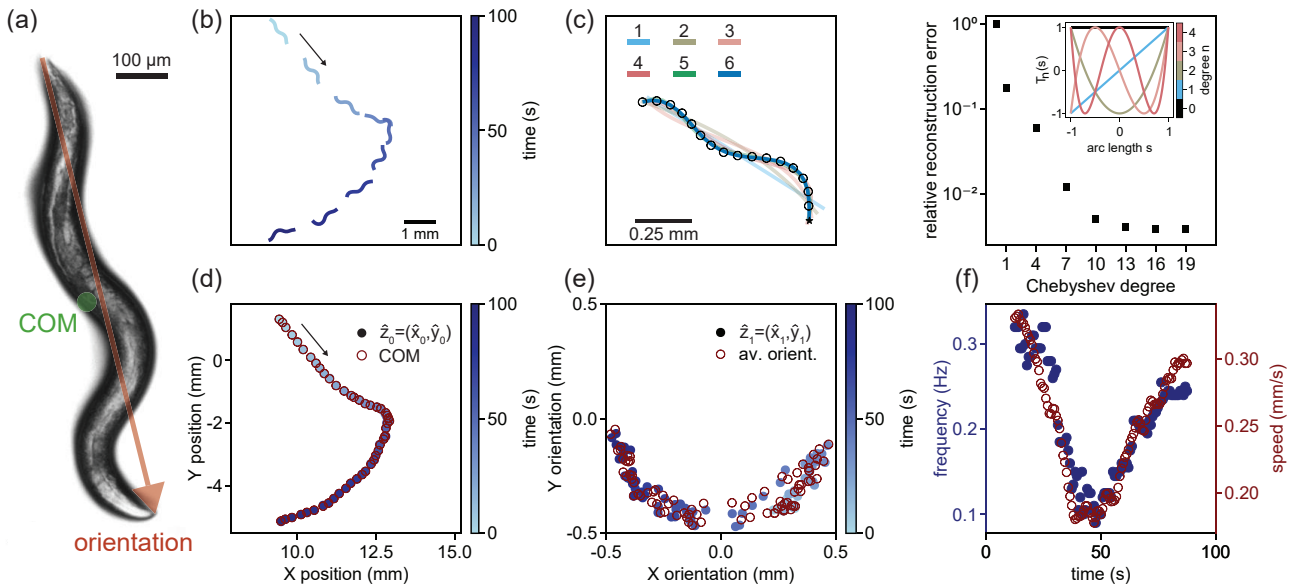

Figure 7.1: Chebyshev mode representation enables an efficient and interpretable low-dimensional description of undulatory locomotion across species and model systems. (a) Experimental image of *C. elegans* worm from the Flavell lab with center of mass (COM) and mean orientation overlayed. (b) Tracked centerline of worm over 100 seconds. The arrow indicates the direction of motion. (c) A small number of Chebyshev polynomials suffices to reconstruct the worm shape (left) accurately. Faint colored lines correspond to centerline reconstructions at different polynomial degrees. Reconstruction error (right) decays rapidly as the Chebyshev degree n increases. (d) The zeroth-order Chebyshev coefficients closely follow the worm's geometric COM, illustrating the physical interpretability of the Chebyshev mode representation. (e) Similarly, the first-order Chebyshev coefficients represent the tail-to-head worm orientation. (f) The mode-averaged dominant frequency of Chebyshev mode oscillations correlates closely with the locomotion speed of the worm.

video data, provided by the Flavell lab [3], for *C. elegans* [Figure 7.1(a) and (b)], a widely studied model organism with 95 body wall muscle cells, 302 neurons, and a rich set of behavioral states and corresponding locomotion patterns [218]. The real and imaginary parts of $\hat{z}_0(t) = \hat{x}_0(t) + i\hat{y}_0(t)$, (6.2), describe the w-weighted Chebyshev center of mass (CCOM) of the moving worm, which follows closely the geometric center of mass [Figure 7.1(d)]. The degree-1 coefficient $\hat{z}_1(t)$, (6.3), represents the mean orientation of the worm [Figure 7.1(e)]. Similarly, the Chebyshev coefficients $\hat{z}_k$ of degree $k \geq 2$ encode curvature and higher deformation modes [Figure 7.1(c), inset]. The average dominant frequency across the mode oscillations closely matches the speed of the worm in real space [Figure 7.1(f)].

Equation (7.1b) describes how the CCOM dynamics couples to the body oscillations through $\boldsymbol{h}_0$. Equation (7.1c), which governs the shape dynamics, is mathematically equivalent to a Schrödinger equation with Hamiltonian H [204].

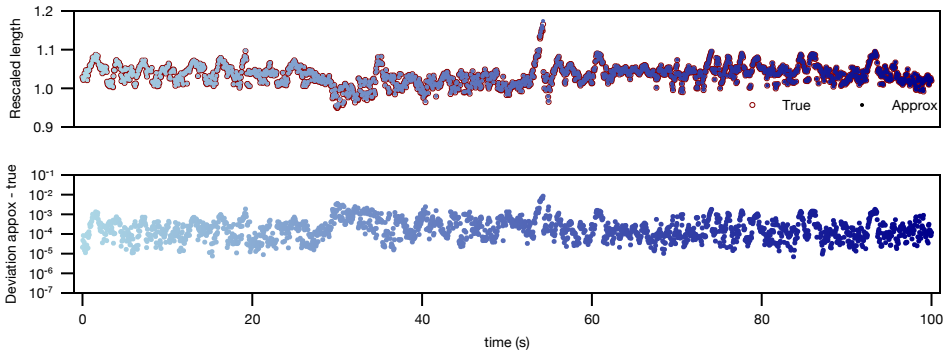

Figure 7.2: (top) True length ℓ and approximate length ℓ_a calculated using (6.12) show close agreement. (bottom) Deviation between $\ell_a - \ell$ shows that the true deviation is much lower than the bound 0.13. Also the deviation is always positive which means ℓ_a provides a close upper bound on ℓ.

7.3 Modeling undulatory locomotion in mode space

Equipped with this representation, we model the dynamics of the undulatory motion in mode space. The model should incorporate rotational and translational invariance. An additional biophysical constraint for undulatory motion is that the length of the centerline $\ell(t) = \int_{-1}^{1} \mathrm{d}s\, |\partial_s z|$, remains approximately constant. Following Chapter 6 we instead keep $\tilde{\ell}$ constant; for the worm dynamics considered here the approximation (6.12) agrees very well with the true value of ℓ (Figure 7.2). Demanding constant $\tilde{\ell}$ corresponds to an energetic penalty against contracting or lengthening and ensures ℓ remains approximately constant and bounded. The class of permissible linear models is, therefore, of the form (6.18), repeated below for clarity,

$$1 = \boldsymbol{\psi}^\dagger \boldsymbol{\psi} \tag{7.1a}$$

$$\dot{\boldsymbol{\psi}}_0 = \boldsymbol{h}_0^\dagger \boldsymbol{\psi} \tag{7.1b}$$

$$i\dot{\boldsymbol{\psi}} = H\boldsymbol{\psi}, \tag{7.1c}$$

where $\boldsymbol{h}_0$ is a complex vector and H is a complex Hermitian matrix with real eigenvalues.

7.3.1 Inferring models directly from data

To confirm that (7.1) can indeed describe and distinguish the undulatory dynamics of *C. elegans* worms [3] and other organisms and systems, we implemented an inference framework, described in Section 6.2.7, for estimating the propulsion vector $\boldsymbol{h}_0(t)$ and the shape Hamil-

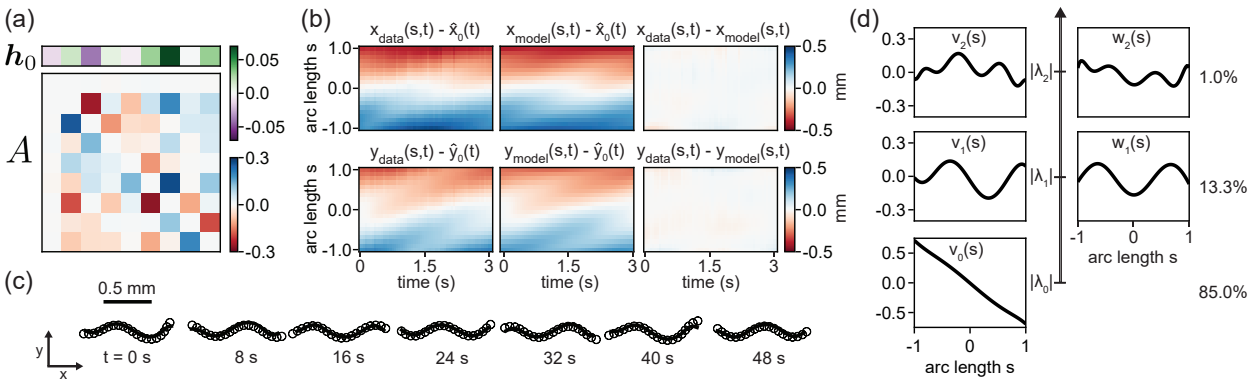

Figure 7.3: Inferred Schrödinger dynamics replicate stereotypical *C. elegans* locomotion (data provided by the Flavell lab). (a) Representative real propulsion vector $\boldsymbol{h}_0$ and Hamiltonian $H = S + iA$ for a minimal periodic straight-motion model (7.1), with $S = 0$ and equidistant spectrum of H, fitted to data from a single oscillation period ($\tau = 3.05$ s). (b) Kymographs of $x(s,t)$ and $y(s,t)$ coordinate fields for observed data (left) and model prediction (middle) show little deviation (right), confirming that (7.1) can accurately capture undulatory shape dynamics of *C. elegans*. (c) Real-space dynamics predicted by the Schrödinger model (line) is consistent with the observed worm dynamics (circles). Experimental data has been periodically extended for visualization to avoid overlapping body segments. (d) Real-space shape functions [Eq. (7.2)] corresponding to the three smallest magnitude eigenvalues, $\lambda_k^{\pm} = \pm k\lambda$ for $k = 0, 1, 2$, account for $> 98\%$ of the shape dynamics, enabling a generalizable low-rank description. More complex turning dynamics can be described using time-varying Hamiltonians with unconstrained spectra (Figure 7.5).

tonian $H(t)$ from experimental data for short straight-motion segments (Figures 7.3, 7.4) as well as longer trajectories that include turning events (Figure 7.5). Recall that any Hermitian matrix H can be decomposed into $H = S + iA$, where S is real symmetric and A is real skew-symmetric. In the present context, S encodes turning behavior whereas A governs straight locomotion: For straight motions, x- and y-modes do not couple significantly, so that $\boldsymbol{h}_0$ is real and $S \approx 0$ and, hence, $H \approx iA$ in this case [Figure 4(a)].

Generally, both $\boldsymbol{h}_0$ and H can be efficiently determined from tracked centerlines via a physics-informed dynamic mode decomposition [19, 155] that exploits matrix structure [203]. Since H is Hermitian, it permits the decomposition $H = U\Lambda U^{\dagger}$, where U is unitary and Λ is a real diagonal matrix. This structure leaves n^2 parameters in U and Λ plus $2n$ in $\boldsymbol{h}_0$ to be estimated from data. If available data is limited, the number of parameters can be reduced further by imposing additional constraints on the spectrum of H. To avoid numerical differentiation of noisy data, our inference scheme compares numerically integrated predictions from the model (7.1) directly to the experimental data. Our algorithm sequentially optimizes U, Λ, and $\boldsymbol{h}_0$ by minimizing deviations from both real space body shapes and mode space trajectories to balance shape matching with model generalizability and prevent overfitting.

Minimization is performed using gradient-based optimization [133, 242, 243] with forward mode automatic differentiation through the ODE solver [37, 181, 202]. This scheme makes it possible to infer the instantaneous shape Hamiltonians $H(t)$ and the propulsion vectors $\boldsymbol{h}_0(t)$ from just a single oscillation period for straight motions (Figures 7.3 and 7.4) as well as from longer curved trajectories (Figure 7.5). For *C. elegans* data, provided by the Flavell lab, (Figure 7.3) as well as for previously proposed neuro-mechanical worm models [236], *C. occipitalis* snake data, provided by the Goldnam lab [240], snake robots, and *L. forficatus* centipede data, provided by the Goldman lab, (Figure 7.4), the best-fit straight-motion models based on (7.1) with $H = iA$ accurately capture the undulatory dynamics.

7.4 Categorizing undulatory locomotion

Since the Hamiltonian H encodes the shape dynamics, we can use its eigenstates to compare and classify undulatory motion across species and systems [15]. Indeed, for straight motions, it suffices to study the eigenstates of A. Considering $n = 9$ as before, A has one zero eigenvalue $\lambda_0 = 0$ corresponding to the zero-mode eigenvector $\boldsymbol{\phi}_0$, and 4 distinct pairs of opposite sign eigenvalues $\lambda_{k\geq 1}^{\pm}$ with complex conjugate eigenvectors $\boldsymbol{\phi}_k^{\pm}$, where $\boldsymbol{\phi}_k^{+} = (\boldsymbol{\phi}_k^{-})^*$. We define two real orthogonal mode space vectors $\boldsymbol{v}_k = \mathrm{Re}(\boldsymbol{\phi}_k^{+})$ and $\boldsymbol{w}_k = \mathrm{Im}(\boldsymbol{\phi}_k^{+})$ that span the eigenspace of $\boldsymbol{\phi}_k^{\pm}$. The real space shape functions corresponding to the real mode space vectors are

$$v_k(s) = \ell(L^{-1}\mathbf{T}(s))^{\dagger}\boldsymbol{v}_k, \quad w_k(s) = \ell(L^{-1}\mathbf{T}(s))^{\dagger}\boldsymbol{w}_k, \tag{7.2}$$

where $\mathbf{T}(s) = [T_1(s), T_2(s), \ldots, T_n(s)]$ is a vector of Chebyshev functions. Time-varying linear combinations of $v_k(s)$ and $w_k(s)$ give the instantaneous centerline reconstruction. We find that the zero-function $v_0(s)$ is close to the best fit straight line through the motion, accounting for 85% of the time-averaged centerline reconstruction while most of the oscillations are accounted for by the first excited-states $v_1(s)$ and $w_1(s)$ corresponding to the smallest magnitude non-zero eigenvalues (13.3%). Since most ($> 98\%$) of the dynamics is captured by the zero-state and first excited states, one can, in fact, further reduce the complexity of the Schrödinger model by approximating A through its projection $\hat{A}$ on the eigenspaces corresponding to the first two distinct eigenvalues. This additional low-rank approximation also further reduces the risk of overfitting and hence improves model generalizability, similar to sparsity promotion in other dynamical inference methods [27].

The compact low-rank characterization of the undulatory shape dynamics makes it possible to compare the locomotion behaviors of *C. elegans*, previously proposed neuro-mechanical worm models [236], *C. occipitalis* snakes [240], robotic toy snakes, and centipedes, by mea-

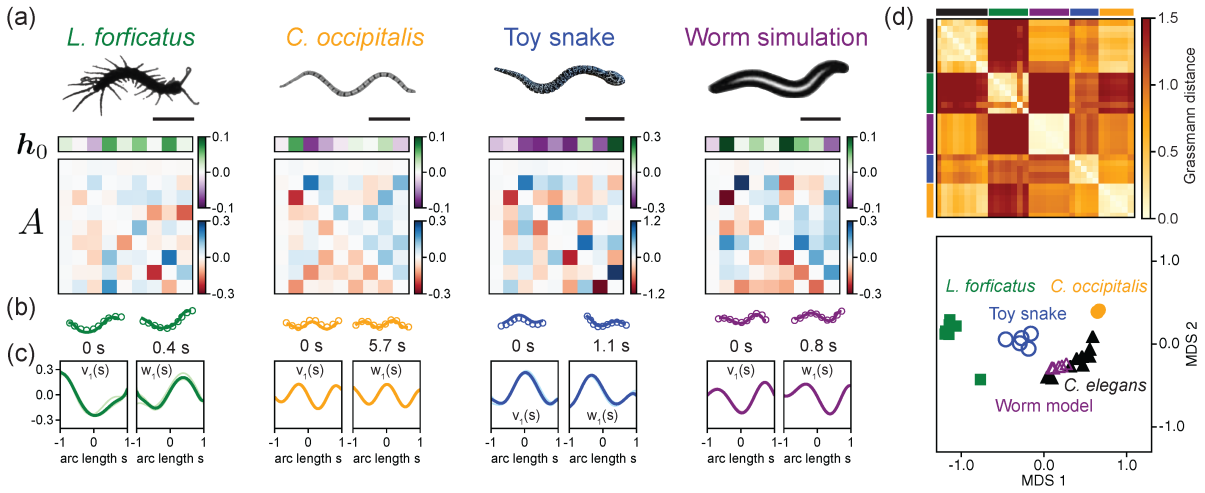

Figure 7.4: Mode-space Hamiltonians provide a compact dynamical description of undulatory motion across different species and model systems. (a) Living and nonliving systems [236, 240] analyzed here and representative straight-motion Hamiltonians $H = iA$ inferred from a single oscillation period. The eigenspaces of the Hamiltonians enable the comparison and classification of undulation dynamics in panel (d). Scale bars are 8 mm [centipede (data provided by the Goldman lab)], 10 cm [snake (data provided by the Goldman lab)], 10 cm (toy snake), and 0.25 mm (worm model). (b) Inferred Schrödinger model dynamics (line) provide an accurate description of the observed dynamics (circles). Models were fitted on a single period $\tau = 0.19$ s (centipede), 0.33 s (snake), 0.45 s (toy snake), 2.2 s (worm model). Experimental data has been periodically extended for visualization to avoid overlapping body segments. (c) The dominant shape eigenvectors $v_1(s)$ and $w_1(s)$ are consistent within each species and capture differences between species. (d) Pairwise Grassmann distances between subspaces spanned by first excited eigenstates of the Hamiltonians (top) and its 2D planar embedding (bottom, constructed by a multidimensional scaling) capture the similarities and differences between undulatory locomotion in organisms, model simulations, and robots. Each point corresponds to a different trajectory.

suring the Grassmann distance [244] between the dominant eigenspaces of $\hat{A}$. As most of the variation of the oscillatory dynamics is contained in the first excited-states $\boldsymbol{v}_1$ and $\boldsymbol{w}_1$, we determined the pairwise Grassmann distances between the eigenspaces spanned by $\boldsymbol{v}_1$ and $\boldsymbol{w}_1$ for the various systems. The Grassmann distance between two subspaces can be calculated by

$$d_G(A, B) = \sqrt{\sum_i \theta_i^2}, \tag{7.3}$$

where A and B are two matrices whose columns are an orthonormal basis of their respective subspaces and θ_i are the principal angles between A and B [244]. The principal angles can be calculated through an SVD, where the singular values of $A^\top B$ are $\sigma_i = \cos(\theta_i)$. Both the distance matrix and a corresponding 2D phase diagram constructed by multidimensional scaling reveal that the neuro-mechanical worm model [236] succeeds in reproducing key dynamical aspects of *C. elegans* locomotion, whereas the robotic toy snake used in our experiments is equally far from real snake or worm locomotion (Figure 7.4d).

7.5 Time varying linear models

Beyond inter-species comparisons, the above framework enables us to characterize behavioral transitions by borrowing concepts from quantum mechanics, such as Berry phases and adiabatic approximations [241]. To illustrate this, we focus on a longer *C. elegans* trajectory during which the worm performs a turn [Figure 7.5(a)] after briefly reversing its motion due to a change in neuro-mechanical activity [245]. By reconstructing the time-dependent Hamiltonian $H(t) = S(t) + iA(t)$ along the path, we observe a significant increase in $||S(t)||$ at the turn, whereas $A(t)$ remains approximately constant throughout. When the worm switches on S to facilitate a turn, the instantaneous eigenvectors of $H(t)$ change, signaled by a rapid change of the Berry phase [blue curve in Figure 7.5(b)]. Furthermore, while the locomotion dynamics before the turn is well described by an adiabatic approximation, this approximation becomes inaccurate during the turn [red curve in Figure 7.5(b)].

7.6 Conclusions

From a practical perspective, the above results show how symmetry-constrained mode representations can facilitate a low-dimensional description and efficient classification of biophysical dynamics. The underlying inference framework is directly applicable to diagnosing and quantifying the effects of genetic or chemical perturbations on animal locomotion within and across species. From a theoretical perspective, the fact that translational and rotational

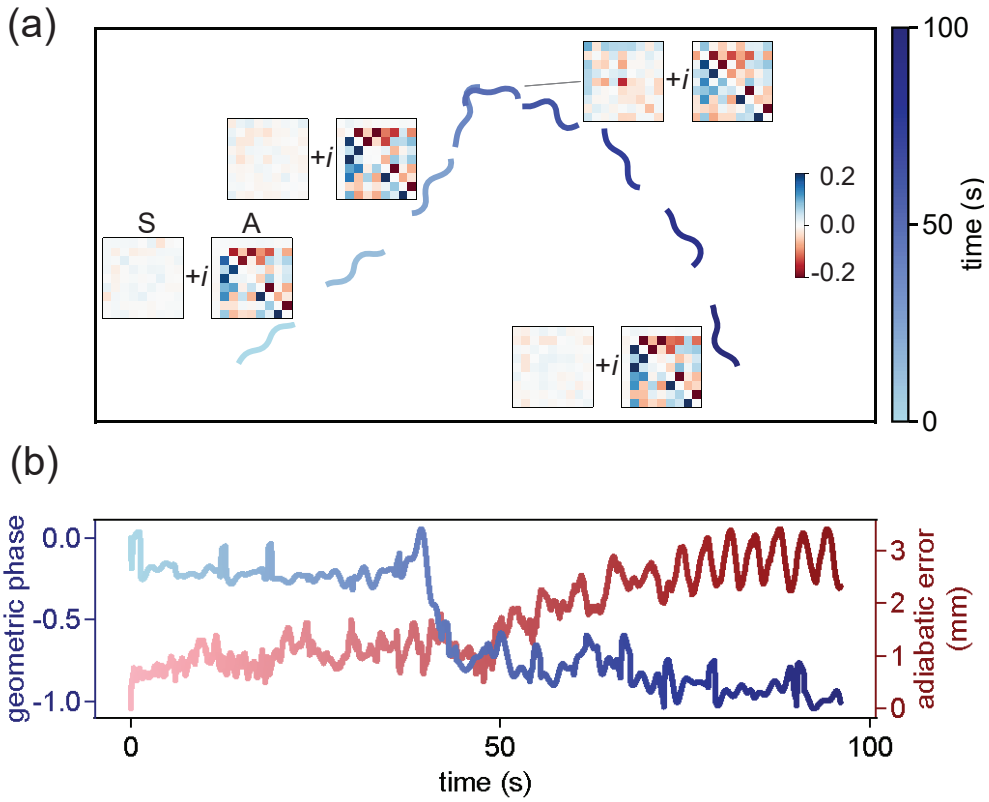

Figure 7.5: Breakdown of adiabaticity during reversal turning behavior of *C. elegans*. (a) The turning part $S(t)$ of the Hamiltonian $H(t) = S(t) + iA(t)$ becomes switched on at the turn. (b) The turn is signaled by a sudden change in the geometric Berry phase (blue) of the dominant eigenvector, and the RMS reconstruction error of the adiabatic approximation increases noticeably after the turn.

invariance combined with a quadratic integral constraint generically lead to a Schrödinger equation [204] in mode space, promises advances in the quantitative understanding of biological systems, as the comprehensive toolbox of quantum physics [246, 247] now becomes available to characterize and predict behavioral dynamics.

Chapter 8

Discovering dynamics and parameters of nonlinear oscillatory and chaotic systems from partial observations

This chapter and the accompanying appendices are adapted from the preprint G. Stepaniants, A. D. Hastewell*, D. J. Skinner, J. F. Totz, and J. Dunkel, "Discovering dynamics and parameters of nonlinear oscillatory and chaotic systems from partial observations," arXiv preprint arXiv:2304.04818, 2023.*

This chapter demonstrates extensions of the model inference optimization framework we presented previously to incorporate nonlinear dynamics and account for unobserved variables. The modified optimization procedure is integrated with a model selection procedure that can account for additional hidden symmetries introduced by partial observations developed by G. Stepaniants in collaboration with A. D. Hastewell and D. J Skinner. We apply the resulting hidden dynamics inference framework to simulated data, previously published recordings of giant squid axon recordings from the SGAMP database, and chemical reaction experiments performed by J. F. Totz.

8.1 Motivation

Significant advances in live-imaging and fluorescence labeling techniques over the last decades have made it possible to record extensive time-series data of neuronal [13, 248] and other cellular activity [11, 249] at high temporal resolution. Nevertheless, notwithstanding such progress, for many complex biophysical and biochemical systems, direct measurements are limited to a single experimentally accessible observable [250] while essential components of

the underlying dynamical circuit stay hidden [251]. As shown in Chapter 4, nonlinear dynamics generally lead to nonlinear ODEs in mode space (4.4). These realities pose an additional challenge in identifying valid theoretical nonlinear models and estimating their parameters from an incomplete set of experimentally accessible time series, such as time-varying mode amplitudes. As a proof of concept, we consider nonlinear oscillations and chaos, which are ubiquitous in natural and artificial systems [152], from neurons [252, 253] and biochemical networks [254] to power grids [255, 256], lasers [257], and the Earth's climate [258]. Limited observability has led to the emergence of competing theoretical models for neuronal [259] and gene-regulatory networks [251], and identifying valid models and their parameters from incomplete data remains a central challenge. Here, we combine sensitivity methods [37] for differential equations with ranked choice voting [260, 261] to construct a hidden dynamics inference (HDI) framework that can discover predictive nonlinear dynamical models for both observable and latent variables from noise-corrupted incomplete data in oscillatory and chaotic systems.

Driven by the rapidly advancing data acquisition techniques, dynamical model inference is becoming increasingly more important [27, 112] in climate physics [262–264], fluid mechanics [265, 266] and biophysics [84, 181, 267]. Time-delay embeddings [268, 269], recurrent neural networks [270] and autoencoders [271] have successfully been used to estimate hidden dimensions and forecast complex dynamics [272]. However, such 'equation-free' approaches often cannot reveal coupling mechanisms and their dependencies on experimental conditions. Complementary equation-based approaches [27, 273] have shown promise in learning interpretable dynamical models from partially observed data using physics-informed neural networks [113, 115, 274–276], manifold methods [277], or data assimilation [114], enabling prediction of nonlinear and chaotic dynamics in mechanical, electrical, and hydrodynamic systems. Despite such substantial progress, however, applications to experimental data from nonlinear biophysical and biochemical systems still face many open problems, as existing methods require long time series recordings with low noise (for example, to construct time-delay embeddings or train neural networks) and do not ensure the stability of learned models, as seen in the previous linear model inference approaches.

The HDI framework introduced here overcomes these challenges by integrating the robustness of sensitivity methods [37] and ranked-choice model selection [260, 261] with traditional library-based learning methods [27, 147]. This enables us to learn physically interpretable models for partially hidden nonlinear systems from short, highly noisy data trajectories in a manner that ensures correct long-time dynamics. Since the hidden-variable dynamical equation discovered from partial observations may not be unique, we develop a systematic algebraic procedure for comparing learned models. After validating the HDI framework on

strongly noise-corrupted simulations of the FitzHugh-Nagumo oscillator, we apply our approach to experimental measurements of squid neuron spike trains and video observations of Belousov-Zhabotinsky chemical reactions, demonstrating how HDI can be used to measure model parameters as a function of external experimental conditions.

Our HDI methodology falls into the class of physically interpretable methods, and we learn ODE models as a sparse combination of polynomial or trigonometric basis functions. However, HDI uses a novel robust procedure for model selection by sampling the space of possible ODE models, forming a cluster of the best-fit models, and keeping those basis terms in our final model that have the least variation in their coefficients across all model fits. This allows HDI to robustly select a few important model terms from libraries containing on the order of 10-100 candidate terms, thus exploring a broad range of potential ODE models. HDI's core model fitting procedure uses gradient-based sensitivity methods that are robust to high levels of noise in the data, ensure the learned models accurately capture the experimental dynamics, and are stable over long-time integration.

8.2 Description of the HDI framework

A canonical example of a nonlinear oscillator is the FitzHugh-Nagumo (FHN) model [259]

$$\dot{v} = v - \frac{v^3}{3} - w + I, \quad \dot{w} = \tau\Big(v + a - bw\Big), \tag{8.1}$$

a simplified model of a firing neuron where the membrane voltage v rapidly increases before being diminished by the slow recovery variable w [278]. The rapid spiking and slow recovery arise from a separation in time scales $\tau \ll 1$ between variables. FHN has become a prototypical model of neuron spike trains, as it is stable and parsimonious, relying only on a small number of polynomial terms. The HDI framework aims to learn models of this type from limited noisy recordings of a single variable, for example, the v-coordinate of FHN [Fig. 8.1(a)]. This motivates us to define the following class of models, with observed variables $x_1, \ldots, x_m$ and hidden variables $h_{m+1}, \ldots, h_M$, given by

$$\dot{x}_k = \tau_k \sum_{|\boldsymbol{\alpha}| \le d_k} c^k_{\boldsymbol{\alpha}} x_1^{\alpha_1} \ldots h_M^{\alpha_M}, \quad 1 \le k \le m \tag{8.2a}$$

$$\dot{h}_k = \tau_k \sum_{|\boldsymbol{\alpha}| \le d_k} c^k_{\boldsymbol{\alpha}} x_1^{\alpha_1} \ldots h_M^{\alpha_M}, \quad m < k \le M \tag{8.2b}$$

which encompass a broad range of nonlinear oscillatory dynamics. Here, we only use polynomial terms on the right-hand side of the equation, although this can be extended to any other

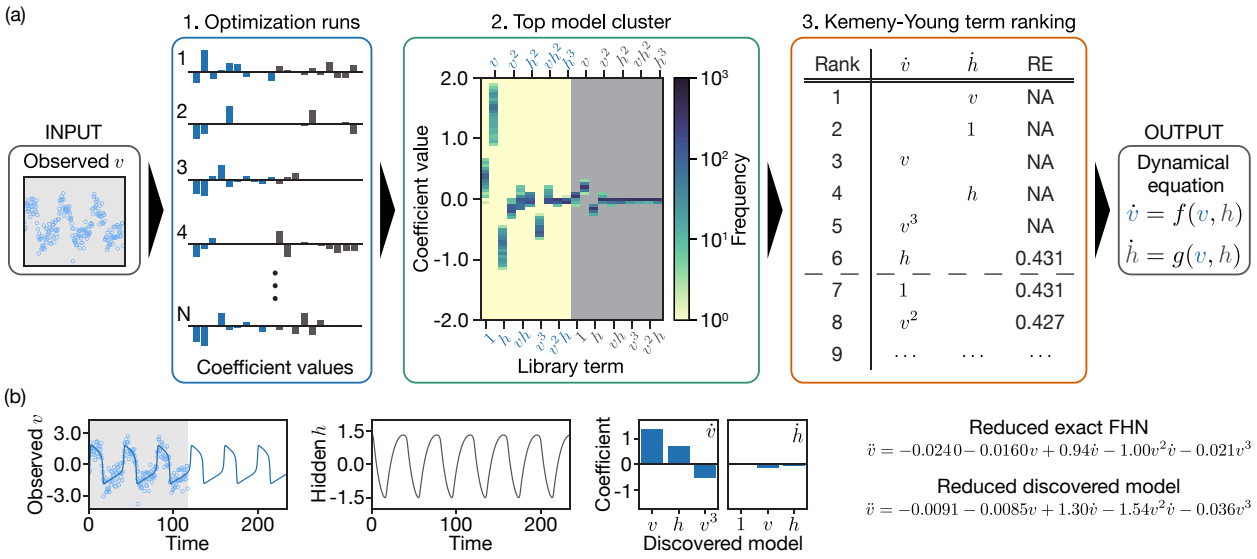

Figure 8.1: General HDI framework illustrated for strongly noise-corrupted FHN simulation data. (a) Algorithm flow chart: (1) ODE sensitivity optimization [37] yields $N \sim 20,000$ candidate models by tuning 20 parameters of dense two-field cubic observed (blue) and hidden (dark-gray) variable equations from random initializations. Models are filtered for stability and fit quality. (2) The remaining $\sim$ 4000 models are hierarchically clustered using the cosine similarity between their parameter vectors. Histograms of parameters in the largest cluster are used to rank the terms based on their coefficient of variation. (3) Kemeny-Young ranking produces a list of candidate models of decreasing sparsity. Models are refit at each sparsity level, and the user can select the model that best balances sparsity and relative error (RE). (b) Using data from only the v time series corrupted by 50% noise, HDI correctly discovers a sparse first-order system that reduces to the same second-order form as the FHN model.

nonlinearities, such as trigonometric functions. To avoid scaling ambiguities between τ_k and $\mathbf{c}^k = \{c^k_{\boldsymbol{\alpha}}\}$ in (8.2) we enforce that each $\mathbf{c}^k$ has unit norm. HDI models are determined by a parameter vector $\mathbf{p}$ containing the initial conditions of the variables $\{x^0_k\}_{k=1}^m$, $\{h^0_k\}_{k=m+1}^M$, time scales $\{\tau_k\}_{k=1}^M$, and polynomial coefficients $\{\mathbf{c}^k\}_{k=1}^M$. While time-delay embeddings can be used to provide lower bound estimates on the number of hidden variables $M - m$, here we restrict to periodic models with $M = 2$ variables or chaotic models with $M = 3$ variables, which we find sufficient to explain the experimental data.

To demonstrate the HDI framework (Fig. 8.1), we consider noise-corrupted observations $y_{i1} = v(t_i) + \xi_{i1}$ of the v-coordinate of the FHN model (8.1) [Fig. 8.1(a, INPUT)]. HDI repeatedly fits hidden two-variable models (x_1, h_1) by minimizing the mean square error on the observed variables

$$\mathrm{MSE}(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}(x_k(t_i, \mathbf{p}) - y_{ik})^2 \tag{8.3}$$

plus a regularization term $\mathrm{Reg}(\mathbf{p})$ that enforces the unit norm constraint on $\mathbf{c}^k$ and promotes sparsity, favoring lower-order terms that often lead to more stable dynamics. The full objective function $L(\mathbf{p}) = \mathrm{MSE}(\mathbf{p}) + \mathrm{Reg}(\mathbf{p})$ is minimized using ODE sensitivities and gradient descent methods [243, 279] from random initializations [Fig. 8.1(a, 1)]. Noise robustness in our approach comes from using the full ODE solution $x_k(t_i, \mathbf{p})$ in the objective function, which avoids numerically differentiating noisy time series data, a typically ill-posed problem [117], and enforces that models learned are stable over the time-span of the training data. Multiple fits are required to sufficiently sample multiple local minima of the complex non-convex loss landscape $L(\mathbf{p})$: the model described by $\mathbf{p}$ is not unique since model symmetries – linear, $h \mapsto \alpha h + \beta$, and nonlinear transformations of the hidden variables – produce new models with identical x_k dynamics.

To select a single candidate model, a set of models is optimized from random initialization (see Appendix C for further details). Model quality is measured using the relative error $\mathrm{RE}(\mathbf{p}) = \sqrt{\mathrm{MSE}(\mathbf{p})/\mathrm{Var}(\{y_{i1}\}_{i=1}^n)}$, given here for a single ($m = 1$) observable, where Var is the uncorrected sample variance. Outlier models with incorrect dynamics (nonperiodic, nonchaotic, or divergent) or large RE values are removed in an automated manner, and the remaining models are hierarchically clustered using the cosine similarity between their parameter vectors, taking into account possible linear hidden variable transformations. Measurement noise and regularization break many of the symmetry ambiguities, resulting in a dominant largest cluster; on FHN data corrupted by 50% noise, we start by optimizing $20,000$ dense two-variable cubic models resulting in 4,006 filtered models, of which 427 models form the dominant cluster [Fig. 8.1(a, 2)]. Sparse models are identified from the dominant cluster by ranking each term by its coefficient of variation in the cluster, the interquartile range divided by the median. Rankings are aggregated over a range of clustering thresholds using the Kemeny-Young method to provide a robust ordering of terms. Based on this ordering, a list of candidate models of decreasing sparsity containing the top-ranked s terms can be refit [Fig. 8.1(a, 3)]. From this list, practitioners can determine a suitable model sparsity that balances the trade-off between a model's complexity and RE based on their scientific judgment [Fig. 8.1(a, OUTPUT)].

From just three noisy oscillations of the FHN v-coordinate, we learn a list of two-variable HDI models that at sparsities six and seven recover models which are equivalent to FHN [Fig. 8.1(a, 3)]. The seven-term model matches the sparsity pattern of FHN while the six-term model is equivalent under the shift $w \mapsto w - I$ in the FHN dynamics (8.1). Indeed by taking the true FHN model (8.1), solving for w in terms of $v, \dot{v}$ and substituting into the $\dot{w}$

equation, we obtain a second-order *reduced model* solely in v

$$\ddot{v} = -3cv^2\dot{v} + \left(1 - \frac{b}{\tau}\right)\dot{v} - \frac{bc}{\tau}v^3 + \frac{b-1}{\tau}v + \frac{bI-a}{\tau}. \tag{8.4}$$

Performing a similar reduction of our learned six-term model, we see that it has the same structure and coefficients as the true FHN model [Fig. 8.1(b)], confirming that HDI has recovered a two-variable model that is equivalent to ground-truth FHN. Full details of the optimization and model selection procedure can be found in [280].

At this point, one might hope to avoid using hidden variables and their associated ambiguities by learning the reduced higher-order equation in the observed variable directly [281, 282]. However, even simple multivariate systems can give rise to complex reduced higher-order equations that are often less sparse, implicit and contain fractional powers [283–291]; for example, $\dot{x} = xy^3$, $\dot{y} = x$ reduces to $x\ddot{x} = \dot{x}^2 + 3x^{7/3}\dot{x}^{2/3}$. Working with reduced models directly would require learning dense implicit ODEs with more candidate terms [142, 292], a challenging approach that can be ill-posed [293]. A general advantage of 'first-order' HDI is that it robustly learns multivariate explicit ODE models that are sparse and integrable, avoiding the above complications. Next, we apply HDI to identify quantitative models from experimental data for neuron activity and chemical reactions.

8.3 Application of HDI to experimental nonlinear oscillators

Figure 8.2(a) shows experimental measurements [294, 295] of the membrane potential v in the giant axon of the North Atlantic longfin inshore squid (*Loligo pealeii*) in response to noisy stimulus input currents. Following previous spike train model formulations [278, 297–299], we apply HDI to the time series data for v to learn a sparse two-variable model [Fig. 8.2(a,b)]. Consistent with prior descriptions of neuron dynamics [259], the phase portrait of the discovered seven-term model is governed by a homoclinic orbit [Fig. 8.2(c)]. Notably, the model generalizes to describe recordings from different squids, yielding consistent coefficients across all samples [Fig. 8.2(d,e)].

For a second, more challenging HDI application, we use Belousov-Zhabotinsky (BZ) reaction experiments performed by J. F. Totz [300]. Over the course of the reaction, a substrate species is slowly consumed, fueling the periodic rise and decay of intermediary reagents far from thermodynamic equilibrium. The basic reaction scheme [301] involves more than 20 chemical species and 40 reaction steps. Many different chemical models have been developed that capture the BZ reaction qualitatively [296, 302–305]. In our experiments, the

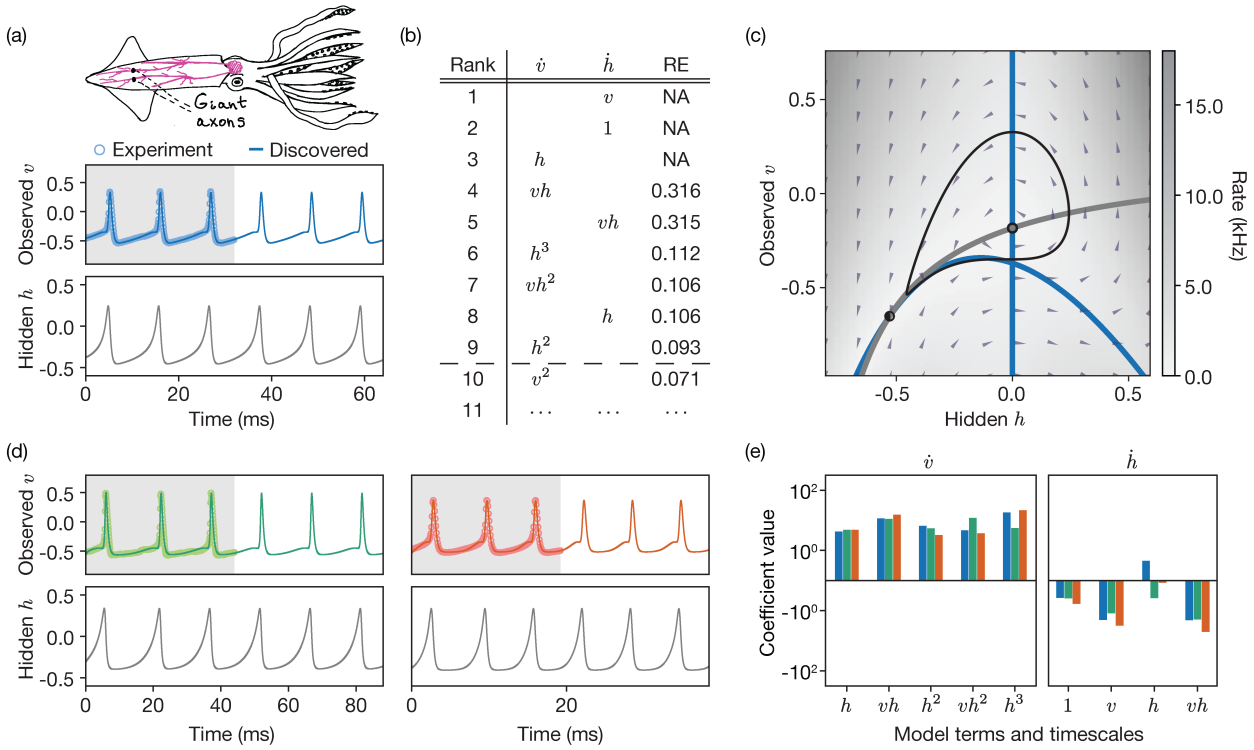

Figure 8.2: HDI framework learns a parsimonious two-variable model from an experimental recording of the membrane potential in a squid giant axon and reproduces the dynamics in additional squid giant axons from the SGAMP database [294, 295]. (a) North Atlantic longfin inshore squid (*Loligo pealeii*) with a sketch of the nervous system and position of giant axons (top). The learned two-variable HDI model with nine terms accurately fits the membrane potential v (center, line) of an experimental squid giant axon (open circles) in response to a noisy stimulus input current. The hidden variable h (bottom) acts as a slow recovery variable. (b) Polynomial model terms in $\dot{v}$ and $\dot{h}$ equations ranked from most to least important based on their coefficient of variation in the largest model cluster. Training data losses of sparse models containing only top s ranked terms are shown, and a model with sparsity nine is chosen. (c) Limit cycle and fixed points (black) of the learned model are consistent with prior models of regular spiking neurons [259] where the proximity of the saddle fixed point to the orbit likely arises from a homoclinic bifurcation. Nullclines of v, h are plotted in blue and gray, respectively. (d) The Selected nine-term model (line) generalizes to two additional squid axon recordings (open circles). (e) Coefficients of the nine-term model align across all three trains and test squid axon experiments.

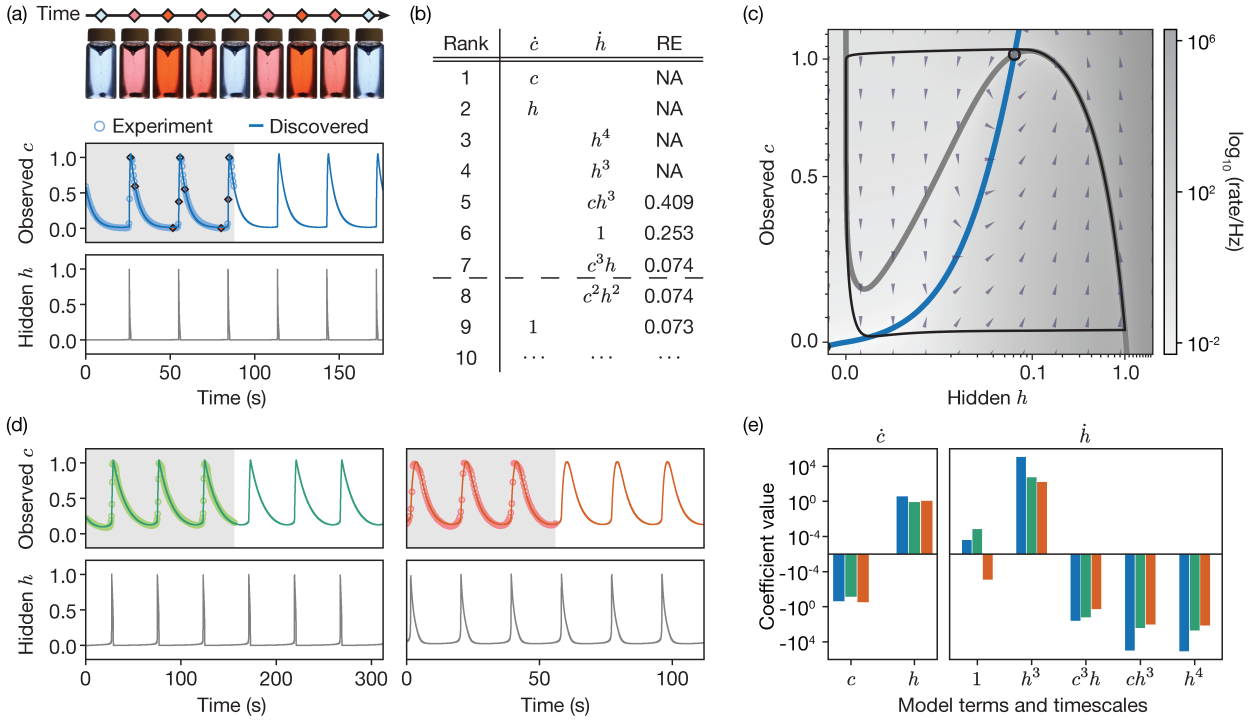

Figure 8.3: HDI applied to our experimental BZ reaction data from J. F. Totz learns a two-variable linear-quartic model that generalizes under catalyst variations. (a) Experimental snapshots of the BZ reaction showing periodic color oscillations (top). Input data (open circles) and observed and hidden variables (solid line) integrated from the learned polynomial ODE model. Using data from three oscillations, the learning framework finds that a seven-term ODE can accurately describe the dynamics. (b) Polynomial ODE terms appearing in $\dot{c}$ and $\dot{h}$ equations ranked from most to least important based on their coefficient of variation. Model terms are added one at a time in order of importance, with the seventh term leading to a drop in the training loss. (c) Phase plane diagram of learned seven-term ODE from the previous panel contains crucial features found in most two-variable BZ models [296]. The limit cycle contains an unstable fixed point (black) with a monotonic x-nullcline (blue) and an h-nullcline (dark gray) in the form of a "cubic" curve as found in the FHN, Rovinsky, and ZBKE models. (d) The Resulting seven-term model (solid line) accurately fits the dynamics of the chemical solution's color (open circles) in two new BZ experiments. (e) Coefficients of the model remain consistent across all three experimental BZ reactions. Chemical concentrations: 0.20 M H_2SO_4, 0.11 M $NaBrO_3$, 0.05 M $CH_2(COOH)_2$, 0.03 M NaBr, 0.3 mM ferroin (blue), 0.41 M H_2SO_4, 0.17 M $NaBrO_3$, 0.03 M $CH_2(COOH)_2$, 0.02 M NaBr, 0.3 mM ferroin (green), 0.51 M H_2SO_4, 0.10 M $NaBrO_3$, 0.03 M $CH_2(COOH)_2$, 0.02 M NaBr, 0.3 mM ferroin (red).

repeated oxidation and reduction of the metal catalyst ferroin produces a periodic change in the color of the solution from red to light blue [Fig. 8.3(a)]. The recorded average color of the solution follows a 1D curve in color space, which we map to our single observed coordinate $c(t)$ (Appendix B). Working with polynomial approximations consistent with established two-variable BZ models [300], we optimize overall two-variable ODEs that are linear in the $\dot{c}$ equation and quartic in the $\dot{h}$ equation. Using this library, HDI discovers a seven-term model that accurately fits the color dynamics $c(t)$ for BZ reactions [Fig. 8.3(a, d)] with parameters that vary smoothly across the different reactant concentrations in each experiment [Fig. 8.3(e)]. Furthermore, the phase portrait of the learned model correctly captures the dynamical properties of the BZ reaction [300], showing an unstable fixed point enclosed in a stable limit cycle with a typical cubic-shaped nullcline $\dot{h} = 0$ [Fig. 8.3(c)].

8.4 Application of HDI to chaotic systems

HDI straightforwardly extends to higher-dimensional nonlinear systems. For example, when only given observations of the x and y coordinates of the 3D Lorenz system, with $\sigma = 10, \rho = 28, \beta = 8/3$,

$$\dot{x} = \sigma(y - x) \tag{8.5a}$$

$$\dot{y} = x(\rho - z) - y \tag{8.5b}$$

$$\dot{z} = xy - \beta z., \tag{8.5c}$$

for one or two lobe transitions [gray-shaded in Fig. 8.4(a)], a HDI search over all polynomial three-variable ODEs in (x, y, z) with quadratic interactions recovers the exact Lorenz equations with correct coefficient values (modulo a trivial scaling of the hidden z variable) [Fig. 8.4(b)]. The learned model has the correct attractor dynamics and can predict the x, y, and z dynamics substantially beyond the training interval [Fig. 8.4(a,c)]. Further analysis shows that, even when only given observations of x, HDI learns a predictive model for Lorenz dynamics, albeit with reduced predictive power [280].

8.5 Conclusions

By combining sensitivity methods and ranked-choice voting, HDI can discover parsimonious predictive models from partial noisy observations of oscillatory and chaotic dynamics without extensive preprocessing of time-series data. The above framework can be directly applied to experimental observations of biophysical, ecological, and other systems, for which ODE

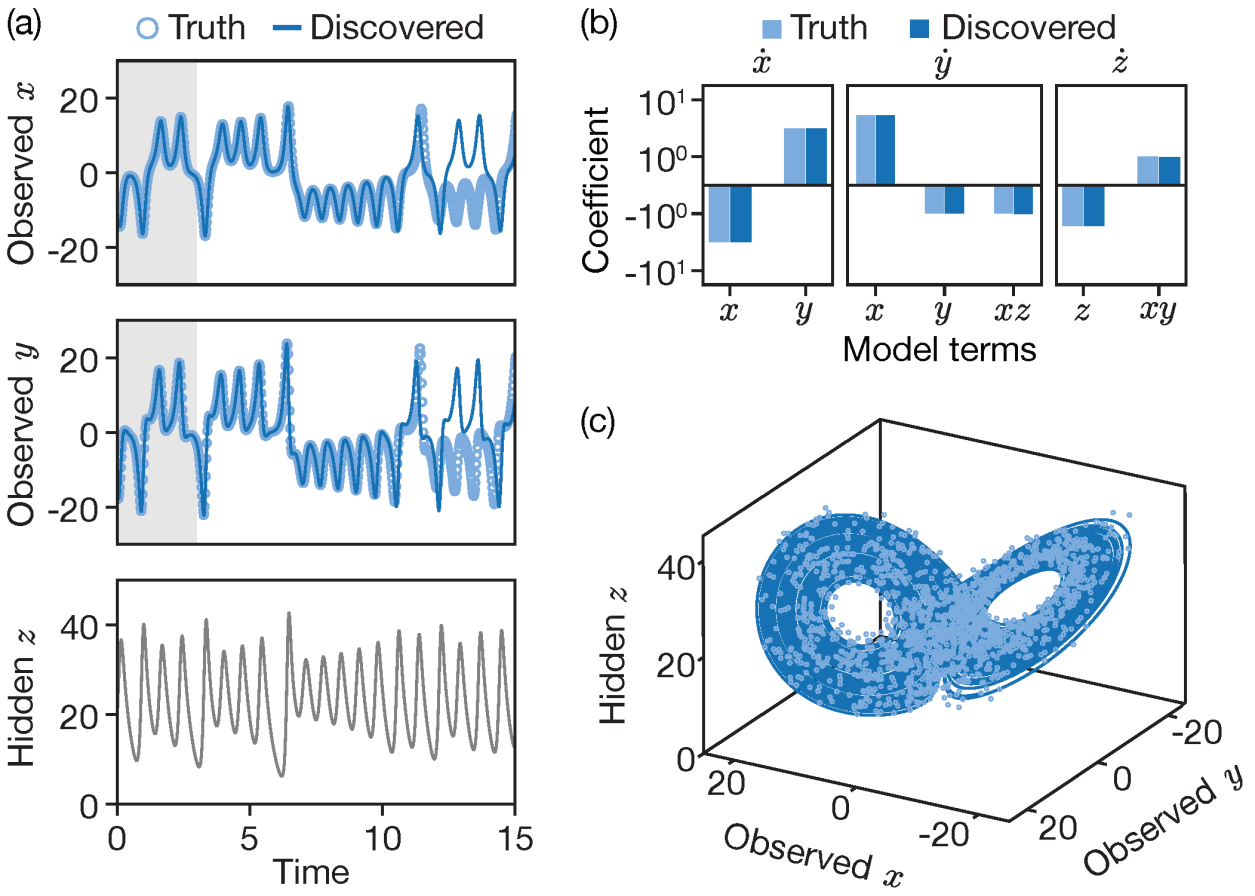

Figure 8.4: HDI discovers true Lorenz system from observations of x and y coordinates. (a) Given observations of only the x and y coordinates (gray region), the learned model predicts the evolution for several additional Lyapunov time scales. (b) Lorenz model terms and coefficients are discovered exactly by an HDI search based solely on data in the gray region of the previous panel. (c) The reconstructed attractor of the learned model closely agrees with the true Lorenz attractor.

models can inform the prediction, control, and optimal perturbations [306] of dynamical behavior. By mapping time series to ODE model coefficients, HDI can help facilitate the clustering of dynamical data, as those appearing in health [307] and climate [258] studies.

Chapter 9

Conclusions and outlook

Spectral representations are a powerful tool for the analysis and modeling of biological data. In the first part of this thesis, we showed how spectral methods are applicable across a broad range of experimental systems, from high-resolution microscopy data to tracked single-particle trajectories to gene expression profiles. In all cases, the spectral projections gave a noise-robust low-dimensional representation of the data. Using these low-dimensional representations, we could extract quantitative measurements from experimental data and compare experiments and theory directly. The methods presented here readily apply to data that live on simple geometries, like rectangles and spheres where basis functions are known ahead of time or other simple geometries where analytic Gram-Schmidt orthogonalization is possible. However, many experimental systems exist in more general geometries – active matter systems may form arbitrarily shaped clusters [308] or during morphogenesis tissues can form complicated time-evolving surfaces [9] – requiring the extension of the presented methods to more complex domains. Developing more general construction methods for basis functions, domain mapping techniques, and robust and efficient algorithms to extract representation coefficients is an ongoing research direction. Furthermore, in the work presented here, we assume that the underlying geometry of the system is identifiable independently from a specific model or dataset. This assumption is frequently valid for physical dynamics. However, for biological signaling data, such as neural activity or gene expression, this assumption may no longer hold. Methods that can extract the correct geometry and spectral representation to use for such data will be critical for incorporating additional fields into the framework and for building fuller pictures of the multi-scale, multi-modal underlying dynamics of biological systems.

Mode space dynamical models have many practical and theoretical advantages. The reduction of potentially complex real space dynamics to a coupled set of ordinary differential equations is a powerful simplification that loses no predictive power due to the orthogonal-

ity of the transform. In the second half of this thesis, we demonstrate the power of mode space models in capturing complex biophysical phenomena. By considering linear models of morphogenesis, we found that mode space models can effectively capture symmetry-breaking during the early stages of development. By studying the resulting linear dynamical system, we can also identify dynamic similarities between the early stages of development and active Brownian motion. Additionally, we considered physically constrained models of planar curve dynamics and applied the resulting model framework to undulatory locomotion. We show how linear mode space models can accurately distinguish different types of undulatory locomotion across species and behavioral states. Taken together, these results demonstrate the power of linear mode space models to act as a tool to characterize complex biophysical phenomena. The simplicity of the underlying linear dynamics enables efficient inference procedures directly from experimental data, and by directly comparing integrated predictions of the linear model, we can enforce stability and gain robustness to noise.

Despite the power of linear mode space models, most biological dynamics are not linear. In the last chapter, we show some initial extensions of the underlying inference framework to partially observed dynamics and nonlinear models. The sensitivity approach introduced here, coupled with sparse model selection, enables inference of predictive models even when variables are unobserved and in the presence of noise. The application of the nonlinear inference framework to mode space dynamics is straightforward. However, the number of model terms multiplies as the number of modes increases and quickly becomes infeasible. Therefore, constraining the number of terms in the nonlinear models will be necessary. This requirement poses the interesting theoretical challenge of determining what model structures, physical symmetries, biological constraints, and basis choice can impose on nonlinear dynamical systems. From a practical perspective, developing algorithms that can then use this additional structure to speed up the inference will also be crucial for the broader applicability of these ideas.

As high-resolution, multi-scale, and multi-modal data increasingly becomes available for biological dynamics, new mathematical analysis tools will become increasingly important to further our understanding of the complex interplay between physics, chemistry, and biology that governs the dynamics. Spectral representations provide a powerful, extensible, and robust framework for developing such tools.

Appendix A

Derivation of planar curve length constraint in Chebyshev basis

Making use of the following four Chebyshev polynomial identities,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}T_0(s) &= 0 \\
\frac{\mathrm{d}}{\mathrm{d}s}T_k(s) &= kU_{k-1}(s) \\
U_n(s)U_m(s) &= \sum_{k=0}^{2\min(n,m)} U_{|n-m|+2k}(s) \\
\int_{-1}^{1} \mathrm{d}s\, U_n(s) &= \left.\frac{T_{n+1}}{n+1}\right|_{-1}^{1} = \frac{1}{n+1} - \frac{(-1)^{n+1}}{n+1} = \frac{1+(-1)^n}{n+1}
\end{aligned}$$

we can derive the values of $W_{m,n}$ (6.15b) explicitly for the Chebyshev basis,

$$\begin{aligned}
W_{m,n} &= \int_{-1}^{1} \mathrm{d}s\, \frac{\mathrm{d}T_n}{\mathrm{d}s}(s)\frac{\mathrm{d}T_m}{\mathrm{d}s}(s) \\
&= nm \int_{-1}^{1} \mathrm{d}s\, U_{n-1}(s)U_{m-1}(s) \\
&= nm \int_{-1}^{1} \mathrm{d}s \sum_{k=0}^{\min(n-1,m-1)} U_{|n-m|+2k}(s) \\
&= nm \sum_{k=0}^{\min(n-1,m-1)} \int_{-1}^{1} \mathrm{d}s\, U_{|n-m|+2k}(s) \\
&= nm \sum_{k=0}^{\min(n-1,m-1)} \frac{1+(-1)^{|n-m|+2k}}{|n-m|+2k+1}
\end{aligned}$$

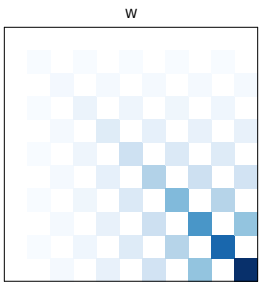

Figure A.1: W matrix for $n = 10$

$$= nm(1 + (-1)^{|n-m|}) \sum_{k=0}^{\min(n,m)-1} \frac{1}{|n-m| + 2k + 1}$$

$$= \begin{cases} 0 & \text{if } m - n \text{ odd} \\ 2nm \sum_{k=|n-m|+1}^{n+m-1} \frac{1}{k} & \text{if } m - n \text{ even} \end{cases}$$

which shows that W has a checkerboard pattern and is diagonally dominant (Figure A.1).

Appendix B

Belousov-Zhabotinsky reaction trajectory extraction

We provide details on the extraction of a single variable from optical measurements of the Belousov-Zhabotinsky reaction performed by J. F. Totz.

The Belousov-Zhabotinsky (BZ) reaction is the paradigmatic nonlinear chemical oscillator. It involves more than 30 chemical species and 40 elementary reactions. Over the course of the net reaction an organic substrate such as malonic acid (MA) is consumed:

$$3\,\mathrm{MA} + 2\,\mathrm{BrO_3^-} + 2\,\mathrm{H^+} \longrightarrow 2\,\mathrm{BrMA} + 3\,\mathrm{CO_2} + 4\,\mathrm{H_2O}\,. \tag{B.1}$$

However, intermediary species ($\mathrm{Br^-}$, $\mathrm{BrO_3^+}$, $[\mathrm{Fe(phen)_3}]^{3+}$) are periodically built up and expended, leading to a periodically changing consumption rate of the organic substrate. One of the intermediary species, the oxidized form of the catalyst ferroin $[\mathrm{Fe(phen)_3}]^{3+}$, can be readily observed optically using spectrophotometry (Figure B.1(a) to track the chemical oscillation state.

The chemical oscillation state is optically observable because the absorption spectrum of the solution depends on the ratio of reduced to oxidized catalyst [Figure B.1(b)]. In the reduced state the catalyst absorbs blue and green wavelengths 450 nm to 550 nm, so that one observes a red coloration of the solution. In the oxidized state the catalyst weakly absorbs red wavelengths around 600 nm, leading to a faint blue coloration of the BZ solution.

A single trajectory was extracted from the movie Figure B.2. First the movie was cropped to a rectangle containing only the BZ solution (dashed box). Next the frame with with the highest average blue channel value over the rectangle was chosen as a reference frame and the distance between this color and the average color was calculated using the Euclidean distance between the colors in the Lab colorspace. The distance was then normalized to lie

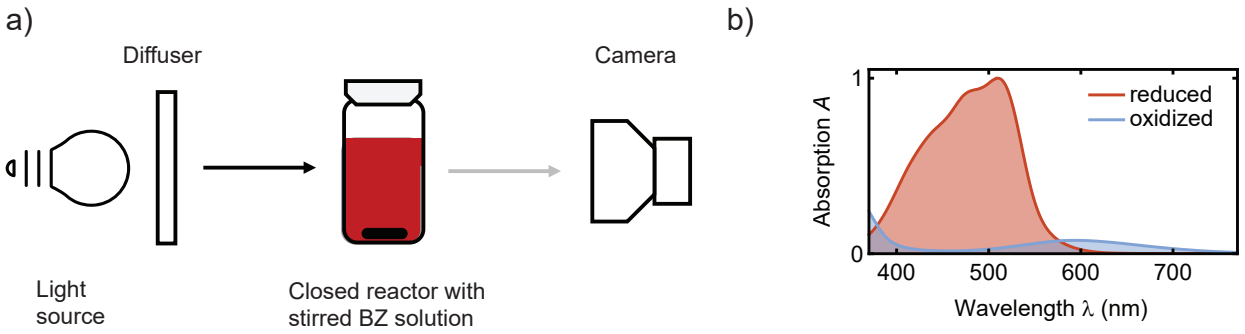

Figure B.1: Optical measurement of periodic concentration changes in the oscillating Belousov-Zhabotinsky reaction performed by J. F. Totz. (a) The experimental setup consists of a spatially homogenized broadband light-source supplying the illumination that is absorbed by the reagents in the closed stirred chemical. The transmitted light is captured by a camera. (b) During chemical oscillations the ferroin reagent cyclically changes its oxidation state affecting its corresponding absorption spectrum: reduced catalyst Fe^{2+} (red) and oxidized catalyst Fe^{3+} (blue). This allows for tracking the chemical oscillation state optically.

in $[0, 1]$ and inverted so that the reference frame is a peak at 1.

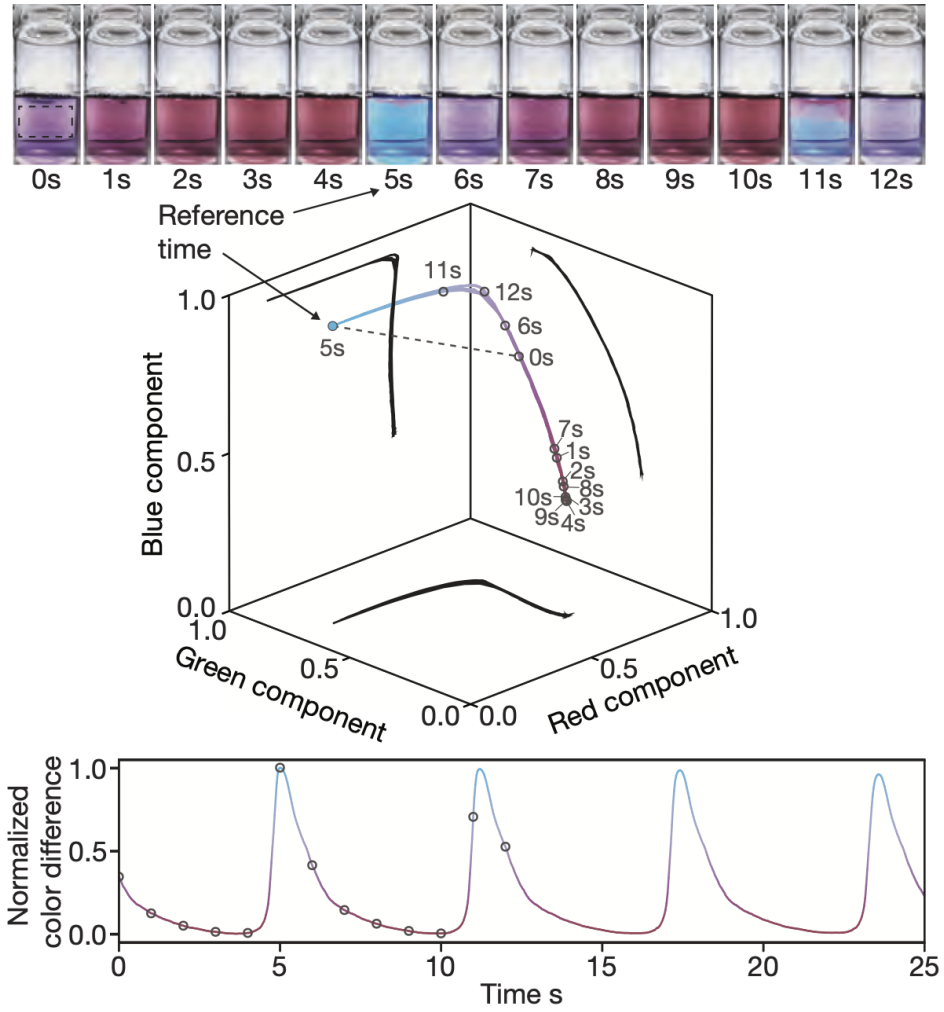

Figure B.2: (Top) Snapshots of BZ reaction. (Middle) Average color in cropped region (dashed box top) plotted in RGB color space shows that the reaction follows a 1 dimensional curve in color space. (Bottom) A single trajectory is extracted from the BZ movies by first cropping the movies to a rectangle (dashed box top row) and then calulating the color difference between the average color of a frame and a reference frame (5s) using Euclidean distance in the Lab colorspace. The resulting trajectory is rescaled to lie between $[0, 1]$.

Appendix C

Additional details on the nonlinear optimization procedure

Here we describe the initial optimization procedure that is used to fit models to experimental data for use in the larger model selection framework. Our aim is to fit ODE models consisting of M scalar variables, such that the first m variables $\mathbf{x}(t) = \{x_k(t)\}_{k=1}^m$ fit observed data $\{(t_i, \mathbf{y}_i)\}_{i=1}^n$ where $\mathbf{y}_i = \{y_{ik}\}_{k=1}^m$ and the next $M-m$ variables $\{h_k(t)\}_{k=m+1}^M$ are hidden.

Given a recorded time series, we select a window of training data points $\{(t_i, \mathbf{y}_i)\}_{i=1}^n$ where $t_1 < t_2 < \ldots < t_n$. If the time series dynamics behaves periodically, then we choose our training window $[t_1, t_n]$ to contain three periods of oscillation. This heuristic ensures that our optimization converges to a model with periodic behavior from most random initializations. Choosing too few periods in the training window results in learned models with only transient periodic behavior which then either diverge or converge to a fixed point. On the other hand, choosing too many time periods (or generally a large training window) leads the HDI optimization to ODE models that converge to a fixed point close to the mean of the data, a common feature in sensitivity-based ODE inference methods. If the time series dynamics behaves chaotically, as with the Lorenz system, we choose our training window $[t_1, t_n]$ to contain two or three branch switches which helps our algorithm converge to a chaotic model more frequently. Providing too few branch switches leads to models that eventually fall on a periodic limit cycle. As before, we do not include more than three branch switches as long time windows lead to models which converge to fixed points.

We define an M-variable polynomial HDI model with degree combination $(d_1, \ldots, d_M)$ as

$$\dot{x}_k = \tau_k \sum_{|\boldsymbol{\alpha}| \leq d_k} c_{\boldsymbol{\alpha}}^k x_1^{\alpha_1} \ldots h_M^{\alpha_M}, \quad x_k(0) = b_k, \quad 1 \leq k \leq m \tag{C.1a}$$

$$\dot{h}_k = \tau_k \sum_{|\boldsymbol{\alpha}| \leq d_k} c^k_{\boldsymbol{\alpha}} x_1^{\alpha_1} \dots h_M^{\alpha_M}, \quad h_k(0) = b_k, \quad m < k \leq M \tag{C.1b}$$

where the free parameters are the initial conditions b_k, polynomial coefficient vectors $\mathbf{c}_k = \{c^k_{\boldsymbol{\alpha}}\}_{|\boldsymbol{\alpha}| \leq d_k}$, and time scales τ_k for all $1 \leq k \leq M$. By definition, the degree of any polynomial term in equation k does not exceed d_k. The power of each monomial term above is expressed in multi-index notation $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_M)$ where each entry $\alpha_r \in \mathbb{N}$ denotes the power of x_r and we write $|\boldsymbol{\alpha}| = \sum_{k=1}^M \alpha_k$ to denote the total degree of the monomial term.

In a model sweep, each HDI model is initialized randomly by setting the initial conditions of its observed variables to $b_k = y_{1k}$ for $1 \leq k \leq m$ and of its hidden variables to $b_k = 1$ for $m + 1 \leq k \leq M$. Lastly, we choose the M coefficient vectors $\{\mathbf{c}_k\}_{k=1}^M$ independently at random as uniformly distributed vectors on the unit sphere and set the time scales of all equations to $\tau_k = 0$ for $1 \leq k \leq M$.

All initialized HDI models are dense which means they contain all polynomial terms up to degree d_k in the kth equation for $1 \leq k \leq M$. Each randomly initialized model is parameterized by the stacked vector

$$\mathbf{p} = (\{b_k\}_{k=1}^M, \{\mathbf{c}_k\}_{k=1}^M, \{\tau_k\}_{k=1}^M) \tag{C.2}$$

of its initial conditions, coefficients, and time scales and is trained for one round of optimization to minimize the objective

$$L(\mathbf{p}) = \text{MSE}(\mathbf{p}) + \text{Reg}(\mathbf{p}) \tag{C.3}$$

which is a sum of the mean squared error

$$\text{MSE}(\mathbf{p}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m (x_j(t_i, \mathbf{p}) - y_{ij})^2 \tag{C.4}$$

and a regularization term. The regularization term in the objective function is given by

$$\text{Reg}(\mathbf{p}) = \lambda \sum_{k=1}^M \sum_{|\boldsymbol{\alpha}| \leq d_k} \sqrt{1 + |\boldsymbol{\alpha}|} |c^k_{\boldsymbol{\alpha}}| + \gamma \sum_{k=1}^M (\|\mathbf{c}_k\|^2 - 1)^2 \tag{C.5}$$

where the first term penalizes the sparsity of the learned model, since polynomial terms $x_1^{\alpha_1} ... x_M^{\alpha_M}$ in the model with a higher total degree $|\boldsymbol{\alpha}| = \alpha_1 + \ldots + \alpha_M$ are more actively down-weighted by the factor $\sqrt{1 + |\boldsymbol{\alpha}|}$. The second part of the regularization function above enforces that the coefficients $\mathbf{c}_k$ of the kth equation have a unit norm, which allows us to

avoid scaling ambiguities between the timescales τ_k and model coefficients $\mathbf{c}_k$. The unit norm penalty on the coefficients is always set to the large value of $\gamma = 5 \times 10^4$, forcing the unit norm constraints to be satisfied almost exactly.

Each model is optimized for 100,000 AdaBelief [243] iterations followed by 50,000 iterations of BFGS [279] (e.g. large enough for our optimization to converge). The derivatives of the loss functions are calculated using the by solving the forward sensitivity equations to calculate the derivative through the ODE solution [37, 135]. In general, our method is not overly sensitive to how these values are set.

References

1. Pereira, T. D. *et al.* Fast animal pose estimation using deep neural networks. *Nature Methods* **16,** 117–125 (2019).
2. Pereira, T. D., Shaevitz, J. W. & Murthy, M. Quantifying behavior to understand the brain. *Nature Neuroscience* **23,** 1537–1549 (2020).
3. Cermak, N. *et al.* Whole-organism behavioral profiling reveals a role for dopamine in state-dependent motor program coupling in C. elegans. *eLife* **9,** e57093 (2020).
4. Tan, T. H. *et al.* Topological Turbulence in the Membrane of a Living Cell. *Nature Physics* **16,** 657–662 (2020).
5. Mitchell, N. P. *et al.* Morphodynamic atlas for Drosophila development. *bioRxiv,* 2022–05 (2022).
6. Stelzer, E. H. K. Light-sheet fluorescence microscopy for quantitative biology. *Nature Methods* **12,** 23–26 (2015).
7. Power, R. M. & Huisken, J. A guide to light-sheet fluorescence microscopy for multi-scale imaging. *Nature Methods* **14,** 360–373 (2017).
8. Shah, G. *et al.* Multi-scale imaging and analysis identify pan-embryo cell dynamics of germlayer formation in zebrafish. *Nature Communications* **10,** 5753 (2019).
9. Mitchell, N. P. & Cislo, D. J. TubULAR: tracking in toto deformations of dynamic tissues via constrained maps. *Nature Methods* **20,** 1980–1988 (2023).
10. Hartmann, R. *et al.* Emergence of three-dimensional order and structure in growing biofilms. *Nature Physics* **15,** 251–256 (2019).
11. Jeckel, H. *et al.* Learning the space-time phase diagram of bacterial swarm expansion. *Proceedings of the National Academy of Sciences* **116,** 1489–1494 (2019).
12. Jeckel, H. *et al.* Simultaneous spatiotemporal transcriptomics and microscopy of Bacillus subtilis swarm development reveal cooperation across generations. *Nature Microbiology* **8,** 2378–2391 (2023).
13. Atanas, A. A. *et al.* Brain-wide representations of behavior spanning multiple timescales and states in C. elegans. *Cell* **186,** 4134–4151 (2023).
14. Kantsler, V. & Goldstein, R. E. Fluctuations, dynamics, and the stretch-coil transition of single actin filaments in extensional flows. *Physical Review Letters* **108,** 038103 (2012).

15. Griffiths, D. J. & Schroeter, D. F. *Introduction to quantum mechanics* (Cambridge University Press, 2018).

16. Riedel-Kruse, I. H., Hilfinger, A., Howard, J. & Jülicher, F. How molecular motors shape the flagellar beat. *HFSP journal* **1,** 192–208 (2007).

17. Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L. & Kutz, J. N. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics* **1,** 391 (2014).

18. Schmid, P. J. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics* **656,** 5–28. ISSN: 1469-7645, 0022-1120. (2021) (2010).

19. Kutz, J. N., Brunton, S. L., Brunton, B. W. & Proctor, J. L. *Dynamic mode decomposition: data-driven modeling of complex systems* (SIAM, 2016).

20. Berkooz, G., Holmes, P. & Lumley, J. L. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics* **25,** 539–575 (1993).

21. Driscoll, T. A., Hale, N. & Trefethen, L. N. *Chebfun guide* 2014.

22. Burns, K. J., Vasil, G. M., Oishi, J. S., Lecoanet, D. & Brown, B. P. Dedalus: A flexible framework for numerical simulations with spectral methods. *Physical Review Research* **2,** 023068 (2020).

23. Boyd, J. P. *Chebyshev and Fourier spectral methods* (Dover Publications, 2001).

24. Olver, S., Slevinsky, R. M. & Townsend, A. Fast algorithms using orthogonal polynomials. *Acta Numerica* **29,** 573–699 (2020).

25. Henshaw, R. J., Martin, O. G. & Guasto, J. S. Dynamic mode structure of active turbulence. *Physical Review Fluids* **8,** 023101 (2023).

26. Smits, C. M., Dutta, S., Jain-Sharma, V., Streichan, S. J. & Shvartsman, S. Y. Maintaining symmetry during body axis elongation. *Current Biology* **33,** 3536–3543 (2023).

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

28. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics* **378,** 686–707 (2019).

29. Bär, M., Hegger, R. & Kantz, H. Fitting partial differential equations to space-time dynamics. *Physical Review E* **59,** 337–342 (1999).

30. Vallette, D. P., Jacobs, G. & Gollub, J. P. Oscillations and spatiotemporal chaos of one-dimensional fluid fronts. *Physical Review E* **55,** 4274–4287 (1997).

31. Rudy, S. H., Brunton, S. L., Proctor, J. L. & Kutz, J. N. Data-driven discovery of partial differential equations. *Science Advances* **3,** e1602614 (2017).

32. Maddu, S., Cheeseman, B. L., Sbalzarini, I. F. & Müller, C. L. Stability selection enables robust learning of differential equations from limited noisy data. *Proceedings of the Royal Society A* **478,** 20210916 (2022).

33. Reinbold, P. A., Gurevich, D. R. & Grigoriev, R. O. Using noisy or incomplete data to discover models of spatiotemporal dynamics. *Physical Review E* **101,** 010203(R) (2020).

34. Schmidt, M. & Lipson, H. Distilling Free-Form Natural Laws. *Science* **324,** 81–86 (2009).

35. Bongard, J. & Lipson, H. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **104,** 9943–9948 (2007).

36. Champion, K., Lusch, B., Nathan Kutz, J. & Brunton, S. L. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences* **116,** 22445–22451 (2019).

37. Rackauckas, C. *et al. Universal differential equations for scientific machine learning* 2021.

38. Battles, Z. & Trefethen, L. N. An extension of MATLAB to continuous functions and operators. *SIAM Journal on Scientific Computing* **25,** 1743–1770 (2004).

39. Olver, S. & Townsend, A. A fast and well-conditioned spectral method. *siam REVIEW* **55,** 462–489 (2013).

40. Boyd, J. P. Computing zeros on a real interval through Chebyshev expansion and polynomial rootfinding. *SIAM Journal on Numerical Analysis* **40,** 1666–1682 (2002).

41. Frigo, M. & Johnson, S. G. The design and implementation of FFTW3. *Proceedings of the IEEE* **93,** 216–231 (2005).

42. Wright, G. B., Javed, M., Montanelli, H. & Trefethen, L. N. Extension of Chebfun to periodic functions. *SIAM Journal on Scientific Computing* **37,** C554–C573 (2015).

43. Trefethen, L. N. & Weideman, J. The exponentially convergent trapezoidal rule. *SIAM review* **56,** 385–458 (2014).

44. Mason, J. C. & Handscomb, D. C. *Chebyshev polynomials* (CRC press, 2002).

45. Trefethen, L. N. *Approximation theory and approximation practice, extended edition* (SIAM, 2019).

46. Smith, F. J. An algorithm for summing orthogonal polynomial series and their derivatives with applications to curve-fitting and interpolation. *Mathematics of Computation* **19,** 33–36 (1965).

47. Boyd, J. P. Defeating the Runge phenomenon for equispaced polynomial interpolation via Tikhonov regularization. *Applied mathematics letters* **5,** 57–59 (1992).

48. Phillips, G. M. *Interpolation and approximation by polynomials* (Springer, 2006).

49. Barnett, A. H., Magland, J. & af Klinteberg, L. A parallel nonuniform fast Fourier transform library based on an “exponential of semicircle" kernel. *SIAM Journal on Scientific Computing* **41,** C479–C504 (2019).

50. Aurentz, J. L. & Trefethen, L. N. Chopping a Chebyshev Series. *ACM Transactions on Mathematical Software* **43** (2017).

51. Naik, R. A. Failure Analysis of Woven and Braided Fabric Reinforced Composites. *Journal of Composite Materials* **29,** 2334–2363 (1995).

52. Thiffeault, J.-L. Measuring Topological Chaos. *Physical Review Letters* **94,** 084502 (2005).

53. Stern, A. & Lindner, N. H. Topological Quantum Computation—From Basic Concepts to First Experiments. *Science* **339,** 1179–1184 (2013).

54. Rossmanith, H., Doebroenti, M., Albach, M. & Exner, D. Measurement and Characterization of High Frequency Losses in Nonideal Litz Wires. *IEEE Transactions on Power Electronics* **26,** 3386–3394 (2011).

55. Kauffman, L. H. & Lomonaco, S. J. Braiding, Majorana Fermions, Fibonacci Particles and Topological Quantum Computing. *Quantum Information Processing* **17,** 201 (2018).

56. Budišić, M. & Thiffeault, J.-L. Finite-Time Braiding Exponents. *Chaos* **25,** 087407 (2015).

57. Bauer, B. *et al.* Topologically Protected Braiding in a Single Wire Using Floquet Majorana Modes. *Physical Review B* **100,** 041102 (2019).

58. Tkalec, U., Ravnik, M., Čopar, S., Žumer, S. & Muševič, I. Reconfigurable Knots and Links in Chiral Nematic Colloids. *Science* **333,** 62–65 (2011).

59. Čopar, S., Tkalec, U., Muševič, I. & Žumer, S. Knot Theory Realizations in Nematic Colloids. *Proceedings of the National Academy of Sciences* **112,** 1675–1680 (2015).

60. Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Das Sarma, S. Non-Abelian Anyons and Topological Quantum Computation. *Reviews of Modern Physics* **80,** 1083–1159 (2008).

61. Freedman, M., Kitaev, A., Larsen, M. & Wang, Z. Topological Quantum Computation. *Bulletin of the American Mathematical Society* **40,** 31–38 (2003).

62. Lechleiter, J., Girard, S., Peralta, E. & Clapham, D. Spiral Calcium Wave Propagation and Annihilation in Xenopus Laevis Oocytes. *Science* **252,** 123–126 (1991).

63. Davidenko, J. M., Pertsov, A. V., Salomonsz, R., Baxter, W. & Jalife, J. Stationary and Drifting Spiral Waves of Excitation in Isolated Cardiac Muscle. *Nature* **355,** 349–351 (1992).

64. Gray, R. A., Pertsov, A. M. & Jalife, J. Spatial and Temporal Organization during Cardiac Fibrillation. *Nature* **392,** 75–78 (1998).

65. Bursac, N., Aguel, F. & Tung, L. Multiarm Spirals in a Two-Dimensional Cardiac Substrate. *Proceedings of the National Academy of Sciences* **101,** 15530–15534 (2004).

66. Christoph, J. *et al.* Electromechanical Vortex Filaments during Cardiac Fibrillation. *Nature* **555,** 667–672 (2018).

67. Huang, X. *et al.* Spiral Wave Dynamics in Neocortex. *Neuron* **68,** 978–990 (2010).

68. Copenhagen, K., Alert, R., Wingreen, N. S. & Shaevitz, J. W. Topological Defects Promote Layer Formation in Myxococcus Xanthus Colonies. *Nature Physics* **17,** 1–5 (2021).

69. Saw, T. B. *et al.* Topological Defects in Epithelia Govern Cell Death and Extrusion. *Nature* **544,** 212–216 (2017).

70. Liu, J. *et al.* Topological braiding and virtual particles on the cell membrane. *Proceedings of the National Academy of Sciences* **118,** e2104191118 (2021).

71. Piekny, A., Werner, M. & Glotzer, M. Cytokinesis: Welcome to the Rho Zone. *Trends in Cell Biology* **15,** 651–658 (2005).

72. Fukata, Y., Kaibuchi, K., Amano, M. & Kaibuchi, K. Rho–Rho-Kinase Pathway in Smooth Muscle Contraction and Cytoskeletal Reorganization of Non-Muscle Cells. *Trends in Pharmacological Sciences* **22,** 32–39 (2001).

73. Etienne-Manneville, S. & Hall, A. Rho GTPases in Cell Biology. *Nature* **420,** 629–635 (2002).

74. Bement, W. M. *et al.* Activator–Inhibitor Coupling between Rho Signalling and Actin Assembly Makes the Cell Cortex an Excitable Medium. *Nature Cell Biology* **17,** 1471–1483 (2015).

75. Wigbers, M. C. *et al.* A Hierarchy of Protein Patterns Robustly Decodes Cell Shape Information. *Nature Physics* **17,** 578–584 (2021).

76. Aranson, I. S. & Kramer, L. The World of the Complex Ginzburg-Landau Equation. *Reviews of Modern Physics* **74,** 99–143 (2002).

77. Marple, L. Computing the Discrete-Time "Analytic" Signal via FFT. *IEEE Trans. Signal Process.* **47,** 2600–2603 (1999).

78. Shaebani, M. R., Wysocki, A., Winkler, R. G., Gompper, G. & Rieger, H. Computational models for active matter. *Nature Reviews Physics* **2,** 181–199 (2020).

79. Qin, B. *et al.* Cell position fates and collective fountain flow in bacterial biofilms revealed by light-sheet microscopy. *Science* **369,** 71–77 (2020).

80. Hartmann, R. *et al.* Quantitative image analysis of microbial communities with BiofilmQ. *Nature Microbiology* (2021).

81. Marchetti, M. C. *et al.* Hydrodynamics of soft active matter. *Reviews of Modern Physics* **85,** 1143–1189 (3 2013).

82. Jülicher, F., Grill, S. W. & Salbreux, G. Hydrodynamic theory of active matter. *Reports on Progress in Physics* **81,** 076601 (2018).

83. Maddu, S., Cheeseman, B. L., Müller, C. L. & Sbalzarini, I. F. Learning physically consistent differential equation models from data using group sparsity. *Physical Review E* **103,** 042310 (2021).

84. Supekar, R. *et al.* Learning hydrodynamic equations for active matter from particle simulations and experiments. *Proceedings of the National Academy of Sciences* **120,** e2206994120 (2023).

85. Zhang, A., Yang, B. & Huang, L. *Feature Extraction of EEG Signals Using Power Spectral Entropy* in *2008 International Conference on BioMedical Engineering and Informatics* **2** (2008), 435–439.

86. Pan, Y. N., Chen, J. & Li, X. L. Spectral entropy: A complementary index for rolling element bearing performance degradation assessment. *Proc. Inst. Mech. Eng. C* **223,** 1223–1231 (2009).

87. Geyer, D., Morin, A. & Bartolo, D. Sounds and hydrodynamics of polar active fluids. *Nature Materials* **17,** 789–793 (2018).

88. Bruno, O. & Hoch, D. Numerical differentiation of approximated functions with limited order-of-accuracy deterioration. *SIAM Journal on Numerical Analysis* **50,** 1581–1603 (2012).

89. Walter, T. & Couzin, I. D. TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. *eLife* **10,** e64000 (2021).

90. Flemming, H.-C. & Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. *Nature Reviews Microbiology* **17,** 247–260 (2019).

91. Kearns, D. B. A field guide to bacterial swarming motility. *Nature reviews microbiology* **8,** 634–644 (2010).

92. Yan, J., Monaco, H. & Xavier, J. B. The ultimate guide to bacterial swarming: an experimental model to study the evolution of cooperative behavior. *Annual review of microbiology* **73,** 293–312 (2019).

93. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology* **14,** 20–32 (2016).

94. Earle, K. A. *et al.* Quantitative imaging of gut microbiota spatial organization. *Cell host & microbe* **18,** 478–488 (2015).

95. Flemming, H.-C. *et al.* Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology* **14,** 563–575 (2016).

96. Partridge, J. D., Ariel, G., Schvartz, O., Harshey, R. M. & Be'er, A. The 3D architecture of a bacterial swarm has implications for antibiotic tolerance. *Scientific reports* **8,** 15823 (2018).

97. Harshey, R. M. & Partridge, J. D. Shelter in a swarm. *Journal of molecular biology* **427,** 3683–3694 (2015).

98. Evans, C. R., Kempes, C. P., Price-Whelan, A. & Dietrich, L. E. Metabolic heterogeneity and cross-feeding in bacterial multicellular systems. *Trends in microbiology* **28,** 732–743 (2020).

99. Mark Welch, J. L., Hasegawa, Y., McNulty, N. P., Gordon, J. I. & Borisy, G. G. Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice. *Proceedings of the National Academy of Sciences* **114,** E9105–E9114 (2017).

100. Hellweger, F. L., Clegg, R. J., Clark, J. R., Plugge, C. M. & Kreft, J.-U. Advancing microbial sciences by individual-based modelling. *Nature Reviews Microbiology* **14,** 461–471 (2016).

101. Nadell, C. D., Drescher, K. & Foster, K. R. Spatial structure, cooperation and competition in biofilms. *Nature Reviews Microbiology* **14,** 589–600 (2016).

102. Trefethen, L. N. & Bau, D. *Numerical linear algebra* (SIAM, 2022).

103. Michna, R. H., Commichau, F. M., Tödter, D., Zschiedrich, C. P. & Stülke, J. Subti wiki–a database for the model organism Bacillus subtilis that links pathway, interaction and expression information. *Nucleic Acids Research* **42,** D692–D698 (2014).

104. Pedreira, T., Elfmann, C. & Stülke, J. The current state of Subti Wiki, the database for the model organism Bacillus subtilis. *Nucleic Acids Research* **50,** D875–D882 (2022).

105. Schubert, E. & Rousseeuw, P. J. *Faster k-medoids clustering: improving the PAM, CLARA, and CLARANS algorithms* in *Similarity Search and Applications: 12th International Conference, SISAP 2019, Newark, NJ, USA, October 2–4, 2019, Proceedings 12* (2019), 171–187.

106. Borg, I. & Groenen, P. J. *Modern multidimensional scaling: Theory and applications* (Springer Science & Business Media, 2005).

107. Fedotov, S. Single integrodifferential wave equation for a Lévy walk. *Physical Review E* **93,** 020101 (2016).

108. Othmer, H. G., Dunbar, S. R. & Alt, W. Models of dispersal in biological systems. *Journal of mathematical biology* **26,** 263–298 (1988).

109. Othmer, H. G. & Hillen, T. The diffusion limit of transport equations II: Chemotaxis equations. *SIAM Journal on Applied Mathematics* **62,** 1222–1250 (2002).

110. Chen, L., Painter, K., Surulescu, C. & Zhigun, A. Mathematical models for cell migration: a non-local perspective. *Philosophical Transactions of the Royal Society B* **375,** 20190379 (2020).

111. Box, G. & Coutie, G. Application of digital computers in the exploration of functional relationships. English. *Proceedings of the IEE* **103,** 100–107(7). ISSN: 2050-2656 (1S 1956).

112. Aguirre, L. A. & Letellier, C. Modeling nonlinear dynamics and chaos: a review. *Mathematical Problems in Engineering* **2009** (2009).

113. Bakarji, J., Champion, K., Kutz, J. N. & Brunton, S. L. Discovering Governing Equations from Partial Measurements with Deep Delay Autoencoders. *arXiv:2201.05136* (2022).

114. Ribera, H., Shirman, S., Nguyen, A. & Mangan, N. Model selection of chaotic systems from data with hidden variables using sparse data assimilation. *Chaos* **32,** 063101 (2022).

115. Raissi, M., Yazdani, A. & Karniadakis, G. E. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science* **367,** 1026–1030 (2020).

116. Heinonen, M., Yildiz, C., Mannerström, H., Intosalmi, J. & Lähdesmäki, H. *Learning unknown ODE models with Gaussian processes* in *International Conference on Machine Learning* (2018), 1959–1968.

117. Van Breugel, F., Kutz, J. N. & Brunton, B. W. Numerical differentiation of noisy data: A unifying multi-objective optimization framework. *IEEE Access* **8,** 196865–196877 (2020).

118. Varah, J. M. A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific and Statistical Computing* **3,** 28–46 (1982).

119. Messenger, D. A. & Bortz, D. M. Weak SINDy for partial differential equations. *Journal of Computational Physics,* 110525 (2021).

120. Brown, R., Rulkov, N. F. & Tracy, E. R. Modeling and synchronizing chaotic systems from experimental data. *Physics Letters A* **194,** 71–76. ISSN: 0375-9601 (1994).

121. Dattner, I., Ship, H. & Voit, E. O. Separable nonlinear least-squares parameter estimation for complex dynamic systems. *Complexity* **2020,** 1–11 (2020).

122. Cohen, A. E., Hastewell, A. D., Pradhan, S., Flavell, S. W. & Dunkel, J. Schrödinger Dynamics and Berry Phase of Undulatory Locomotion. *Physical Review Letters* **130,** 258402 (2023).

123. Law, K., Stuart, A. & Zygalakis, K. Data assimilation. *Cham, Switzerland: Springer* **214,** 52 (2015).

124. Rudy, S. H., Kutz, J. N. & Brunton, S. L. Deep learning of dynamics and signal-noise decomposition with time-stepping constraints. *Journal of Computational Physics* **396,** 483–506 (2019).

125. Fullana, J., Rossi, M. & Zaleski, S. Parameter identification in noisy extended systems: a hydrodynamic case. *Physica D: Nonlinear Phenomena* **103,** 564–575 (1997).

126. Abarbanel, H. D., Brown, R. & Kadtke, J. B. Prediction in chaotic nonlinear systems: Methods for time series with broadband Fourier spectra. *Physical Review A* **41,** 1782 (1990).

127. Breeden, J. L. & Hübler, A. Reconstructing equations of motion from experimental data with unobserved variables. *Physical Review A* **42,** 5817 (1990).

128. Baake, E., Baake, M., Bock, H. G. & Briggs, K. M. Fitting ordinary differential equations to chaotic data. *Physical Review A* **45,** 5524–5529 (8 1992).

129. Krueger, J. M. *Parameter estimation methods for ordinary differential equation models with applications to microbiology* PhD thesis (Virginia Tech, 2017).

130. Aguirre, L. A., Teixeira, B. O. S. & Tôrres, L. A. B. Using data-driven discrete-time models and the unscented Kalman filter to estimate unobserved variables of nonlinear systems. *Physical Review E* **72,** 026226 (2 2005).

131. Levine, M. & Stuart, A. A framework for machine learning of model error in dynamical systems. *Communications of the American Mathematical Society* **2,** 283–344 (2022).

132. Schneider, T., Stuart, A. M. & Wu, J.-L. Ensemble Kalman inversion for sparse learning of dynamical systems from time-averaged data. *Journal of Computational Physics* **470,** 111559 (2022).

133. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980* (2017).

134. Liu, D. C. & Nocedal, J. On the limited memory BFGS method for large scale optimization. *Mathematical programming* **45,** 503–528 (1989).

135. Cao, Y., Li, S., Petzold, L. & Serban, R. Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution. *SIAM Journal on Scientific Computing* **24,** 1076–1089 (2003).

136. Villaverde, A. F., Barreiro, A. & Papachristodoulou, A. Structural identifiability of dynamic systems biology models. *PLoS computational biology* **12,** e1005153 (2016).

137. Massonis, G., Villaverde, A. F. & Banga, J. R. Distilling identifiable and interpretable dynamic models from biological data. *bioRxiv,* 2023–03 (2023).

138. Farmer, J. D. & Sidorowich, J. J. Predicting chaotic time series. *Physical review letters* **59,** 845 (1987).

139. Costa, A. C., Ahamed, T. & Stephens, G. J. Adaptive, locally linear models of complex dynamics. *Proceedings of the National Academy of Sciences* **116,** 1501–1510 (2019).

140. Crutchfield, J. P. & McNamara, B. Equations of Motion from a Data Series. *Complex systems* **1,** 417–452 (1987).

141. Corrêa, M. V., Aguirre, L. A. & Mendes, E. M. Modeling chaotic dynamics with discrete nonlinear rational models. *International Journal of Bifurcation and Chaos* **10,** 1019–1032 (2000).

142. Mangan, N. M., Brunton, S. L., Proctor, J. L. & Kutz, J. N. Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications* **2,** 52–63 (2016).

143. Wei, H.-L. & Billings, S. A. Identification and reconstruction of chaotic systems using multiresolution wavelet decompositions. *International Journal of Systems Science* **35,** 511–526 (2004).

144. Suontausta, J., Ruoppila, V. & Koivo, H. Modelling of non-linear systems using radial basis function networks. *IFAC Proceedings Volumes* **27,** 627–632 (1994).

145. Albano, A., Passamante, A., Hediger, T. & Farrell, M. E. Using neural nets to look for chaos. *Physica D: Nonlinear Phenomena* **58,** 1–9 (1992).

146. Bakker, R., Schouten, J. C., Giles, C. L., Takens, F. & Van den Bleek, C. M. Learning chaotic attractors by neural networks. *Neural Computation* **12,** 2355–2383 (2000).

147. Fasel, U., Kutz, J. N., Brunton, B. W. & Brunton, S. L. Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proc. Math. Phys. Eng. Sci.* **478,** 20210904 (2022).

148. Small, M. & Tse, C. K. Minimum description length neural networks for time series prediction. *Physical Review E* **66,** 066701 (2002).

149. Dong, X., Bai, Y.-L., Lu, Y. & Fan, M. An improved sparse identification of nonlinear dynamics with Akaike information criterion and group sparsity. *Nonlinear Dynamics* **111,** 1485–1510 (2023).

150. Wang, Z., Wu, B., Garikipati, K. & Huan, X. A perspective on regression and Bayesian approaches for system identification of pattern formation dynamics. *Theoretical and Applied Mechanics Letters* **10,** 188–194 (2020).

151. Wang, Z., Huan, X. & Garikipati, K. Variational system identification of the partial differential equations governing the physics of pattern-formation: Inference under varying fidelity and noise. *Computer Methods in Applied Mechanics and Engineering* **356,** 44–74 (2019).

152. Strogatz, S. H. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering* (CRC press, 2018).

153. Linderman, S., Nichols, A., Blei, D., Zimmer, M. & Paninski, L. Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in C. elegans. *BioRxiv,* 621540 (2019).

154. Askham, T. & Kutz, J. N. Variable projection methods for an optimized dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems* **17,** 380–416 (2018).

155. Baddoo, P. J., Herrmann, B., McKeon, B. J., Nathan Kutz, J. & Brunton, S. L. Physics-informed dynamic mode decomposition. *Proceedings of the Royal Society A* **479,** 20220576 (2023).

156. Nocedal, J. & Wright, S. *Numerical optimization* (Springer, 2006).

157. Gilbert, S. F. & Barresi, M. J. F. *Developmental biology* (Oxford University Press, Oxford, 2016).

158. Solnica-Krezel, L. Conserved patterns of cell movements during vertebrate gastrulation. *Current Biology* **15,** R213–R228. ISSN: 0960-9822. (2021) (2005).

159. Collinet, C. & Lecuit, T. Programmed and self-organized flow of information during morphogenesis. *Nature Reviews Molecular Cell Biology* **22,** 245–265 (2021).

160. Levy, B. *Laplace-Beltrami eigenfunctions towards an algorithm that "understands" geometry* in *IEEE International Conference on Shape Modeling and Applications 2006 (SMI'06)* (2006), 13–13.

161. Kac, M. Can one hear the shape of a drum? *The American Mathematical Monthly* **73,** 1–23 (1966).

162. Goldenfeld, N. & Woese, C. Life is physics: Evolution as a collective phenomenon far from equilibrium. *Annual Review of Condensed Matter Physics* **2,** 375–399 (2011).

163. Bhaduri, B., Yessenov, M. & Abouraddy, A. F. Anomalous refraction of optical spacetime wave packets. *Nature Photonics* **14,** 416–421 (2020).

164. Soong, T. T. & Grigoriu, M. Random vibration of mechanical and structural systems. *NASA STI/Recon Technical Report A* **93,** 14690 (1993).

165. Heydari, N., Diplas, P., Nathan Kutz, J. & Sadeghi Eshkevari, S. Modal analysis of turbulent flow near an inclined bank–longitudinal structure junction. *Journal of Hydraulic Engineering* **147,** 04020100 (2021).

166. Fortunato, D., Hale, N. & Townsend, A. The ultraspherical spectral element method. *Journal of Computational Physics* **436,** 110087 (2021).

167. Slater, J. C. & Koster, G. F. Simplified LCAO method for the periodic potential problem. *Physical Review* **94,** 1498–1524. (2021) (1954).

168. Jaynes, E. T. & Cummings, F. W. Comparison of quantum and semiclassical radiation theories with application to the beam maser. *Proceedings of the IEEE* **51,** 89–109 (1963).

169. Kolmogorov, A. N. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* **434,** 9–13 (1991).

170. Wang, X., Shih, H.-Y. & Goldenfeld, N. *Emergence of puffs, weak and strong slugs from a stochastic predator-prey model for transitional turbulence with stream-wise shear interactions* in *Bulletin of the American Physical Society* (American Physical Society, 2021).

171. Kraichnan, R. H. & Montgomery, D. Two-dimensional turbulence. *Reports on Progress in Physics* **43,** 547–619 (1980).

172. Pope, S. B. in, 182–263 (Cambridge University Press, 2000).

173. Dunkel, J. *et al.* Fluid dynamics of bacterial turbulence. *Physical Review Letters* **110,** 228102 (2013).

174. Bratanov, V., Jenko, F. & Frey, E. New class of turbulence in active fluids. *Proceedings of the National Academy of Sciences* **112,** 15048–15053 (2015).

175. Ramaswamy, R. & Jülicher, F. Activity induces traveling waves, vortices and spatiotemporal chaos in a model actomyosin layer. *Scientific Reports* **6,** 20838 (2016).

176. Alert, R., Joanny, J.-F. & Casademunt, J. Universal scaling of active nematic turbulence. *Nature Physics* **16,** 682–688. (2021) (2020).

177. Krieg, M. *et al.* Tensile forces govern germ-layer organization in zebrafish. *Nature Cell Biology* **10,** 429–436 (2008).

178. Morita, H. *et al.* The physical basis of coordinated tissue spreading in zebrafish gastrulation. *Developmental Cell* **40,** 354–366.e4. ISSN: 1534-5807 (2017).

179. Rohde, L. A. & Heisenberg, C.-P. in *International Review of Cytology* 159–192 (Academic Press, 2007). (2021).

180. Kobitski, A. Y. *et al.* An ensemble-averaged, cell density-based digital model of zebrafish embryo development derived from light-sheet microscopy data with single-cell resolution. *Scientific Reports* **5,** 8601 (2015).

181. Romeo, N., Hastewell, A., Mietke, A. & Dunkel, J. Learning developmental mode dynamics from single-cell trajectories. *eLife* **10,** e68679 (2021).

182. Stein, E. M. & Shakarchi, R. *Fourier Analysis: An Introduction* en. ISBN: 978-1-4008-3123-4 (Princeton University Press, 2011).

183. Arfken, G., Weber, H. & Harris, F. *Mathematical Methods for Physicists: A Comprehensive Guide* (Elsevier, 2013).

184. Sandberg, V. D. Tensor spherical harmonics on S2 and S3 as eigenvalue problems. *Journal of Mathematical Physics* **19,** 2441–2446. ISSN: 0022-2488. (2021) (1978).

185. Sknepnek, R. & Henkes, S. Active swarms on a sphere. *Physical Review E* **91,** 022306 (2 2015).

186. Fily, Y., Baskaran, A. & Hagan, M. F. Active Particles on Curved Surfaces. *arXiv:1601.00324* (2016).

187. Castro-Villarreal, P. & Sevilla, F. J. Active motion on curved surfaces. *Physical Review E* **97,** 052605 (5 2018).

188. Hannezo, E. & Heisenberg, C.-P. Mechanochemical feedback loops in development and disease. *Cell* **178,** 12–25 (2019).

189. Münster, S. *et al.* Attachment of the blastoderm to the vitelline envelope affects gastrulation of insects. *Nature* **568,** 395–399 (2019).

190. Etournay, R. *et al.* Interplay of cell dynamics and epithelial tension during morphogenesis of the *Drosophila* pupal wing. *eLife* **4,** e07090 (2015).

191. Hannezo, E., Dong, B., Recho, P., Joanny, J.-F. & Hayashi, S. Cortical instability drives periodic supracellular actin pattern formation in epithelial tubes. *Proceedings of the National Academy of Sciences* **112,** 8620–8625 (2015).

192. Streichan, S. J., Lefebvre, M. F., Noll, N., Wieschaus, E. F. & Shraiman, B. I. Global morphogenetic flow is accurately predicted by the spatial distribution of myosin motors. *eLife* **7** (ed Jülicher, F.) e27454 (2018).

193. Barrera, R. G., Estevez, G. A. & Giraldo, J. Vector spherical harmonics and their application to magnetostatics. *European Journal of Physics* **6,** 287–294 (1985).

194. Keller, P. J. *et al.* Fast, high-contrast imaging of animal development with scanned light sheet–based structured-illumination microscopy. *Nature Methods* **7,** 637–642 (2010).

195. Royer, L. A. *et al.* Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms. *Nature Biotechnology* **34,** 1267–1278 (2016).

196. Lenne, P. F. *et al.* Roadmap on multiscale coupling of biochemical and mechanical signals during development. *Physical Biology* (2020).

197. Balakrishnan, R. & Dandoloff, R. The Schrödinger equation as a moving curve. *Physics Letters A* **260,** 62–67 (1999).

198. Blumenfeld, R. Planar Curve Representation of Many-Body Systems and Dynamics. *Physical Review Letters* **78,** 1203 (1997).

199. Hu, D. L., Nirody, J., Scott, T. & Shelley, M. J. The mechanics of slithering locomotion. *Proceedings of the National Academy of Sciences* **106,** 10081–10085 (2009).

200. Goldstein, R. E. & Petrich, D. M. The Korteweg–de Vries hierarchy as dynamics of closed curves in the plane. *Physical Review Letters* **67,** 3203 (1991).

201. Goldschmidt, A., Kaiser, E., Dubois, J. L., Brunton, S. L. & Kutz, J. N. Bilinear dynamic mode decomposition for quantum control. *New Journal of Physics* **23,** 033035 (2021).

202. Ma, Y., Dixit, V., Innes, M. J., Guo, X. & Rackauckas, C. *A comparison of automatic differentiation and continuous sensitivity analysis for derivatives of differential equation solutions* in *2021 IEEE High Performance Extreme Computing Conference (HPEC)* (2021), 1–9.

203. Sorber, L., Van Barel, M. & De Lathauwer, L. Structured data fusion. *IEEE Journal of Selected Topics in Signal Processing* **9,** 586–600 (2015).

204. Schrödinger, E. An undulatory theory of the mechanics of atoms and molecules. *Physical Review* **28,** 1049 (1926).

205. Hardy, G. H., Littlewood, J. E. & Pólya, G. *Inequalities* (Cambridge University Press, 1952).

206. Gray, J. Undulatory propulsion. *Journal of Cell Science* **3,** 551–578 (1953).

207. Cohen, N. & Boyle, J. H. Swimming at low Reynolds number: a beginners guide to undulatory locomotion. *Contemporary Physics* **51,** 103–123 (2010).

208. Niebur, E. & Erdös, P. Theory of the locomotion of nematodes: dynamics of undulatory progression on a surface. *Biophysical Journal* **60,** 1132–1146 (1991).

209. Majmudar, T., Keaveny, E. E., Zhang, J. & Shelley, M. J. Experiments and theory of undulatory locomotion in a simple structured medium. *Journal of The Royal Society Interface* **9,** 1809–1823 (2012).

210. Kudrolli, A. & Ramirez, B. Burrowing dynamics of aquatic worms in soft sediments. *Proceedings of the National Academy of Sciences* **116,** 25569–25574 (2019).

211. Sznitman, J., Purohit, P. K., Krajacic, P., Lamitina, T. & Arratia, P. E. Material properties of Caenorhabditis elegans swimming at low Reynolds number. *Biophysical Journal* **98,** 617–626 (2010).

212. Shen, X. N. & Arratia, P. E. Undulatory Swimming in Viscoelastic Fluids. *Physical Review Letters* **106,** 208101 (20 2011).

213. Smits, A. J. Undulatory and oscillatory swimming. *Journal of Fluid Mechanics* **874,** P1 (2019).

214. Thewissen, J. G. & Fish, F. Locomotor evolution in the earliest cetaceans: functional model, modern analogues, and paleontological evidence. *Paleobiology* **23,** 482–490 (1997).

215. Maladen, R. D., Ding, Y., Umbanhowar, P. B., Kamor, A. & Goldman, D. I. Mechanical models of sandfish locomotion reveal principles of high performance subsurface sand-swimming. *Journal of The Royal Society Interface* **8,** 1332–1345 (2011).

216. Chong, B., Wang, T., Erickson, E., Bergmann, P. J. & Goldman, D. I. Coordinating tiny limbs and long bodies: Geometric mechanics of lizard terrestrial swimming. *Proceedings of the National Academy of Sciences* **119,** e2118456119 (2022).

217. Guo, Z. & Mahadevan, L. Limbless undulatory propulsion on land. *Proceedings of the National Academy of Sciences* **105,** 3179–3184 (2008).

218. Flavell, S. W., Raizen, D. M. & You, Y.-J. Behavioral states. *Genetics* **216,** 315–332 (2020).

219. Hong, W. *et al.* Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proceedings of the National Academy of Sciences* **112,** E5351–E5360 (2015).

220. Pokala, N. & Flavell, S. W. in *C. elegans* 357–373 (Springer, 2022).

221. Shipley, F. B., Clark, C. M., Alkema, M. J. & Leifer, A. M. Simultaneous optogenetic manipulation and calcium imaging in freely moving C. elegans. *Frontiers in Neural Circuits* **8,** 28 (2014).

222. Leifer, A. M., Fang-Yen, C., Gershow, M., Alkema, M. J. & Samuel, A. D. Optogenetic manipulation of neural activity in freely moving Caenorhabditis elegans. *Nature Methods* **8,** 147–152 (2011).

223. Prevedel, R. *et al.* Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nature Methods* **11,** 727–730 (2014).

224. Bozek, K., Hebert, L., Portugal, Y., Mikheyev, A. S. & Stephens, G. J. Markerless tracking of an entire honey bee colony. *Nature Communications* **12,** 1733 (2021).

225. Liu, M., Sharma, A. K., Shaevitz, J. W. & Leifer, A. M. Temporal processing and context dependency in Caenorhabditis elegans response to mechanosensation. *eLife* **7,** e36419 (2018).

226. Hebert, L., Ahamed, T., Costa, A. C., O'Shaughnessy, L. & Stephens, G. J. WormPose: Image synthesis and convolutional networks for pose estimation in C. elegans. *PLOS Computational Biology* **17,** e1008914 (2021).

227. Mathis, M. W. & Mathis, A. Deep learning tools for the measurement of animal behavior in neuroscience. *Current Opinion in Neurobiology* **60,** 1–11 (2020).

228. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* **21,** 1281–1289 (2018).

229. Stephens, G. J., Johnson-Kerner, B., Bialek, W. & Ryu, W. S. Dimensionality and dynamics in the behavior of C. elegans. *PLoS Computational Biology* **4,** e1000028 (2008).

230. Ahamed, T., Costa, A. C. & Stephens, G. J. Capturing the continuous complexity of behaviour in Caenorhabditis elegans. *Nature Physics* **17,** 275–283 (2021).

231. Hosoi, A. & Goldman, D. I. Beneath our feet: strategies for locomotion in granular media. *Annual Review of Fluid Mechanics* **47,** 431–453 (2015).

232. Ozkan-Aydin, Y., Goldman, D. I. & Bhamla, M. S. Collective dynamics in entangled worm and robot blobs. *Proceedings of the National Academy of Sciences* **118,** e2010542118 (2021).

233. Berman, G. J., Choi, D. M., Bialek, W. & Shaevitz, J. W. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface* **11,** 20140672 (2014).

234. Lauga, E. & Powers, T. R. The hydrodynamics of swimming microorganisms. *Reports on Progress in Physics* **72,** 096601 (2009).

235. Astley, H. C. *et al.* Surprising simplicities and syntheses in limbless self-propulsion in sand. *Journal of Experimental Biology* **223,** jeb103564 (2020).

236. Boyle, J. H., Berri, S. & Cohen, N. Gait modulation in C. elegans: an integrated neuromechanical model. *Frontiers in Computational Neuroscience* **6,** 10 (2012).

237. Sarma, G. P. *et al.* OpenWorm: overview and recent advances in integrative biological simulation of Caenorhabditis elegans. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373,** 20170382 (2018).

238. Marvi, H. *et al.* Sidewinding with minimal slip: Snake and robot ascent of sandy slopes. *Science* **346,** 224–229 (2014).

239. Aguilar, J. *et al.* A review on locomotion robophysics: the study of movement at the intersection of robotics, soft matter and dynamical systems. *Reports on Progress in Physics* **79,** 110001 (2016).

240. Schiebel, P. E. *et al.* Mechanical diffraction reveals the role of passive dynamics in a slithering snake. *Proceedings of the National Academy of Sciences* **116,** 4798–4803 (2019).

241. Berry, M. V. Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **392,** 45–57 (1984).

242. Wright, S., Nocedal, J., *et al.* Numerical optimization. *Springer* **35,** 7 (1999).

243. Zhuang, J. *et al.* Adabelief optimizer: Adapting stepsizes by the belief in observed gradients. *Advances in neural information processing systems* **33,** 18795–18806 (2020).

244. Ye, K. & Lim, L.-H. Schubert varieties and distances between subspaces of different dimensions. *SIAM Journal on Matrix Analysis and Applications* **37,** 1176–1197 (2016).

245. Gray, J. M., Hill, J. J. & Bargmann, C. I. A circuit for navigation in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences* **102,** 3184–3191 (2005).

246. Goldstein, R. E. & Petrich, D. M. Solitons, Euler's equation, and vortex patch dynamics. *Physical Review Letters* **69,** 555 (1992).

247. Hasimoto, H. A soliton on a vortex filament. *Journal of Fluid Mechanics* **51,** 477–485 (1972).

248. Ling, T. *et al.* High-speed interferometric imaging reveals dynamics of neuronal deformation during the action potential. *Proceedings of the National Academy of Sciences* **117,** 10278–10285 (2020).

249. Alvelid, J., Damenti, M., Sgattoni, C. & Testa, I. Event-triggered STED imaging. *Nature Methods* **19,** 1268–1275 (2022).

250. Xie, S. & Martin, A. C. Intracellular signalling and intercellular coupling coordinate heterogeneous contractile events to facilitate tissue folding. *Nature Communications* **6,** 7161 (2015).

251. Hilfinger, A., Norman, T. M., Vinnicombe, G. & Paulsson, J. Constraints on Fluctuations in Sparsely Characterized Biological Systems. *Physical Review Letters* **116,** 058101 (5 2016).

252. Stiefel, K. M. & Ermentrout, G. B. Neurons as oscillators. *Journal of Neurophysiology* **116,** 2950–2960 (2016).

253. Buzsaki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304,** 1926–1929 (2004).

254. Srinivas, N., Parkin, J., Seelig, G., Winfree, E. & Soloveichik, D. Enzyme-free nucleic acid dynamical systems. *Science* **358,** eaal2052 (2017).

255. Filatrella, G., Nielsen, A. H. & Pedersen, N. F. Analysis of a power grid using a Kuramoto-like model. *The European Physical Journal B* **61,** 485–491 (2008).

256. Rohden, M., Sorge, A., Timme, M. & Witthaut, D. Self-organized synchronization in decentralized power grids. *Physical Review Letters* **109,** 064101 (2012).

257. Marty, G., Combrié, S., Raineri, F. & De Rossi, A. Photonic crystal optical parametric oscillator. *Nature Photonics* **15,** 53. ISSN: 1749-4893 (2021).

258. Vettoretti, G., Ditlevsen, P., Jochum, M. & Rasmussen, S. O. Atmospheric CO2 control of spontaneous millennial-scale ice age climate oscillations. *Nature Geoscience* **15,** 300. ISSN: 1752-0908 (2022).

259. Izhikevich, E. M. *Dynamical Systems in Neuroscience* (MIT Press, 2007).

260. Kemeny, J. G. Mathematics without numbers. *Daedalus* **88,** 577–591 (1959).

261. Young, P. Optimal voting rules. *Journal of Economic Perspectives* **9,** 51–64 (1995).

262. Talagrand, O. & Courtier, P. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society* **113,** 1311–1328 (1987).

263. Courtier, P. & Talagrand, O. Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Quarterly Journal of the Royal Meteorological Society* **113,** 1329–1347 (1987).

264. Majda, A. J., Franzke, C. & Crommelin, D. Normal forms for reduced stochastic climate models. *Proceedings of the National Academy of Sciences* **106,** 3649–3653 (2009).

265. Raissi, M. & Karniadakis, G. E. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics* **357,** 125–141 (2018).

266. Kochkov, D. *et al.* Machine learning-accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences* **118,** e2101784118 (2021).

267. Yang, S., Wong, S. W. & Kou, S. Inference of dynamic systems from noisy and sparse data via manifold-constrained Gaussian processes. *Proceedings of the National Academy of Sciences* **118** (2021).

268. Takens, F. in *Dynamical systems and turbulence, Warwick 1980* 366–381 (Springer, 1981).

269. Sugihara, G. *et al.* Detecting causality in complex ecosystems. *Science* **338,** 496–500 (2012).

270. Haehne, H., Casadiego, J., Peinke, J. & Timme, M. Detecting hidden units and network size from perceptible dynamics. *Physical Review Letters* **122,** 158301 (2019).

271. Chen, B. *et al.* Discovering state variables hidden in experimental data. *arXiv:2112.10755* (2021).

272. Hewamalage, H., Bergmeir, C. & Bandara, K. Recurrent neural networks for time series forecasting: Current status and future directions. *Int. J. of Forecast.* **37,** 388–427 (2021).

273. Wang, W.-X., Yang, R., Lai, Y.-C., Kovanis, V. & Grebogi, C. Predicting catastrophes in nonlinear dynamical systems by compressive sensing. *Physical review letters* **106,** 154101 (2011).

274. Lu, P. Y., Ariño Bernad, J. & Soljačić, M. Discovering sparse interpretable dynamics from partial observations. *Commun. Phys.* **5,** 206 (2022).

275. Ouala, S. *et al.* Learning latent dynamics for partially observed chaotic systems. *Chaos* **30,** 103121 (2020).

276. Ayed, I., de Bézenac, E., Pajot, A., Brajard, J. & Gallinari, P. Learning dynamical systems from partial observations. *arXiv:1902.11136* (2019).

277. Cenedese, M., Axås, J., Bäuerlein, B., Avila, K. & Haller, G. Data-driven modeling and prediction of non-linearizable dynamics via spectral submanifolds. *Nature Communications* **13,** 872 (2022).

278. FitzHugh, R. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal* **1,** 445–466 (1961).

279. Fletcher, R. *Practical methods of optimization* (John Wiley & Sons, 2013).

280. Stepaniants, G., Hastewell, A. D., Skinner, D. J., Totz, J. F. & Dunkel, J. Discovering dynamics and parameters of nonlinear oscillatory and chaotic systems from partial observations. *arXiv preprint arXiv:2304.04818* (2023).

281. Lainscsek, C., Letellier, C. & Gorodnitsky, I. Global modeling of the Rössler system from the z-variable. *Physics Letters A* **314,** 409–427 (2003).

282. Somacal, A. *et al.* Uncovering differential equations from data with hidden variables. *Physical Review E* **105,** 054209 (5 2022).

283. Gouesbet, G. & Letellier, C. Global vector-field reconstruction by using a multivariate polynomial L 2 approximation on nets. *Physical Review E* **49,** 4955 (1994).

284. Gottlieb, H. Question# 38. What is the simplest jerk function that gives chaos? *American Journal of Physics* **64,** 525–525 (1996).

285. Sprott, J. Simplest dissipative chaotic flow. *Physics Letters A* **228,** 271–274 (1997).

286. Linz, S. J. Nonlinear dynamical models and jerky motion. *American Journal of Physics* **65,** 523–526 (1997).

287. Eichhorn, R., Linz, S. J. & Hänggi, P. Transformations of nonlinear dynamical systems to jerky motion and its application to minimal chaotic flows. *Physical Review E* **58,** 7151 (1998).

288. Eichhorn, R., Linz, S. J. & Hänggi, P. Classes of dynamical systems being equivalent to a jerky motion. *Applied Mathematics and Mechanics* (1999).

289. Eichhorn, R., Linz, S. J. & Hänggi, P. Simple polynomial classes of chaotic jerky dynamics. *Chaos Solitons Fractals* **13,** 1–15 (2002).

290. Letellier, C., Aguirre, L. A. & Maquet, J. Relation between observability and differential embeddings for nonlinear dynamics. *Physical Review E* **71,** 066213 (2005).

291. Mendes, E. M., Lainscsek, C. & Letellier, C. Diffeomorphical equivalence vs topological equivalence among Sprott systems. *Chaos* **31,** 083126 (2021).

292. Kaheman, K., Kutz, J. N. & Brunton, S. L. SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proc. R. Soc. A* **476,** 20200279 (2020).

293. Kunkel, P. & Mehrmann, V. *Differential-algebraic equations: analysis and numerical solution* (European Mathematical Society, 2006).

294. Paydarfar, D., Forger, D. B. & Clay, J. R. Noisy inputs and the induction of on–off switching behavior in a neuronal pacemaker. *Journal of Neurophysiology* **96,** 3338–3348 (2006).

295. Goldberger, A. L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **101,** e215–e220 (2000).

296. Tyson, J. J. Scaling and Reducing the Field-Koros-Noyes Mechanism of the Belousov-Zhabotinskii Reaction. *The Journal of Physical Chemistry* **86,** 3006–3012 (1982).

297. Morris, C. & Lecar, H. Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal* **35,** 193–213 (1981).

298. Hodgkin, A. L. & Huxley, A. F. Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo. *Journal of Physiology* **116,** 449 (1952).

299. Gerstner, W., Kistler, W. M., Naud, R. & Paninski, L. *Neuronal dynamics: From single neurons to networks and models of cognition* (Cambridge University Press, 2014).

300. Epstein, I. R. & Pojman, J. A. *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos: Oscillations, Waves, Patterns, and Chaos* (Oxford University Press, 1998).

301. Taylor, A. F. Mechanism and Phenomenology of an Oscillating Chemical Reaction. *Prog. React. Kinet. Mech.* **27,** 247–325 (2002).

302. Field, R. J., Koros, E. & Noyes, R. M. Oscillations in Chemical Systems. II. Thorough Analysis of Temporal Oscillation in the Bromate-Cerium-Malonic Acid System. *J. Am. Chem. Soc.* **94,** 8649–8664 (1972).

303. Rovinskii, A. B. & Zhabotinskii, A. M. Mechanism and Mathematical Model of the Oscillating Bromate-Ferroin-Bromomalonic Acid Reaction. *The Journal of Physical Chemistry* **88,** 6081–6084 (1984).

304. Zhabotinsky, A. M., Buchholtz, F., Kiyatkin, A. B. & Epstein, I. R. Oscillations and Waves in Metal-Ion-Catalyzed Bromate Oscillating Reactions in Highly Oxidized States. *The Journal of Physical Chemistry* **97,** 7578–7584 (1993).

305. Ren, L. *et al.* Experimental, Numerical, and Mechanistic Analysis of the Nonmonotonic Relationship between Oscillatory Frequency and Photointensity for the Photosensitive Belousov–Zhabotinsky Oscillator. *Chaos* **25,** 064607 (2015).

306. Stepaniants, G., Brunton, B. W. & Kutz, J. N. Inferring causal networks of dynamical systems through transient dynamics and perturbation. *Physical Review E* **102,** 042309 (2020).

307. Umetani, K., Singer, D. H., McCraty, R. & Atkinson, M. Twenty-Four Hour Time Domain Heart Rate Variability and Heart Rate: Relations to Age and Gender Over Nine Decades. *Journal of the American College of Cardiology* **31,** 593–601. ISSN: 0735-1097 (1998).

308. Tan, T. H. *et al.* Odd dynamics of living chiral crystals. *Nature* **607,** 287–293 (2022).