Adversarial Dynamical Systems Reveal Limits and Rules for Trustworthy Data-Driven Learning

Matthew J. Colbrook[1*], Igor Mezić[2], Alexei Stepanenko[1]

[1] DAMTP, University of Cambridge, Cambridge, CB3 0WA, UK.

[2] University of California, Santa Barbara, CA 93106, USA. AIMdyn, Inc, Santa Barbara, CA 93101.

Abstract

Modern natural and technological systems are often too complex for analytical modelling, motivating efforts to extract their fundamental dynamics from data. While promising, data-driven methods often fail to converge or generalize, prompting a central question: *When can we reliably learn system behaviors from data, and when is it impossible?* We establish the first impossibility results for problems in data-driven learning of dynamical systems, showing that adversarial systems—constructed to obstruct algorithmic learning—can occur with high probability. In such cases, no algorithm can reliably succeed, regardless of data quality. These examples are given for Koopman operators, a leading technique for transforming nonlinear dynamics into linear representations for control and analysis. By analyzing these limitations, we identify conditions under which learning is possible and introduce optimal algorithms with provable convergence and certification, overcoming key shortcomings (e.g., non-convergence) of existing methods. Our methods provide the first general convergence guarantees for Koopman operators, validated on classical oscillators, chaotic fluid flows, and Arctic sea ice forecasting. In the latter, we uncover hidden modes, deliver long-range forecasts with geographic error bounds, and outperform state-of-the-art dynamical and deep learning models—all at a fraction of the computational cost, enabling real-time deployment on standard hardware. This framework offers a unified, rigorous classification of the capabilities and limits of data-driven system learning, with broad implications for science and engineering.

Models across science often involve systems that evolve over time, known as dynamical systems. These systems have long been used to understand, predict, and control complex behaviour across physics, chemistry, biology, and medicine. Yet in many fields such as climate science, neuroscience, and robotics, systems are often too complex for direct analysis, or their governing equations are unknown. Machine learning (ML) has transformed the analysis of complex data [1], with breakthroughs in protein structure prediction [2] (see also the 2024 Nobel Prize in Chemistry [3]), image classification [4], and drug/material discovery [5]. The emerging field of *data-driven dynamical systems* seeks to combine ML with time-series data to uncover underlying structure and principles without requiring explicit models [6–16].

Current ML techniques often struggle to converge or generalize, with limited guarantees of their trustworthiness. This hinders their effectiveness in critical applications and poses a central challenge: *When can we reliably learn system behaviors from data, and when is it impossible?* We address this with:

• **Adversarial dynamical systems:** We present the first examples in data-driven dynamical systems for which no sequence of learning algorithms—probabilistic or otherwise—can solve, even with unlimited data. By carefully altering a system's behavior in a way that respects both its structure and the data, we design adversarial systems that block reliable learning. These are not rare edge cases: success is fundamentally limited to 50% and they arise in well-studied classes. Even for smooth systems on simple low-dimensional surfaces, tasks such as learning finite-dimensional linearized representations (e.g., via autoencoders) remain unsolvable.

This parallels adversarial attacks in ML, where small perturbations expose vulnerabilities and motivate robustness. Likewise, our adversarial systems reveal fundamental learning limits in broad system classes and offer principles for trustworthy methods. They may also help explain phenomena like hallucinations in large language models (LLMs).

• **Optimal algorithms with learning guarantees:** Building on insights from these adversarial systems (see challenges **(C1)** and **(C2)** below), we develop provably optimal algorithms with guaranteed convergence and error bounds under broad conditions. Unlike traditional approaches that rely on sequences of algorithms, our methods reach fundamental limits and enable reliable extrapolation, crucial for trustworthy AI. Our models are also trained on CPUs at a fraction of the cost of deep learning approaches, far exceeding the scales at which recent claims of efficiency have been made [17].

• **A universal framework for classification:** We establish a rigorous yet practical mathematical foundation that clarifies when and why learning succeeds or fails. Matching lower and upper bounds reveal the core challenges and offer a comprehensive classification of problem complexity.

These results offer a unified, rigorous, and practical basis for understanding when data-driven models can or cannot succeed, advancing the broader goal of trustworthy ML. They are applicable across fields (see Discussion)—from climate science and neuroscience to engineering and control—where reliable prediction and mechanistic insight are essential.

We apply this general framework to Koopman operators, a major research focus that addresses nonlinearity by acting on an *infinite-dimensional* space of measurements rather than the system's state. Introduced nearly a century ago by Koopman and von Neumann [18, 19], Koopman operators now play a central role in data-driven dynamical systems [7, 12, 20–22]. Their spectral properties (e.g., eigenfunctions and eigenvalues) decompose complex behavior into simpler components like trends, oscillations, or decay, allowing for the use of linear methods in prediction, estimation, and control. This enables explainable, robust, and cost-efficient ML. Notable successes include robot control [23], climate analysis [24], neural network training [25], disease modeling [26], brain analysis [27], non-autonomous systems [28], and interpretable AI [29].

However, Koopman theory faces major practical challenges. Spectral approximation in infinite-dimensional spaces is often non-convergent—even with perfect data [11, 22, 30]. The most widely used method, Dynamic Mode Decomposi-
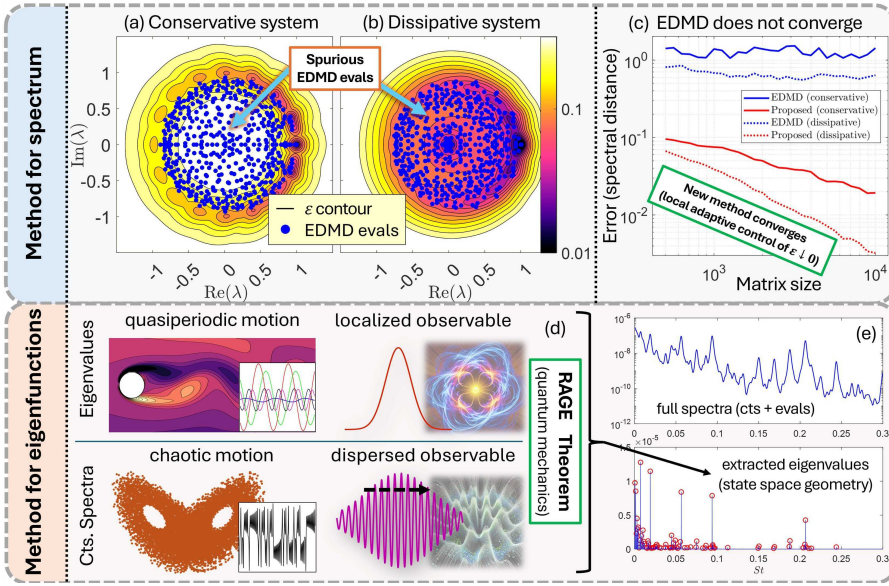
Figure 1: Convergent general methods for Koopman learning. *Top (Example 1.1):* We test our new method (Algorithm 1 in S.I.) on the classical Duffing oscillator in both conservative (panel (a)) and dissipative (panel (b)) regimes. The state-of-the-art EDMD method fails to converge (panel (c)), producing spurious eigenvalues (blue markers). In contrast, our approach converges reliably by adaptively computing observables (scalar functions of the state) ϕ_ε and scalars λ (that approximate the spectrum) that satisfy the temporal coherence relation $\phi_\varepsilon(x_n) = \lambda^n \phi_\varepsilon(x_0) + \mathcal{O}(n\varepsilon)$ (here x_n is the state of the system at time n). The contours of ε show where approximate, near-eigenvalue behavior occurs—providing a more robust picture of dynamics. The tolerance ε is locally adjusted based on the amount of data and the number of observables used. This also allows error bounds to be computed. The error shown is the maximum value of ε across the outputs (averaged over 10 independent runs due to randomly sampled trajectories). This quantity measures how closely the computed spectrum approximates the true spectrum. It is estimated using a method that converges uniformly from above as the matrix size increases, thereby providing rigorous error bounds. *Bottom (Example 1.2):* The new method (Algorithm 6 in S.I.) is used to analyze high Reynolds number cavity flow, extracting eigenvalues of the Koopman operator embedded in continuous spectra. This separation is achieved by tracking which observables are localized or dispersed in time (panel (d)). These behaviors are reflected in the spectral distribution and extracted eigenvalues (panel (e)).

tion (DMD) [31], and its variants, including extended DMD (EDMD) [32], produce spurious eigenvalues and often fail to converge (Figure 1, top). In particular, Koopman operators are generally non-self-adjoint (non-Hermitian) and often have a continuous rather than discrete spectrum, so much of the classical toolkit for computing spectra [33, 34] does not apply. While recent advances [14, 16, 35, 36] address specific spectral properties under constrained conditions, they require multiple limiting processes, lacking a unified approach for convergence in all cases. Hence, critical evaluation is essential for theoretical insight and reliable application.

Koopman theory thus provides an ideal setting to explore our central question. We overcome its limitations by introducing the first general-purpose, convergent methods for learning Koopman spectral properties (Algorithms 1–5 in the S.I., see top panel of Figure 1), which avoid spurious eigenvalues by explicitly locally minimizing spectral errors. We further adapt tools from quantum mechanics to disentangle spectral components with distinct physical signatures (Algorithms 6–7 in the S.I., see bottom panel of Figure 1). These methods perform well across both low- and high-dimensional systems—including challenging cases where the system's behavior spans a continuous range of frequencies, rather than distinct repeating patterns. This includes practical applications such as forecasting Arctic sea ice (Figures 2 to 6).

This final example is motivated by Arctic amplification, where near-surface Arctic temperatures are rising faster than the global average [37]. Sea ice loss has major impacts on polar bear habitats, Indigenous communities, shipping, and the Atlantic Meridional Overturning Circulation (AMOC). Concurrently, extreme weather events (e.g., wildfires, floods, heatwaves, and severe winters) have intensified in recent decades, affecting billions worldwide. The link between Arctic sea ice loss and Northern Hemisphere extreme weather remains an active area of research and debate [38–40]. While regional effects appear likely, identifying geographically significant regions and their influence is particularly challenging. Forecasting Arctic sea ice beyond two months remains a major challenge [41]

Our algorithms uncover a family of hidden Koopman modes linked to sea ice decline and identify the associated geographic regions with error bounds. These patterns support accurate long-term predictions and reveal how different parts of the system are connected. By building data-driven decompositions from these and other validated modes, we achieve state-of-the-art Arctic sea ice forecasts (Figures 5 and 6). Notably, such modes can influence tipping behavior, including patterns related to the AMOC [42, 43].

The adversarial dynamics we construct may also shed light on hallucinations in LLMs. These models generate trajectories (sentences) over a state space (words) via one-step-ahead prediction, a process linked to Koopman operators [25, 44, 45]. The adversarial systems we construct that impede learning often have Koopman operators with a continuous spread of frequencies, characteristic of chaotic dynamics. This enables short-term accuracy but causes long-term unpredictability due to sensitivity to initial conditions, mirroring how small prompt changes in LLMs can cause divergent outputs. Our results may therefore help explain inherent limitations of autoregressive architectures in maintaining long-term accuracy.

Multiple limits: Bad, good, or insufficient data?

Learning systems from data requires two elements: quantities we can measure using sensors and their corresponding time series data. In ML, one often studies sequences of algorithms indexed by n, where n might represent the size of the dataset or the complexity of the model (for example, the width or depth of a neural network). A simple illustration is estimating the average energy E of an oscillating system based on measurements of its instantaneous energy $e(n)$ at each time step n. If the system is ergodic—meaning that averages along a single long-term trajectory reflect the behavior of the entire system—then E can be estimated iteratively by:

$$\Gamma_{n+1} = \frac{e(1) + \cdots + e(n+1)}{n+1} = \frac{n}{n+1}\Gamma_n + \frac{1}{n+1}e(n+1). \quad (1)$$

This running average Γ_n becomes more accurate as more data is collected. A classical result known as Birkhoff's ergodic theorem [46] guarantees that Γ_n converges to the true average energy E as $n \to \infty$, i.e., as we collect more data.

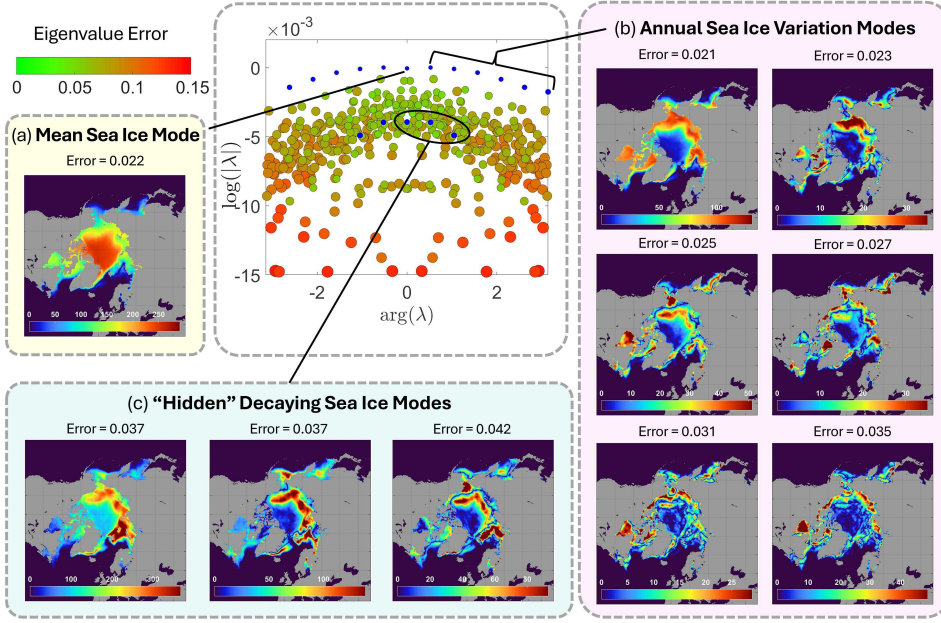Learning often assumes that increasing the amount of data

Figure 2: Verified eigenvalues and Koopman modes of Arctic sea ice (1979–2021) and detection of new hidden modes for long-time forecasts. For each EDMD eigenvalue, our proposed method computes an associated error bound (colorbar). The displayed size of each eigenvalue is proportional to this error. While many EDMD eigenvalues are spurious, we identify 17 reliable ones (shown in blue) with small associated errors. The Koopman modes (the $\mathbf{g}_j$ in Equation (6), where we plot their absolute value) are categorized into three groups: (a) the mean sea ice concentration, corresponding to a stationary mode with eigenvalue $\lambda = 1$; (b) annual variation in sea ice concentration, with $\lambda = \exp(m\pi i/6)$ for $m = \pm 1, \pm 2, \ldots, \pm 11, 12$, representing periodic variation across the months; and (c) "hidden" decaying modes with $|\lambda| < 1$ (revealed by our error bounds), representing long-term sea ice loss. The spatial structure of each mode indicates the geographic regions where these behaviors occur, and the corresponding Koopman eigenfunction time series shows a clear trend over the analysis interval. The hidden modes with nonzero $\arg(\lambda)$ can be interpreted as seasonal patterns (group (b)) modulated by the decaying mode with zero argument, providing insight into evolving Arctic sea ice dynamics.
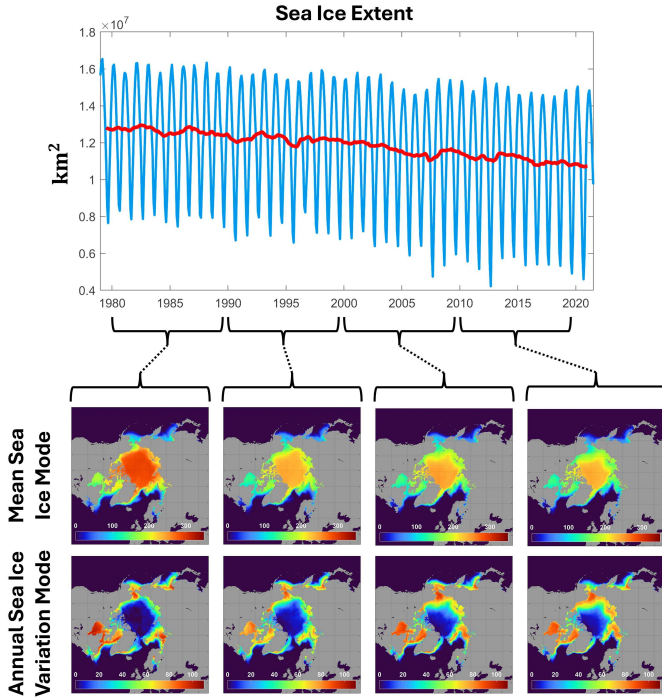

Figure 3: Sea ice decay and time-windowed Koopman modes. Top: Sea ice extent over the past several decades. The red curve shows the moving 12-month mean. Bottom: The absolute value of Koopman modes corresponding to the mean sea ice and annual variations over ten year windows. The maximum relative error of all of these modes is 0.048. Mean and annual modes reveal declining sea ice, reduced winter extent, and amplified seasonal contrast in marginal seas, consistent with a slow decaying mode (Figure 2).

Figure 4: Forecast error for sea ice concentration at every grid point. Relative error of forecasted anomalies (relative to the periodic monthly average). We consider ten-year forecasts initialized at each month from 2005 to 2010 and plot the average error for each lead time. The proposed method consistently outperforms DMD.
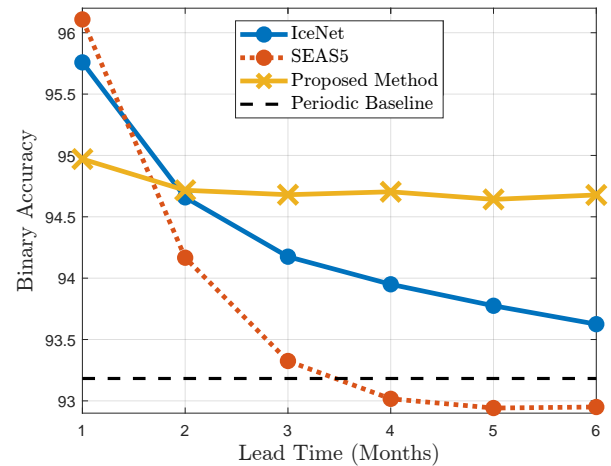
Figure 5: Comparison with machine learning and statistical benchmarks. Mean binary accuracy (see text for definition) over the test years 2012–2020, shown for IceNet, SEAS5, and our proposed method that avoids spurious Koopman eigenvalues. Our proposed method achieves better accuracy for lead times greater than one month, with very little increase in errors at larger lead times. Moreover, this is achieved using orders of magnitude fewer trainable parameters and substantially less computational cost.

or model complexity will improve performance. However, our results show that for many key problems in dynamical systems, this assumption is false. In fact, no algorithm Γ_n can succeed by taking a single limit, regardless of how that limit is defined (e.g., sample size, model complexity, simulation length).

Instead, some problems only become solvable when multiple data limits are applied in a specific order. For example, computing long-term structures in a system, such as ergodic partitions [47], requires not just averaging over time (as in Equation (1)) but also subsequently averaging across a growing number of measured quantities (instead of just the energy
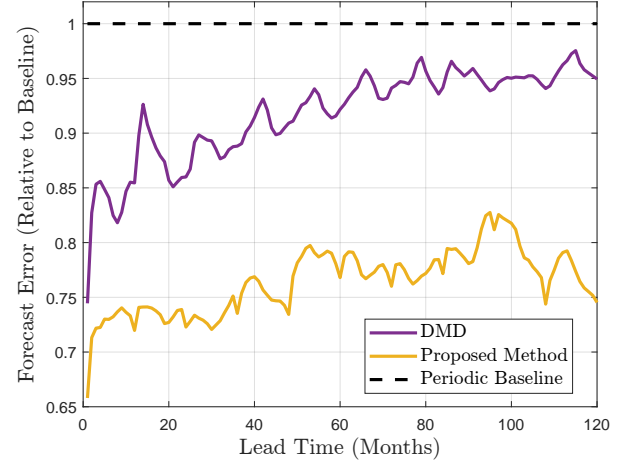
Figure 6: Binary accuracy across months (same experiment as Figure 5). Left: The mean accuracy of our proposed method over all lead times and years. Right: The improvements over IceNet and SEAS5.
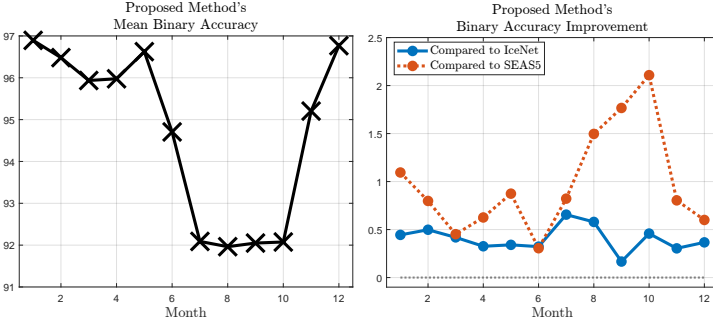
e). These two types of data increase, longer observations and richer measurements, typically cannot be combined or chosen adaptively from the data itself. This leads to a hierarchy of difficulty, where each level corresponds to the number and type of data limits needed for reliable learning.

Dynamical setup

A dynamical system describes how a system changes over time. Suppose we track a collection of variables, denoted x, that represent the system's state at a given moment. The set of all possible states is called the state space, written as $\mathcal{X}$, so that $x \in \mathcal{X}$. The system evolves in discrete time steps, meaning the next state x_{n+1} depends on the current state x_n via a rule:

$$x_{n+1} = F(x_n), \qquad n = 0, 1, 2, \ldots. \tag{2}$$

Here, $F : \mathcal{X} \to \mathcal{X}$ is an *unknown* function that may come from sampling a continuous-time process.

To study a system, we measure its physical properties like temperature or velocity, represented by functions $g : \mathcal{X} \to \mathbb{C}$, called "observables". One should think of $g(x)$ as a quantity we can measure from state x. The *Koopman operator*, $\mathcal{K}_F$ (or $\mathcal{K}$), captures how these measurements evolve. It acts on an observable g by composing it with the system's evolution:

$$[\mathcal{K}g](x) = [g \circ F](x) = g(F(x)), \quad x \in \mathcal{X}. \tag{3}$$

This means $\mathcal{K}g$ gives the value of the observable one step into the future: $[\mathcal{K}g](x_n) = g(x_{n+1})$. The key property of $\mathcal{K}$ is *linearity*: for any observables f and g and scalars α and β, $\mathcal{K}(\alpha f + \beta g) = \alpha \mathcal{K}(f) + \beta \mathcal{K}(g)$. Linearity is powerful since it allows us to analyze the system through spectral properties of $\mathcal{K}$ (e.g., eigenvalues and eigenfunctions). The trade-off for this global linearization is that $\mathcal{K}$ acts on an *infinite-dimensional* space of observables. One can think of $\mathcal{K}$ as an infinite matrix, corresponding to the infinite number of observables g.

The goal is to learn the spectral properties of the Koopman operator from *snapshot data* – discrete sample pairs of the system's behavior:

$$\left\{ \left(x^{(m)}, y^{(m)} = F(x^{(m)}) \right) : m = 1, \ldots, M \right\}. \tag{4}$$

Here, each $y^{(m)}$ represents the state one time step ahead of $x^{(m)}$. Such data can arise from experiments or simulations, and observations of either long or short trajectories. We shall see examples of each of these throughout the paper.

Nonlinear separation of variables and spectra

Spectral properties of Koopman operators contain valuable information about the system. For example, a complex number $\lambda \in \mathbb{C}$ is called an almost eigenvalue if, for a tolerance $\varepsilon > 0$, there exists a normalized observable ϕ_ε with $\|\mathcal{K}\phi_\varepsilon - \lambda\phi_\varepsilon\| \leq \varepsilon$. (Here, $\|\cdot\|$ measures an observable's energy.) These observables, called *approximate eigenfunctions*, are physically relevant because they exhibit approximate temporal coherence [12, 48]:

$$\phi_\varepsilon(x_n) = \lambda^n \phi_\varepsilon(x_0) + \mathcal{O}(n\varepsilon) \quad \text{for } n = 1, 2, \ldots. \tag{5}$$

That is, λ approximately describes the oscillatory behavior and decay (or growth) of the measurement $\phi_\varepsilon(x)$ over time through its powers λ^n. Smaller values of ε correspond to longer timescales where this approximation is valid. For $\varepsilon = 0$, ϕ_ε becomes an exact eigenfunction of $\mathcal{K}$ with eigenvalue λ.

Approximate eigenfunctions also encode key dynamical features of the system, such as the global stability of equilibria. The contours (or level sets) of these functions highlight key structures in the system's dynamics, such as regions that behave independently over long times, pathways along which the system evolves, and surfaces where states settle into long-term behavior at the same rate [20, 49]. The *approximate point spectrum* of $\mathcal{K}$, denoted $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K})$, is the set of all scalars λ for which ε can be made arbitrarily small, and forms the most fundamental spectral property of Koopman operators.

Just as a matrix is diagonalized by its eigenvalues and eigenvectors, a nonlinear system can be "diagonalized" by its Koopman spectra. Koopman eigenfunctions act as fundamental components of behavior, revealing persistent patterns (like oscillations or trends) in the system. For a vector of observations $\mathbf{g} \in \mathbb{C}^N$, the Koopman mode associated with an eigenvalue is the projection of $\mathbf{g}$ onto the corresponding eigenspace. Under suitable conditions, this yields a spectral expansion for the time evolution of the observables [7]:

$$\mathbf{g}(x_n) = \sum_{j=1}^{\infty} \lambda_j^n \phi_j(x_0) \mathbf{g}_j. \tag{6}$$

Here, $\mathbf{g}_j \in \mathbb{C}^N$ is the jth Koopman mode, associated with eigenvalue λ_j and eigenfunction ϕ_j, which one may think of as an expansion coefficient. This decomposition is conceptually and operationally similar to separation of variables: the eigenvalues describe time evolution through the powers λ_j^n, while the Koopman modes capture how each pattern is expressed in space and the regions where this dynamical behaviour occurs.

When the observable $\mathbf{g}$ is real-valued, the eigenvalues, eigenfunctions, and Koopman modes in Equation (6) appear in complex-conjugate pairs. To illustrate how this affects time evolution, consider one such pair: $\lambda_1^n \phi_1(x_0)\mathbf{g}_1 + \overline{\lambda_1}^n \overline{\phi_1(x_0)} \overline{\mathbf{g}_1}$. Writing $\lambda_1 = re^{i\theta}$ and $\phi_1(x_0)\mathbf{g}_1 = \mathbf{R}e^{i\mathbf{\Theta}}$, the pair becomes

$$\lambda_1^n \phi_1(x_0)\mathbf{g}_1 + \overline{\lambda_1}^n \overline{\phi_1(x_0)} \overline{\mathbf{g}_1} = 2\mathbf{R}r^n \cos(n\theta + \mathbf{\Theta}).$$

This expression reveals key dynamical features: r^n controls exponential growth ($r > 1$), decay ($r < 1$), or neutral evolution ($r = 1$); the angle θ sets the oscillation frequency; and the Koopman mode (e.g., via $\mathbf{R}$) encodes spatial structure.

Methods for learning $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K})$ and spectral expansions from data face significant challenges, including spurious eigenvalues (Figure 1), missing critical spectral components, and numerical instabilities. Recent advances have helped mitigate some of these issues [36], but accurately computing $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K})$ from trajectory data remains a fundamental open problem.

There is a critical need to develop a deeper theoretical

understanding of the conditions under which such spectral computations are feasible. Equally important is identifying scenarios where these computations are fundamentally impossible. We shall see that addressing these questions guides the development of more robust and reliable methods.

Results

We first present three illustrative examples that demonstrate the reliability, advantages, and broad applicability of our general-purpose learning algorithms: (i) a classic oscillator where previous methods diverge but ours converges, (ii) a fluid flow with continuous spectrum where we extract meaningful eigenvalues, and (iii) an Arctic sea ice dataset where we uncover physically interpretable modes with error bounds and achieve state-of-the-art forecasts. We then examine the underlying theoretical foundations, showing that our algorithms optimally match fundamental limits revealed by adversarial dynamical systems (Figure 7). Indeed, analyzing the reasons behind these fundamental limitations enabled us to pinpoint the essential algorithmic properties required for reliable convergence. Finally, we provide a unified classification of the complexity of these learning problems, clarifying precisely when learning is possible and when it is fundamentally impossible.

Example 1.1 (**Overcoming lack of convergence in current methods**). We begin with a simple low-dimensional system. Figure 1 (top) examines the classical Duffing oscillator in two distinct regimes: the undamped case (conservative, panel (a)) and the damped case (dissipative, panel (b)).

EDMD, widely regarded as the state-of-the-art method, constructs a finite $N \times N$ matrix approximation of the Koopman operator by projecting onto N trial functions (a 'dictionary' of observables). We use a common and effective choice of dictionary: radial basis functions centered via k-means clustering [32]. In contrast, our algorithm searches for approximate eigenfunctions where ε in Equation (5) is locally minimized, with guaranteed convergence to the spectrum as N increases. The number of observables N grows with the number of snapshots M. Details of the experimental setup are in the Methods.

Panel (c) shows that EDMD fails to converge as the matrix size N increases, producing many spurious eigenvalues This is due to the truncation from infinite dimensions to a finite-dimensional subspace of observables. In contrast, our algorithm, using the same data and observables, converges reliably, demonstrating its robustness across different dynamical regimes. We can control the truncation error to a finite set of observables and validate whether the dictionary is a good choice, which is crucial in practice since only a finite number of observables can be used in a computation. Figure 9 shows the same behavior over a range of examples.

Example 1.2 (**Eigenvalue extraction in a high-dimensional chaotic flow**). Figure 1 (bottom) illustrates the application of our algorithms to a high Reynolds number cavity flow, which has eigenvalues embedded within continuous spectra. Hence, there is an integral over continuous spectra in addition to the sum in Equation (6) [7, 50]. In fluid flows, discrete Koopman eigenvalues typically signal quasiperiodic motion, while a broadband spectrum indicates chaotic dynamics [51]. The resulting structure of the spectrum reveals the geometry of the attractor in state space. In systems like this flow with discrete and continuous spectra, the attractors are

known as skew-periodic [52]: one part oscillates with a fixed period, whilst the other evolves chaotically but is influenced by the phase of the periodic part.

The new algorithm (see Methods) successfully extracts eigenvalues and separates spectral components using a two-limit approach: adaptively increasing a time lag for autocorrelations to detect localized observables and expanding projections onto increasing finite-dimensional subspaces. This is grounded in linking localized quasiperiodic dynamics and the corresponding eigenvalues (the so-called RAGE theorem [53], see Equation (10)). This foundation ensures that the separation of the continuous spectra and eigenvalues is both mathematically rigorous and practically optimal (see Figure 8).

Physically, the evolution of Koopman spectra with increasing Reynolds number (Figure 10) illustrates a transition to chaos consistent with Ruelle–Takens theory [54], in which a chaotic state arises after one or two bifurcations (sudden qualitative changes in the system's behaviour) from a stable steady flow. The Koopman spectrum offers a powerful tool for analyzing bifurcations, quantifying energy in both quasiperiodic and continuous components, and links to the geometry of the state space and the flow domain (see S.I. Figs. 6–8).

Example 1.3 (**Arctic sea ice: Physical modes and state-of-the-art forecasting**). We consider monthly Arctic sea ice concentration satellite data from 1979-2021. The data is formed on a 432×432 grid of 625 km^2 cells. Details on data collection and processing are provided in the Methods. As discussed in the introduction, Arctic sea ice is a critical component of the climate system, notorious for its complex dynamics. In this setting, Koopman eigenvalues reveal the time scales of changes in sea ice cover, while the corresponding modes indicate the spatial pattern of these changes. Notably, Koopman methods require no model—measurements of the concentration suffice to compute the eigenvalues and modes.

Koopman spectra and modes: We first analyze the entire data set and approximate Koopman eigenvalues using EDMD (Figure 2). Similar to Figure 1, our method also provides error bounds for eigenpairs, shown in the plot. While many EDMD eigenvalues are spurious, several (shown in blue) have small errors and correspond to key dynamics: the annual mean sea ice ($\lambda = 1$); yearly growth-melt cycle captured by a fundamental monthly mode ($\lambda \approx \exp(\pi i/6)$); and decaying modes ($|\lambda| < 1$). The small errors of the hidden modes indicate strong coherency and forecasting power (see Equation (5)).

The decaying modes, "hidden" behind a sea of spurious modes but nonetheless revealed by our error bounds, are particularly interesting as they reflect dissipative dynamics. The corresponding Koopman eigenfunction time series shows a clear trend over the 511 month analysis interval (see S.I. Fig. 9), providing strong evidence of a connection to sea ice loss from climate change. Hidden modes with nonzero $\arg(\lambda)$ can be interpreted as seasonal patterns modulated by this long-term trend. In other words, a seasonal oscillation that is gradually decaying over decades – our modes capture both the oscillation and the long-term decay. If ϕ_{decay} is the decaying eigenfunction with zero complex argument and ϕ_{var} an annual variation mode with argument $\pi/6$, then other decaying eigenfunctions are approximately given by $\phi_{\text{decay}} \times [\phi_{\text{var}}]^j$. This structure arises from the multiplicative property of Koopman eigenfunctions: if $\mathcal{K}g_1 = \lambda_1 g_1$ and $\mathcal{K}g_2 = \lambda_2 g_2$, then $\mathcal{K}(g_1 g_2) = \lambda_1 \lambda_2 g_1 g_2$, which follows from Equation (3). For

example, the canonical correlations between these decaying modes and the underlying modulated monthly dynamics are 0.9807 and 0.9798 for modes with arguments $\pi/6$ and $\pi/3$, respectively. In the climate science literature, such seasonally modulated product modes are sometimes referred to as "combination modes" [24, 55]. The mean decay time, $-1/\log(|\lambda|)$, for these modes is 233 months, consistent with decay times observed in the Antarctic region [56]. Notably, this decay rate was not observed in the Arctic region in [56], which we attribute to the challenges of extracting reliable eigenpairs in the absence of error bounds. Several studies indicate nonlinear trends in the decline of sea ice [57], which further advocates the use of Koopman operator techniques to disentangle the complex nonlinear dynamics.

We compute Koopman modes ($\mathbf{g}_j$ in Equation (6) where $\mathbf{g}$ is the vector of sea ice concentrations), which capture spatial and temporal patterns in sea ice concentration not easily discernible by conventional methods. Although the modes share units with the input data, they may take values above 100% due to the non-orthogonal expansion in Equation (6). Each mode highlights geographic regions where sea ice exhibits oscillatory, growing, or decaying behavior, as indicated by its corresponding eigenvalue. For example, modes with one-year oscillations reflect the seasonal cycle, eigenvalues near one capture the mean state, and multi-year or slowly varying modes indicate long-term trends, with the associated modes highlighting the regions where these changes occur.

The hidden decaying mode is concentrated in the Barents and Kara Seas, implying that the decrease in sea ice concentration over these time scales is localized to those regions. Indeed, there are links between sea ice reduction in these regions and extreme weather, such as severe winters in central Eurasia [58]. Importantly, our method provides error estimates, giving confidence that these identified modes are real and not artifacts (unlike many spurious modes given by EDMD).

These modes also capture seasonally-modulated reemergence of correlations (see S.I. Fig. 10), revealing 'memory' in the climate system, whereby sea ice anomalies occurring during the growth season reemerge in the following melt season despite a loss of correlation in the intervening winter months [59]. For studies of this phenomenon using kernel methods, see [60, 61]. Such long-lived modes and their geographic foci are critical, as they could influence large-scale climate patterns (e.g., ocean circulation changes) and have practical implications for shipping routes and climate resilience. Further work is needed to perform data-driven prediction of changes due to tipping points such as the greater mixing between the Barents Sea and North Atlantic [62]. Modes in sea ice can significantly influence the onset or prevention of tipping behavior of AMOC patterns [42, 43].

Sea ice decay is commonly assessed by sea ice extent, the area covered by grid cells with sea ice concentration exceeding 15%, shown in Figure 3. We compute the Koopman modes corresponding to the mean sea ice concentration and the annual variation over ten-year periods (1980–1989, 1990–1999, 2000–2009, 2010–2019) and plot their absolute values. The mean mode reveals a clear decline in overall sea ice and reduced winter extent. Annual modes show a geographic shift in seasonal variability, with marginal seas—particularly the Beaufort, Kara, and nearby coastal regions—exhibiting increased amplitude, indicating stronger seasonal contrast in these re-

gions. The overall decline in concentration suggests a slow decaying mode, consistent with the spectrum in Figure 2.

Forecasting: We now address the problem of forecasting, focusing first on the challenging task of reconstructing sea ice concentration at every grid point. Forecasts are initialized monthly from 2005 to 2010, each with a 10-year horizon. Training data consists of observations from 1979 up to the month preceding initialization, yielding 72 distinct forecast trajectories. To obtain accurate forecasts, we truncate the Koopman mode decomposition using spectral approximations and error bounds computed by our algorithm (Equation (11) in Methods). We compare our method to EDMD with delay embedding, which has previously shown strong performance in Arctic sea ice forecasting [56], and also benchmark against the monthly climatology at each grid point, referred to as the periodic baseline. We measure the forecast error relative to the periodic baseline at each grid point (details in Methods) and average over the trajectories. Figure 4 shows the results. Our approach outperforms EDMD, particularly at long lead times, consistent with its explicit minimization of ε in Equation (3).

We next consider a binary classification problem and compare our approach to IceNet [63], a deep learning model that provides state-of-the-art six-month sea ice forecasts. A point is classified as open water if the sea ice concentration is below 15%, the standard threshold for defining the ice edge. Following the setup in [63], we evaluate monthly forecasts from 2012 to 2020, with lead times from 1 to 6 months (chosen to align with the lead times for which IceNet produces forecasts). Binary accuracy is defined as the percentage of predicted classes matching the observations. We also compare against SEAS546 [64], a leading dynamical model from the European Centre for Medium-Range Weather Forecasts. Figure 5 shows the average error across lead times. Beyond one-month forecasts, the proposed verified Koopman model consistently outperforms both benchmarks, with significantly greater accuracy as lead time increases. This robustness arises from the method's direct minimization of ε, yielding longer coherency and predictive time scales (see Equation (5)).

This improvement is achieved at a fraction of the computational cost and with significantly fewer parameters than deep learning approaches. IceNet uses 4.4×10^7 trainable weights and requires over a day to train on a Nvidia Quadro P4000 GPU, while the Koopman approach relies on an observable space of at most 510 dimensions (260100 parameters) and is trained on a laptop in under a second.

Interestingly, we found no clear link between prediction accuracy and El Niño events. Instead, Figure 6 shows average accuracy by calendar month and the improvement offered by our method. Accuracy drops noticeably during the summer, reflecting the well-known "spring predictability barrier," which affects all forecasting models due to the influence of melt-season ice thickness. These months also show the greatest performance gains over SEAS5. Finally, we note that even as $\varepsilon \downarrow 0$ with more data, predictability remains fundamentally limited by atmospheric chaos and observational noise [65].

These practical advantages motivate a return to our central question: *When can we reliably learn system behaviors from data, and when is it impossible?* To address this, we present our theoretical results, with full proofs provided in the S.I.

Adversarial systems reveal challenges

We first construct adversarial dynamical systems that expose two fundamental challenges in computing Koopman spectra. These show *lower bounds* on problem complexity. Even with infinite data, no algorithm can guarantee learning certain system behaviors, a first example of inherent limits in data-driven system learning.

While it is sometimes possible to design algorithms that work for a single known system, this does not reflect the goals of general-purpose learning or complexity classification. Instead, our focus is on identifying the minimal assumptions under which an algorithm can reliably learn across a broad class of systems, a core aim in both computational complexity and ML. To illustrate, let $\Omega_{\mathbb{D}}$ be the class of systems that are continuous, measure-preserving[1], and invertible on the unit disk in two dimensions. Let $\Omega_{[0,1]}$ be the class of smooth, invertible F on the interval $[0,1]$ with uniform bounds on their derivatives (not necessarily measure-preserving).

Our results (Theorems 2.3 and 2.5 of S.I.) show that for either of these classes—denoted collectively as Ω—no deterministic algorithms Γ_n exist that, using snapshot data, converge to $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ for all F in the class Ω as $n \to \infty$. Furthermore, for any probabilistic learning algorithms, the probability of convergence cannot exceed 50%. These impossibility results are *universal*, applying to any type of algorithm and regardless of what n represents. Hence, for *any* algorithm, simply increasing the number of data points $M \to \infty$ will not lead to convergence, as this would correspond to an instance of the sequence Γ_n. These constructive results reveal fundamental challenges that occur across data-driven dynamical systems:

(C1): For systems in $\Omega_{\mathbb{D}}$, the challenge lies in determining when enough data has been collected to approximate the action of $\mathcal{K}$ on a given observable, e.g., by approximating the averages in Equation (1). The convergence rate is problem dependent [68]: no universal rate exists [69, page 14].

(C2): For systems in $\Omega_{[0,1]}$, the difficulty stems from the non-normality of $\mathcal{K}$ (non-orthogonality of its eigenfunctions). This is a well-known challenge in spectral approximation more generally [70], and manifests itself in the Koopman context as the difficulty in distinguishing data corresponding to transient dynamics from post-transient dynamics.

Moreover, these challenges:

- cover randomized algorithms, e.g., random trajectory sampling, or training with probability distributions over data, as in stochastic gradient descent and other ML methods;

- hold whatever the distribution of data;

- hold for any type of computer, e.g., digital computation (Turing machines) or exact arithmetic (BSS machines);

- hold even if we consider smoother F and allow our algorithms to sample the derivatives of F as well as Equation (4).

The system classes for which we construct adversaries include widely studied examples such as Hamiltonian flows and smooth interval exchange maps. The mechanisms can also be embedded in higher dimensions and other state spaces: the impossibility result holds for any class of systems satisfying either of **(C1)** and **(C2)**. These mechanisms are not limited to Koopman spectral estimation and reflect general challenges in data-driven dynamical systems. Failures are neither rare nor pathological, and no algorithm can succeed with probability more than 50%, highlighting fundamental limits of black-box spectral learning—even with large datasets or randomness.

A universal algorithm with learning guarantees

Learning from the above results, we now show that Koopman spectra can be computed from trajectory data, provided two key conditions are satisfied. Together with the other algorithms discussed below, these show *upper bounds* on problem complexity. This result also resolves the fundamental open problem of data-driven computation of $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$.

To address challenge **(C1)**, we assume the system is measure-preserving—this can be relaxed in many dissipative cases. To address **(C2)**, we require some control over the smoothness of the dynamical map F. Specifically, we assume F has a known modulus of continuity α. This function controls the distance between $F(x)$ and $F(y)$ by the distance between states x and y. Although such a function always exists, the impossibility result above shows that without knowledge of α, one cannot compute Koopman spectra in a single limit.

Let $\Omega_{\mathcal{X}}^{\alpha,m}$ be the class of systems satisfying both conditions. For such systems, we have developed deterministic learning algorithms Γ_n that reliably approximate the system's dynamics using snapshot data (Theorem 2.1 and Algorithm 1 of the S.I.). These algorithms converge to $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ for all F in the class $\Omega_{\mathcal{X}}^{\alpha,m}$ as $n \to \infty$. They also provide explicit error bounds that verify the accuracy of the approximation.

Our analysis of **(C1)** and **(C2)** directly leads to this provably convergent algorithm, in contrast to EDMD, which does not converge. The core idea of the algorithm (see Methods) is to use the modulus of continuity to adaptively select the dictionary size N based on the available data M in an averaging procedure similar to Equation (1). Unlike EDMD, however, we do not compute eigenvalues of a finite matrix. Instead, we use this adaptive strategy to compute a correlation matrix that allows us to compute the error metric shown in Figure 1. We search for local minimizers of this function and exploit the measure-preserving property of the system to relate this metric to the distance between a point $z \in \mathbb{C}$ and the spectrum. This distance is computed alongside an associated approximate eigenfunction satisfying the coherency condition Equation (5) with ε equal to this distance.

Example 1.1 demonstrated the convergence of this algorithm applied to the Duffing oscillator, while Figure 9 confirms its performance on other systems. Moreover, the error bounds achieved by the algorithm enabled improved sea ice forecasts in Example 1.3 (e.g., Figure 5).

To infinity and beyond

Surprisingly, spectral properties can still be learned in the presence of **(C1)** and **(C2)** by adjusting the approach. Instead of requiring a single data limit as $n \to \infty$, we consider separate *successive* limits for key parameters. Each is tied to a different type of data—such as increasing dataset size, measurement resolution, or dictionary complexity.

For instance, for EDMD (see Methods) without the modulus of continuity, one must first take the number of data sam-

---

[1] A system is measure-preserving if it preserves a volume on the state space $\mathcal{X}$ during its evolution, e.g., an idealized frictionless system. Such systems are ubiquitous, including classical Hamiltonian systems [66], physical systems in equilibrium [67], and the post-transient behavior of many general systems [7].

ples M to infinity to approximate correlations and only then the number of observables N to infinity (even then, EDMD may fail to converge). The convergence rate of time-averaged quantities such as correlations depends sensitively on the system's mixing properties and the regularity of the observable.[2] While polynomial rates hold for certain strongly mixing systems [68, 72], Birkhoff's theorem guarantees only convergence without any uniform rate and time averages can converge arbitrarily slowly [69, Chapter 2]. This non-uniformity prevents the formulation of a universal convergence rate or a general-purpose $N(M)$ strategy, unless we can control (**C1**).

To further understand how assumptions about a system's structure influence our ability to learn from data, we consider three scenarios (see Theorem 2.1 of the S.I.).

First, let $\Omega_{\mathcal{X}}^{\alpha}$ denote systems whose dynamics have a known smoothness quantified by a modulus of continuity α. For these systems, there exist learning algorithms that depend on two parameters, n_1 and n_2, and reliably use snapshot data to approximate the dynamics. These algorithms converge in two successive data limits ($n_1 \to \infty$ and then $n_2 \to \infty$) to the spectrum $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ for all systems F in $\Omega_{\mathcal{X}}^{\alpha}$.

Second, let $\Omega_{\mathcal{X}}^{m}$ denote measure-preserving systems with continuous dynamics. Again, algorithms exist that achieve convergence to $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ in two successive data limits for all systems in this class. The necessity of using two successive limits arises naturally. In Equations (7) and (8), challenge (**C1**) implies that convergence rates of the underlying sums cannot be uniformly controlled [69], a fact well known for mixing systems, among others. Thus, one must first take the large-data limit ($M \to \infty$), then the limit of dictionary size ($N \to \infty$); these limits cannot, in general, be combined or reversed. An implication of our analysis is that, for *any* algorithm (not just EDMD), no universal rule linking M and N guarantees convergence when both grow together.

Finally, consider the most general scenario, $\Omega_{\mathcal{X}}$, consisting of all continuous systems without further assumptions.[3] Remarkably, reliable algorithms still exist, but they require three successive data limits to convergence to the spectrum $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ for all $\Omega_{\mathcal{X}}$: increasing the number of snapshots ($M \to \infty$), expanding the dimensionality of the subspaces ($N \to \infty$), and refining coherence estimates (take the regularization parameter $\varepsilon \downarrow 0$ in Equation (5)) into spectra. Without assumptions on the underlying system, no simpler two-step algorithm can achieve guaranteed convergence (Theorem 2.9 of the S.I.)). (**C1**) and (**C2**) each cost a limit, illustrating the intrinsic difficulty of learning general dynamical systems.

Unified complexity classifications of learning

The Solvability Complexity Index (SCI) [73, 74] formalizes how many limits are needed to learn properties from data. For the full class $\Omega_{\mathcal{X}}$, computing $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ requires three separate limits, which cannot be reduced or combined: no method can succeed with only two. For structured subclasses like $\Omega_{\mathcal{X}}^{\alpha,m}$, $\Omega_{\mathcal{X}}^{\alpha}$, or $\Omega_{\mathcal{X}}^{m}$, the above shows that fewer limits suffice.

This framework provides a systematic way to assess the complexity of data-driven problems. Applied to Koopman op-
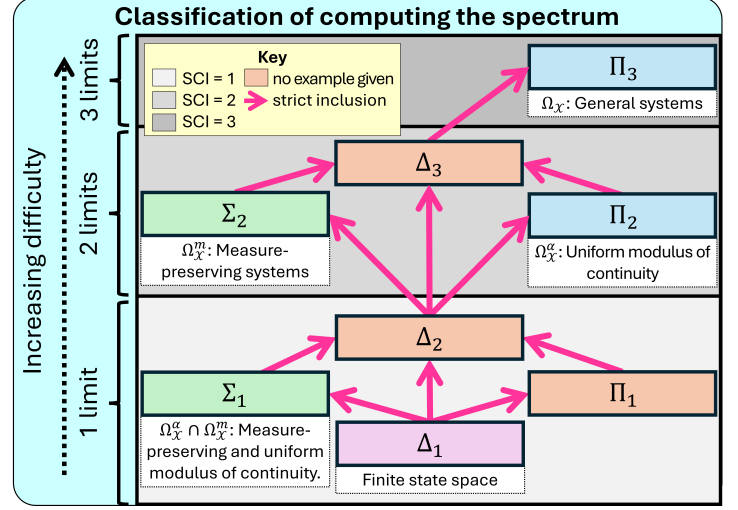
---

[2] As another example, Bandtlow et al. [71] construct an EDMD-based method that provably converges with $N = N(M)$ for analytic expanding maps, using analyticity and spectral gaps to unify the two limits.

[3] While the continuity of F can be relaxed, it is often assumed because discontinuities can lead to pathologies [46].

Figure 7: Classifications for learning Koopman spectra from trajectory data. Each SCI level indicates that solving the problem requires more complex, layered procedures—each 'limit' reflecting an extra data refinement. Our results establish both upper bounds (convergent algorithms) and lower bounds (established by constructing adversarial dynamical systems) for the complexity of these problems. Δ_{m+1} are problems with SCI $\leq m$. The Σ and Π classes provide a framework for understanding how verification is achieved in the final limit when learning Koopman operator spectra. The case of finite state space is Δ_1 since, in this case, the Koopman operator simplifies to a finite-dimensional matrix, making it possible to compute the spectrum directly using a single convergent algorithm with known convergence rates.

erators, Figure 7 summarizes the difficulty of learning their spectra based on our results. Each classification includes both an upper bound (a convergent algorithm), and a lower bound, established by constructing adversarial dynamical systems to prove that fewer limits are insufficient. We can also relate existing algorithms to the SCI hierarchy by summarizing convergence results from the Koopman literature and their corresponding implicit upper bounds in Table 1. Each algorithm relies on specific system assumptions and, in some cases, uses more data limits than necessary for convergence.

Beyond Table 1, Ulam's method is widely used to approximate eigenvalues of transfer operators, which describe the evolution of probability densities and are dual to Koopman operators. It typically requires two successive limits: one for Monte Carlo approximation and another for increasing matrix size [79], but convergence is not always reliable [80, Section 2.6]. Adding a third limit via noise smoothing can improve convergence [80], though adaptive noise selection sometimes reduces this back to two. Similar SCI classifications apply to other data-driven dimensionality reduction methods [81]. For multiple limits in control theory, see [82, Theorem 3].

These examples illustrate that multiple-limit phenomena are central to many data-driven methods in dynamical systems. While methods provide SCI upper bounds, a key challenge is determining whether these bounds are optimal. This raises fundamental questions: *Can convergence be achieved with fewer limits? If not, what assumptions make it easier?* To address these, we have derived lower bounds that show how system properties and the quantities being computed shape optimal algorithm design. When upper and lower bounds match, the algorithm is provably *optimal* for the problem.

In short, we now have a complete characterization of when data-driven spectral learning can succeed and when it cannot.

Table 1: Convergence results for Koopman operators in the SCI hierarchy. This table summarizes convergence results from the Koopman literature, interpreted through the lens of the Solvability Complexity Index (SCI). "N/C" denotes non-convergence without additional strong assumptions (e.g., requiring observables to lie within a finite-dimensional invariant subspace), "n/a" indicates the method is not applicable to the spectral problem and "m.p." stands for measure-preserving systems. A superscript * indicates the SCI bound improves by one if errors in approximations of correlations (see Equations (7) and (8)) are controlled, e.g., through known variational bounds on F. This reduction depends on system properties [68]. The column $\mathcal{K}g$ refers to approximating $\mathcal{K}$'s action on observables g. Note that computing the spectral measure (distribution of the system's behavior across different frequencies) may not yield the spectrum due to spurious eigenvalues. Upper bounds typically assume access to a dictionary with projections converging strongly to the identity; this construction is either specified (e.g., the entries "compactification methods" and "diffusion maps" have the distinct advantage of learning a well-conditioned dictionary) or required as an additional input by the user (e.g., EDMD and generator EDMD) which may or may not increase the SCI. In our upper bounds in Figure 7, we construct a dictionary. *N.B. These results are upper bounds on the SCI. Many of these bounds are not sharp, meaning they overestimate the number of limits required.*

Algorithm	Comments/Assumptions	Spectral Problem's Corresponding SCI Upper Bound			
		$\mathcal{K}g$	Spectrum	Spectral Measure (if m.p.)	Spectral Type (if m.p.)
Extended DMD [32]	general L^2 spaces	SCI $\leq 2^*$	N/C	N/C	n/a
Residual DMD [36]	general L^2 spaces	SCI $\leq 2^*$	SCI $\leq 3^*$	SCI $\leq 2^*$	varies, see [75] e.g., a.c. density: SCI $\leq 2^*$
Measure-preserving EDMD [76]	m.p. systems	SCI ≤ 1	N/C	SCI $\leq 2^*$ (general) SCI ≤ 1 (delay-embedding)	n/a
Hankel DMD [77]	m.p. ergodic systems	SCI $\leq 2^*$	N/C	N/C	n/a
Christoffel–Darboux kernel [35]	m.p. ergodic systems	SCI ≤ 3	n/a	SCI ≤ 2	e.g., a.c. density: SCI ≤ 2
Generator EDMD [78]	cts.-time, samples ∇F (otherwise additional limit)	SCI ≤ 2	N/C	SCI ≤ 2	n/a
Compactification [16]	cts.-time, m.p. ergodic systems	SCI ≤ 4	N/C	SCI ≤ 4	n/a
Resolvent compactification [48]	cts.-time, m.p. ergodic systems	SCI ≤ 5	N/C	SCI ≤ 5	n/a
Diffusion maps [9]	cts.-time, m.p. ergodic systems	SCI ≤ 3	n/a	n/a	n/a

Learning eigenpairs and latent spaces is hard

As a final problem, we consider the complexity of determining the spectral type, i.e., eigenvalues versus continuous components (see bottom panel of Figure 1). Spectral types distinguish between recurrent patterns (periodic or quasiperiodic oscillations) and more chaotic or mixing behavior, based on how the system spreads energy over frequencies [83, page 45]. They play a key role in applications such as fluid mechanics [84], anomalous transport [85], and analysis of trajectory invariants and exponents [86], and are particularly important in reduced-order modeling [7, 87]. We discussed their role in the context of our cavity flow problem in Example 1.2.

Identifying eigenvalues and eigenfunctions of Koopman operators ($\varepsilon = 0$ in Equation (5)) reveals a coordinate system in which complex dynamics appear linear, enabling the expansion in Equation (6). Non-unit eigenvalues ($\lambda \neq 1$) are especially important, as they capture time-changing features. Let Ω_p be the class of smooth, invertible systems on a torus (ring-shaped surface) that are measure-preserving with uniform bounds on their derivatives. This is a subclass of $\Omega_{\mathcal{X}}^{\alpha,m}$, so our universal one-limit algorithm applies to computing the full spectrum without distinguishing spectral type.

Surprisingly, even for this well-structured class of systems, there are no single-limit learning algorithms – whether deterministic or probabilistic with a success probability greater than 50% – using trajectory data, that can determine if the Koopman operator $\mathcal{K}_F$ has a non-unit eigenvalue for F in Ω_p. Similarly, no single-limit learning algorithms, deterministic or probabilistic with success probability greater than 50%, can converge to the set of eigenvalues (Theorem 2.7 of the S.I.).

However, both problems can be computed in two interdependent data limits (Algorithms 6 and 7 of the S.I.): adjusting the time lag for auto-correlations adaptively based data size; and increasing projections onto finite-dimensional subspaces defined by a dictionary (e.g., see Figure 10). Example 1.2 applied this approach to a complex fluid flow, see also the Methods. This two-limit approach is optimal, as no single-limit method can solve these problems. Figure 8 summarises the SCI classifications of computing spectral type.

This difficulty helps explain the challenges in finding finite-dimensional representations, such as autoencoders or latent spaces, where dynamics appear linear [29]. It is fundamen-

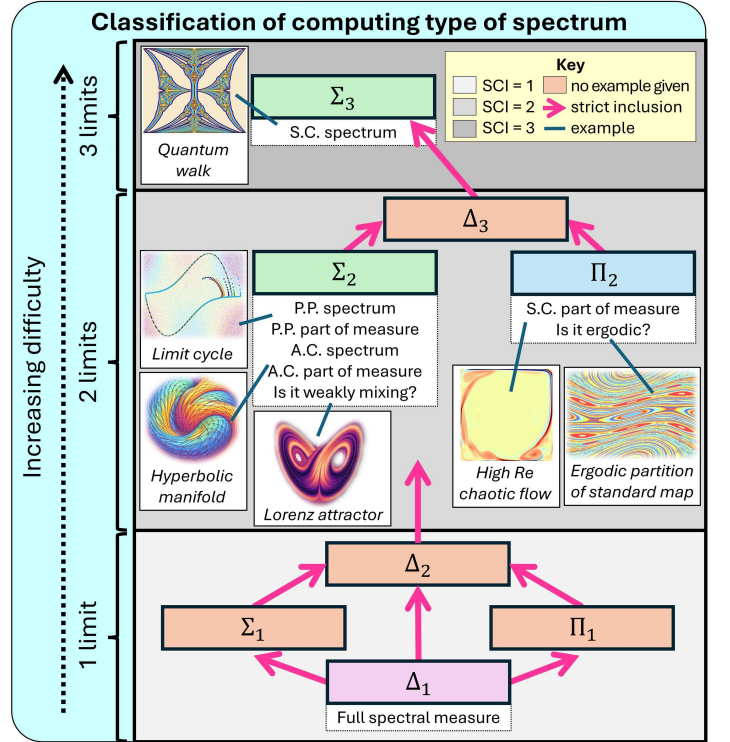**Classification of computing type of spectrum**

Figure 8: Classifications for learning spectral types (eigenvalues (p.p.), absolutely continuous (a.c.) and singular continuous (s.c.)) of Koopman operators for measure-preserving invertible systems from trajectory data. Each classification comprises of an upper bound (convergent algorithms) and lower bound (established by constructing adversarial dynamical systems). The SCI is the number of data limits needed to solve a problem. For each learning objective, we provide a representative system to emphasize that the classification fundamentally depends on the nature of the underlying dynamics. These examples illustrate how different dynamical behaviors influence the complexity of learning spectral types from data.

tally impossible to determine or verify the existence of eigenvalues for Koopman operators using only a single limit. A more practical alternative is to compute approximate eigenfunctions corresponding to $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$. This requires only a single limit and can be performed with verification (Figure 7), making it a more implementable approach in practice. We applied this to Arctic sea ice in Example 1.3.

Discussion

We have developed a powerful technique—namely, the construction of adversarial dynamical systems (see §*Adversarial*

systems and Figure 11)—to establish the first impossibility results in the field of data-driven dynamical systems. We identified conditions ((**C1**) and (**C2**)) for which no sequence of randomized algorithms (e.g., randomly sampling trajectories) can succeed with probability greater than 50%. These failures are not rare, but represent fundamental barriers. These barriers provided insights, which we used to develop a suite of new algorithms (Algorithms 1–7 in the S.I.) that achieve optimal performance and are provably convergent. These algorithms produce trustworthy, verifiable outputs, and include what we term 'multi-data-limit' methods (see Figures 7 and 8). By linking dynamical systems theory with the foundations of computation, we established a computational complexity framework for data-driven dynamical systems. This framework clarifies the inherent limitations of using finite data to analyze complex dynamics while also highlighting new algorithmic possibilities with broad applications. By combining a new theoretical complexity framework with practical algorithmic advances, our work provides a rare instance in ML of a synthesis between rigorous theory and application.

We demonstrated the power of our framework by solving a long-standing problem: computing Koopman spectra from data without spurious eigenvalues or missing components. Our algorithms succeeded where existing methods (e.g., EDMD) struggle, recovering accurate spectra in both low- and high-dimensional systems, including cases with continuous spectra and a real-world application to Arctic sea ice.

Arctic amplification has accelerated sea ice loss, with major consequences for ecosystems, local communities, and extreme weather. A key challenge is identifying the geographic patterns driving these changes. Our algorithms uncover hidden Koopman modes linked to sea ice decline (Figure 2), offering dynamic and spatial insight supported by *verification*. These decaying modes are physically meaningful and may guide future measurements. Using spectral approximations and error bounds to truncate spurious components, we built forecast models that achieve state-of-the-art accuracy at a fraction of the cost of deep learning models (Figures 5 and 6 and surrounding discussion). Applications include optimizing shipping routes, reducing environmental risks, and informing early-warning systems. Understanding sea ice loss is also vital given potential links to extreme events such as wildfires, floods, heatwaves, and cold spells. The Koopman framework further enables systematic comparisons across climate models and supports the construction of interpretable response formulas and reaction coordinates near critical transitions, with broad implications for climatology [88].

Our approach extends beyond Koopman operators, provided appropriate domain-specific "sudden change" lemmas can be established (see Figure 11). In each case, an analog of the "ball of learnability" and the classifications in Figures 7 and 8 applies. This perspective is relevant to a range of methods—including SINDy [10], neural ODEs [89], Fourier neural operators [90], LSTMs [91], and PDE-net [92]—as well as to other areas of scientific computing with ML. For example, recent work has shown that linear elliptic PDEs can be learned from input-output pairs [93], analogous to snapshot-based learning. Whether similar approaches extend to hyperbolic or nonlinear PDEs remains open, but our proof techniques may offer insight into this challenge.

This paper initiates a broader effort to explore the limits of robust learning and to develop a theory of necessary and sufficient conditions. Several promising directions remain. First, we have assumed access to full-state observations (see snapshot data in Equation (4)), whereas many applications involve only partial measurements of the state x. Extending our upper and lower bounds to such settings is a natural next step; results like Takens' embedding theorem [94], which allows reconstruction of attractors from a single time series, may be particularly useful. Second, we focused on discrete-time systems, reflecting how data is typically collected. It would be valuable to study continuous-time dynamics and the sampling conditions needed for reliable learning. Our results should carry over under generic time discretizations, but a formal analysis remains open. Third, Koopman-based methods have shown promise in control problems across domains such as power grids [95], robotics [23], fluid dynamics [96], chemistry [97], and biology [98], where the linearity of $\mathcal{K}$ enables tractable optimization. Our development of provably convergent, error-bounded algorithms for Koopman spectral properties opens the door to significant advances in nonlinear control. Fourth, it is natural to seek lower bounds complementing the upper bounds in Table 1 for problems that do not rely on spectral computations. These may have different lower bounds than those we established, but we anticipate that the adversarial dynamical systems framework can be extended to cover them.

ML in data-driven dynamical systems is skyrocketing, and this momentum shows no signs of slowing. Across nearly every area, key challenges are being reexamined through big data and deep learning. With this surge of interest and innovation, it is crucial for the community to grasp not only what is possible but also what is fundamentally impossible. This prevents the pursuit of unattainable algorithms or methods, safeguards against potentially catastrophic errors, and reveals the conditions under which learning is feasible: upper and lower bounds inform and sharpen one another. This process led us to our convergent algorithms. Such classifications are essential if we are to fully harness the power of ML in dynamical systems.

Methods

Computing Koopman modes and EDMD

For a dictionary of observables $\{g_j\}_{j=1}^N$ and snapshot pairs $(x^{(m)}, y^{(m)})$ in Equation (4), EDMD approximates two correlation matrices $G_{ij} = \langle g_j, g_i \rangle$ and $A_{ij} = \langle \mathcal{K}g_j, g_i \rangle$ ($i, j = 1, \ldots, N$) from data averages using $\mathcal{K}g_j(x^{(m)}) = g_j(y^{(m)})$, analogously to Equation (1):

$$G_{ij} \approx \hat{G}_{ij} = \frac{1}{M} \sum_{m=1}^M g_j(x^{(m)})g_i(x^{(m)}), \quad i,j = 1, \ldots, N, \quad (7)$$

$$A_{ij} \approx \hat{A}_{ij} = \frac{1}{M} \sum_{m=1}^M g_j(y^{(m)})g_i(x^{(m)}), \quad i,j = 1, \ldots, N. \quad (8)$$

(Here, the L^2 inner product $\langle \cdot, \cdot \rangle$ is a generalized dot product that measures how strongly two observables are correlated.) The Koopman approximation is the $N \times N$ matrix $\hat{G}^{-1}\hat{A}$ and EDMD computes its eigenvalues.

We compute the Koopman modes $\mathbf{g}_j$ appearing in Equation (6) as follows. Given a collection of n approximate eigenfunctions (computed, e.g., by EDMD or our algorithms), let

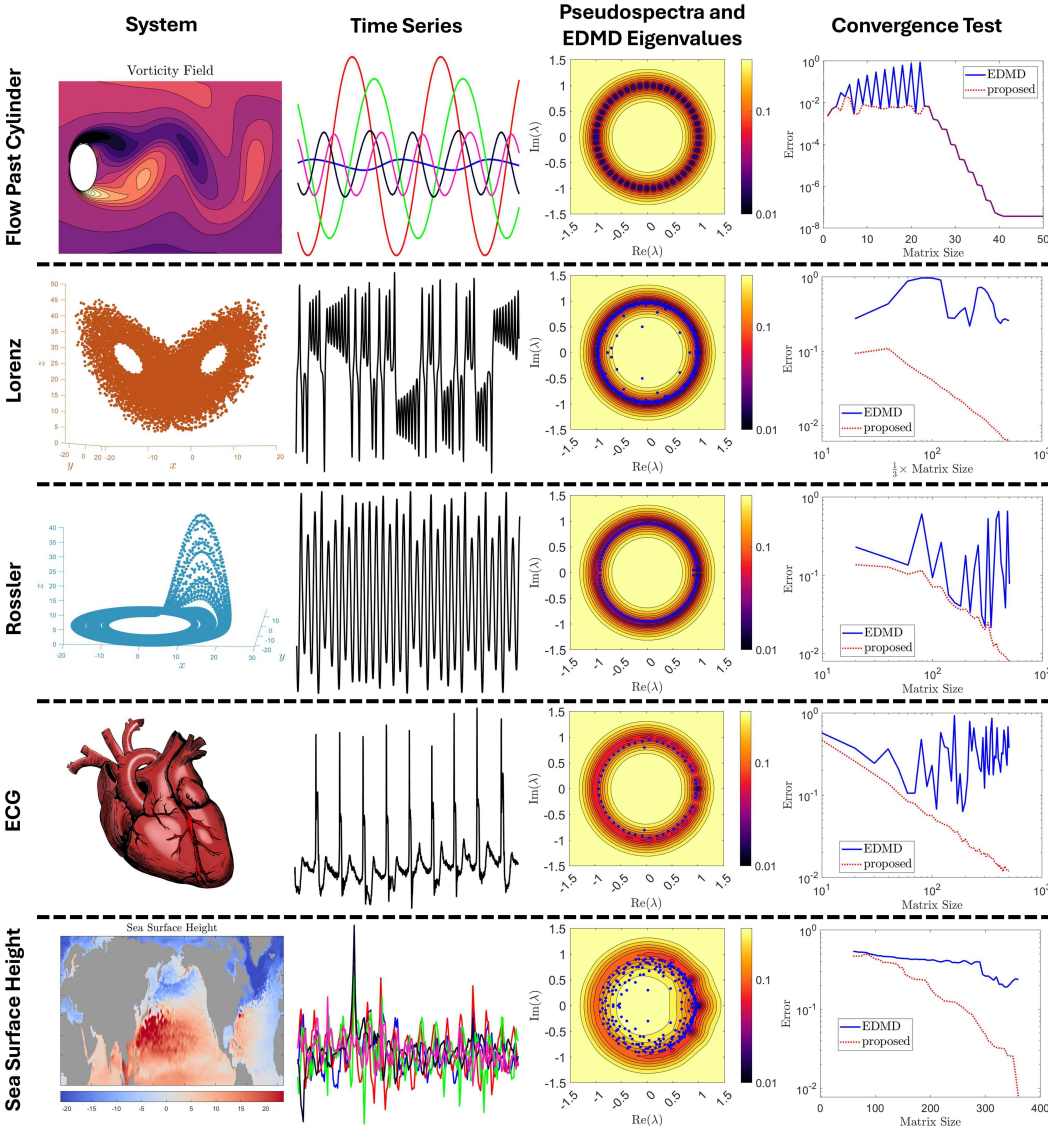|	System	Time Series	Pseudospectra and EDMD Eigenvalues	Convergence Test

Figure 9: Spectral analysis of a range of analytic and real-world dynamical systems using our proposed algorithms. Each row corresponds to a different system: (i) periodic flow past a cylinder, (ii) Lorenz system, (iii) Rössler system, (iv) electrocardiogram (ECG) data, and (v) monthly mean sea surface height in the Northern Hemisphere (1950–present). Columns show: the system or dataset, a representative time series of observables, the computed level curves of ε for ε-approximate eigenfunctions (color scale) termed "pseudospectra" overlaid with EDMD eigenvalues (blue dots), and a convergence comparison between EDMD (blue) and our method (red, using Algorithm 1 from the S.I.). The pseudospectrum shows where approximate, near-eigenvalue behavior occurs—providing a more robust picture of dynamics. In this final column, the error metric is the same as in Figure 1. Except for the periodic flow, EDMD yields spurious eigenvalues and lacks convergence. In contrast, our approach produces qualitatively accurate and convergent spectral approximations. Notably, the level curves of ε reveal distinct spectral structures across systems: continuous spectra for chaotic systems (Lorenz, Rössler), spectral clustering near $\lambda = 1$ for ECG, and non-normal features and seasonal modes around $\lambda \approx \exp(\pi i/6)$ in sea surface height data. Further experimental details are provided in the S.I.

$\Psi \in \mathbb{C}^{M \times n}$ denote the matrix corresponding to their time series across the training data $x^{(m)}$ ($m = 1, \ldots, M$) in Equation (4). For a vector of observations $\mathbf{g} \in \mathbb{C}^N$, let $\mathbf{O} \in \mathbb{C}^{M \times N}$ be their collected data over the same training data. The Koopman modes $\mathbf{g}_j$ are computed by

$$\begin{pmatrix} \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_n \end{pmatrix}^\top = [\Psi^\dagger]\mathbf{O},$$

where $\dagger$ denotes the Moore–Penrose inverse (solution of least squares problem) and $\top$ the matrix transpose. In Figures 2 and 3, we have plotted the absolute values of Koopman modes since these describe physical locations of the dynamics described by eigenvalues (as explained in the discussion surrounding Equation (6)). The error bars for these Koopman modes and eigenvalues correspond to $\|\mathcal{K}\phi_\varepsilon - \lambda\phi_\varepsilon\|$ (computed using the function in Equation (9)) for corresponding approximate eigenfunction ϕ_ε. These provide an error bound for coherency in Equation (5).

Upper bounds: A suite of convergent algorithms

There are several algorithms, each corresponding to different classes of dynamical systems in Figure 7, and full pseudocode is provided in the S.I. Recall from Example 1.1 that we consider a dictionary of observables $\{g_j\}_{j=1}^N$ and that we have access to the snapshot data in Equation (4).

We first describe the algorithm for $\Omega_{\mathcal{X}}^{\alpha,m}$ (Algorithm 1 in S.I.). Traditional methods for analyzing the dynamics of complex systems (e.g., EDMD) often rely on estimating the Koopman operator by computing eigenvalues of finite-dimensional matrices. However, these approximations can be unreliable and fail to converge, as shown in Figures 1 and 9. Our approach avoids this pitfall by taking a different route. Instead of directly computing eigenvalues of an approximate matrix, we use trajectory data from the system to construct *three* matrices that capture the essential correlations between observables and their evolution. These matrices approximate inner products involving the Koopman operator $\mathcal{K}_F$ and its adjoint $\mathcal{K}_F^*$ (which encodes how observables change in *reverse time*).

We form a correlation matrix based on how different observables behave across the dataset:

$$\hat{L}_{ij} = \frac{1}{M} \sum_{m=1}^M g_j(y^{(m)})\overline{g_i(y^{(m)})}, \quad i,j = 1, \ldots, N,$$

which estimates how the time-evolved observables overlap. It approximates $\langle \mathcal{K}g_j, \mathcal{K}g_i \rangle$ in an analogous fashion to how Equations (7) and (8) approximate $\langle g_j, g_i \rangle$ and $\langle \mathcal{K}g_j, g_i \rangle$, respectively. Using these matrices, we define a function

$$h_N(z, F) = \sqrt{\sigma_{\inf}(\hat{L} - \bar{z}\hat{A} - z\hat{A}^* + |z|^2\hat{G})}. \tag{9}$$

11

Here, σ_{inf} denotes the smallest singular value of a matrix. This method has a key advantage: If F has a known modulus of continuity, we adaptively select N based on the number of data points M (amounting to controlling errors in the approximations of the correlations) to ensure that

$$\lim_{N\to\infty} h_N(z,F) = \sigma_{\text{inf}}(\mathcal{K}_F - zI),$$

with convergence from above. Here, I denotes the identity operator. Moreover, this specific construction guarantees convergence, in contrast to approaches using $\sigma_{\text{inf}}(\hat{A} - z\hat{G})$, which lack this property. It establishes the rigorous error bounds demonstrated, for example, in Figures 1 and 9.

We evaluate this function over an initial grid of z-points that is adaptively refined with increasing N (a good choice is $\{z \in \frac{1}{n}\mathbb{Z} + \frac{i}{n}\mathbb{Z} : |z| \leq n\}$). The computed function $h_N(z,F)$ yields an upper bound $d_{N,z}$ on $\text{dist}(z, \text{Sp}_{\text{ap}}(\mathcal{K}_F))$, provided the non-normality of $\mathcal{K}_F$ is suitably controlled (e.g., if the system is measure-preserving, i.e., where statistical properties do not change over time). We then find local minimizers of $h_N(z,F)$ in a radius $d_{N,z}$ neighborhood of each grid point. These minimizers provide approximations to the spectrum near each point. The union of these local approximations is guaranteed to converge to the true spectrum as $N \to \infty$. By computing the singular vectors corresponding to the smallest singular values, we simultaneously obtain ε-approximation of eigenfunctions (with $\varepsilon = \sigma_{\text{inf}}(\mathcal{K}_F - \lambda I)$ for a spectral parameter λ), called approximate eigenfunctions ϕ_ε, which satisfy the condition $\|\mathcal{K}_F \phi_\varepsilon - \lambda \phi_\varepsilon\| \leq \varepsilon$ and hence the coherency bound in Equation (5). In addition to providing error bounds, the algorithm is local, trivially parallelizable, and stable.

For the class $\Omega_{\mathcal{X}}^\alpha$, we cannot directly convert $h_N(z,F)$ into the distance bound $d_{N,z}$ as described above. This limitation reflects the challenge (C2) and requires an additional limiting process to achieve the conversion (Algorithm 3 in the S.I.). Similarly, for the class $\Omega_{\mathcal{X}}^m$, we are unable to adaptively choose N based on the number of data points M, due to the challenge (C1). As a result, we must evaluate $h_N(z,F)$ using another limiting procedure (Algorithm 4 in the S.I.). In the most general case, $\Omega_{\mathcal{X}}$, both challenges arise simultaneously, necessitating three successive limits (Algorithm 5 in the S.I.).

To separate eigenvalues from continuous spectra, we use a mathematical result known as the RAGE theorem. It states for unitary $\mathcal{K}$ that we can isolate the contribution of eigenfunctions by averaging time-evolved observables after projecting onto an increasing sequence of finite-dimensional subspaces. Specifically, if $\{\mathcal{P}_n\}$ are finite-rank orthogonal projections converging to the identity, then for any observable g, the projection $\mathcal{P}_{\text{pp}}$ onto the eigenspaces of $\mathcal{K}$ satisfies:

$$\|\mathcal{P}_{\text{pp}}g\|^2 = \lim_{n\to\infty}\lim_{L\to\infty} \frac{1}{2L+1} \sum_{\ell=-L}^{L} \left\|\mathcal{P}_n \mathcal{K}^\ell g\right\|^2. \qquad (10)$$

This formula separates observables into parts that remain localized over time and parts that disperse. The projections $\mathcal{P}_n$ retain the localized portion, which corresponds to persistent patterns described by eigenfunctions. This provides a principled way to identify and extract the parts of an observable associated with eigenfunctions. The procedure is implemented in Algorithms 6 and 7 of the S.I.

Equation (10) involves two consecutive limits: one over the increasing rank of the projections or number of observables ($n \to \infty$) and another over the time window length ($L \to \infty$). This method enables accurate spectral computations using only finite data, without requiring an explicit approximation of the Koopman operator (which must be formed for EDMD). We also point out that this technique is substantially different to partitions of the full spectral measure (distribution of the system's behavior across different frequencies) over intervals, which have previously been computed using the ergodic theorem, e.g., [35]. Moreover, the RAGE theorem does not require the system to be ergodic.

We assume the state space is a compact metric space $(\mathcal{X}, d_{\mathcal{X}})$ and consider observables g in $L^2(\mathcal{X}, \omega)$, the space of functions whose squared values can be averaged (integrated) over $\mathcal{X}$ using a probability measure ω, which defines how different parts of the state space are weighted. We also assume that F is nonsingular with respect to ω (if $\omega(E) = 0$ then $\omega(\{x : F(x) \in E\}) = 0$) and $\mathcal{K}_F$ is bounded. This is the standard setting in Koopman analysis, often referred to as the *spectral study* of the system. Notably, our methods apply to *any* compact metric space $\mathcal{X}$ and probability measure ω. To assess how well algorithms capture the spectrum of the Koopman operator, we use the Hausdorff metric, which measures how far two spectral sets are from each other by quantifying the greatest distance one must travel from a point in one set to reach the other. This metric ensures that the computed spectra converge accurately, avoiding errors such as including false eigenvalues or missing important spectral regions.

Duffing oscillator

The Duffing oscillator is the system of equations:

$$\frac{dx}{dt} = y, \quad \frac{dy}{dt} = -\gamma y + x(1 - x^2).$$

The state $\mathbf{x} = (x,y) \in \mathbb{R}^2$ evolves in a two-dimensional state space. To analyze this system in discrete time, we use a time-step of $\Delta t = 0.3$. We consider two cases based on the parameter γ. *Conservative case ($\gamma = 0$):* This choice corresponds to a Hamiltonian system, where the dynamics conserve a volume in state space, reflecting energy conservation. *Dissipative case ($\gamma = 0.3$):* In this scenario, the system has energy dissipation, leading to two stable spirals at $(\pm 1, 0)$ and a saddle point at the origin. These two cases highlight distinct dynamical behaviors. In the Hamiltonian case, trajectories exhibit long-term behavior without dissipation, while in the dissipative case, trajectories eventually converge to stable attractors. EDMD struggles to reliably approximate the Koopman operator spectrum in both cases, illustrating its non-convergence under varying dynamical conditions.

We first generate an initial dataset by uniformly sampling 10^4 initial conditions $\mathbf{x}_0 = (x_0, y_0)$ in the square $[-2,2]^2$ and recording trajectories of length 5 for each sample. To construct a dictionary of observables, we apply k-means clustering to this initial dataset and use the resulting cluster centres $\{c_j\}_{j=1}^N$ to define N radial basis functions:

$$g_j(\mathbf{x}) = \exp(-2\|\mathbf{x} - c_j\|_{l^2}/\sigma), \quad j = 1, \ldots, N.$$

The trajectory length 5 is chosen to reduce the condition number of the resulting basis (see Fig. 4 in the S.I.) and k-means clustering is a standard way to ensure that the centers are

well distributed, preventing over-concentration and improving the representativeness and conditioning of the dictionary. The scaling parameter σ is set to the average l^2-norm of the snapshot data after it is shifted to mean zero, which we have found empirically to work well across a range of examples. Other dictionary choices are certainly possible. The only requirement is that, as we use more basis functions, the projections acting on any fixed observable converging to that observable.

Once the dictionary is constructed, we sample points uniformly at random from $[-2, 2]^2$ to create M snapshots of the system. This corresponds to Monte Carlo integration, where the snapshots approximate the correlations required in our analysis. Fig. 5 of the S.I. illustrates the convergence of this approach as the number of snapshots increases, as well as the convergence of $h_N(z, F)$ with increasing N.

The level curves of ε in Figure 1 were computed using $N = 500$ basis functions and $M = 50000$ snapshots, and Algorithm 2 in the S.I. To compute the spectrum, we used the above adaptive procedure (Algorithm 1 in the S.I). We repeated experiments 10 times with different random seeds for the random trajectory data to ensure robustness.

Further examples comparing EDMD and algorithms

Our algorithms are further applied to a variety of analytic and real-world systems, as shown in Figure 9, using trajectory data with qualitatively different characteristics. Further details about each system and experiment are provided in the S.I., and code for all examples is publicly available. The examples span a broad range of dynamics, including periodic flow past a cylinder and canonical chaotic systems such as the Lorenz and Rössler attractors, which are two of the simplest systems that exhibit chaotic motion. To demonstrate applicability to more realistic scenarios, we also include data collected from an electrocardiogram (ECG), and monthly mean sea surface height in the Northern Hemisphere from 1950 to the present. Data sources are listed in the S.I. and in all cases, our algorithms and EDMD use the same data and the same dictionary to ensure a fair comparison.

In each case, we computed both the pseudospectra (level curves of ε for approximate eigenfunctions) using Algorithm 2 in the S.I. and the EDMD eigenvalues (blue dots). Except for the periodic flow case, EDMD produces spurious eigenvalues, as it did for the Duffing oscillator in Figure 1. Figure 9 demonstrates this phenomenon over a range of dictionary choices, data collections, and types of dynamical systems.

The pseudospectra exhibit different structures. For the cylinder flow, the multiplicative structure of the eigenvalues (see the discussion of combination modes in the Arctic sea ice example) is clearly visible in the first row. The Lorenz and Rössler systems display continuous spectra concentrated on the unit circle. The ECG data yields a tight spectral cluster near $\lambda = 1$ and near the frequency of the recorded heartbeat. The pseudospectra of the sea surface height data show strong non-normal features (i.e., the pseudospectra differ from spectral distances), capturing transient dynamics and nonstationary trends in data, along with dominant spectral regions corresponding to seasonal variations around $\lambda \approx \exp(\pi i/6)$.

In all cases, our spectral approximation algorithm (Algorithm 1 in the S.I.) converges as expected, whereas EDMD does not converge (except for the periodic flow case).

High Reynolds number fluid flows

As an example of detecting eigenfunctions using an optimal two-limit procedure (Algorithm 6 in the S.I.), we consider two-dimensional lid-driven cavity flow. This system involves the motion of an incompressible, viscous fluid at a high Reynolds number (Re), which characterizes the dominance of inertial forces over viscous forces, leading to complex flow patterns. This setup provides a challenging example for identifying eigenfunctions of the underlying dynamics.

The physical domain is the cavity $[-1, 1] \times [-1, 1]$, with stationary solid boundaries on all sides except the top. The top boundary moves with a regularized velocity profile $u_{\text{top}} = (1 - x^2)^2$. This standard boundary condition ensures both continuity and incompressibility, even at the corners of the top boundary. Using the streamfunction ψ, the incompressible Navier–Stokes equations for this flow can be reformulated as:

$$\frac{\partial}{\partial t}\nabla^2\psi + \frac{\partial\psi}{\partial y}\frac{\partial}{\partial x}\nabla^2\psi - \frac{\partial\psi}{\partial x}\frac{\partial}{\partial y}\nabla^2\psi = \frac{1}{Re}\nabla^4\psi,$$

$$\psi|_{\partial B} = 0, \ \frac{\partial\psi}{\partial n}(y{=}{-}1) = \frac{\partial\psi}{\partial n}(x{=}{\pm}1) = 0, \ \frac{\partial\psi}{\partial n}(y{=}1) = u_{\text{top}}.$$

These equations have a unique solution, and the flow dynamics converge to a universal attractor as time progresses [99] As discussed in Example 1.2, the spectral structure of the Koopman operator tells us about the geometry of this attractor. To compute ψ, we use a Chebyshev spectral collocation method with an adaptive grid resolution [100], which depends on Re, ensuring accurate computation for a range of flow conditions. For our analysis, we utilize $M = 20000$ snapshots of the flow sampled at time intervals of 0.1s to capture its dynamics, after an initial burn-in time to ensure data is collected from the attractor.

We apply Equation (10) to analyze the mean-subtracted total kinetic energy as our observable g. To construct the projections $\mathcal{P}_n$, we use time-delay embedding, with n time delays. This approach captures the temporal structure of the system by embedding the observable in a higher-dimensional space. Specifically, the terms $\|\mathcal{P}_n\mathcal{K}^\ell g\|^2$ are calculated as follows: First, the Koopman operator $\mathcal{K}^\ell$ is applied to the trajectory data of g, which acts by shifting the time series by ℓ steps. Then, the Moore–Penrose inverse is used to apply the projection $\mathcal{P}_n$, ensuring consistency with the chosen time-delay embedding. Finally, the squared norm is computed by averaging as in Equation (8), which allows us to extract long-term statistical averages from the trajectory data.

Figure 10 shows the results for various choices of Re. We observe the double limit $\lim_{n\to\infty}\lim_{L\to\infty}$ at play and the structure of the spectrum reveals a sequence of bifurcations as the Reynolds number increases. For $Re \leq 10000$, the flow converges to a steady laminar solution, corresponding to a fixed point in the state space. Just above $Re = 10000$, this steady solution becomes unstable, and the system transitions to a time-periodic flow, which remains stable up to $Re = 15000$. The fundamental frequency of the periodic flow decreases with increasing Reynolds number. At $Re \geq 15000$, a second bifurcation occurs, and the flow becomes quasiperiodic. The basic frequencies of the quasiperiodic motion also decrease with Reynolds number, until around $Re = 18000$, where a third bifurcation is observed: the portion of kinetic energy associ-
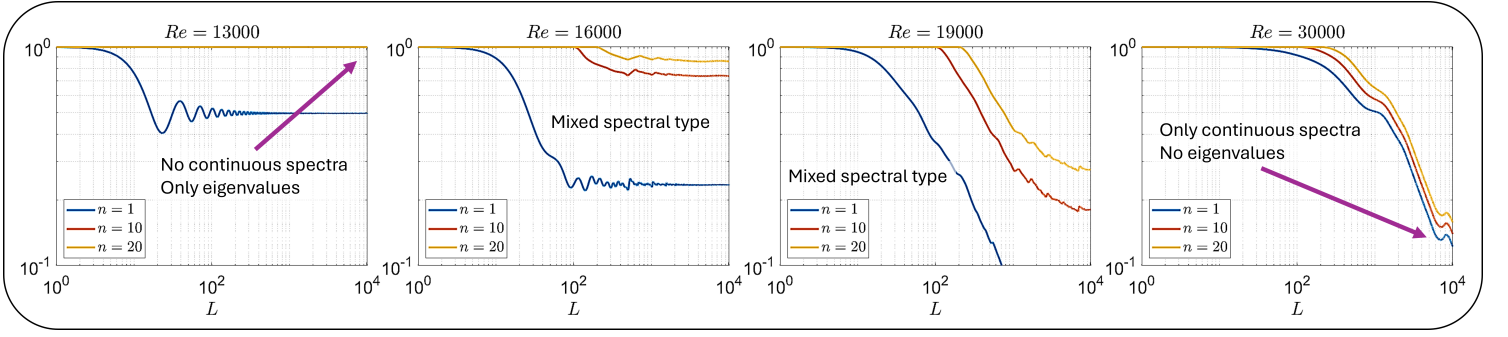
Figure 10: Application of Equation (10) to the kinetic energy of a lid-driven cavity flow. The plots show $\frac{1}{2L+1}\sum_{\ell=-L}^{L}\|\mathcal{P}_n\mathcal{K}^\ell g\|^2$ (for normalized g), which in the double limit $\lim_{n\to\infty}\lim_{L\to\infty}$ converge to the fraction of g made up of eigenfunctions (see Algorithm 6 in the S.I.). This double limit procedure is used to prove upper bounds. As we move from left to right, the Reynolds number of the flow increases, and the spectral type becomes more dominated by continuous spectra. The structure of the spectrum reveals a sequence of bifurcations described in the main text, providing a picture of the transition to chaos that is consistent with the theory of Ruelle and Takens [54].

ated with the continuous spectrum rises sharply to a few percent. This continuous component continues to grow, and by $Re \geq 22000$, the state space dynamics exhibit no quasiperiodic structure: the Koopman spectrum consists solely of continuous components, indicating fully chaotic dynamics.

The bottom of Figure 1 illustrates the extraction of the eigenvalues for $Re = 19000$. The top part of panel (e) shows the smoothed spectral measure (distribution of the kinetic energy across different frequencies) with a second-order smoothing kernel and smoothing parameter $\epsilon = 10^{-3}$ [50]. The bottom part of panel (e) displays the extracted eigenvalues, computed using Equation (10) (Algorithm 6 in the S.I.). These are plotted against the Strouhal number, a dimensionless quantity that measures how often vortices are shed from a body relative to the speed of the flow and the size of the body.

Arctic sea ice

Satellites have measured sea ice for decades using passive microwave sensors, which are then used to compute sea ice concentration using retrieval algorithms. We used data from the European Organisation for the Exploitation of Meteorological Satellites' (EUMETSAT) Ocean and Sea Ice Satellite Application Facilities (OSI-SAF) data record, comprising retrieval algorithms OSI-450 (1979–2015) and OSI-430-b (2016 onwards). These algorithms have been shown to be more accurate than other retrieval algorithms. Some portions of the data surrounding the North Pole are missing, and we use bilinear interpolation to fill these gaps. There are also missing data in three months in the 1980s due to satellite malfunctions, and we use linear extrapolation to fill these gaps. These data-points were not used in the testing of forecasts in Figures 4 and 5.

To construct our approximation of the Koopman operator, we consider observables $\mathbf{x} \in \mathbb{R}^{97877}$ representing sea ice concentration at each sea grid point. We then apply time-delay embedding: with $\tau - 1$ delays, the augmented system state at time n is given by (we use $(\cdot)$ to indicate time dependence instead of a subscript since $\mathbf{x}$ is a vector):

$$\widehat{\mathbf{x}}(n) = \begin{pmatrix} \mathbf{x}(n) \\ \mathbf{x}(n-1) \\ \vdots \\ \mathbf{x}(n-\tau+1) \end{pmatrix} \in \mathbb{R}^{97877 \cdot \tau}.$$

That is, we include the sea ice concentration at the current and $\tau - 1$ previous time steps. We choose $\tau = 6$ to help capture semiannual patterns, though other choices also yield good

performance. As our dictionary, we use Gaussian radial basis functions of the form $\exp(-\|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}^{(m)}\|_{l^2}^2 / \sigma)$, where the centres $\widehat{\mathbf{x}}^{(m)}$ correspond to the M snapshot data of the augmented system state (in contrast to the Duffing oscillator where we used k-means to choose centers). As in the Duffing oscillator example, the scaling parameter σ is set to the average l^2-norm of the snapshot data after centering it to have zero mean.

A key strength of our algorithm is its rigorous error bounds (e.g., Figures 1 and 9), which enable direct evaluation of dictionary performance by tracking the error—without needing held-out data or forecasts. This allows us to verify that our chosen dictionary yields small errors when approximating the Koopman spectrum. Future work will explore integrating neural network embeddings [29] with our error-bound framework.

To compute the periodic benchmark in Figures 4 and 5, let I_k denote the time indices for month k ($= 1, 2, \ldots, 12$) in the training data when we make a forecast. For forecasting month k, $\mathbf{x}_{\mathrm{per}}$ is the average of $\mathbf{x}$ over the times in I_k. The difference $\mathbf{x} - \mathbf{x}_{\mathrm{per}}$ then represents the sea ice anomaly. Letting $\mathbf{x}$ be the true sea ice concentration, $\mathbf{x}_{\mathrm{rec}}$ the forecast and $\mathbf{x}_{\mathrm{per}}$ the periodic baseline, we define the relative error as Error $= \|\mathbf{x} - \mathbf{x}_{\mathrm{rec}}\|_{l^2}^2 / \|\mathbf{x} - \mathbf{x}_{\mathrm{per}}\|_{l^2}^2$. The denominator isolates the anomaly by removing the trivially predictable seasonal cycle, allowing for a more meaningful assessment of skill.

To compute forecasts, we get rid of spurious modes in the decomposition in Equation (6) (where $\mathbf{g}$ is the vector of sea ice concentrations $\mathbf{x}$). The evolution is predicted forward in time for $\mathbf{x}$ using approximate eigenfunctions $\phi_\varepsilon^{(j)}$ for errors ε below a threshold ε_0:

$$\mathbf{x}(n) = \sum_{\varepsilon \leq \varepsilon_0} \lambda_j^n \phi_\varepsilon^{(j)}(0)\mathbf{g}_j. \tag{11}$$

Here, the parameter j ranges from 1 to M (corresponding to the space generated by the radial basis functions). The errors associated with the approximate eigenfunctions g_ε are ordered by ε, and we identify the "elbow" in the error curve to determine a principled truncation point (see S.I. Fig. 12). The Koopman modes $\mathbf{g}_j$ correspond to the vector $\mathbf{x}$ of observables. After this model has been built, forecasts are produced by increasing n. The DMD forecast in Figure 5 is computed in the same manner, but now does not get rid of spurious modes and uses the augmented state space $\widehat{\mathbf{x}}$ as the dictionary of observables. IceNet and SEAS5 data in Figure 5 is taken directly from [63].
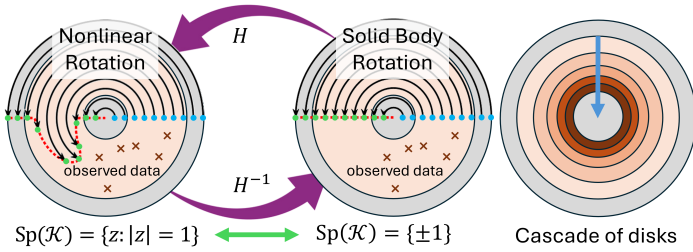
14

Figure 11: Proof idea of the impossibility result for $\Omega_\mathbb{D}$. At each stage, we modify the system consistently with the observed data ("x"), ensuring it is related to a rotation, thereby drastically altering the spectrum (see green arrow). This alteration is executed such that the cascade of dynamical systems converge to an underlying limit, providing the adversarial family of systems.

Lower bounds: The method of adversarial systems

We establish each lower bound (impossibility result) using simple examples of state spaces $\mathcal{X}$ (e.g., disk, interval, or torus) where ω represents the standard Lebesgue measure. The techniques are general and can be extended to other state spaces $\mathcal{X}$ and function spaces. Snapshot data is subject to noise and finite precision. To establish robust results, we use a measurement device $\mathcal{T}_F$ that allows arbitrarily accurate sampling:

$$\mathcal{T}_F = \left\{ \hat{y}_{j,n} \in \mathcal{X} : d_\mathcal{X}(F(\hat{x}_j), \hat{y}_{j,n}) \leq 2^{-n} \right\},$$

where $\{\hat{x}_j\}_{j=1}^\infty$ is a dense subset of the metric space $(\mathcal{X}, d_\mathcal{X})$, corresponding physically to "measurement points". This formulation assumes that measurements can approximate the mapping F with arbitrarily high precision. This is a strong assumption, and hence it allows us to derive correspondingly strong impossibility results, highlighting fundamental limits even under idealized conditions. Limitations under idealized conditions imply they hold under more realistic ones. (It can also be significantly relaxed for our upper bounds.)

To establish the lower bounds, we construct families of adversarial dynamical systems. These systems are carefully designed to embed sudden changes in the spectral properties of the Koopman operator $\mathcal{K}_F$ directly into the dynamics, while remaining consistent with the observed trajectory data.

Figure 11 illustrates this idea for the class $\Omega_\mathbb{D}$. The construction uses a homeomorphism (a continuous, reversible transformation that stretches or bends a space without tearing, gluing, or creating holes) to deform the map whilst preserving the sampled data. This sudden change lemma (Lemma 2 in the S.I.) alters the spectral behavior of the Koopman operator in a controlled way. By applying this construction recursively on a cascade of nested disks, we effectively mislead any proposed algorithm into failing to converge. Each adversarial construction is tailored to a specific sudden change lemma, with full details provided in the S.I. This strategy bridges computational techniques with classical ergodic theory, offering a new framework for analyzing dynamical systems. Moreover, the method is flexible and can be adapted to tackle a broader range of problems beyond those considered in this study.

To prove a learning problem cannot be solved in one limit (i.e., it has SCI > 1), we assume, for contradiction, that a convergent sequence of algorithms exists. We then construct an adversarial family that causes the algorithm to fail, ensuring convergence occurs with probability no greater than 50%. This involves tricking the algorithm into oscillating between two different outputs as more trajectory data is collected (see the green arrow in Figure 11). To prove that a problem cannot be solved in two limits (i.e., SCI > 2), we embed complex combinatorial problems into the system's dynamics. These problems, involving sets of numbers with an inherent complexity in their description, are embedded into the dynamics, lifting the lower bound from combinatorics to dynamics.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[2] J. Jumper et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[3] L. Abriata. The Nobel Prize in Chemistry: past, present, and future of AI in biology. *Commun. Biol.*, 7(1):1409, 2024.

[4] Kaiming H., X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016.

[5] A. Zhavoronkov et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.*, 37(9):1038–1040, 2019.

[6] C. William Gear et al. Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.*, 1(4):715–762, 2003.

[7] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.*, 41(1-3):309–325, 2005.

[8] M. Schmidt and H. Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

[9] T. Berry, D. Giannakis, and J. Harlim. Nonparametric forecasting of low-dimensional dynamical systems. *Phys. Rev. E*, 91(3):032915, 2015.

[10] S. Brunton, J. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci.*, 113(15):3932–3937, 2016.

[11] J. N. Kutz, S. Brunton, B. Brunton, and J. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM, 2016.

[12] D. Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.*, 47(2):338–396, 2019.

[13] S. Brunton, B. Noack, and P. Koumoutsakos. Machine learning for fluid mechanics. *Annu. Rev. Fluid Mech.*, 52(1):477–508, 2020.

[14] D. Giannakis, A. Henriksen, J. Tropp, and R. Ward. Learning to forecast dynamical systems from streaming data. *SIAM J. Appl. Dyn. Syst.*, 22(2):527–558, 2023.

[15] T. Bury et al. Deep learning for early warning signals of tipping points. *Proc. Natl. Acad. Sci.*, 118(39), 2021.

[16] S. Das, D. Giannakis, and J. Slawinska. Reproducing kernel Hilbert space compactification of unitary evolution groups. *Appl. Comput. Harmon. Anal.*, 54:75–136, 2021.

[17] E. Gibney. China's cheap, open AI model DeepSeek thrills scientists. *Nature*.

[18] B. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proc. Natl. Acad. Sci.*, 17(5):315–318, 1931.

[19] B. Koopman and J. von Neumann. Dynamical systems of continuous spectra. *Proc. Natl. Acad. Sci.*, 18(3):255–263, 1932.

[20] M. Budišić, R. Mohr, and I. Mezić. Applied Koopmanism. *Chaos*, 22(4):047510, 2012.

[21] S. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz. Modern Koopman theory for dynamical systems. *SIAM Review*, 64(2):229–340, 2022.

[22] M. Colbrook. The multiverse of dynamic mode decomposition algorithms. *Handbook of Numerical Analysis*, pages 127–230, 2024.

[23] D. Haggerty et al. Control of soft robots with inertial dynamics. *Science Robotics*, 8(81), 2023.

[24] G. Froyland, D. Giannakis, B. Lintner, M. Pike, and J. Slawinska. Spectral analysis of climate dynamics with operator-theoretic approaches. *Nat. Commun.*, 12(1), 2021.

[25] A. Orvieto et al. Resurrecting recurrent neural networks for long sequences. In *International Conference on Machine Learning*, pages 26670–26698. PMLR, 2023.

[26] J. Proctor and P. Eckhoff. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International Health*, 7(2):139–145, 2015.

[27] B. Brunton, L. Johnson, J. Ojemann, and J. N. Kutz. Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J. Neurosci. Methods*, 258:1–15, 2016.

[28] G. Froyland, D. Giannakis, E. Luna, and J. Slawinska. Revealing trends and persistent cycles of non-autonomous systems with autonomous operator-theoretic techniques. *Nat. Commun.*, 15(1):4268, 2024.

[29] B. Lusch, J. N. Kutz, and S. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nat. Commun.*, 9(1):1–10, 2018.

[30] M. Korda and I. Mezić. On convergence of extended dynamic mode decomposition to the Koopman operator. *J. Nonlinear Sci.*, 28(2):687–710, 2018.

[31] P. Schmid. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, 656:5–28, 2010.

[32] M. Williams, I. Kevrekidis, and C. Rowley. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *J. Nonlinear Sci.*, 25(6):1307–1346, 2015.

[33] Ivo Babuška and John Osborn. Eigenvalue problems. In *Handbook of Numerical Analysis*, volume 2, pages 641–787. Elsevier, 1991.

[34] Françoise Chatelin. *Spectral Approximation of Linear Operators*. Society for Industrial and Applied Mathematics, 2011.

[35] M. Korda, M. Putinar, and I. Mezić. Data-driven spectral analysis of the Koopman operator. *Appl. Comput. Harmon. Anal.*, 48(2):599–629, 2020.

[36] M. Colbrook and A. Townsend. Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. *Commun. Pure Appl. Math.*, 77(1):221–283, 2023.

[37] Mika Rantanen et al. The Arctic has warmed nearly four times faster than the globe since 1979. *Commun. Earth Environ.*, 3(1):168, 2022.

[38] J. Cohen et al. Divergent consensuses on Arctic amplification influence on midlatitude severe winter weather. *Nat. Clim. Change*, 10(1):20–29, 2020.

[39] J. Cohen et al. Linking arctic variability and change with extreme winter weather in the United States. *Science*, 373(6559):1116–1121, 2021.

[40] D. Coumou et al. The influence of Arctic amplification on midlatitude summer circulation. *Nat. Commun.*, 9(1):2959, 2018.

[41] N. Wayand, C. Bitz, and E. Blanchard-Wrigglesworth. A year-round subseasonal-to-seasonal sea ice prediction portal. *Geophys. Res. Lett.*, 46(6):3298–3307, 2019.

[42] J. Lohmann and V. Lucarini. Melancholia states of the Atlantic Meridional Overturning Circulation. *Phys. Rev. Fluids*, 9:123801, Dec 2024.

[43] J. Lohmann, H. Dijkstra, M. Jochum, V. Lucarini, and P. Ditlevsen. Multistability and intermediate tipping of the Atlantic Ocean circulation. *Science advances*, 10(12):eadi4253, 2024.

[44] I. Mezic. Operator is the model. *arXiv:2310.18516*, 2023.

[45] Oussama Zekri, Ambroise Odonnat, Abdelhakim Benechehab, Linus Bleistein, Nicolas Boullé, and Ievgen Redko. Large language models as markov chains. *arXiv preprint arXiv:2410.02724*, 2024.

[46] R. Mañé. *Ergodic theory and differentiable dynamics*, volume 8. 2012.

[47] I. Mezić. *On the geometrical and statistical properties of dynamical systems: Theory and applications*. PhD thesis, 1994.

[48] D. Giannakis and C. Valva. Consistent spectral approximation of Koopman operators using resolvent compactification. *Nonlinearity*, 37(7):075021, 2024.

[49] A. Mauroy, I. Mezić, and J. Moehlis. Isostables, isochrons, and Koopman spectrum for the action–angle representation of stable fixed point dynamics. *Physica D: Nonlinear Phenomena*, 261:19–30, 2013.

[50] M. Colbrook, C. Drysdale, and A. Horning. Rigged dynamic mode decomposition. *SIAM J. Appl. Dyn. Syst.*, 24(2):1150–1190, 2025.

[51] H. Swinney and J. Gollub. The transition to turbulence. *Physics today*, 31(8):41–49, 1978.

[52] H. Broer and F. Takens. Mixed spectra and rotational symmetry. *Archive for rational mechanics and analysis*, 124:13–42, 1993.

[53] J. Fillman and D. Ong. Purely singular continuous spectrum for limit-periodic CMV operators with applications to quantum walks. *J. Funct. Anal.*, 272(12):5107–5143, 2017.

[54] D. Ruelle and F. Takens. On the nature of turbulence. *Les rencontres physiciens-mathématiciens de Strasbourg-RCP25*, 12:1–44, 1971.

[55] M. Stuecker, F. Jin, and A. Timmermann. El Niño- southern oscillation frequency cascade. *Proc. Natl. Acad. Sci.*, 112(44):13490–13495, 2015.

[56] J. Hogg, M. Fonoberova, and I. Mezić. Exponentially decaying modes and long-term prediction of sea ice concentration using Koopman mode decomposition. *Scientific reports*, 10(1):16313, 2020.

[57] J. Stroeve, M. Serreze, M. Holland, J. Kay, J. Malanik, and A. Barrett. The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Climatic change*, 110:1005–1027, 2012.

[58] M. Mori, M. Watanabe, H. Shiogama, J. Inoue, and M. Kimoto. Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades. *Nat. Geosci.*, 7(12):869–873, 2014.

[59] E. Blanchard-Wrigglesworth, K. Armour, C. Bitz, and E. DeWeaver. Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. *J. Climate*, 24(1):231–250, 2011.

[60] M. Bushuk and D. Giannakis. The seasonality and interannual variability of Arctic sea ice reemergence. *J. Climate*, 30(12):4657–4676, 2017.

[61] M. Bushuk, D. Giannakis, and A. Majda. Reemergence mechanisms for North Pacific sea ice revealed through nonlinear Laplacian spectral analysis. *J. Climate*, 27(16):6265–6287, 2014.

[62] S. Lind, R. Ingvaldsen, and T. Furevik. Arctic warming hotspot in the northern Barents Sea linked to declining sea-ice import. *Nat. Clim. Change*, 8(7):634–639, 2018.

[63] T. Andersson, J Scott Hosking, María Pérez-Ortiz, Brooks Paige, Andrew Elliott, Chris Russell, Stephen Law, Daniel C Jones, Jeremy Wilkinson, Tony Phillips, et al. Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nat. Commun.*, 12(1):5124, 2021.

[64] Stephanie Johnson et al. SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3):1087–1117, 2019.

[65] D. Olonscheck, T. Mauritsen, and D. Notz. Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations. *Nat. Geosci.*, 12(6):430–434, 2019.

[66] V. I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer New York, 1989.

[67] T. Hill. *An introduction to statistical thermodynamics*. Courier Corporation, 1986.

[68] A. Kachurovskii. The rate of convergence in ergodic theorems. *Russian Mathematical Surveys*, 51(4):653–703, 1996.

[69] U. Krengel. *Ergodic theorems*. Walter de Gruyter, 2011.

[70] L. N. Trefethen and M. Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. PUP, 2005.

[71] Oscar F. Bandtlow, Wolfram Just, and Julia Slipantschuk. EDMD for expanding circle maps and their complex perturbations. *Appl. Comput. Harmon. Anal.*, to appear.

[72] Igor Mezić and Fotis Sotiropoulos. Ergodic theory and experimental visualization of invariant sets in chaotically advected flows. *Physics of Fluids*, 14(7):2235–2243, 2002.

[73] A. Hansen. On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators. *Journal of the American Mathematical Society*, 24(1):81–124, 2011.

[74] J. Ben-Artzi, M. Colbrook, A. Hansen, O. Nevanlinna, and M. Seidel. Computing spectra - On the solvability complexity index hierarchy and towers of algorithms. *arXiv*, 2020.

[75] M. Colbrook. Computing spectral measures and spectral types. *Commun. Math. Phys.*, 384(1):433–501, 2021.

[76] M. Colbrook. The mpEDMD algorithm for data-driven computations of measure-preserving dynamical systems. *SIAM J. Numer. Anal.*, 61(3):1585–1608, 2023.

[77] Hassan Arbabi and Igor Mezić. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator. *SIAM J. Appl. Dyn. Syst.*, 16(4):2096–2126, 2017.

[78] S. Klus et al. Data-driven approximation of the Koopman generator: Model reduction, system identification, and control. *Physica D: Nonlinear Phenomena*, 406:132416, 2020.

[79] G. Froyland. On Ulam approximation of the isolated spectrum and eigenfunctions of hyperbolic maps. *Discrete and Continuous Dynamical Systems*, 17(3):671–689, 2007.

[80] Michael Blank, Gerhard Keller, and Carlangelo Liverani. Ruelle–Perron–Frobenius spectrum for Anosov maps. *Nonlinearity*, 15(6):1905–1973, 2002.

[81] S. Klus et al. Data-driven model reduction and transfer operator approximation. *J. Nonlinear Sci.*, 28(3):985–1010, 2018.

[82] S. Peitz and S. Klus. Koopman operator-based model reduction for switched-system control of PDEs. *Automatica*, 106:184–191, 2019.

[83] Y. Katznelson. *An Introduction to Harmonic Analysis*. CUP, 2004.

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

[85] G. Zaslavsky. Chaos, fractional kinetics, and anomalous transport. *Physics Reports*, 371(6):461–580, 2002.

[86] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. CUP, Cambridge, 2006.

[87] I. Mezić and A. Banaszuk. Comparison of systems with complex behavior. *Physica D: Nonlinear Phenomena*, 197(1-2):101–133, 2004.

[88] V. Lucarini and M. Chekroun. Theoretical tools for understanding the climate crisis from Hasselmann's programme and beyond. *Nature Reviews Physics*, 5(12):744–765, 2023.

[89] R. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[90] Zongyi L. et al. Fourier neural operator for parametric partial differential equations. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

[91] P. Vlachas, W. Byeon, Z. Wan, T. Sapsis, and P. Koumoutsakos. Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proceedings of the Royal Society A*, 474(2213):20170844, 2018.

[92] Z. Long, Y. Lu, X. Ma, and B. Dong. Pde-net: Learning PDEs from data. In *International conference on machine learning*, pages 3208–3216. PMLR, 2018.

[93] N. Boullé, D. Halikias, and A. Townsend. Elliptic PDE learning is provably data-efficient. *Proc. Natl. Acad. Sci.*, 120(39), 2023.

[94] F. Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence*, pages 366–381. Springer, 2006.

[95] M. Netto and L. Mili. A robust data-driven Koopman Kalman filter for power systems dynamic state estimation. *IEEE Transactions on Power Systems*, 33(6):7228–7237, November 2018.

[96] H. Arbabi, M. Korda, and I. Mezić. A data-driven Koopman model predictive control framework for nonlinear partial differential equations. In *2018 IEEE CDC*, pages 6409–6414, 2018.

[97] A. Narasingam and J. Kwon. Koopman Lyapunov-based model predictive control of nonlinear chemical process systems. *AIChE Journal*, 65(11):e16743, August 2019.

[98] A. Hasnain, N. Boddupalli, S. Balakrishnan, and E. Yeung. Steady state programming of controlled nonlinear systems via deep dynamic mode decomposition. In *IEEE ACC*, pages 4245–4251, 2020.

[99] R. Temam. *Infinite-dimensional dynamical systems in mechanics and physics*, volume 68 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2012.

[100] L. N. Trefethen. *Spectral methods in MATLAB*. SIAM, 2000.