Limits and Powers of Koopman Learning

Matthew J. Colbrook* Igor Mezić† Alexei Stepanenko‡

Abstract

Dynamical systems provide a comprehensive way to study complex and changing behaviors across various sciences. Many modern systems are too complicated to analyze directly or we do not have access to models, driving significant interest in learning methods. Koopman operators have emerged as a dominant approach because they allow the study of nonlinear dynamics using linear techniques by solving an infinite-dimensional spectral problem. However, current algorithms face challenges such as lack of convergence, hindering practical progress. This paper addresses a fundamental open question: *When can we robustly learn the spectral properties of Koopman operators from trajectory data of dynamical systems, and when can we not?* Understanding these boundaries is crucial for analysis, applications, and designing algorithms. We establish a foundational approach that combines computational analysis and ergodic theory, revealing the first fundamental barriers – universal for any algorithm – associated with system geometry and complexity, regardless of data quality and quantity. For instance, we demonstrate well-behaved smooth dynamical systems on tori where non-trivial eigenfunctions of the Koopman operator cannot be determined by any sequence of (even randomized) algorithms, even with unlimited training data. Additionally, we identify when learning is possible and introduce optimal algorithms with verification that overcome issues in standard methods. These results pave the way for a sharp classification theory of data-driven dynamical systems based on how many limits are needed to solve a problem. These limits characterize all previous methods, presenting a unified view. Our framework systematically determines when and how Koopman spectral properties can be learned.

Keywords – dynamical systems, machine learning, Koopman operators, spectra, SCI hierarchy

Dynamical systems are fundamental across the sciences for understanding complex behaviors of systems that evolve over time. Over the past century, they have enabled scientists to model, predict, and control phenomena in fields ranging from physics and chemistry to biology and medicine. However, systems in many applications, such as climate science, neuroscience, robotics, and epidemiology, are increasingly too complicated to describe analytically, or our knowledge of their evolution may be incomplete. The recent success of deep learning has revolutionized our ability to analyze complex data and predict outcomes [1], leading to breakthroughs such as the prediction of protein structures in biology [2], image classification and inverse problems [3], and the development of novel materials [4] and drugs [5]. The emerging field of data-driven dynamical systems aims to extend this success to the study of dynamical systems through trajectory training data; see [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20] for a very small sample. Data-driven approaches to dynamical systems harness the power of machine learning (ML) to extract patterns and principles directly from observations, bypassing the need for explicit model formulations.

A major difficulty encountered in this endeavor is *nonlinearity*. Linear systems are fully characterized by their spectral (eigen) decomposition, but the principle of linear superposition fails for nonlinear dynamical systems. Koopman operators have emerged as a powerful tool to deal with nonlinearity through lifting to an *infinite-dimensional* space of observables, thereby allowing linear tools based on spectral decompositions. In essence, Koopman operators provide a diagonalization of nonlinear dynamical systems. Although Koopman operators are almost a century old going back to Koopman and von Neumann [21, 22], modern Koopman theory has led to a renaissance of data-driven dynamical systems over the past decade [7, 8, 23, 24, 25, 26, 27, 28]. Notable successes include control of robots [29, 30], extracting coherent behavior of climate variability [31, 32], state-of-the-art performance and training of recurrent neural networks [33], discovering patterns in disease spreads [34], analyzing neural recordings in the brain [35], and interpretable neural networks [36].

Despite its potential, practical challenges with infinite-dimensional spectral problems often limit the effectiveness of Koopman theory [27]. For instance, the projection used in dynamic mode decomposition (DMD) methods [37, 38], while serving as a catalyst for algorithms, generally fails to converge and can be unstable, even with perfect data [17, 28] (see Example 1.3 and Figure 2). While some techniques effectively learn certain spectral properties of Koopman operators [7, 39, 40, 41, 42, 43, 44, 45, 46], others like EDMD [47] seem effective but convergence is weak and achieved only along subsequences [48]. This raises a fundamental open question about when and how the robust computation of spectral properties can be ensured:

Question: *When can we robustly learn the spectral properties of Koopman operators from trajectory data of dynamical systems, and when can we not?*

To address this question, we develop a foundational program for Koopman operator learning that links the disparate fields of computations on infinite-dimensional Hilbert spaces, particularly the *Solvability Complexity Index (SCI) hierarchy* [49, 50, 51], and ergodic/dynamical systems theory. The SCI hierarchy classifies the difficulty of computational problems and proves algorithmic optimality by counting the number of successive limits required in a computation, known as *towers of algorithms*. These were first introduced in dynamical systems theory by Smale, who explored the dynamics of iterative rational maps for computing polynomial roots, addressing the issue that algorithms such as Newton's method need not converge. Smale posed the question [52], *"Is there any purely iterative generally convergent algorithm for polynomial zero finding?"*. McMullen [53, 54] answered affirmatively for

*DAMTP, University of Cambridge, Cambridge, CB3 0WA, UK.
†University of California, Santa Barbara, CA 93106, USA.
‡DAMTP, University of Cambridge, Cambridge, CB3 0WA, UK.
Corresponding author: m.colbrook@damtp.cam.ac.uk.
Author contributions: M.J.C. and I.M. designed research and connected SCI and ergodic theory; M.J.C. designed and proved theorems; M.J.C. wrote the S.M. Appendix and developed its results; M.J.C. performed experiments; M.J.C., I.M. and A.S. proved lemmas B.8, B.12, B.14; and M.J.C. and I.M. wrote the paper.

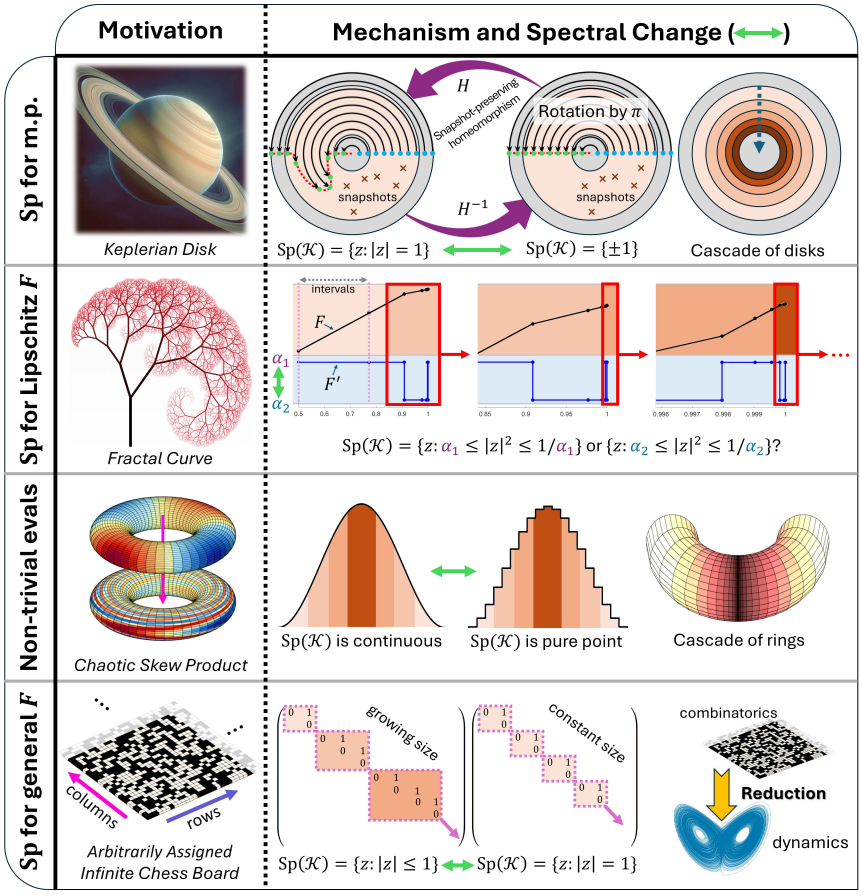

Figure 1: The various changes in the spectrum (green arrows) as we alter the system's geometric properties, which we use to prove lower bounds in this paper. Each row corresponds to a theorem in the paper and a phase transition lemma in the Appendix. The strategy is to embed this behavior into families of dynamical systems so that a learning algorithm cannot converge to the desired spectral property. One can view this as creating an adversarial family of dynamical systems that prevent convergence with probability greater than $1/2$ (i.e., no better than coin flipping).
First row (lower bound for measure-preserving systems in Theorem 2.1): We consider measure-preserving dynamics on the unit disk. At each stage, we alter the dynamical system consistently with the observed data so that it is related to a rotation by a measure-preserving homeomorphism, thus drastically changing the spectrum. This alteration is executed such that the cascade of homeomorphisms and ensuing dynamical systems converge to an underlying limit, providing the adversarial family of dynamical systems.
Second row (lower bound for smooth systems in Theorem 2.1): We consider (smoothed) interval exchange maps, whose derivatives switch in a self-similar manner at the endpoints of the interval. This switching causes the approximated spectrum to alternate between different annuli.
Third row (lower bound for extremely nice systems in Theorem 2.5): We examine skew products on tori, where the spectral type depends on the existence of neighborhoods with constant cross-sections. The spectral change is centered around constructing a skew product through iterations of locally constructed piecewise constant and smooth approximations.
Fourth row (lower bound for discrete-space systems in Theorem 2.4): Here, the spectral change is more abstract since we are proving that three limits are needed. Using an infinite matrix of zeros and ones, we embed the combinatorial problem of deciding whether a matrix contains only finitely many columns with finitely many 1's into the dynamics.

degree 3, but negatively for higher degrees. Doyle and McMullen later strikingly found that the problem could be solved for degrees 4 and 5 using towers of algorithms (successive iterations taken to infinity), but not for degree 6 or higher, no matter the tower's height [55]. This is further detailed in McMullen's Fields Medal citation [56]. Towers of algorithms have also resolved the "classical computational spectral problem" of computing spectra from infinite matrix coefficients [49], such as in discrete Schrödinger operators, a problem dating back to Szegő's work on finite section approximations [57] and Schwinger's foundational studies on Schrödinger operators [58]. We introduce new techniques that extend the SCI hierarchy and expand its applicative scope to data-driven dynamical systems.

Our program tackles upper and lower bounds within the SCI hierarchy. For *upper bounds*, we introduce the first learning algorithms that guarantee convergence under broad conditions, and include verification processes. Our techniques mitigate spurious eigenvalues from DMD methods by locally minimizing spectral distances. We also adapt Fourier tools from mathematical physics (the famous RAGE theorem) into towers of algorithms, introducing novel methods to the Koopman framework. For *lower bounds*, we establish for the first time that certain problems are unsolvable by any learning algorithms, even probabilistic ones with unlimited data (see Figure 1). This involves embedding abrupt spectral changes tied to the system's geometry into dynamics, effectively creating families of adversarial systems where no algorithm can reliably compute spectral properties, performing no better than random chance (probability of convergence $\leq 1/2$).

Therefore, we precisely identify the barriers to robust Koopman learning and classify the complexity of these problems. We lay the groundwork for a classification theory - a toolbox - to determine which dynamical systems and Koopman spectral properties can be learned. Below are some examples of the questions we explore.

Example 1.1. Consider the rotation $\theta \mapsto R(\theta) = \theta + 2\pi\sigma$, on the circle $S^1 = \{e^{i\theta} : \theta \in [-\pi, \pi)\}$, where $\sigma \in \mathbb{R}$. One might ask: Is R ergodic? In other words, does

$$\lim_{n_1\to\infty} \frac{1}{n_1} \sum_{j=0}^{n_1-1} g(T^j(\theta)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\theta)\, \mathrm{d}\theta \;\forall g \in L^2(S^1)? \quad (1)$$

If we know the map symbolically, we can determine the answer (R is ergodic if and only if σ is irrational) in a single limit by testing the equality of σ with rational numbers. However, when relying solely on data from a set of observables (functions $g \in L^2(S^1)$), we require two limits. We need the time limit $n_1 \to \infty$. Then we need a basis of functions $\{g_{n_2}\}_{n_2\in\mathbb{N}}$ on S^1 to verify if Equation (1) holds [59], taking the function limit $n_2 \to \infty$. Importantly, Birkhoff's ergodic theorem allows us to make this claim (without additional tools like RKHS), despite L^2 being a space of equivalence classes of functions, without additional tools like RKHS. ■

Example 1.2. We adapt our discussion to Arnold's circle map: $\theta \mapsto A(\theta) = \theta + 2\pi\sigma + \epsilon \sin(\theta)$, where $\epsilon \in [0, 2\pi)$. When $\epsilon = 0$, we recover our previous example, where the map is not

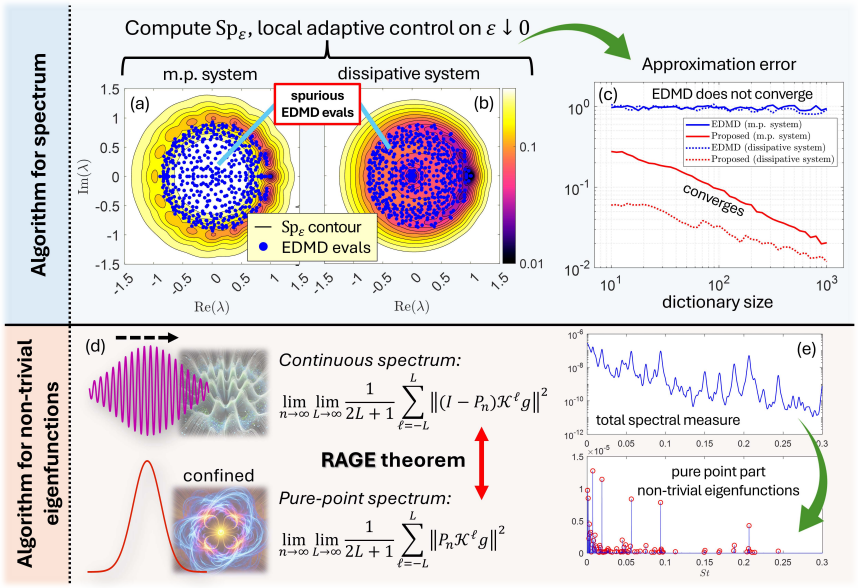

Figure 2: *Top (c.f. Example 2.2):* The algorithm from the proof of Theorem 2.1 for a measure-preserving system (panel (a)) and dissipative system (panel (b)). The proposed algorithm converges, whereas EDMD (the standard workhorse algorithm) does not (panel (c)). The approximation error is measured using pseudospectra techniques, which allows us to compute the spectrum with error control. For a Koopman operator $\mathcal{K}_\mathcal{F}$, the idea of the proposed algorithm is to locally compute pseudospectra $\mathrm{Sp}_\epsilon(\mathcal{K}_\mathcal{F}) = \{\lambda \in \mathbb{C} : \|(\mathcal{K}_\mathcal{F} - \lambda)^{-1}\|^{-1} \leq \epsilon\}$ with a locally adaptive ϵ, effectively computing local minimizers of the residual in Equation (5). *Bottom (c.f. Example 2.6):* The algorithm from Theorem 2.5, which extracts non-trivial eigenfunctions of the Koopman operator, is applied to a high Reynolds number cavity flow with mixed spectral type. The approximation uses the RAGE theorem (panel (d)), which separates continuous and point spectra based on their dynamical behavior using Fourier techniques. This process implicitly involves two limits, which we prove to be optimal. Panel (e) shows the full spectral measure and extracted pure point spectrum. Experimental details for this figure are given in the methods section and Appendix.

only ergodic but uniformly ergodic [60], indicating that any initial condition demonstrates ergodicity. However, for certain values of ϵ, the map A loses its ergodicity for any σ. Here, one might explore the ergodic partition [59], which divides the space into invariant sets where dynamics remain ergodic. This analysis would necessitate a third successive limit $n_3 \to \infty$, where an increasing number of initial conditions are sampled according to an invariant measure. ■

Example 1.3. In our final example on the unit circle, we examine the doubling map: $\theta \mapsto D(\theta) = 2\theta$. This map is measure-preserving but not invertible, and the Koopman spectrum is the unit disk. Trying to compute this spectrum via EDMD requires two successive limits: a large data limit $n_1 = M \to \infty$ followed by a large subspace limit $n_2 \to \infty$. However, using a Fourier basis results in any EDMD matrix post-first limit $n_1 \to \infty$ being a direct sum of Jordan matrices and zeroes [39], yielding a spectrum of $\{0\}$. Thus, significant regions of the spectrum are missed, and EDMD becomes exponentially unstable as $n_2 \to \infty$. Figure 2 shows another warning example with spurious eigenvalues of EDMD. ■

Summary of Main Results

We consider discrete-time dynamical systems

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

Here, the statespace is a compact metric space $(\mathcal{X}, d_\mathcal{X})$, $x \in \mathcal{X}$ is the system's state, $F : \mathcal{X} \to \mathcal{X}$ is an *unknown* continuous function, and n indexes time. The *Koopman operator* is defined via the composition formula:

$$[\mathcal{K}_F g](x) = [g \circ F](x) = g(F(x)), \qquad g \in L^2(\mathcal{X}, \omega), \tag{3}$$

where ω is a finite positive Borel measure. For our upper bounds, ω is general, but in our lower bounds, unless otherwise stated, ω will be the standard Lebesgue measure. Hence, $[\mathcal{K}_F g](x_n) = g(F(x_n)) = g(x_{n+1})$ projects state observations (with "observable" g) forward by one time step, capturing the dynamic progression of the system. The key property of $\mathcal{K}_F$ is its *linearity*, enabling analysis based on spectral decompositions. The cost of this linearization is that $\mathcal{K}_F$ acts on an *infinite-dimensional* function space ($L^2(\mathcal{X}, \omega)$ here, but others are certainly possible). Our goal is to learn spectral properties of $\mathcal{K}_F$ from *snapshot data*

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

Data may originate from long or short trajectories in experiments or simulations. This setting is central to the growing interest in using Koopman operator theory to drive practical applications of trajectory data across diverse fields.

Our results elucidate the conditions for robust Koopman learning, identify fundamental barriers to learning algorithms, and lay the foundation for a classification theory. We show (see Appendix for proofs and further problems):

- **Theorem 2.1:** If F preserves ω (this can be weakened) and its variation is uniformly controlled, there exists a deterministic learning algorithm Γ_n that computes the spectrum of $\mathcal{K}_F$ from snapshots as $n \to \infty$. Unlike EDMD, which generally fails to converge (see Example 1.3 and Figure 2), Γ_n is not based on eigenvalues of a matrix discretization. However, if either of these properties of F is dropped, then unless extreme assumptions are made (e.g., access to a finite-dimensional invariant subspace), no learning algorithm – deterministic or probabilistic – can converge using a single limit.

  NB: Common in ML and scientific computation, we analyze sequences of algorithms indexed by n, which might represent the amount of data $M = f(n)$ used or the number of neural network layers. Our impossibility results are far broader, covering any type of algorithm - n can index anything.

- **Theorem 2.3:** Nevertheless, when dropping the measure-preserving and variation assumptions, it is possible to compute spectra by taking *successive* parameters to infinity. Convergence is achieved as $\lim_{n_k\to\infty}\cdots\lim_{n_1\to\infty}\Gamma_{n_k,\ldots,n_1}$ for learning algorithms $\Gamma_{n_k,\ldots,n_1}$. This approach lays the

groundwork for a classification theory that identifies the necessary number of limits for optimal Koopman learning algorithms. Using these tools, we show that one can robustly learn the spectra of Koopman operators in three limits.

- **Theorem 2.4:** For general systems, computing spectra of $\mathcal{K}_F$ from snapshots requires three successive limits, irrespective of the algorithm used. Generically, this cannot be achieved in two limits. Hence, Theorem 2.3's algorithm is optimal.

- **Theorem 2.5:** Finding or determining the existence of nontrivial eigenfunctions of $\mathcal{K}_F$ in a single limit is impossible, even if the system is measure-preserving, invertible, and F and its inverse are smooth with known bounded Lipschitz constants. However, this can be accomplished with two successive limits. This issue ties directly to the challenge of identifying finite-dimensional coordinate systems and embeddings that appear to linearize the dynamics.

The impossibility results hold even if we allow the algorithm to sample F as many times as it wants to any accuracy it wants. These results are universal; no algorithm, including clever reparametrizations of a convergent procedure $\lim_{n_k\to\infty}\cdots\lim_{n_1\to\infty}\Gamma_{n_k,\ldots,n_1}$ or enhanced neural networks, can circumvent them. As detailed in our methods section and illustrated in Figure 1, we embed sudden changes to the spectrum into the system dynamics to prevent algorithmic convergence. We establish each lower bound for basic choices of $\mathcal{X}$ (disk, interval, torus), and the techniques extend to other $\mathcal{X}$ and function spaces.

The use of successive limits $\lim_{n_k\to\infty}\cdots\lim_{n_1\to\infty}$ may surprise the reader. However, unless one has very strong assumptions about the system, *every* convergent method for Koopman learning in the literature depends on several parameters taken successively to infinity (as it must, see Table 1). For example, $n_1\to\infty$ could correspond to an increasing amount of snapshot data, while $n_2\to\infty$ might correspond to an increasing number of observables g. Key questions are how many limits are required and how the answer depends on the properties of the dynamical system.

2 Fundamental Barriers

In the real world, we never have access to exact snapshot data in Equation (4), which is affected by noise, measurement errors, or finite precision storage. To strengthen our impossibility results, we consider a *perfect measurement device*, whereby data can be approximated and stored to arbitrary precision. We assume access to the training data

$$\mathcal{T}_F=\left\{(x,y_n)\in\mathcal{X}\times\mathcal{X}: x\in\mathcal{X}, d_{\mathcal{X}}(F(x),y_n)\le 2^{-n}\right\}.$$

This means an algorithm can sample any of the points $\{F(x): x\in\mathcal{X}\}$ to any desired accuracy.

Can we learn Koopman spectra?

The most fundamental spectral property of $\mathcal{K}_F$, is its *approximate point spectrum*:

$$\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)=\left\{\lambda\in\mathbb{C}:\exists\, g_n \text{ s.t. } \|g_n\|=1, \lim_{n\to\infty}\|(\mathcal{K}_F-\lambda I)g_n\|=0\right\}.$$

An observable g with

$$\|g\|=1 \quad\text{and}\quad \|(\mathcal{K}_F-\lambda I)g\|\le\epsilon \tag{5}$$

for a scalar $\lambda\in\mathbb{C}$ is known as ϵ-pseudoeigenfunction and is physically relevant since

$$\|\mathcal{K}_F^n g-\lambda^n g\|=\mathcal{O}(n\epsilon)\quad\forall n\in\mathbb{N}. \tag{6}$$

Hence, λ describes an approximate coherent oscillation and decay/growth of the observable g with time. (Pseudo)eigenfunctions and $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ encode information about the underlying dynamical system [61]. For example, they characterize the global stability of equilibria [62], and their level sets determine ergodic partitions [23], invariant manifolds [63], isostables [64].

Methods for learning $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ (such as EDMD) face problems like spurious eigenvalues, missing parts of the spectrum and instabilities. Recent work shows how to avoid spurious eigenvalues by computing pseudospectra [44]. However, it remains an open problem whether $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ can be computed for general systems from trajectory data. We measure convergence in the Hausdorff metric, which captures convergence without spurious eigenvalues or missing spectral regions.

Our theorems solve this problem. First, we show that we can compute $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ from trajectory data if F is measure-preserving (m.p.) (but not necessarily invertible) and we have control on its variation. We set

$$\Omega_{\mathcal{X}}^{\alpha,m}=\{F:\mathcal{X}\to\mathcal{X} \text{ s.t. } F \text{ is m.p., has a mod. of cty. } \alpha\}.$$

Here, $\alpha:\mathbb{R}_{\ge0}\to\mathbb{R}_{\ge0}$ is an increasing continuous function with $\alpha(0)=0$, known as a *modulus of continuity*, that controls the variation of F: $d_{\mathcal{X}}(F(x),F(y))\le\alpha(d_{\mathcal{X}}(x,y))$ for all $x,y\in\mathcal{X}$. Since F is uniformly continuous ($\mathcal{X}$ is compact), there is always such a modulus of continuity α, but when defining $\Omega_{\mathcal{X}}^{\alpha,m}$ we are assuming that a suitable α is known across the whole considered class of dynamical systems. Theorem 2.1 also shows that if we drop these assumptions, computing $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ is no longer possible. In particular, Theorem 2.1 shows a sharp boundary in assumptions needed for Koopman learning. We use $B_{2^{-n}}(0)\subset\mathbb{C}$ and $\mathbb{D}$ to denote the closed ball of radius 2^{-n} center 0 and the open unit disk, respectively.

Theorem 2.1. *There exists a sequence of deterministic learning algorithms $\{\Gamma_n\}$ using $\mathcal{T}_F$ such that:*

Convergence: $\lim_{n\to\infty}\Gamma_n(F)=\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)\ \forall F\in\Omega_{\mathcal{X}}^{\alpha,m}$;

Error control: $\Gamma_n(F)\subset\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)+B_{2^{-n}}(0)\ \forall F\in\Omega_{\mathcal{X}}^{\alpha,m}$.

However, if we drop either assumption (m.p. or uniform modulus of continuity α), there is no sequence of convergent algorithms. For example, let $\Omega=\Omega_{\mathbb{D}}$ or $\Omega_{[0,1]}$, where

$$\Omega_{\mathbb{D}}=\left\{F:\overline{\mathbb{D}}\to\overline{\mathbb{D}} \text{ s.t. } F \text{ cts., m.p., invertible}\right\},$$
$$\Omega_{[0,1]}=\left\{F:[0,1]\to[0,1] \text{ s.t. } F \text{ smooth, } F \,\&\, F^{-1} \text{ Lip. } 2\right\}.$$

- *There **does not** exist deterministic learning algorithms $\{\Gamma_n\}$ using $\mathcal{T}_F$ with $\lim_{n\to\infty}\Gamma_n(F)=\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)\ \forall F\in\Omega$.*

- *For any probabilistic learning algorithms $\{\Gamma_n\}$ using $\mathcal{T}_F$,*

$$\inf_{F\in\Omega}\mathbb{P}\left(\lim_{n\to\infty}\Gamma_n(F)=\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)\right)\le\frac{1}{2}.$$

I.e., performance is no better than random chance.

Theorem 2.1 highlights fundamental barriers regarding the existence of algorithms for Koopman learning. Despite widespread optimism and the undeniable success of applied Koopmanism, Theorem 2.1 underscores the necessity of understanding the assumptions required for the dynamical system. Our results immediately open up a classification theory on which dynamical systems and which problems allow Koopman learning. Several remarks are worth making:

Error control, AI, and computer-assisted proofs: The convergence in Theorem 2.1 includes error control, allowing us to bound how far the output is from the spectrum. This capability is crucial as it assures the *reliability* of the output for applications, including computer-assisted proofs. The significance of this field is rapidly increasing, as illustrated by the recent computer-assisted proof of the blow-up of the 3D Euler equation with smooth initial data [65], considered one of the major open problems in nonlinear PDEs, and the emerging use of neural networks in this area [66].

Example 2.2 (Example of robust Koopman learning)**.** As an example, the top panel of Figure 2 shows the convergence of the algorithm from the proof of Theorem 2.1 for an undamped and damped Duffing oscillator. In contrast, and in general, methods such as EDMD do not converge. ■

Beyond measure-preserving systems: In the upper bound, the measure-preserving assumption can be replaced by systems where we can control how fast the resolvent norm $\|(\mathcal{K}_F - zI)^{-1}\|$ blows up as z approaches the spectrum. Quite general estimates of this type can be obtained for finite rank operators and compact perturbations of self-adjoint and unitary operators, among others [67, 68]. These estimates can be applied to spectral Koopman operators [69, Section 2]. For example, systems with a global (Milnor) attractor of zero Lebesgue measure, where the non-unitary part acts on a domain-truncated modulated Fock space [70].

Randomized algorithms and training: The impossibility results in Theorem 2.1 cover randomized algorithms, including scenarios where trajectories are sampled randomly (e.g., Monte Carlo). In ML, employing a probability distribution over training data and using randomized training algorithms, such as stochastic gradient descent, is standard practice. Theorem 2.1 encompasses all these situations, and no computable probability distribution can negate these limitations.

Probabilistic phase transitions at $1/2$: The phase transition at $1/2$ of the probability of success is expected in *discrete* mathematics. For example, if we wanted to answer a decision problem such as "Is the system ergodic?", we could achieve a probability of success of $1/2$ by flipping a coin to decide "yes" or "no". For such problems, if one could achieve a probability greater than $1/2$, the probability of success can be made arbitrarily close to 1 through repeated trials (essentially, as one would detect a biased coin). We prove that this also occurs for arbitrary computational problems (including those in *continuous* mathematics) in the Appendix.

Learning statistics is not enough: The classes $\Omega_{\mathbb{D}}$ and $\Omega_{[0,1]}$ are well-behaved and allow Monte Carlo integration of integrals involving observables [71]. Hence, Theorem 2.1 shows that, even when we are capable of computing statistical properties of the dynamical system, we may not be able to learn Koopman spectra.

Model of computation: The algorithms can be restricted to Turing machines (digital computation) [72] or BSS machines (real number computation) [73]. However, our results are stronger than those obtained by Turing's techniques. The impossibility results in Theorem 2.1 hold in any model of computation, even for any randomized Turing or BSS machine that can solve the halting problem.

Smoother F and sampling derivatives do not help: The impossibility results hold even if we consider smoother F and allow our algorithms to sample derivatives of F.

Towers of algorithms and classifications

Despite the impossibility results of Theorem 2.1, we can learn spectra in these cases if we drop the requirement of a convergent sequence $\{\Gamma_n\}$. Let

$$\begin{aligned}\Omega_{\mathcal{X}}^{\alpha} &= \{F : \mathcal{X} \to \mathcal{X} \text{ s.t. } F \text{ has has a mod. of cty. } \alpha\},\\ \Omega_{\mathcal{X}}^{m} &= \{F : \mathcal{X} \to \mathcal{X} \text{ s.t. } F \text{ is m.p.}\},\\ \Omega_{\mathcal{X}} &= \{F : \mathcal{X} \to \mathcal{X} \text{ s.t. } F \text{ is continuous}\}.\end{aligned}$$

Theorem 2.3 shows that we can learn $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ using algorithms that converge in *multiple successive limits*. We call these *towers of algorithms*. They allow us to build a classification theory that lays the foundation and mathematical language for robust Koopman learning.

Theorem 2.3. ***Two limits for $\Omega_{\mathcal{X}}^{\alpha}$:*** *There exist deterministic learning algorithms $\{\Gamma_{n_2,n_1}^{\alpha}\}_{n_2,n_1\in\mathbb{N}}$ using $\mathcal{T}_F$ with*

$$\begin{aligned}&\lim_{n_2\to\infty}\lim_{n_1\to\infty}\Gamma_{n_2,n_1}^{\alpha}(F) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) \quad \forall F \in \Omega_{\mathcal{X}}^{\alpha}\\ &\textit{and } \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) \subset \lim_{n_1\to\infty}\Gamma_{n_2,n_1}^{\alpha}(F) + B_{2^{-n_2}}(0). \qquad (7)\end{aligned}$$

Two limits for $\Omega_{\mathcal{X}}^{m}$: *There exists deterministic learning algorithms $\{\Gamma_{n_2,n_1}^{m}\}_{n_2,n_1\in\mathbb{N}}$ using $\mathcal{T}_F$ with*

$$\begin{aligned}&\lim_{n_2\to\infty}\lim_{n_1\to\infty}\Gamma_{n_2,n_1}^{m}(F) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) \quad \forall F \in \Omega_{\mathcal{X}}^{m}\\ &\textit{and } \lim_{n_1\to\infty}\Gamma_{n_2,n_1}^{m}(F) \subset \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) + B_{2^{-n_2}}(0). \qquad (8)\end{aligned}$$

Three limits for $\Omega_{\mathcal{X}}$: *There exists deterministic learning algorithms $\{\Gamma_{n_3,n_2,n_1}\}_{n_3,n_2,n_1\in\mathbb{N}}$ using $\mathcal{T}_F$ with*

$$\begin{aligned}&\lim_{n_3\to\infty}\lim_{n_2\to\infty}\lim_{n_1\to\infty}\Gamma_{n_3,n_2,n_1}(F) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) \quad \forall F \in \Omega_{\mathcal{X}}\\ &\textit{and } \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) \subset \lim_{n_2\to\infty}\lim_{n_1\to\infty}\Gamma_{n_3,n_2,n_1}(F) + B_{2^{-n_3}}(0).\end{aligned}$$

The final part of Theorem 2.3 shows the remarkable result that we can robustly learn spectra of general Koopman operators for continuous F. The assumption of continuity can be weakened but is natural since discontinuous F can lead to pathologies such as the nonexistence of ergodic partitions [74].

The SCI for Koopman

The phenomenon in Theorem 2.3 is captured by the SCI hierarchy, which classifies the difficulty of computational problems based on the smallest number of limits needed to solve them. For a class Ω of dynamical systems and map $\Xi : \Omega \to (\mathcal{M}, d)$ to a metric space (e.g., $\Xi(F) = \mathrm{Sp}(\mathcal{K}_F)$ mapping to the Hausdorff metric $(\mathcal{M}, d)$), the SCI is the fewest number of limits

needed to learn Ξ from $\mathcal{T}_F$. Hence, the final part of Theorem 2.3 says that computing $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ has SCI ≤ 3. In Theorem 2.4, we show that this is sharp (SCI ≥ 3), even for discrete-space dynamical systems where one does not have to worry about the variation of F. A problem lies in Δ_{m+1} if it has SCI $\leq m$, and it lies in Δ_1 if there exists a sequence of deterministic learning algorithms $\{\Gamma_n\}$ such that $d(\Gamma_n(F), \Xi(F)) \leq 2^{-n}$ for all $F \in \Omega$.

Theorem 2.3 shows more than convergence. The penultimate limit in Equation (7) is *contained in* the desired set, up to a controllable error. We call this Σ_2, with a subscript indicating the number of limits. Similarly, the penultimate limit in Equation (8) *covers* the desired set, up to a controllable error. We call this Π_2. These refined classes capture the notion of verification. For example, the first part of Theorem 2.1 demonstrates Σ_1 convergence. Figure 3 summarizes the SCI hierarchy for computing the spectrum of Koopman operators. The SCI hierarchy also contains a calculus, allowing users to mix and match assumptions. For example, the class $\Omega_{\mathcal{X}}^{\alpha} \cap \Omega_{\mathcal{X}}^{m}$ is in $\Sigma_1^A \cap \Pi_1^A = \Delta_2^A$. We show the stronger result of verification - that the class is in $\Sigma_1^A \subset \Delta_2^A$. Similarly, Figure 4 summarizes the SCI hierarchy for determining the spectral type of systems in $\Omega_{\mathcal{X}}^{\alpha} \cap \Omega_{\mathcal{X}}^{m}$. Our results span various classes but represent just the beginning of a vast classification theory.

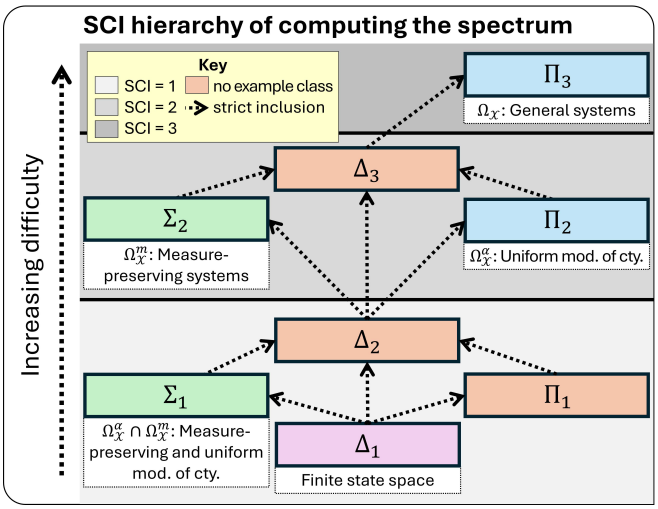

Figure 3: The SCI hierarchy (SCI shaded gray) for the computational problem of learning spectra of Koopman operators from trajectory data. The Σ and Π classes capture complementary versions of verification in the final limit. Results are given in Theorems 2.1, 2.3 and 2.4, which provide both upper and lower bounds and are expanded upon in the Appendix. The case of finite statespace is Δ_1 since the Koopman operator reduces to a finite matrix.

Multiple limit phenomenon in Koopman algorithms

The SCI hierarchy directly connects with the literature on algorithms that learn the spectral properties of Koopman operators. Table 1 contains convergence results in the Koopman literature and the corresponding upper bounds in the SCI hierarchy. Each algorithm comes with assumptions made about the dynamical system. Note that some of these techniques are not sharp, as they use more limits than needed.

Beyond Table 1, Dellnitz and Junge [75] exemplify multiple successive limits in data-driven dynamical systems. Their seminal work on approximating SRB measures involves three successive limits: the first is for Monte Carlo integration, the second increases the dimension of a finite-dimensional subspace to compute spectra of a compact operator, particularly the invariant measure of a randomly perturbed system, and the final limit pertains to a smoothness parameter (see their Theorem 4.4). It is common to use two successive limits when using Ulam's method to approximate isolated eigenvalues of quasicompact Perron–Frobenius (the pre-dual of the Koopman operator), corresponding to Monte Carlo approximation and finite matrix size, respectively [76, 77, 78], see also the Fourier scheme in [79]. Ulam's method need not always converge (see [80, Section 2.6] for an example of spectral pollution), which can often be rectified by smoothing with noise [80]. In some situations, it is also possible to adaptively choose the size of this noise based on the partition diameter to ensure a two-limit process [80, Theorem 4]. As in some of the classes considered in this paper, one can reduce the number of successive limits if one can control the quadrature error. The same remarks for EDMD in Table 1 hold for related data-driven dimension reduction techniques for dynamical systems [81]. For multiple successive limits in control, see [82, Theorem 3].

These examples provide upper bounds. A key question in algorithm design for each method is whether they are optimal. Can convergence be achieved with fewer limits, and if not, what assumptions about the system are necessary to make

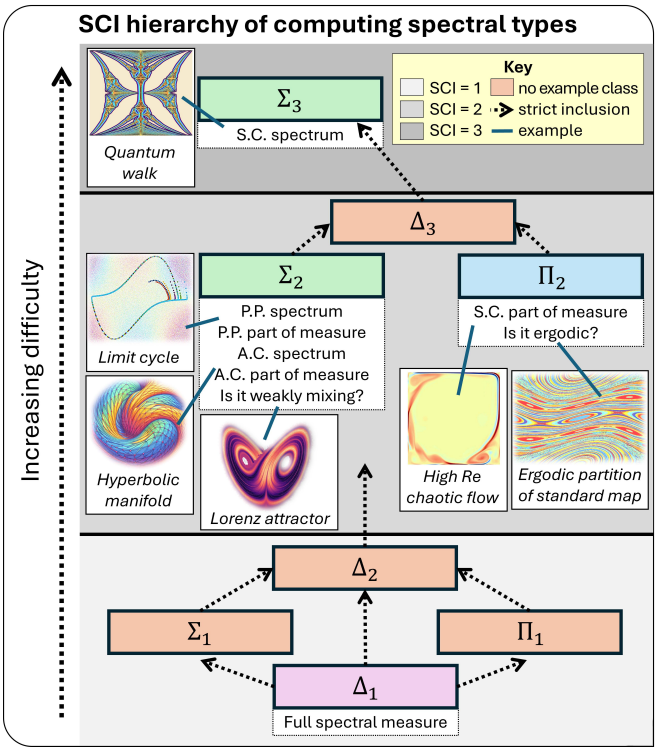

Figure 4: The SCI hierarchy for the computational problem of learning spectral types of Koopman operators for measure-preserving invertible systems (i.e., the unitary parts of Koopman operators for $\Omega_{\mathcal{X}}^{\alpha} \cap \Omega_{\mathcal{X}}^{m}$) from trajectory data. Results are detailed in Theorem 2.5, which provides both upper and lower bounds and are further expanded upon in the Appendix. (The full spectral measure can be computed with error control with respect to the 1-Wasserstein distance that captures weak convergence.) For each problem, we have presented a representative dynamical system to underscore that the classification of these computational problems fundamentally depends on the flavor of dynamics.

Table 1: Convergence results from the Koopman literature interpreted in the context of the SCI hierarchy. Precise statements can be found in the theorems of the cited papers. An entry "N/C" indicates that the method need not converge without additional strong assumptions (such as the observable being inside a finite-dimensional invariant subspace), and an entry "n/a" indicates that the algorithm is not applicable to the spectral problem. A superscript * indicates that the SCI upper bound is reduced by 1 if we can control quadrature errors (e.g., through known variational bounds on F). This also depends on the properties of the system. For example, when sampling a single trajectory, whether the system is uniformly ergodic or mixing [83]. Unless specified otherwise, all examples are for discrete-time systems. Though mpEDMD and the compactification methods need not converge to the spectrum due to spectral pollution, they approximate all of the spectrum.

Algorithm	**Comments/Assumptions**	**Spectral Problem's Corresponding SCI Upper Bound**			
		KMD	*Spectrum*	*Spectral Measure (if m.p.)*	*Spectral Type (if m.p.)*
Extended DMD [47]	general L^2 spaces	SCI $\leq 2^*$	N/C	N/C	n/a
Residual DMD [44]	general L^2 spaces	SCI $\leq 2^*$	SCI $\leq 3^*$	SCI $\leq 2^*$	varies, see [84] e.g., a.c. density: SCI $\leq 2^*$
Measure-preserving EDMD [45]	m.p. systems	SCI ≤ 1	N/C	SCI $\leq 2^*$ (general) SCI ≤ 1 (delay-embedding)	n/a
Hankel DMD [85]	m.p. ergodic systems	SCI $\leq 2^*$	N/C	N/C	n/a
Periodic approximations [86]	m.p. + ω a.c.	SCI ≤ 2	N/C	SCI ≤ 2 (see [87])	a.c. density: SCI ≤ 3
Christoffel–Darboux kernel [40]	m.p. ergodic systems	SCI ≤ 3	n/a	SCI ≤ 2	e.g., a.c. density: SCI ≤ 2
Generator EDMD [88]	cts.-time, samples ∇F (otherwise additional limit)	SCI ≤ 2	N/C	SCI ≤ 2 (see [89])	n/a
Compactification [42]	cts.-time, m.p. ergodic systems	SCI ≤ 4	N/C	SCI ≤ 4	n/a
Resolvent compactification [43]	cts.-time, m.p. ergodic systems	SCI ≤ 5	N/C	SCI ≤ 5	n/a
Diffusion maps [90] (see also [10])	cts.-time, m.p. ergodic systems	SCI ≤ 3	n/a	n/a	n/a

it easier? Addressing this requires proving lower bounds, as we do in this paper. Establishing computational boundaries involves various techniques for constructing convergent algorithms (upper bounds) and proving impossibility results (lower bounds), depending on the type of dynamical system and the quantities we aim to compute.

General systems require three limits

Theorem 2.3 shows that the spectra of general Koopman operators can be computed using three limits (SCI $\leq$ 3). We establish that this is optimal; for general systems, computing $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ inherently requires three limits (SCI $\geq$ 3).

Let $\mathcal{X} = \mathbb{N}$ (or any countable discrete space), equipped with the usual counting measure $\omega = \sum_{j=1}^{\infty} \delta_j$, and consider dynamical systems governed by a nonlinear function $F : \mathbb{N} \to \mathbb{N} \cup \{\infty\}$. Such discrete-space systems are well studied since they are often conjugate to chaotic systems on continuous spaces; hence, we may also translate computational problems and classifications to continuous spaces. The Koopman operator acts on sequences $a : \mathbb{N} \to \mathbb{C}$ via

$$[\mathcal{K}_F a](j) = \begin{cases} a(F(j)), & \text{if } F(j) \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

We consider $\mathcal{K}_F$ as an operator on $L^2(\mathbb{N}, \omega) \cong l^2(\mathbb{N})$, and assume that it is bounded. Let

$$\Omega_{\mathbb{N}} = \{F \text{ s.t. } F : \mathbb{N} \to \mathbb{N} \cup \{\infty\} \text{ and } \mathcal{K}_F \text{ is bounded}\}$$

and consider the training set $\mathcal{T}_{\mathbb{N}} = \{S_{i,j} : S_{i,j}(F) = \delta_{F(j),i}\}$, corresponding to discrete samples of F.

Theorem 2.4. *The problem of learning* $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ *from* $\mathcal{T}_{\mathbb{N}}$ *for* $F \in \Omega_{\mathbb{N}}$ *lies in* Π_3 *and has* SCI $= 3$*. That is, we can learn* $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ *in three successive limits but not two.*

Learning non-trivial eigenpairs is very hard, even for simple well-behaved systems

As a final problem, we consider whether we can learn *non-trivial* eigenvalues and eigenfunctions. This question is often the focus of research in applied Koopmanism and corresponds to finding a coordinate system in which the dynamics are (or appear) linear. The constant function is an eigenfunction with eigenvalue $\lambda = 1$ but is trivial and contains no dynamical information. An eigenfunction $g \in \mathrm{span}\{1\}^{\perp}$ is non-trivial. A system is weakly mixing precisely when there are no non-trivial eigenfunctions. We let $\mathrm{Sp}_{\mathrm{pp}}(\mathcal{K}_F)$ denote the closure of the set of eigenvalues of $\mathcal{K}_F$, i.e., the so-called pure point spectrum.

Theorem 2.5 shows that computing non-trivial eigenfunctions is impossible in one limit, even with randomized methods, if F and F^{-1} are smooth functions with a known Lipschitz constant, and the statespace is as simple as a torus.

Theorem 2.5. *Consider the torus* $\mathcal{X} = [-\pi, \pi]^2_{\mathrm{per}}$ *and*

$$\Omega_p = \{F : \mathcal{X} \to \mathcal{X} \textit{ s.t. } F \textit{ m.p., } F \,\&\, F^{-1} \textit{ smooth, Lip. 2}\}.$$

There ***are no*** *learning algorithms (deterministic or probabilistic with success probability* $> 1/2$*)* Γ_n *using* $\mathcal{T}_{\mathbb{N}}$ *with*

$$\lim_{n\to\infty} \Gamma_n(F) = \begin{cases} 1, & \textit{if } \exists \textit{ a non-trivial eigenfunction}, \\ 0, & \textit{otherwise}, \end{cases} \quad \forall F \in \Omega_p.$$

There ***are no*** *learning algorithms (deterministic or probabilistic with success probability* $> 1/2$*)* Γ_n *using* $\mathcal{T}_{\mathbb{N}}$ *with*

$$\lim_{n\to\infty} \Gamma_n(F) = \mathrm{Sp}_{\mathrm{pp}}(\mathcal{K}_F) \quad \forall F \in \Omega_p.$$

However, both problems lie in Σ_2 *(two limits with verification).*

This result immediately explains why obtaining finite-dimensional coordinate systems and embeddings (e.g., autoencoders and latent space representation) in which the dynamics are linear is a considerable challenge. Instead, one must settle for approximate pseudoeigenfunctions as in $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ and Equation (6), for which the problem is Σ_1 (see Figure 3). Our Σ_2-tower depends on two parameters: a time lag for auto-correlations (the number of data points depends adaptively on this) and a sequence of increasing projections onto finite-dimensional subspaces (according to a dictionary). Theorem 2.5 proves this is optimal - no one limit procedure for detecting or computing non-trivial eigenvalues exists.

Example 2.6. The bottom panel of Figure 2 shows the application of the Σ_2-tower for extracting non-trivial eigenvalues of a high Reynolds number cavity flow with mixed spectral type. The two limits are motivated by the so-called discrete-time RAGE theorem from mathematical physics [91], which is a consequence of Wiener's classical lemma, which relates the asymptotic behavior of Fourier coefficients of a measure to its atomic part. ■

3 Discussion

We have developed a framework that reveals both the potential and the limitations of applying Koopman operators in the data-driven analysis of dynamical systems. Our results include upper bounds (algorithms with guaranteed convergence and verification) and, for the first time, lower bounds (impossibility theorems) for robust Koopman learning. Through several key theorems, we reveal the conditions under which the spectral properties of Koopman operators can be robustly learned, and we identify scenarios where learning is inherently impossible. Our lower bounds and techniques can easily be generalized to other dynamical systems. Specifically, we uncover the exact barriers to learning spectral properties, such as detecting non-trivial eigenfunctions.

This study introduces successive limits and the SCI hierarchy to applied Koopmanism, offering a classification for the computational and learning challenges present in dynamical systems. The hierarchy is shown in Figures 3 and 4 for computing spectra and spectral type, respectively. This approach not only paves the way for applying Koopman operator theory to real-world problems but also confirms that the spectral properties of these operators can be learned under suitable conditions. It encompasses all previous convergence results in the literature, now providing a unified language for developing lower bounds and proving optimality.

Our results are only the start of traversing the boundaries of robust Koopman learning and developing a theory of necessary and sufficient conditions. Future problems amenable to our techniques include:

- Dealing with other spaces of observables than $L^2(\Omega, \omega)$. This also useful for transfer operators that act on distributions. A key question is whether the difficulty of a computational problem (and whether it can even be solved) changes with the function space.
- Cases where we only have partial observations of the system instead of the full snapshots in Equation (4).
- Koopman generators for continuous-time systems.
- What are the boundaries of Koopman control? E.g., our upper bounds open the door to verified control.
- What are the boundaries for learning other objects, such as invariant measures of the system?
- Beyond eigenvalues and spectra, what are the boundaries of learning the spectral type of Koopman operators? Some of these problems are extremely difficult - in the Appendix, we give an example where determining whether a system is ergodic has $\mathrm{SCI} = \infty$.

The techniques of this paper may have applications in other areas of scientific computation in ML. For example, recent results show that it is possible to learn elliptic PDEs from training data consisting of pairs of random forcing terms and corresponding solutions [92]. This is similar in spirit to the snapshot setting of learning Koopman operators. It is currently unknown whether hyperbolic PDEs can be learned, and our proof techniques could shed light on this problem.

4 Methods

For *lower bounds* (impossibility results), we embed specific phase transitions of the spectral properties of $\mathcal{K}_F$ into the dynamics, ensuring they are consistent with the sampled training (trajectory) data. To prove $\mathrm{SCI} > 1$, we proceed by contradiction, assuming that a convergent sequence of algorithms exists. We construct an adversarial family of dynamical systems, taking into account the probability of success, to ensure that convergence cannot occur with a probability greater than $1/2$. To prove $\mathrm{SCI} > 2$, we embed specific combinatorial problems from descriptive set theory into the dynamics. Each lower bound we establish depends on different phase transitions, as detailed in Figure 1, where we also motivate each phase transition. This proof method systematically links the foundations of computation with classical ergodic theory and can be extended to problems beyond those considered in this paper.

We also extend the mathematics behind the SCI hierarchy in several directions, including going beyond the matrix viewpoint to approximation theory and sampling the function F, as well as the allowance for sequences of probabilistic algorithms. (See the concept of a sequence of probabilistic general algorithms we develop and the proven structure theorems in the Appendix.) These tools enable us to prove impossibility results in any model of computation while capturing concepts such as adaptive and probabilistic sampling of the function F.

The construction of algorithms for *upper bounds* depends on the spectral problem. To compute the spectrum, we provide a general scheme for arbitrary compact metric spaces $\mathcal{X}$ and Borel measures ω. This involves building a suitable dictionary and using it to approximate the function $z \mapsto \|(\mathcal{K}_F - zI)^{-1}\|^{-1}$. Similar in spirit to a barrier function (used to guide optimization algorithms), this function vanishes on the spectrum $\mathrm{Sp}(\mathcal{K}_F)$ and increases as z moves further away from $\mathrm{Sp}(\mathcal{K}_F)$. We then locally minimize this function to approximate spectra. A crucial component is generalizing and extending the techniques of ResDMD, which employed a Galerkin approximation of $\mathcal{K}_F^*\mathcal{K}_F$ using trajectory data to approximate pseudospectra (level sets of $\|(\mathcal{K}_F - zI)^{-1}\|^{-1}$). This approach allows us to rigorously approximate $\|(\mathcal{K}_F - zI)^{-1}\|^{-1}$ from above with locally uniform convergence while avoiding issues such as spurious eigenvalues. We show how this leads to the sharp upper bounds shown in Figure 3 for different classes of dynamical systems. In contrast to DMD methods, our techniques do not rely on computing eigenvalues of a matrix discretization of $\mathcal{K}_F$. To compute non-trivial eigenfunctions, we use the RAGE theorem to compute the pure point part of spectral measures of observables. We use these approximations across different observables and a nest of dyadic spectral intervals to filter out the pure point spectrum (eigenvalues).

Appendix of Supplementary Material

Here, we provide proofs of theorems, and detailed explanations of the experimental setup. These results are briefly summarized as follows, and extend upon the theorems presented in the main text:

- **Theorem B.1:** As a warm-up, for continuous, measure-preserving and invertible maps on the unit circle, the SCI of computing the spectrum is 1, the SCI of determining ergodicity is 2, and after dropping the assumption of continuity, the SCI of computing the spectrum is ∞.
- **Theorem B.5:** For general statespace $\mathcal{X}$, the SCI of computing the spectrum of measure-preserving systems with uniformly bounded modulus of continuity is ≤ 1, the SCI of computing the spectrum of measure-preserving systems is ≤ 2, the SCI of computing the spectrum of systems with uniformly bounded modulus of continuity is ≤ 2, the SCI of computing the spectrum of continuous systems is ≤ 3.
- **Theorem B.7:** For continuous, measure-preserving and invertible maps on the unit disk, the SCI of computing the spectrum is 2, i.e., controlling the variability of F cannot be dropped if we want to compute the spectrum in one limit.
- **Theorem B.10:** For smooth, invertible maps on the unit interval, with uniformly bounded derivatives (for F and F^{-1}), the SCI of computing the spectrum is 2, i.e., the measure-perserving condition cannot be dropped if we want to compute the spectrum in one limit.

In all of the above theorems, we also prove the relevant Σ or Π classifications.

A Background

This section provides background information regarding Koopman operators and the Solvability Complexity Index (SCI) for the proofs. We end this section by discussing the paper's setup.

A.1 Koopman operators

Throughout, we consider discrete-time dynamical systems:

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

Here, $x \in \mathcal{X}$ denotes the state of the system, and the metric space $(\mathcal{X}, d_{\mathcal{X}})$ denotes the statespace. Often, $\mathcal{X} \subset \mathbb{R}^d$, though this is not required in what follows (unless explicitly stated). The function $F : \mathcal{X} \to \mathcal{X}$ governs the evolution of the dynamical system and is generally nonlinear. (Here, we mean that F is linear if $\mathcal{X}$ is a vector space and F is linear. Otherwise, we call F nonlinear.) We lift the system into a (typically infinite-dimensional) vector space of observable functions using a Koopman operator to deal with the nonlinearity. A Koopman operator [21, 22] is defined on a Banach space $\mathcal{F}$ of functions $g : \mathcal{X} \to \mathbb{C}$, where the functions g are referred to as *observables* and measure the state of the system. Koopman operators allow us to study the evolution of observables in $\mathcal{F}$ through a linear framework. The Koopman operator is defined via the composition formula:

$$[\mathcal{K}g](x) = [g \circ F](x) = g(F(x)), \qquad g \in \mathcal{D}(\mathcal{K}),$$

were $\mathcal{D}(\mathcal{K}) \subset \mathcal{F}$ is a suitable domain. This definition means that $[\mathcal{K}g](x_n) = g(F(x_n)) = g(x_{n+1})$ represents the measurement of the state one time-step ahead of $g(x_n)$, and hence that $\mathcal{K}$ effectively captures the dynamic progression of the system.

The critical property of the Koopman operator $\mathcal{K}$ is its *linearity*. This linearity holds irrespective of whether the map F in Equation (2) is linear or nonlinear. Consequently, the spectral properties of $\mathcal{K}$ become a powerful tool in analyzing the dynamical system's behavior. The Koopman operator is not defined uniquely by the dynamical system in (2), but fundamentally depends on the space $\mathcal{F}$. Throughout, we focus on the choice

$$\mathcal{F} = \mathcal{D}(\mathcal{K}) = L^2(\mathcal{X}, \omega) \quad \text{with inner product} \quad \langle g_1, g_2 \rangle = \int_{\mathcal{X}} g_1(x)\overline{g_2(x)}\, \mathrm{d}\omega(x) \quad \text{and norm} \quad \|g\| = \sqrt{\langle g, g \rangle},$$

for some positive measure ω.[1] In going from a pointwise definition in (3) to the space $L^2(\mathcal{X}, \omega)$, a little care is needed since $L^2(\mathcal{X}, \omega)$ consists of equivalence classes of functions. We assume that the map F is nonsingular with respect to ω, meaning that

$$\omega(E) = 0 \quad \text{implies that} \quad \omega(\{x : F(x) \in E\}) = 0.$$

This ensures that the Koopman operator is well-defined since $g_1(x) = g_2(x)$ for ω-almost every x implies that $g_1(F(x)) = g_2(F(x))$ for ω-almost every x. The pushforward measure is defined as $F\#\omega(E) = \omega(F^{-1}(S))$, and the fact that F is nonsingular with respect to ω is equivalent to saying that $F\#\omega$ is absolutely continuous with respect to ω. We assume that $\mathcal{K}$ is a bounded linear operator on the Hilbert space $L^2(\mathcal{X}, \omega)$. This assumption is equivalent to saying that the Radon–Nikodym derivative

[1] We do not assume that this measure is invariant. For Hamiltonian systems, a common choice of ω is the standard Lebesgue measure, for which the Koopman operator is unitary on $L^2(\mathcal{X}, \omega)$. For other systems, we can select ω according to the region where we wish to study the dynamics, such as a Gaussian measure. In many applications, ω corresponds to an unknown ergodic measure on an attractor.

$dF\#\omega/d\omega$ lies in $L^\infty(\mathcal{X},\omega)$. The above Hilbert space setting is standard in the Koopman literature, though our results can be extended to other function spaces such as those studied in [70]. Once $(\mathcal{X},\omega)$ are specified, we let $\mathcal{K}_F$ denote the corresponding Koopman operator on the corresponding Hilbert space $L^2(\mathcal{X},\omega)$.

Since $\mathcal{K}_F$ acts on an *infinite-dimensional* function space, we have exchanged the nonlinearity in (2) for an infinite-dimensional linear system. This means that the spectral properties of $\mathcal{K}_F$ can be significantly more complex than those of a finite matrix, making them more challenging to compute. A message of this paper is that, in most cases, unless strong assumptions are made regarding the system, the spectral properties of $\mathcal{K}_F$ are impossible to compute in a single limit, even if we had a perfect measurement device to sample trajectories of the dynamical system.

A.1.1 Koopman spectrum

If $g \in L^2(\mathcal{X},\omega)$ is an *eigenfunction* of $\mathcal{K}_F$ with *eigenvalue* λ, then g exhibits perfect coherence[2] since

$$g(x_n) = [\mathcal{K}_F^n g](x_0) = \lambda^n g(x_0) \quad \forall n \in \mathbb{N}. \tag{9}$$

The oscillation and decay/growth of the observable g are dictated by the complex argument and absolute value of the eigenvalue λ, respectively. In infinite dimensions, the appropriate generalization of the set of eigenvalues of $\mathcal{K}_F$ is the *spectrum*:

$$\mathrm{Sp}(\mathcal{K}_F) = \left\{ z \in \mathbb{C} : (\mathcal{K}_F - zI)^{-1} \text{ does not exist as a bounded operator} \right\} \subset \mathbb{C}.$$

Here, I denotes the identity operator. In contrast to finite matrices, the spectrum $\mathrm{Sp}(\mathcal{K}_F)$ may contain points that are not eigenvalues in addition to eigenvalues. This phenomenon occurs because there are more ways for $(\mathcal{K}_F - zI)^{-1}$ to not exist in infinite dimensions than in finite dimensions. For example, the standard Lorenz system on the Lorenz attractor has a Koopman operator with no nontrivial eigenvalues [93]. The *approximate point spectrum* is

$$\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) = \left\{ \lambda \in \mathbb{C} : \exists \{g_n\}_{n\in\mathbb{N}} \subset L^2(\mathcal{X},\omega) \text{ such that } \|g_n\| = 1,\ \lim_{n\to\infty} \|(\mathcal{K}_F - \lambda I) g_n\| = 0 \right\} \subset \mathrm{Sp}(\mathcal{K}_F) \subset \mathbb{C}$$

and for $\epsilon > 0$, the *approximate point pseudospectrum* is

$$\mathrm{Sp}_{\mathrm{ap},\epsilon}(\mathcal{K}_F) = \left\{ \lambda \in \mathbb{C} : \exists \{g_n\}_{n\in\mathbb{N}} \subset L^2(\mathcal{X},\omega) \text{ such that } \|g_n\| = 1,\ \lim_{n\to\infty} \|(\mathcal{K}_F - \lambda I) g_n\| \leq \epsilon \right\} \subset \mathrm{Sp}_\epsilon(\mathcal{K}_F) \subset \mathbb{C}.$$

An observable g with $\|g\| = 1$ and $\|(\mathcal{K}_F - \lambda I)g\| \leq \epsilon$ for $\lambda \in \mathbb{C}$ is known as ϵ-pseudoeigenfunction. Such observables satisfy

$$\|\mathcal{K}_F^n g - \lambda^n g\| = \mathcal{O}(n\epsilon) \quad \forall n \in \mathbb{N}.$$

In other words, λ describes an approximate coherent oscillation and decay/growth of the observable g with time. The (pseudo)eigenfunctions and $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ encode information about the underlying dynamical system [61]. For example, the level sets of certain eigenfunctions determine ergodic partitions [59, 23], invariant manifolds [63], isostables [64], and the global stability of equilibria [62] can be characterized by pseudoeigenfunctions and $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$.

In this paper, we will focus on the computation of $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$. We anticipate that further foundational results can be proven on the computation of other spectral properties of Koopman operators, such as spectral type. For example, see Theorem B.13 regarding the detection of non-trivial eigenfunctions.

Two special classes of Koopman operators are defined as follows:

- Measure-preserving systems: The dynamical system preserves ω if and only if $\mathcal{K}_F$ is an isometry, that is $\mathcal{K}_F^*\mathcal{K}_F = I$.
- Measure-preserving invertible systems: The dynamical system preserves ω and is invertible modulo ω-null sets [94, Chapter 7] if and only if $\mathcal{K}_F$ is unitary, that is $\mathcal{K}_F^*\mathcal{K}_F = \mathcal{K}_F\mathcal{K}_F^* = I$.

Note that if $\mathcal{K}_F$ is an isometry, but not unitary, then the spectrum of $\mathcal{K}_F$ is the unit disc and $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) = \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. If $\mathcal{K}_F$ is unitary, then the spectrum of $\mathcal{K}_F$ is equal to $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ and is a subset of $\mathbb{T}$.

A.2 The Solvability Complexity Index – classifying the difficulty of problems

We now outline the fundamentals of the Solvability Complexity Index Hierarchy. This tool allows us to precisely classify the difficulty of computational problems and prove that algorithms are optimal, realizing the boundaries of what is possible. We give the definitions of the hierarchy before specializing to the computational setup of this paper, where we give a precise formulation of a *perfect measuring device*.

[2] Coherence here is meant in the sense of an observation for which all the points in state space exhibit the same, (complex) exponential, time-dependence.

A.2.1 Computational problems and general algorithms

Before classifying the difficulty of computational problems, we must precisely define what a computational problem means. Precision here is essential since altering the information an algorithm is permitted to use can significantly affect the difficulty of a problem or even whether the problem can be solved at all. The following definition of a computational problem is deliberately general, designed to encompass all types of problems encountered in computational mathematics. For example, as well as spectral problems, the SCI hierarchy has been applied to other areas of mathematics, including PDEs [95, 96], the limits of AI, and Smale's 18th problem [97], and optimization [98].

Definition A.1 (Computational problem)**.** *The basic objects of a computational problem are:*

- *A primary set, Ω, that describes the input class;*
- *A metric space $(\mathcal{M}, d)$;*
- *A problem function $\Xi : \Omega \to \mathcal{M}$;*
- *An evaluation set, Λ, of functions on Ω.*

The problem function Ξ is the object we want to compute, with the notion of convergence captured by the metric space $(\mathcal{M}, d)$. The evaluation set Λ describes the information that algorithms can read. We require that Λ separates elements of Ω to the degree of separation achieved by Ξ:

$$\text{if } A, B \in \Omega \text{ with } \Xi(A) \neq \Xi(B), \text{ then } \exists f \in \Lambda \text{ with } f(A) \neq f(B). \tag{10}$$

In other words, any $\Xi(A) \in \mathcal{M}$ is uniquely determined by the set of evaluations $\{f(A) : f \in \Lambda\}$ (otherwise it is impossible to recover Ξ from Λ). We refer to the collection $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ as a computational problem.

Example A.2. In the above setting of dynamical systems, we can fix the metric space $(\mathcal{X}, d_{\mathcal{X}})$ and the measure ω. The primary set Ω could be a class of functions $F : \Omega \to \Omega$, each of which induces a dynamical system with bounded $\mathcal{K}_F$ on $L^2(\Omega, \omega)$. The problem function could describe a spectral property of $\mathcal{K}_F$. For example, we could consider the computation of the approximate point spectrum with $\Xi(F) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$. Since $\mathcal{K}_F$ is bounded, $\Xi(F)$ is a compact subset of $\mathbb{C}$. Hence, we let $(\mathcal{M}, d)$ be the Hausdorff metric. The Hausdorf metric space, $(\mathcal{M}_{\mathrm{H}}, d_{\mathrm{H}})$, is the collection of non-empty compact subsets of $\mathbb{C}$ equipped with the Hausdorff metric:

$$d_{\mathrm{H}}(X, Y) = \max\left\{\sup_{x \in X} \inf_{y \in Y} |x - y|, \sup_{y \in Y} \inf_{x \in X} |x - y|\right\}, \quad X, Y \in \mathcal{M}_{\mathrm{H}}.$$

Convergence to the spectrum in this metric means our algorithms converge without spectral pollution (persistent spurious eigenvalues) or spectral invisibility (missing parts of the spectrum). As our evaluation set, we could consider maps $f_x(F) = F(x)$ for $x \in \mathcal{X}$ or a subset of such x. For example, if $\mathcal{X} \subset \mathbb{R}^d$, then f_x is a real vector-valued function. ■

With the definition of a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ established, we now define what we mean by an algorithm. An algorithm is a function $\Gamma : \Omega \to \mathcal{M}$ that, unlike the problem function Ξ, utilizes the evaluation set Λ in some manner. The specifics of how Λ is used (or even which sets of Λ are permitted) depend on the computational model. We adopt a general definition for proving lower bounds (impossibility results). This approach not only yields stronger results but also significantly simplifies the proofs. Specifically, we aim to establish lower bounds that are valid in *any model of computation*.

Definition A.3 (General algorithm)**.** *Given a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$, a general algorithm is a map $\Gamma : \Omega \to \mathcal{M}$ such that $\forall A \in \Omega$,*

1. *There exists a non-empty finite subset of evaluations $\Lambda_\Gamma(A) \subset \Lambda$;*
2. *The action of Γ on A only depends on $\{f(A)\}_{f \in \Lambda_\Gamma(A)}$;*
3. *If $B \in \Omega$ with $f(A) = f(B)$ for every $f \in \Lambda_\Gamma(A)$, then $\Lambda_\Gamma(A) = \Lambda_\Gamma(B)$ and $\Gamma(A) = \Gamma(B)$.*

Definition A.3 outlines the most fundamental properties of any reasonable *deterministic* computational device:

- The first property says that Γ can only use a finite amount of information, though it can adaptively choose this information as it processes the input.
- The second property ensures that the output of Γ depends solely on the information it has accessed.
- The final property ensures that Γ produces outputs and consistently accesses information. Specifically, if Γ sees the same information for two different inputs, it must behave identically for those two inputs.

A general algorithm has no restrictions on the operations allowed. It is more powerful than a Turing machine [72] or BSS machine[3] [73] and serves two main purposes:

[3] Blum–Shub–Smale (BSS) machines are a model of computation designed to work over any ring or field, most notably over the real numbers. This distinguishes them from Turing machines, which are based on discrete values. A BSS machine extends the concept of computation to include computation with real numbers and other continuous data. One should think of a BSS machine as akin to an algorithm that deals with exact arithmetic.

(i) *A focus on what really matters:* Definition A.3 significantly simplifies the process of proving lower bounds. The non-computability results we present stem from the intrinsic non-computability of the problems themselves, not from the type of operations allowed being too restrictive. Specifically, the limitation lies in the algorithmic input Λ being inadequate for solving the problem.

(ii) *Classifications with the strongest possible lower and upper bounds:* The generality of Definition A.3 implies that a lower bound established for general algorithms also applies to any computational model. Furthermore, the algorithms we provide can be executed using only arithmetic operations, both in the Turing and BSS models. Therefore, we derive the strongest possible lower and upper bounds simultaneously.

A.2.2 Towers of algorithms

Having established precise definitions for a computational problem and a general algorithm, we now introduce the concept of a tower of algorithms. This captures the observation in the main text that algorithms for data-driven Koopmanism depend on several parameters that must be taken to successive limits to ensure convergence.

Definition A.4 (Tower of algorithms)**.** *Let $k \in \mathbb{N}$. A tower of algorithms of height k for a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ is a collection of functions*

$$\Gamma_{n_k},\, \Gamma_{n_k,n_{k-1}},\, \ldots,\, \Gamma_{n_k,\ldots,n_1} : \Omega \to \mathcal{M}, \quad n_k, \ldots, n_1 \in \mathbb{N},$$

where $\{\Gamma_{n_k,\ldots,n_1}\}$ are general algorithms (Definition A.3) and for every $A \in \Omega$, the following convergence holds in $(\mathcal{M}, d)$:

$$\Xi(A) = \lim_{n_k \to \infty} \Gamma_{n_k}(A), \quad \Gamma_{n_k}(A) = \lim_{n_{k-1} \to \infty} \Gamma_{n_k,n_{k-1}}(A), \quad \ldots,$$
$$\Gamma_{n_k,\ldots,n_2}(A) = \lim_{n_1 \to \infty} \Gamma_{n_k,\ldots,n_1}(A).$$

We shall use the term "tower" even if $k = 1$.

When we prove upper bounds (i.e., provide algorithms that solve a problem), we can specify the type of tower by imposing conditions on the functions $\{\Gamma_{n_k,\ldots,n_1}\}$ at the lowest level. In essence, the type is the toolbox allowed:

- A *general tower*, denoted by $\alpha = G$, refers to Definition A.4 with no further restrictions.
- An *arithmetic tower*, denoted by $\alpha = A$, refers to Definition A.4 where each $\Gamma_{n_k,\ldots,n_1}(A)$ can be computed using Λ and finitely many arithmetic operations and comparisons. More precisely, if Λ is countable, each output $\Gamma_{n_k,\ldots,n_1}(A)$ is a finite string of numbers (or encoding) that can be identified with an element in $\mathcal{M}$, and the following function is recursive[4]:

$$(n_k, \ldots, n_1, \{f(A)\}_{f\in\Lambda}) \mapsto \Gamma_{n_k,\ldots,n_1}(A)$$

We can now define the Solvability Complexity Index (SCI).

Definition A.5 (Solvability Complexity Index)**.** *A computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ has Solvability Complexity Index $k \in \mathbb{N}$ with respect to type α, written $\mathrm{SCI}(\Xi, \Omega, \mathcal{M}, \Lambda)_\alpha = k$, if k is the smallest integer for which there exists a tower of algorithms of type α and height k that solves the problem. If no such tower exists, then $\mathrm{SCI}(\Xi, \Omega, \mathcal{M}, \Lambda)_\alpha = \infty$. If there exists an algorithm Γ of type α with $\Xi = \Gamma$, then $\mathrm{SCI}(\Xi, \Omega, \mathcal{M}, \Lambda)_\alpha = 0$.*

The SCI induces the SCI hierarchy as follows.

Definition A.6 (SCI hierarchy)**.** *Consider a collection $\mathcal{C}$ of computational problems and let $\mathcal{T}_\alpha$ be the collection of all towers of algorithms of type α. We define the following subclasses of $\mathcal{C}$:*

$$\Delta_0^\alpha = \{\{\Xi, \Omega, \mathcal{M}, \Lambda\} \in \mathcal{C} : \mathrm{SCI}(\{\Xi, \Omega, \mathcal{M}, \Lambda\})_\alpha = 0\},$$
$$\Delta_1^\alpha = \{\{\Xi, \Omega, \mathcal{M}, \Lambda\} \in \mathcal{C} : \exists \{\Gamma_n\}_{n=1}^\infty \in \mathcal{T}_\alpha \text{ s.t. } \forall A \in \Omega, d(\Gamma_n(A), \Xi(A)) \leq 2^{-n}\},$$
$$\Delta_{m+1}^\alpha = \{\{\Xi, \Omega, \mathcal{M}, \Lambda\} \in \mathcal{C} : \mathrm{SCI}(\{\Xi, \Omega, \mathcal{M}, \Lambda\})_\alpha \leq m\}, \quad \text{for } m \in \mathbb{N}.$$

In summary, a Δ_{m+1}^α problem can be computed in m successive limits, and a Δ_1^α problem can be computed in one limit with complete error control. The 2^{-n} in the definition of Δ_1^α is arbitrary; replacing it with any sequence converging to zero and computable from Λ does not alter the definition.

[4]By recursive, we mean the following. If $f(A) \in \mathbb{Q} + i\mathbb{Q}$ for all $f \in \Lambda$, $A \in \Omega$, then $\Gamma_{n_k,\ldots,n_1}(\{f(A)\}_{f\in\Lambda})$ can be executed by a Turing machine that takes $(n_k, \ldots, n_1)$ as input, and that has an oracle tape consisting of $\{f(A)\}_{f\in\Lambda}$. If $f(A) \in \mathbb{C}$ for all $f \in \Lambda$, $A \in \Omega$, then $\Gamma_{n_k,\ldots,n_1}(\{f(A)\}_{f\in\Lambda})$ can be executed by a BSS machine that takes $(n_k, \ldots, n_1)$, as input, and that has an oracle that can access any $f(A)$ for $f \in \Lambda$.

A.2.3 Inexact input

So far, we have only discussed algorithms with exact input from the evaluation set Λ. However, in practice, we may only have access to input of a certain accuracy. Suppose we are given a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ with evaluation set $\Lambda = \{f_j : \Omega \to \mathcal{M}_\Lambda\}_{j\in\mathcal{I}}$, for some index set $\mathcal{I}$ and metric space $(\mathcal{M}_\Lambda, d_\Lambda)$. Obtaining f_j may be a computational task in its own right. For instance, $f_j(A)$ could be the number $\cos(\sin(1))$ or an inner product that is approximated using quadrature. Alternatively, it may be the case that we can only measure $f_j(A)$ to a certain accuracy due to effects such as noise. In the context of Koopman operators, any physical measurement device will have a non-zero measurement error.

Hence, we cannot access $f_j(A)$, but rather $f_{j,n}(A)$ with $\lim_{n\to\infty} f_{j,n}(A) = f_j(A)$. Alternatively, just as for problems high up in the SCI hierarchy, we could need several successive limits to access $f_j(A)$. In particular, we may only have access to functions $f_{j,n_m,\dots,n_1} : \Omega \to \mathcal{M}_\Lambda$ such that

$$\lim_{n_m\to\infty} \cdots \lim_{n_1\to\infty} f_{j,n_m,\dots,n_1}(A) = f_j(A) \quad \forall A \in \Omega, \tag{11}$$

with convergence in $(\mathcal{M}_\Lambda, d_\Lambda)$. We may view the problem of obtaining $f_j(A)$ as a problem in the SCI hierarchy. A Δ_1-classification would correspond to access of $f_{j,n} : \Omega \to \mathcal{M}_\Lambda$ such that

$$d_\Lambda(f_{j,n}(A), f_j(A)) \leq 2^{-n} \quad \forall A \in \Omega. \tag{12}$$

We want algorithms that can handle all possible choices of inexact input. We can make this precise by replacing the class Ω by the class of suitable evaluation functions $\{f_{j,n_m,\dots,n_1}(A)\}_{j,n_m,\dots,n_1\in\mathcal{I}\times\mathbb{N}^m}$ that satisfy Equation (11) or Equation (12). This viewpoint is well-defined since Equation (10) holds.

Definition A.7 (Computational problems with Δ_k-information)**.** *Given $m \in \mathbb{N}$ and a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$, the corresponding computational problem with Δ_{m+1}-information is denoted by $\{\Xi, \Omega, \mathcal{M}, \Lambda\}^{\Delta_{m+1}} = \{\tilde{\Xi}, \tilde{\Omega}, \mathcal{M}, \tilde{\Lambda}\}$, and defined as follows:*

- *The primary set $\tilde{\Omega}$ is the class of tuples $\tilde{A} = \{f_{j,n_m,\dots,n_1}(A) : j \in \mathcal{I}, n_1, \dots, n_m \in \mathbb{N}\}$, where $A \in \Omega$, $\{f_j\}_{j\in\mathcal{I}}$ and Equation (11) holds;*
- *The problem function is $\tilde{\Xi}(\tilde{A}) = \Xi(A)$, which is well-defined by Equation (10);*
- *The evaluation set is $\tilde{\Lambda} = \{\tilde{f}_{j,n_m,\dots,n_1}\}_{j,n_m,\dots,n_1\in\mathcal{I}\times\mathbb{N}^m}$, where $\tilde{f}_{j,n_m,\dots,n_1}(\tilde{A}) = f_{j,n_m,\dots,n_1}(A)$.*

Similarly, the corresponding computational problem with Δ_1-information is denoted by $\{\Xi, \Omega, \mathcal{M}, \Lambda\}^{\Delta_1} = \{\tilde{\Xi}, \tilde{\Omega}, \mathcal{M}, \tilde{\Lambda}\}$, and defined as follows:

- *The primary set $\tilde{\Omega}$ is the class of tuples $\tilde{A} = \{f_{j,n_1}(A) : j \in \mathcal{I}, n_1 \in \mathbb{N}\}$, where $A \in \Omega$, $\{f_j\}_{j\in\mathcal{I}}$ and Equation (12) holds;*
- *The problem function is $\tilde{\Xi}(\tilde{A}) = \Xi(A)$, which is well-defined by Equation (10);*
- *The evaluation set is $\tilde{\Lambda} = \{\tilde{f}_{j,n_1}\}_{j,n_1\in\mathcal{I}\times\mathbb{N}}$, where $\tilde{f}_{j,n_1}(\tilde{A}) = f_{j,n_1}(A)$.*

For any $k \in \mathbb{N}$, the SCI hierarchy given Δ_k-information is then defined in an obvious manner.

The most important case of this definition is Δ_1-information, which captures the notion of algorithms being robust to noise. We can also connect Δ_k-information to the Turing and BSS models. For example, in the Turing case, we could enforce that each $\tilde{f}_{j,n_k,\dots,n_1} \in \tilde{\Lambda}$ maps to $\mathbb{Q} + i\mathbb{Q}$ and view these evaluation functions as an infinite input tape.

A.2.4 Refinements that capture error control

When performing numerical computations, particularly in many spectral applications, determining the accuracy of the results is essential. The importance of error bounds extends beyond science and engineering and holds in pure mathematics, especially when using spectral problems in computer-assisted proofs. For instance, even when a discretization method for computing spectra through eigenvalues converges, typically, only a subset of the numerically computed eigenvalues is reliable. Note that such a problem is not computable in the classical Turing sense but instead verifiable. Most infinite-dimensional spectral problems do not lie in Δ_1 [51], but many lie in the following refinements that capture error control [99, 100]. Classifying when error bounds can or cannot be obtained is a fundamental challenge in dealing with infinite-dimensional spectral problems.

Sufficient structure in $(\mathcal{M}, d)$ enables two types of verification or error control: convergence from above and below. In this paper, there are two metric spaces that we use for computational problems $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$:

- If $\mathcal{M} = \{0, 1\}$ with the discrete topology, we call the problem a decision problem and denote this space by $\mathcal{M}_{\text{dec}}$. For an input $A \in \Omega$, we interpret the output $\Xi(A) = 1$ as "Yes" and the output $\Xi(A) = 0$ as "No".
- The Hausdorff metric space, $(\mathcal{M}_{\text{H}}, d_{\text{H}})$, is suitable for computing spectra of bounded operators. It is the collection of non-empty compact subsets of $\mathbb{C}$ equipped with the Hausdorff metric:

$$d_{\text{H}}(X, Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y}|x - y|, \sup_{y\in Y}\inf_{x\in X}|x - y|\right\}, \quad X, Y \in \mathcal{M}_{\text{H}}. \tag{13}$$

  As noted above, we are interested in the Hausdorff metric since convergence to the spectrum in this metric means that our algorithms converge without spectral pollution or spectral invisibility.

We now define notions of error control for these two metric spaces. When $\mathcal{M}$ is a totally ordered set with relation $\leq$, such as $\mathbb{R}$ or $\mathbb{N}$, convergence from above or below is straightforward to define.

Definition A.8 (Σ and Π classes for totally ordered sets)**.** *Consider a collection $\mathcal{C}$ of computational problems and let $\mathcal{T}_\alpha$ be the collection of all towers of algorithms of type α. Suppose that $\mathcal{M}$ is a totally ordered set. We set $\Sigma_0^\alpha = \Pi_0^\alpha = \Delta_0^\alpha$ and for $m \in \mathbb{N}$, define*

$$\Sigma_m^\alpha = \{\{\Xi, \Omega, \mathcal{M}, \Lambda\} \in \Delta_{m+1}^\alpha : \exists\, \{\Gamma_{n_m,\dots,n_1}\} \in \mathcal{T}_\alpha \text{ s.t. } \Gamma_{n_m}(A) \uparrow \Xi(A) \forall A \in \Omega\},$$
$$\Pi_m^\alpha = \{\{\Xi, \Omega, \mathcal{M}, \Lambda\} \in \Delta_{m+1}^\alpha : \exists\, \{\Gamma_{n_m,\dots,n_1}\} \in \mathcal{T}_\alpha \text{ s.t. } \Gamma_{n_m}(A) \downarrow \Xi(A) \forall A \in \Omega\},$$

where $\uparrow$ and $\downarrow$ denote convergence from below and above, respectively. In other words, we have convergence from below or above in the final limit of the tower of algorithms.

The following two examples discuss $(\mathcal{M}, d) = \mathbb{R}$ and $\mathcal{M}_{\text{dec}}$, respectively.

Example A.9 (Spectral radius of normal operators)**.** Let Ω be the class of bounded normal operators on $l^2(\mathbb{N})$, $\Lambda = \{A \mapsto \langle Ae_j, e_i\rangle : i, j \in \mathbb{N}\}$, and consider the spectral radius problem function $\Xi(A) = \sup_{z\in \text{Sp}(A)} \|z\|$. For normal operators, $\sup_{z\in \text{Sp}(A)} \|z\| = \|A\|$. Hence, we let $\Gamma_n(A)$ be an approximation of $\|\mathcal{P}_n A \mathcal{P}_n^*\|$ to accuracy $1/n$ from below, where $\mathcal{P}_n$ is the orthogonal projection onto $\text{span}\{e_1, \dots, e_n\}$. $\Gamma_n(A) \uparrow \|A\|$ and hence $\{\Xi, \Omega, \mathbb{R}, \Lambda\} \in \Sigma_1^A$. This classification means that for any finite n, we obtain a lower bound for the value $\Xi(A)$. However, we may not know how close $\Gamma_n(A)$ is to $\|A\|$ and one can also show that $\{\Xi, \Omega, \mathbb{R}, \Lambda\} \notin \Delta_1^G$. To see why, suppose for a contradiction that $\{\Xi, \Omega, \mathbb{R}, \Lambda\} \in \Delta_1^G$. Hence, there exists a general algorithm Γ such that $|\Gamma(A) - \|A\|| \leq 1$ for all $A \in \Omega$. We may choose $A = \text{diag}(0, \dots, 0, 3, 3, \dots)$ such that the number of zeros in the diagonal of A ensures that $\Gamma(A) = \Gamma(0)$ (this follows from the consistency requirement in the definition of a general algorithm). But then $1 \geq |\Gamma(A) - 3| = |\Gamma(0) - 3| \geq 3 - |\Gamma(0) - 0| \geq 2$, a contradiction. The point is that we cannot compute an upper bound on $\|A\|$ from a finite amount of information, and hence, we cannot get full error control in the metric space $\mathbb{R}$. ■

Example A.10 (Is the spectral radius larger than one?)**.** Consider the setup of Example A.9, but now let Ξ be the decision problem, 'Is $\sup_{z\in \text{Sp}(A)} \|z\| > 1$?' Let $\Gamma_n(A)$ be as before, then if $\Xi(A) = 1$ (yes), $\Gamma_n(A) > 1$ for some n, otherwise $\Gamma_n(A) \leq 1$ for all n. Note that we have used the fact that $\{\Gamma_n\}$ is a Σ_1^A-tower for the problem in Example A.9. It follows that

$$\tilde{\Gamma}_n(A) = \begin{cases} 1, & \text{if } \Gamma_n(A) > 1, \\ 0, & \text{otherwise.} \end{cases}$$

provides a Σ_1^A-tower for $\{\Xi, \Omega, \mathcal{M}_{\text{dec}}, \Lambda\}$. Again, one can show that $\{\Xi, \Omega, \mathcal{M}_{\text{dec}}, \Lambda\} \notin \Delta_1^G$. If we changed the decision problem to 'Is $\sup_{z\in \text{Sp}(A)} \|z\| \leq 1$?', we would obtain a Π_1^A classification instead. ■

More generally, the classes Σ_1^A and Π_1^A in Definition A.8 allow *verification*. For example, suppose that we have a problem function $\Xi : \Omega \to \mathbb{R}$ and we wish to verify a theorem $\Xi(A) < 0$ for some $A \in \Omega$. If there exists a Π_1^A-tower $\{\Gamma_n\}$ for the problem and the theorem is true for the given A, then $\Gamma_n(A) < 0$ for sufficiently large n. We can compute $\Gamma_n(A)$ for various n and as soon as $\Gamma_n(A) < 0$, we know that $\Xi(A) \leq \Gamma_n(A) < 0$ and have verified the theorem. Note, however, that we cannot use a Π_1^A-tower to negate such a theorem (but can, instead, use a Σ_1^A-tower if it exists). We can only verify one way using a Π_1^A-tower. Similar remarks hold for Σ_1^A.

While Definition A.8 is straightforward, it does not carry over to the more complicated Hausdorff metric. To define convergence of $\Gamma_n(A)$ to $\Xi(A)$ in the Hausdorff metric "from below", a first attempt may be to require that $\Gamma_n(A) \subset \Xi(A)$. However, this is severely restrictive. For example, when computing $\text{Sp}(A) \subset \mathbb{C}$, we can rarely ensure that a point z is exactly in $\text{Sp}(A)$. Nevertheless, we can often ensure that z is close to $\text{Sp}(A)$ and measure how close. Hence, it is natural to relax the condition $\Gamma_n(A) \subset \Xi(A)$ to

$$\sup_{z\in\Gamma_n(A)} \text{dist}(z, \Xi(A)) \leq 2^{-n}.$$

The exact form of the sequence $\{2^{-n}\}$ does not matter. What matters is that we can control the proximity of $\Gamma_n(A)$ to being contained within $\Xi(A)$ using a *known* sequence that converges to zero as $n \to \infty$. Based on this discussion, the following provides the generalization of Definition A.8.

Definition A.11 (Σ and Π classes for Hausdorff metric)**.** *Consider a collection $\mathcal{C}$ of computational problems and let $\mathcal{T}_\alpha$ be the collection of all towers of algorithms of type α. Suppose that $(\mathcal{M}, d)$ is the Hausdorff metric. We set $\Sigma_0^\alpha = \Pi_0^\alpha = \Delta_0^\alpha$ and for $m \in \mathbb{N}$, we define*

$$\Sigma_m^\alpha = \Big\{\{\Xi, \Omega, \mathcal{M}, \Lambda\} \in \Delta_{m+1}^\alpha : \exists\{\Gamma_{n_m,\dots,n_1}\} \in \mathcal{T}_\alpha, \{X_{n_m}(A)\} \subset \mathcal{M} \text{ s.t. } \forall A \in \Omega$$
$$\Gamma_{n_m}(A) \subset X_{n_m}(A), \lim_{n_m\to\infty} \Gamma_{n_m}(A) = \Xi(A), d(X_{n_m}(A), \Xi(A)) \leq 2^{-n_m}\Big\},$$
$$\Pi_m^\alpha = \Big\{\{\Xi, \Omega, \mathcal{M}, \Lambda\} \in \Delta_{m+1}^\alpha : \exists\{\Gamma_{n_m,\dots,n_1}\} \in \mathcal{T}_\alpha, \{X_{n_m}(A)\} \subset \mathcal{M} \text{ s.t. } \forall A \in \Omega$$
$$\Xi(A) \subset X_{n_m}(A), \lim_{n_m\to\infty} \Gamma_{n_m}(A) = \Xi(A), d(X_{n_m}(A), \Gamma_{n_m}(A)) \leq 2^{-n_m}\Big\}.$$

These classes capture convergence from below or above, up to a small error parameter 2^{-n}. It is precisely the classes Σ_1^α and Π_1^α that allow computations with verification, used, for example, in computer-assisted proofs. For example, to build a Σ_1^α algorithm in the case of the Hausdorff metric, it is enough to construct a convergent tower $\{\Gamma_n\}$ such that $\Gamma_n(A) \subset \Xi(A) + B_{E_n}(0)$ with some computable E_n that converges to zero.

A.2.5 Randomized algorithms

We also consider sequences of probabilistic general algorithms, which are more general than a sequence of probabilistic Turing [101, Ch. 7] or probabilistic BSS [73, Ch. 17] machines. It is common to use randomized algorithms in machine learning and optimization.[5] For example, in the context of Koopman operators, it is common to apply Monte Carlo methods that randomly sample the snapshots. We consider the following definition.

Definition A.12 (A sequence of probabilistic general algorithms)**.** *Let $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ be a computational problem. A sequence of probabilistic general algorithms (SPGA) is a sequence $\{\Gamma_n\}$ of general algorithms $\Gamma_n : \Omega \to \mathcal{M}$ using Λ with the additional properties that given an input $F \in \Omega$:*

1. *(Coin flips) Each Γ_n has access to an oracle (viewed as an input appended to Λ) that, when queried, gives the answers 0 or 1 at random, each with probability $1/2$;*
2. *(Always halts) Each Γ_n halts with probability 1;*
3. *(Access to previous outputs) Γ_n has access to the oracle queries that Γ_k produces upon input F for any $k < n$.*

These conditions hold for any standard probabilistic machine (e.g., Turing or BSS) and model machines that flip coins.[6]

Remark A.13 (The reason for part 3 of Definition A.12)**.** It may appear that part 3 of Definition A.12 strengthens the requirements for algorithms, potentially weakening our lower bounds. However, this is not the case. If $\{\Gamma_n\}$ is a sequence of general algorithms that satisfies conditions 1 and 2 of Definition A.12, then there is a corresponding SPGA $\{\tilde{\Gamma}_n\}$ that also satisfies condition 3. Specifically, each $\tilde{\Gamma}_n$ ignores the additional input information in part 3 and simply runs Γ_n. Therefore, part 3 of Definition A.12 enlarges our class of algorithms, strengthening our lower bounds. For example, to illustrate the use of part 3, suppose we have a decision problem (where the answer is either "Yes" or "No"). If each Γ_n flips a coin to output "Yes" or "No", the sequence Γ_n fails to converge with probability 1. However, if Γ_n for $n \geq 2$ is allowed to output the same answer as Γ_1 (as permitted by part 3 of Definition A.12), the probability that $\lim_{n\to\infty} \Gamma_n$ is the correct answer is now $1/2$.

We can now make sense of probabilistic classes in the SCI hierarchy.

Definition A.14. *A computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ does not belong to $\Delta_1^{\mathbb{P}}$ if for any SPGA $\{\Gamma_n\}$,*

$$\inf_{F\in\Omega} \mathbb{P}\left(d(\Gamma_n(F), \Xi(F)) \leq 2^{-n} \text{ for all } n \in \mathbb{N}\right) \leq 1/2.$$

A computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ does not belong to $\Delta_2^{\mathbb{P}}$ if for any SPGA $\{\Gamma_n\}$,

$$\inf_{F\in\Omega} \mathbb{P}\left(\lim_{n\to\infty} \Gamma_n(F) = \Xi(F)\right) \leq 1/2.$$

The following lemma shows that $\notin \Delta_2^{\mathbb{P}}$ implies $\notin \Delta_2^G$. In other words, being limit computable by an SPGA is weaker than being limit computable by a sequence of general algorithms. Similarly, $\notin \Delta_1^{\mathbb{P}}$ implies $\notin \Delta_1^G$. Hence, at various points, we prove statements of the form $\notin \Delta_m^{\mathbb{P}}$, which imply the corresponding results $\notin \Delta_m^G$.

Lemma A.15. *Let $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ be a computational problem. If $\{\Xi, \Omega, \mathcal{M}, \Lambda\} \notin \Delta_1^{\mathbb{P}}$, then $\{\Xi, \Omega, \mathcal{M}, \Lambda\} \notin \Delta_1^G$. If $\{\Xi, \Omega, \mathcal{M}, \Lambda\} \notin \Delta_2^{\mathbb{P}}$, then $\{\Xi, \Omega, \mathcal{M}, \Lambda\} \notin \Delta_2^G$.*

Proof. The proof is trivial. We prove the contrapositive. Suppose first that $\{\Xi, \Omega, \mathcal{M}, \Lambda\} \in \Delta_2^G$ with a Δ_2^G-tower $\{\Gamma_n\}$ solving the problem. Then $\{\Gamma_n\}$ is an SPGA with $\mathbb{P}(\lim_{n\to\infty} \Gamma_n(F) = \Xi(F)) = 1 > 1/2$ for any $F \in \Omega$. This shows the second part of the lemma. The proof of the first part is analogous. □

One may wonder why we chose $1/2$ in Definition A.14. The reason is that if we had chosen any number greater than $1/2$ instead, this number could be made arbitrarily close to 1 when defining the classes $\Delta_1^{\mathbb{P}}$ and $\Delta_2^{\mathbb{P}}$. This is a standard argument for Turing machines and decision problems. However, the following shows (the non-trivial and perhaps surprising result) that it also holds for arbitrary computational problems and SPGAs.

Lemma A.16. *Let $\{\Xi, \Omega, \mathcal{M}, \Lambda\}$ be a computational problem and $\delta_1, \delta_2 \in (1/2, 1)$ with $\delta_1 \leq \delta_2$.*

- *If there exists an SPGA $\{\Gamma_n\}$ with*

$$\inf_{F\in\Omega} \mathbb{P}\left(d(\Gamma_n(F), \Xi(F)) \leq 2^{-n} \text{ for all } n \in \mathbb{N}\right) \geq \delta_1$$

then there exists another SPGA $\{\Gamma_n'\}$ with

$$\inf_{F\in\Omega} \mathbb{P}\left(d(\Gamma_n'(F), \Xi(F)) \leq 2^{-n} \text{ for all } n \in \mathbb{N}\right) \geq \delta_2.$$

[5] For Turing machines, it is unknown whether randomization is beneficial from a complexity class viewpoint [101, Ch. 7]. This is not the case for BSS machines [73, Ch. 17] (but it is an open problem whether any probabilistic BSS machine can be simulated by a deterministic machine having the same machine constants and with only a polynomial slowdown). However, randomization is extremely useful in practice.

[6] One could also consider other, even continuous, probability distributions. In the case of BSS machines, machines that can pick numbers uniformly at random in $[0, 1]$ are no more computationally powerful [73, Section 17.5]. Hence, we do not consider such scenarios, which are also unrealistic in practice.

- *If there exists an SPGA* $\{\Gamma_n\}$ *with*

$$\inf_{F\in\Omega} \mathbb{P}\left(\lim_{n\to\infty} \Gamma_n(F) = \Xi(F)\right) \geq \delta_1$$

then there exists another SPGA $\{\Gamma'_n\}$ *with*

$$\inf_{F\in\Omega} \mathbb{P}\left(\lim_{n\to\infty} \Gamma_n(F) = \Xi(F)\right) \geq \delta_2.$$

In other words, the probability of success can be made arbitrarily close to 1*, and hence, success can be achieved with overwhelming probability.*

Proof. We prove the first statement, and the second is analogous. Suppose then that $\{\Gamma_n\}$ is an SPGA with

$$\inf_{F\in\Omega} \mathbb{P}\left(d(\Gamma_n(F), \Xi(F)) \leq 2^{-n} \text{ for all } n \in \mathbb{N}\right) \geq \delta_1.$$

Given the SPGA $\{\Gamma_n\}$ and input F, we run $2k$ independent copies of Γ_n and label these as $\Gamma_n^{(1)}, \ldots, \Gamma_n^{(2k)}$. For $n \in \mathbb{N}$, we consider the objective function defined for $x \in \mathcal{M}$ as

$$J_{n,F}(x) = \min_{i_1<i_2<\cdots<i_k} \sum_{l=1}^{k} d(x, \Gamma_n^{(i_l)}(F)).$$

Let $S_n(F) \in \mathcal{M}$ be such that $J_{n,F}(S_n(F)) < 2^{-n} + \inf_{x\in\mathcal{M}} J_{n,F}(x)$. We may choose $S_n(F)$ so that $\{S_n\}$ is as SPGA. Let E_k be the event that at least $k+1$ of $\Gamma_n^{(1)}, \ldots, \Gamma_n^{(2k)}$ satisfy

$$d(\Gamma_n(F), \Xi(F)) \leq 2^{-n} \text{ for all } n \in \mathbb{N}.$$

Suppose that E_k occurs and let $n \in \mathbb{N}$. Without loss of generality, we may assume that

$$d(\Gamma_n^{(j)}(F), \Xi(F)) \leq 2^{-n}, \quad \text{for } j = 1, \ldots, k+1.$$

It follows that

$$J_{n,F}(S_n(F)) \leq 2^{-n} + J_{n,F}(\Xi(F)) \leq 2^{-n} + \sum_{j}^{k} d(\Xi(F), \Gamma_n^{(j)}(F)) \leq (k+1)2^{-n}.$$

Moreover, by the pigeonhole principle, there exists $j \leq k+1$ with

$$d(S_n(F), \Gamma_n^{(j)}(F)) \leq J_{n,F}(S_n(F)).$$

It follows that

$$d(S_n(F), \Xi(F)) \leq 2^{-n} + d(S_n(F), \Gamma_n^{(j)}(F)) \leq (k+2)2^{-n}.$$

We can now define $\Gamma'_n(F) = S_{m(n)}(F)$ where $m(n)$ is chosen so that $(k+2)2^{-m(n)} \leq 2^{-n}$. It remains to show that we may choose k sufficiently large so that $\mathbb{P}(E_k) \geq \delta_2$.

Let Q_j be the random variable

$$Q_j = \begin{cases} 1, & \text{if } d(\Gamma_n^{(j)}(F), \Xi(F)) \leq 2^{-n} \text{ for all } n \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

Each Q_j is an independent Bernoulli random variable with parameter $p \geq \delta_1$. Let $\mathcal{Q}_k = Q_1 + \cdots + Q_{2k}$. Then $\mathcal{Q}_k$ is a binomially distributed random variable with parameters $2k$ and p and

$$\mathbb{P}(E_k) = \mathbb{P}(\mathcal{Q}_k \geq k+1), \quad \mathbb{E}(\mathcal{Q}_k) \geq 2k\delta_1.$$

Since $\delta_1 > 1/2$, it follows that we can choose k large so that $\mathbb{P}(E_k) \geq \delta_2$ using standard bounds for the tail of the binomial distribution. □

A.3 The setup of this paper - a perfect measurement device

In this paper, we consider computational problems where:

- Ω is a class of dynamical systems, or F, on some statespace $\mathcal{X}$ as in (2);
- Λ, the evaluation set, will be pointwise evaluations of the function F:

$$\Lambda = \{F : x \mapsto F(x) \in \mathcal{X} \text{ such that } x \in \mathcal{X}_p \subset \mathcal{X}\}.$$

  Here, $\mathcal{X}_p$ is a suitable subset of $\mathcal{X}$ for which we allow pointwise evaluations of F.
- The metric space $(\mathcal{M}, d)$ will be either $\mathcal{M}_{\text{dec}}$, in the case of decision problems, or $\mathcal{M}_{\text{H}}$, when we compute spectral sets.

This setup agrees with the usual "snapshot" setting of data-driven Koopmanism where one is given access to a finite collection of pairs

$$\left\{x^{(m)}, y^{(m)} = F(x^{(m)}) : m = 1, \dots, M\right\}$$

However, our lower bounds become *stronger*. Specifically, this strengthening occurs when we prove lower bounds for a computational problem $\{\Xi, \Omega, \mathcal{M}, \Lambda\}^{\Delta_1}$, which we remind the reader, corresponds to allowing arbitrary precision of the evaluation set according to Definition A.7. Hence, our lower bound holds even if we allow algorithms arbitrarily many point samples of F to arbitrary precision. When proving lower bounds, we will always deal with the case that ω is the natural Lebesgue measure on $\mathcal{X} \subset \mathbb{R}^d$, though many of our results can be easily extended to other measures.

B Statement of Theorems and Results

With the above preliminaries covered, we can now provide precise statements of our theorems.

B.1 Measure-preserving maps on the unit circle

As a simple first example, we consider the well-studied class of measure-preserving maps on the unit-circle [102, Chapter 3, Section 3], which we view as the periodic interval $[-\pi, \pi]_{\mathrm{per}}$ through standard angular coordinates. These classes are of particular importance in dynamical systems theory, and many important nonergodic dynamical systems have phase spaces that split into invariant tori. Consider the primary set

$$\Omega_{[-\pi,\pi]_{\mathrm{per}}} = \{F : [-\pi, \pi]_{\mathrm{per}} \to [-\pi, \pi]_{\mathrm{per}} \text{ such that } F \text{ is continuous, measure and orientation preserving, and invertible}\},$$

where $\mathcal{X} = [-\pi, \pi]_{\mathrm{per}}$ is equipped with the standard (normalized) Lebesgue measure. The Koopman operator associated with any of the dynamical systems in $\Omega_{[-\pi,\pi]_{\mathrm{per}}}$ is unitary. We take $\Lambda = \{F : x \mapsto F(x) : x \in [-\pi, \pi]_{\mathrm{per}}\}$. We consider the problem of computing the spectrum Sp in the Hausdorff metric:

$$\Xi_{\mathrm{Sp}} : \Omega_{[-\pi,\pi]_{\mathrm{per}}} \ni F \mapsto \mathrm{Sp}(\mathcal{K}_F) \in \mathcal{M}_{\mathrm{H}},$$

We also consider the problem of deciding whether the dynamical system is ergodic. This second computation problem has the following problem function

$$\Xi_{\mathrm{erg}} : \Omega_{[-\pi,\pi]_{\mathrm{per}}} \ni F \mapsto \begin{Bmatrix} 1, & \text{if } F \text{ is ergodic} \\ 0, & \text{otherwise} \end{Bmatrix} \in \mathcal{M}_{\mathrm{dec}}.$$

Theorem B.1. *Given the above setup, we have the following classifications:*

(a) The spectrum can be computed in one limit, but full error control is impossible, $\Delta_1^{\mathbb{P}} \not\ni \{\Xi_{\mathrm{Sp}}, \Omega_{[-\pi,\pi]_{\mathrm{per}}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Sigma_1^A$.

(b) Ergodicity can be determined in two limits but not one, $\Delta_2^{\mathbb{P}} \not\ni \{\Xi_{\mathrm{erg}}, \Omega_{[-\pi,\pi]_{\mathrm{per}}}, \mathcal{M}_{\mathrm{dec}}, \Lambda\}^{\Delta_1} \in \Pi_2^A$.

(c) If we remove the condition that F is continuous from $\Omega_{[-\pi,\pi]_{\mathrm{per}}}$ to define

$$\widehat{\Omega}_{[-\pi,\pi]_{\mathrm{per}}} = \{F : [-\pi, \pi]_{\mathrm{per}} \to [-\pi, \pi]_{\mathrm{per}} \textit{ such that } F \textit{ is measure-preserving and invertible}\},$$

then the SCI of the problem is infinite, $\{\Xi_{\mathrm{Sp}}, \widehat{\Omega}_{[-\pi,\pi]_{\mathrm{per}}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \notin \cup_{k=1}^{\infty} \Delta_k^G$.

Proof. Before proving each classification, note that the maps $F \in \Omega_{[-\pi,\pi]_{\mathrm{per}}}$ are precisely those of the form

$$F(\theta) = F_\gamma(\theta) = \theta + 2\pi\gamma, \qquad \theta \in [-\pi, \pi]_{\mathrm{per}},$$

for some $\gamma \in [0, 1]$, where θ denotes the angular coordinate. Furthermore, the Fourier basis diagonalizes $\mathcal{K}_{F_\gamma}$. The spectrum of $\mathcal{K}_{F_\gamma}$ is pure point and is the closure of the set of eigenvalues of $\mathcal{K}_{F_\gamma}$:

$$\mathrm{Sp}(\mathcal{K}_{F_\gamma}) = \mathrm{Cl}\left(\{e^{2\pi i n\gamma} : n \in \mathbb{Z}\}\right).$$

In particular, if γ is rational, the spectrum consists of a finite number of points on the unit circle, whereas if γ is irrational, the spectrum is the entire unit circle. The conditions for ergodicity are analogous, with [103, Theorem 1.8]

$$\Xi_{\mathrm{erg}}(F) = \begin{cases} 1, & \text{if } \gamma \text{ is irrational,} \\ 0, & \text{otherwise.} \end{cases}$$

With these basic facts, we can now prove the classifications in Theorem B.1.

Step 1: $\{\Xi_{\mathrm{Sp}}, \Omega_{[-\pi,\pi]_{\mathrm{per}}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \notin \Delta_1^{\mathbb{P}}$. Suppose for a contradiction that $\{\Gamma_n\}$ is an SPGA for $\{\Xi_{\mathrm{Sp}}, \Omega_{[-\pi,\pi]_{\mathrm{per}}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1}$ such that

$$\inf_{F \in \Omega_{[-\pi,\pi]_{\mathrm{per}}}} \mathbb{P}\left(d_{\mathrm{H}}(\Gamma_n(F), \mathrm{Sp}(\mathcal{K}_F)) \leq 2^{-n} \text{ for all } n \in \mathbb{N}\right) = \epsilon > 1/2.$$

For $\theta \in [-\pi,\pi]_{\rm per}$, let $f(\cdot;\theta)$ be the map

$$\Omega_{[-\pi,\pi]_{\rm per}} \ni F \mapsto f(F;\theta) = F(\theta)$$

and let $f_n(\cdot;\theta)$ denote a collection of maps such that

$$|f_n(F;\theta) - f(F;\theta)| \leq 2^{-n} \qquad \forall F \in \Omega_{[-\pi,\pi]_{\rm per}}.$$

The evaluation set Λ is $\{f(\cdot;\theta) : \theta \in [-\pi,\pi]_{\rm per}\}$, whereas the evaluation set with Δ_1-inexactness, $\tilde{\Lambda}$, is the set of maps $f_n(\cdot;\theta)$, $\theta \in [-\pi,\pi]_{\rm per}$, $n \in \mathbb{N}$. In particular, given $\delta > 0$, from Definition A.12 of an SPGA and Definition A.7 of Δ_1-information, there exists a finite subset $\{f_{m_1}(\cdot,\theta_1),\dots,f_{m_N}(\cdot,\theta_N)\}$ such that

$$\mathbb{P}\left(\tilde{\Lambda}_{\Gamma_1(F_0)} \subset \{f_{m_1}(F_0,\theta_1),\dots,f_{m_N}(F_0,\theta_N)\}\right) \geq 1-\delta.$$

We may choose sufficiently small irrational γ such that we may also choose f_n so that

$$f_{m_j}(F_\gamma,\theta_j) = f_{m_j}(F_0,\theta_j) \quad j = 1,\dots,N.$$

We consider running Γ_1 for the two inputs F_0 and F_γ (with independent calls to the oracles). Let $X_1 = \Gamma_1(F_0)$ and $X_2 = \Gamma_1(F_\gamma)$ be the corresponding outputs, which are random variables. Note that $\mathrm{Sp}(\mathcal{K}_{F_0}) = \{1\}$, whereas $\mathrm{Sp}(\mathcal{K}_{F_\gamma}) = \mathbb{T}$.

For $\alpha = 0$ or γ, let E_α be the event $\tilde{\Lambda}_{\Gamma_1(F_\alpha)} \subset \{f_{m_1}(F_\alpha,\theta_1),\dots,f_{m_N}(F_\alpha,\theta_N)\}$. By part (iii) of Definition A.3 and by Definition A.12, $\mathbb{P}(E_\gamma) = \mathbb{P}(E_0) \geq 1-\delta$. It follows that $\mathbb{P}(E_0 \cap E_\gamma) \geq 1-2\delta$. Let B_1 and B_2 be the events $d_{\rm H}(\Gamma_1(F_\alpha),\mathrm{Sp}(\mathcal{K}_{F_\alpha})) \leq 2^{-1}$ for $\alpha = 0$ and γ, respectively. If $E_0 \cap E_\gamma$ occurs, then the law of X_1 and X_2 are the same. Hence,

$$\mathbb{P}(B_1 \cap [E_0 \cap E_\gamma]) + \mathbb{P}(B_2 \cap [E_0 \cap E_\gamma]) = \mathbb{P}(\{d_{\rm H}(X_1,\{1\}) \leq 2^{-1}\} \cap [E_0 \cap E_\gamma]) + \mathbb{P}(\{d_{\rm H}(X_1,\mathbb{T}) \leq 2^{-1}\} \cap [E_0 \cap E_\gamma]).$$

Since $d_{\rm H}(\{1\},\mathbb{T}) > 1$, it follows from the triangle inequality that

$$\mathbb{P}(\{d_{\rm H}(X_1,\{1\}) \leq 2^{-1}\} \cap [E_0 \cap E_\gamma]) + \mathbb{P}(\{d_{\rm H}(X_1,\mathbb{T}) \leq 2^{-1}\} \cap [E_0 \cap E_\gamma]) \leq 1.$$

Hence,

$$1 < 2\epsilon \leq \mathbb{P}(B_1) + \mathbb{P}(B_2) \leq 4\delta + \mathbb{P}(B_1 \cap [E_0 \cap E_\gamma]) + \mathbb{P}(B_2 \cap [E_0 \cap E_\gamma]) \leq 1 + 4\delta.$$

This is a contradiction for sufficiently small δ.

Step 2: $\{\Xi_{\rm Sp},\Omega_{[-\pi,\pi]_{\rm per}},\mathcal{M}_{\rm H},\Lambda\}^{\Delta_1} \in \Sigma_1^A$. An arithmetic algorithm with access to the point values of F_γ may compute γ to any specified accuracy. In particular, we let $\Gamma_n(F_\gamma)$ be an approximation of the set

$$S_n^{(\gamma)} = \{\exp(2\pi i j\gamma) : j = -n,\dots,n\} \subset \mathrm{Sp}(\mathcal{K}_{F_\gamma}) \quad \text{such that} \quad d_{\rm H}(\Gamma_n(F_\gamma),S_n^{(\gamma)}) \leq 2^{-n}.$$

Such a set $\Gamma_n(F_\gamma)$ can be computed using finitely many arithmetic operations and comparisons using the specified Δ_1-information. Note that $\Gamma_n(F) \subset \mathrm{Sp}(\mathcal{K}_{F_\gamma}) + B_{2^{-n}}(0)$ and that $\lim_{n\to\infty}\Gamma_n(F) = \mathrm{Sp}(\mathcal{K}_{F_\gamma})$.

Step 3: $\{\Xi_{\rm erg},\Omega_{[-\pi,\pi]_{\rm per}},\mathcal{M}_{\rm dec},\Lambda\}^{\Delta_1} \notin \Delta_2^{\mathbb{P}}$. Suppose for a contradiction that $\{\Gamma_n\}$ is an SPGA for $\{\Xi_{\rm erg},\Omega_{[-\pi,\pi]_{\rm per}},\mathcal{M}_{\rm dec},\Lambda\}^{\Delta_1}$ such that

$$\inf_{F\in\Omega_{[-\pi,\pi]_{\rm per}}} \mathbb{P}\left(\lim_{n\to\infty}\Gamma_n(F) = \Xi_{\rm erg}(F)\right) = \epsilon > 1/2.$$

For a sequence of integers $\{l_j\}_{j\in\mathbb{N}}$ with $l_{j+1} > l_j$ and $l_0 = 1$, define the quantities

$$\gamma = \gamma(\{l_j\}) = \sum_{n=0}^{\infty} 10^{-\sum_{j=0}^n l_j}, \quad \gamma_N = \sum_{n=0}^{N} 10^{-\sum_{j=0}^n l_j} \quad \text{for} \quad N \in \mathbb{N}.$$

In other words,

$$\gamma(\{l_j\}) = 0.1\underbrace{0\cdots01}_{l_1\text{ digits}}\underbrace{0\cdots01}_{l_2\text{ digits}}\underbrace{0\cdots01}_{l_3\text{ digits}}\cdots.$$

The number $\gamma(\{l_j\})$ has a non-terminating and non-repeating (since $l_{j+1} > l_j$) decimal expansion, and hence it is irrational so that $\Xi_{\rm erg}(F_\gamma) = 1$. On the other hand, $\Xi_{\rm erg}(F_{\gamma_N}) = 0$ for all $N \in \mathbb{N}$. We will choose the $\{l_j\}_{j\in\mathbb{N}}$ inductively to gain a contradiction, thus proving the result.

Suppose that we have defined $\{l_j\}_{j=1}^N$. By our assumption,

$$\mathbb{P}\left(\lim_{n\to\infty}\Gamma_n(F_{\gamma_N}) = \Xi_{\rm erg}(F_{\gamma_N}) = 0\right) \geq \epsilon > 1/2.$$

It follows that there exists $n_N \in \mathbb{N}$ such that

$$\mathbb{P}\left(\Gamma_{n_N}(F_{\gamma_N}) = 0\right) \geq \frac{\epsilon}{2} + \frac{1}{4} > 1/2.$$

Since Γ_{n_N} only has Δ_1-information, given any $\delta > 0$, there exists $M_N(\delta) \in \mathbb{N}$ with $M_N \geq \sum_{j=0}^{N} l_j$ such that the law of the output $\Gamma_{n_N}(F_{\gamma_N})$ depends only on the first M_N digits of γ_N with probability at least $1-\delta$. Choosing δ sufficiently small, we can ensure that

$$\mathbb{P}\left(\Gamma_{n_N}(F_{\gamma_N}) = 0 \text{ and } \Gamma_{n_N}(F_{\gamma_N}) \text{ depends only on the first } M_N \text{ digits of } \gamma_N\right) \geq \frac{\epsilon}{4} + \frac{3}{8} > 1/2.$$

Let $l_{N+1} \in \mathbb{N}$ be large enough so that γ agrees with γ_N in the first M_N decimal places, and let X_N be the random variable $\Gamma_{n_N}(F_\gamma)$. By Definition A.12 of an SPGA and Definition A.3,

$$\mathbb{P}(X_N = 0) \geq \mathbb{P}\left(\Gamma_{n_N}(F_{\gamma_N}) = 0 \text{ and } \Gamma_{n_N}(F_{\gamma_N}) \text{ depends only on the first } M_N \text{ digits of } \gamma_N\right) \geq \frac{\epsilon}{4} + \frac{3}{8} > 1/2.$$

We may assume without loss of generality that $n_N \to \infty$. Now let A_m be the event $\{X_N = 1 \text{ for all } N \geq m\}$, and note that $\mathbb{P}(A_m) \leq 1 - \mathbb{P}(X_m = 0) < 1/2$. By our initial assumption,

$$\mathbb{P}\left(\cup_{m=1}^{\infty} A_m\right) \geq \mathbb{P}\left(\lim_{n\to\infty} \Gamma_n(F_\gamma) = \Xi_{\mathrm{erg}}(F) = 1\right) \geq \epsilon > 1/2.$$

Since $A_1 \subset A_2 \subset A_3 \subset \cdots$, there must exist some A_M with $\mathbb{P}(A_M) > 1/2$, the required contradiction.

Step 4: $\{\Xi_{\mathrm{erg}}, \Omega_{[-\pi,\pi]_{\mathrm{per}}}, \mathcal{M}_{\mathrm{dec}}, \Lambda\}^{\Delta_1} \in \Pi_2^G$. We may list the rationals in $[0,1]$ without repetition as $b_1, b_2, \ldots$. Given an input F_γ, we let $\hat{\Gamma}_n$ be an arithmetic algorithm with Δ_1-information such that $|\hat{\Gamma}_n(F_\gamma) - \gamma| \leq 1/n$. We then define

$$a_{n_2,n_1} = \min_{1\leq j\leq n_2} |\hat{\Gamma}_{n_1}(F_\gamma) - b_j|$$

and set

$$\Gamma_{n_2,n_1}(F_\gamma) = \begin{cases} 1, & \text{if } a_{n_2,n_1} > 1/n_1, \\ 0, & \text{otherwise.} \end{cases}$$

It is clear that

$$\lim_{n_1\to\infty} \Gamma_{n_2,n_1}(F_\gamma) = \Gamma_{n_2}(F_\gamma) = \begin{cases} 1, & \text{if } \gamma \notin \{b_1, \ldots, b_{n_2}\}, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, since $\Xi_{\mathrm{erg}}(F_\gamma) = 1$ if and only if γ is irrational, $\Gamma_{n_2}(F_\gamma)$ converges to $\Xi_{\mathrm{erg}}(F_\gamma)$ from above.

Step 5: $\{\Xi_{\mathrm{Sp}}, \widehat{\Omega}_{[-\pi,\pi]_{\mathrm{per}}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \notin \cup_{k=1}^{\infty} \Delta_k^G$. We prove this result for exact input instead of Δ_1-information (which also implies the result for Δ_1-information). Suppose for a contradiction that $\{\Gamma_{n_k,\ldots,n_1}\}$ is a height-k tower of general algorithms for $\{\Xi_{\mathrm{Sp}}, \widehat{\Omega}_{[-\pi,\pi]_{\mathrm{per}}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}$. For the input F_0, the base algorithms $\Gamma_{n_k,\ldots,n_1}$ only ever point sample F_0 at a countable number of points Θ as $n_k, \ldots, n_1$ vary. We can define

$$F(\theta) = \begin{cases} \theta, & \text{if } \theta \in \Theta, \\ \theta + 2\pi\gamma, & \text{otherwise,} \end{cases}$$

where γ is irrational. Note that this map is measure-preserving and invertible modulo $\omega-$null sets. It follows from part (iii) of Definition A.3 that $\Gamma_{n_k,\ldots,n_1}(F_0) = \Gamma_{n_k,\ldots,n_1}(F_\gamma)$. In particular,

$$\{1\} = \mathrm{Sp}(\mathcal{K}_{F_0}) = \lim_{n_k\to\infty} \cdots \lim_{n_1\to\infty} \Gamma_{n_k,\ldots,n_1}(F_0) = \lim_{n_k\to\infty} \cdots \lim_{n_1\to\infty} \Gamma_{n_k,\ldots,n_1}(F_\gamma) = \mathrm{Sp}(\mathcal{K}_{F_\gamma}) = \mathbb{T},$$

which is absurd. □

Remark B.2. Interestingly, a constructive computational procedure exists for part (b) of Theorem B.1. The constructive ergodic partition theorem first proven in [59] and expanded on in [104] and [7] applies to continuous dynamical systems on compact metric spaces. That result implies that if every element of a basis of continuous functions in L^2 (see below for the existence of such a basis) has a constant time average almost everywhere, then the dynamical system is ergodic. Two limits are necessary for this computation: the limit of the number of iterations of the dynamical system going to infinity and the limit of the number of functions in the basis going to infinity.

B.2 A general computational result providing upper bounds

In this section, we let $(\mathcal{X}, d_{\mathcal{X}})$ be a compact metric space and ω a finite Borel measure on $(\mathcal{X}, d_{\mathcal{X}})$. Under these assumptions, there are two key properties of $L^2(\mathcal{X}, \omega)$ that we shall make use of:

- $L^2(\mathcal{X}, \omega)$ is a separable Hilbert space [105, Proposition 3.4.5];
- The space of continuous functions on $\mathcal{X}$, denoted $C(\mathcal{X})$, is dense in $L^2(\mathcal{X}, \omega)$ [106, Proposition 7.9].

In particular, by the Gram–Schmidt process, there exists an orthonormal basis $\{g_1, g_2, \ldots\} \subset C(\mathcal{X})$ of $L^2(\mathcal{X}, \omega)$.

Definition B.3. *Let $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be an increasing continuous function with $\alpha(0) = 0$. We say that a continuous function $F : \mathcal{X} \to \mathcal{X}$ has a modulus of continuity α if*

$$d_{\mathcal{X}}(F(x), F(y)) \leq \alpha(d_{\mathcal{X}}(x, y)) \quad \forall x, y \in \mathcal{X}. \tag{14}$$

Since $\mathcal{X}$ is compact, any continuous function $F : \mathcal{X} \to \mathcal{X}$ is uniformly continuous and hence has a choice of α so that Equation (14) holds. However, there is no α such that Equation (14) holds universally for all continuous functions. We set

$$\begin{aligned}
\Omega_{\mathcal{X}} &= \{F : \mathcal{X} \to \mathcal{X} \text{ such that } F \text{ is continuous and non-singular, } \mathcal{K}_F \text{ is bounded}\},\\
\Omega_{\mathcal{X}}^{\alpha} &= \{F : \mathcal{X} \to \mathcal{X} \text{ such that } F \text{ has a modulus of continuity } \alpha \text{ and is non-singular, } \mathcal{K}_F \text{ is bounded}\},\\
\Omega_{\mathcal{X}}^{m} &= \{F : \mathcal{X} \to \mathcal{X} \text{ such that } F \text{ is measure-preserving}\},\\
\Omega_{\mathcal{X}}^{\alpha,m} &= \Omega_{\mathcal{X}}^{\alpha} \cap \Omega_{\mathcal{X}}^{m}.
\end{aligned}$$

We consider the problem functions $\Xi_{\mathrm{Sp}_{\mathrm{ap}}}(F) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$ and $\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}(F) = \mathrm{Sp}_{\mathrm{ap},\epsilon}(\mathcal{K}_F)$ for any $\epsilon > 0$, and take $\Lambda = \{F : x \mapsto F(x) : x \in \mathcal{X}\}$.

Example B.4 (EDMD does not work for $\Omega_{\mathcal{X}}^{\alpha,m}$)**.** Let $\mathcal{X} = \mathbb{T}$ (the unit circle), equipped with the usual measure ω, and consider the doubling map $F(x) = x^2$. To apply the algorithm EDMD, we use the Fourier basis

$$\psi_j(x) = \frac{x^j}{\sqrt{2\pi}}, \quad j \in \mathbb{Z}.$$

Note that $\mathcal{K}_F\psi_j = \psi_{2j}$ and $\mathrm{Sp}(\mathcal{K}_F) = \{z \in \mathbb{C} : |z| \leq 1\}$ ($\mathcal{K}_F$ is an isometry whose range is a strict subspace of $L^2(\mathcal{X}, \omega)$). We may split the space $L^2(\mathcal{X}, \omega)$ into invariant subspaces as follows. Let $n \in \mathbb{N}$ be odd, then $\mathcal{K}_F$ acts as a unilateral shift on $\mathrm{span}\{\psi_{n2^k} : k = 0, 1, \ldots\}$ and $\mathrm{span}\{\psi_{-n2^k} : k = 0, 1, \ldots\}$. Hence, $\mathcal{K}_F$ acts as a direct sum of unilateral shifts. If we use a finite number of Fourier basis functions as our dictionary, the large data limit of EDMD is a direct sum of finite sections of unilateral shifts. These finite matrices have spectrum $\{0\}$, and hence, we completely miss regions of the spectrum. ■

The above example shows that methods such as EDMD do not converge, even for the class $\Omega_{\mathcal{X}}^{\alpha,m}$. This kind of argument can also be extended to systems in $\Omega_{\mathcal{X}}^{\alpha,m}$ whose Koopman operator is not just an isometry, but also unitary. Nevertheless, part of the following theorem says we can ensure convergence using a different algorithm.

Theorem B.5. *Given the above setup, for $\epsilon > 0$, we have the following classifications:*

$$\begin{aligned}
\{\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}, \Omega_{\mathcal{X}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} &\in \Sigma_2^G, \qquad \{\Xi_{\mathrm{Sp}_{\mathrm{ap}}}, \Omega_{\mathcal{X}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Pi_3^G\\
\{\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}, \Omega_{\mathcal{X}}^{\alpha}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} &\in \Sigma_1^G, \qquad \{\Xi_{\mathrm{Sp}_{\mathrm{ap}}}, \Omega_{\mathcal{X}}^{\alpha}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Pi_2^G\\
\{\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}, \Omega_{\mathcal{X}}^{m}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} &\in \Sigma_2^G, \qquad \{\Xi_{\mathrm{Sp}_{\mathrm{ap}}}, \Omega_{\mathcal{X}}^{m}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Sigma_2^G\\
\{\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}, \Omega_{\mathcal{X}}^{\alpha,m}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} &\in \Sigma_1^G, \qquad \{\Xi_{\mathrm{Sp}_{\mathrm{ap}}}, \Omega_{\mathcal{X}}^{\alpha,m}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Sigma_1^G.
\end{aligned}$$

Proof. **Step 1:** Classifications for $\Omega_{\mathcal{X}}^{\alpha}$. Given $F \in \Omega_{\mathcal{X}}^{\alpha}$, we consider the *infinite* matrices A and L acting on $l^2(\mathbb{N})$ with

$$A_{i,j} = \langle \mathcal{K}_F g_j, g_i\rangle = \int_{\mathcal{X}} g_j(F(x))\overline{g_i(x)}\,\mathrm{d}\omega(x), \quad L_{i,j} = \langle \mathcal{K}_F^*\mathcal{K}_F g_j, g_i\rangle = \langle \mathcal{K}_F g_j, \mathcal{K}_F g_i\rangle = \int_{\mathcal{X}} g_j(F(x))\overline{g_i(F(x))}\,\mathrm{d}\omega(x), \quad i,j \in \mathbb{N}.$$

We first claim that if $F \in \Omega_{\mathcal{X}}^{\alpha}$, then given any $i, j \in \mathbb{N}$ and $\delta > 0$, there exists a general algorithm using Δ_1-information that computes an approximation of $A_{i,j}$ and $L_{i,j}$ within an error bounded by δ. We show this for $A_{i,j}$, and the case of $L_{i,j}$ is similar.

Since $\mathcal{X}$ is a compact metric space, given any $\eta > 0$, there exists a finite subset $\{x_{1,\eta}, \ldots, x_{N_\eta,\eta}\} \subset \mathcal{X}$ and continuous functions $\rho_{p,\eta} : \mathcal{X} \to [0, 1]$ such that

$$\sum_{p=1}^{N_\eta} \rho_{p,\eta}(x) = 1 \quad \forall x \in \mathcal{X}, \quad \mathrm{supp}(\rho_{p,\eta}) \subset \{x \in \mathcal{X} : d_{\mathcal{X}}(x, x_{p,\eta}) < \eta\} \text{ for } p = 1, \ldots, N_\eta.$$

Let $y_{p,\eta}$ be an approximation of $F(x_{p,\eta})$ to accuracy η, which can be computed form the given Δ_1-information. We then approximate the integral $A_{i,j}$ by

$$\Gamma_\eta(F) = \sum_{p=1}^{N_\eta} \int_{\mathcal{X}} \rho_{p,\eta}(x) g_j(y_{p,\eta})\overline{g_i(x)}\,\mathrm{d}\omega(x).$$

To bound the error in this approximation, note that if $x \in \mathrm{supp}(\rho_{p,\eta})$, then $d_{\mathcal{X}}(x, x_{p,\eta}) < \eta$. Recall that α is a modulus of continuity for F. Let α_j be a modulus of continuity for g_j, then

$$\begin{aligned}
|g_j(F(x)) - g_j(y_{p,\eta})| &\leq |g_j(F(x)) - g_j(F(x_{p,\eta}))| + |g_j(F(x_{p,\eta})) - g_j(y_{p,\eta})|\\
&\leq \alpha_j(d_{\mathcal{X}}(F(x), F(x_{p,\eta}))) + \alpha_j(d_{\mathcal{X}}(F(x_{p,\eta}), y)) \leq \alpha_j(\alpha(\eta)) + \alpha_j(\eta).
\end{aligned}$$

It follows that

$$|A_{i,j} - \Gamma_\eta(F)| \leq \int_{\mathcal{X}} \sum_{p=1}^{N_\eta} \rho_{p,\eta}(x)|g_j(F(x)) - g_j(y_{p,\eta})||g_i(x)|\,\mathrm{d}\omega(x) \leq [\alpha_j(\alpha(\eta)) + \alpha_j(\eta)] \int_{\mathcal{X}} |g_i(x)|\,\mathrm{d}\omega(x).$$

We can make this bound smaller than a given $\delta > 0$ by choosing η sufficiently small.

For a given $n \in \mathbb{N}$, let $\mathcal{P}_n$ be the orthogonal projection onto the span of the first n canonical basis vectors of $l^2(\mathbb{N})$. We define the function

$$h_n(z, F) = \sigma_{\inf}((A - zI)\mathcal{P}_n^*), \quad \text{where} \quad \sigma_{\inf}(T) = \inf\{\|Tx\| : \|x\| = 1\},$$

and we view $(A - zI)\mathcal{P}_n^*$ as an operator from the range of $\mathcal{P}_n$ to $l^2(\mathbb{N})$. Since $A^*A = L$, we can rewrite h_n as

$$h_n(z, F) = \sqrt{\sigma_{\inf}(\mathcal{P}_n(A - zI)^*(A - zI)\mathcal{P}_n^*)} = \sqrt{\sigma_{\inf}(\mathcal{P}_n[L - \overline{z}A - zA^* + |z|^2 I]\mathcal{P}_n^*)}.$$

In particular, the operator $\mathcal{P}_n[L - \bar{z}A - zA^* + |z|^2 I]\mathcal{P}_n^*$ is built from a finite matrix truncation of A and L. It follows that we may compute $h_n(z, F)$ to any desired accuracy using finitely many evaluations of F to a given precision. By Dini's theorem, $h_n(z, F)$ converges locally uniformly to the function $\sigma_{\inf}(A - zI)$. We may now apply the general construction of [100] to see that $\{\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}, \Omega_{\mathcal{X}}^{\alpha}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Sigma_1^G$. Namely, we let $\tilde{h}_n$ be an approximation of h_n computed to accuracy $1/n$ and set

$$\gamma_n(F) = \left\{ z \in \frac{1}{n}(\mathbb{Z} + i\mathbb{Z}) \cap B_n(0) : \tilde{h}_n(z, F) + \frac{1}{n} < \epsilon \right\} \subset \mathrm{Sp}_{\mathrm{ap},\epsilon}(\mathcal{K}_F),$$

which converges to $\mathrm{Sp}_{\mathrm{ap},\epsilon}(\mathcal{K}_F)$ as $n \to \infty$. Note that $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) \subset \mathrm{Sp}_{\mathrm{ap},\epsilon}(\mathcal{K}_F)$ and that $\lim_{\epsilon \downarrow 0} \mathrm{Sp}_{\mathrm{ap},\epsilon}(\mathcal{K}_F) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$. It follows from the classification for $\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}$ that $\{\Xi_{\mathrm{Sp}_{\mathrm{ap}}}, \Omega_{\mathcal{X}}^{\alpha}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Pi_2^G$.

Step 2: Classifications for $\Omega_{\mathcal{X}}$. The proof is similar to the $\Omega_{\mathcal{X}}^{\alpha}$ case, but we do not assume access to α, a modulus of continuity for F. It follows that we can compute any matrix element $A_{i,j}$ or $L_{i,j}$ in one limit without error control. In particular, we let $h_{n_2,n_1}(z, F)$ be functions that we compute with $\lim_{n_1 \to \infty} h_{n_2,n_1}(z, F) = h_{n_2}(z, F)$. However, the set

$$\left\{ z \in \frac{1}{n_2}(\mathbb{Z} + i\mathbb{Z}) \cap B_{n_2}(0) : h_{n_2,n_1}(z, F) < \epsilon \right\}$$

need not converge as $n_1 \to \infty$, since the convergence $\lim_{n_1 \to \infty} h_{n_2,n_1}(z, F) = h_{n_2}(z, F)$ need not be monotonic. To fix this, we define $\Gamma_{n_2,n_1}(F)$ as follows. Let $z \in \frac{1}{n_2}(\mathbb{Z} + i\mathbb{Z}) \cap B_{n_2}(0)$ and consider the separated intervals

$$I_{n_2}^1(\epsilon) = [0, \epsilon - 1/n_2], \quad I_{n_2}^2(\epsilon) = [\epsilon + 1/(2n_2), \infty).$$

Given $h_{n_2,j}(z, F)$ for $j = 1, \ldots, n_2$, let k be the largest such j with $h_{n_2,j}(z, F) \in I_{n_2}^1(\epsilon) \cup I_{n_2}^2(\epsilon)$. If such a k exists with $h_{n_2,k}(z, F) \in I_{n_2}^1(\epsilon)$, then $z \in \Gamma_{n_2,n_1}(F)$. Otherwise, $z \notin \Gamma_{n_2,n_1}(F)$. Since the sequence $h_{n_2,j}(z, F)$ cannot visit both intervals $I_{n_2}^1(\epsilon)$ and $I_{n_2}^2(\epsilon)$ infinitely often as $n_1 \to \infty$, it follows that the limit $\lim_{n_1 \to \infty} \Gamma_{n_2,n_1}(F) = \Gamma_{n_2}(F)$ exists. Moreover,

$$\left\{ z \in \frac{1}{n_2}(\mathbb{Z} + i\mathbb{Z}) \cap B_{n_2}(0) : h_{n_2}(z, F) < \epsilon - \frac{1}{2n_2} \right\} \subset \Gamma_{n_2}(F) \subset \left\{ z \in \frac{1}{n_2}(\mathbb{Z} + i\mathbb{Z}) \cap B_{n_2}(0) : h_{n_2}(z, F) < \epsilon \right\} \subset \mathrm{Sp}_{\mathrm{ap},\epsilon}(\mathcal{K}_F).$$

Hence, $\{\Gamma_{n_2,n_1}\}$ is a Σ_2^G-tower for $\{\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}, \Omega_{\mathcal{X}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1}$. Again by taking $\epsilon \downarrow 0$, we see that $\{\Xi_{\mathrm{Sp}_{\mathrm{ap}}}, \Omega_{\mathcal{X}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Pi_3^G$.

Step 3: Classifications for $\Omega_{\mathcal{X}}^m$. Since $\Omega_{\mathcal{X}}^m \subset \Omega_{\mathcal{X}}$, we need only prove the classification for the problem function $\Xi_{\mathrm{Sp}_{\mathrm{ap}}}$. Note that if $F \in \Omega_{\mathcal{X}}^m$, then $\mathcal{K}_F$ is an isometry. It follows from the Wold–von Neumann decomposition [107, Theorem I.1.1] that $\mathcal{K}_F$ can be written as a direct sum of copies of the unilateral shift and a unitary operator. In particular,

$$\sigma_{\inf}(\mathcal{K}_F - zI) = \mathrm{dist}(z, \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)), \quad \mathrm{Sp}_{\mathrm{ap},\epsilon}(\mathcal{K}_F) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) + B_\epsilon(0).$$

We can convert the Σ_2^G tower for $\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}$ to a Σ_2^G tower for $\Xi_{\mathrm{Sp}_{\mathrm{ap}}}$ following the same argument for self-adjoint operators in [50].

Step 4: Classifications for $\Omega_{\mathcal{X}}^{\alpha,m}$. Again, if $F \in \Omega_{\mathcal{X}}^m$, then

$$\sigma_{\inf}(\mathcal{K}_F - zI) = \mathrm{dist}(z, \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)), \quad \mathrm{Sp}_{\mathrm{ap},\epsilon}(\mathcal{K}_F) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F) + B_\epsilon(0).$$

We can argue as in step 1 for $\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}$. We can convert the Σ_1^G tower for $\Xi_{\mathrm{Sp}_{\mathrm{ap},\epsilon}}$ to a Σ_1^G tower for $\Xi_{\mathrm{Sp}_{\mathrm{ap}}}$ as in [100]. □

Remark B.6. It is of interest that there is a constructive computational procedure for $\{\Xi_{\mathrm{Sp}_{\mathrm{ap}}}, \Omega_{\mathcal{X}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1}$ that involves three successive limits. The first limit is the large data limit, which collects more snapshots of the dynamical system. The second limit is the large subspace limit, observing the action of the Koopman operator on more observables. The final limit is the computation of spectra through pseudospectra, which is also how the classical computational spectral problem was solved [49]. This phenomenon of several successive limits occurs in all algorithms for Koopman operators that provably converge. In particular, the above argument using the matrices L is a generalization of the ResDMD algorithm [44, 108]. We shall see below that several successive limits are necessary unless we can control the large data limit (e.g., using a modulus of continuity) and the pseudospectra limit (e.g., by assuming that the system is measure-preserving so that $\sigma_{\inf}(\mathcal{K}_F - zI) = \mathrm{dist}(z, \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F))$).

B.3 Measure-preserving maps on the unit disk

In Theorem B.5, we saw that the approximate point spectrum of Koopman operators associated with measure-preserving dynamical systems can be computed in one limit if we have a bound on the modulus of continuity of F, and two limits otherwise. We now show that this classification is sharp by proving a lower bound.

We consider invertible, measure-preserving dynamical systems on the closed unit disk $\mathcal{X} = \mathrm{Cl}(\mathbb{D}) \subset \mathbb{R}^2$, equipped with the Euclidean metric and standard Lebesgue measure. We set

$$\Omega_{\mathbb{D}} = \{F : \mathrm{Cl}(\mathbb{D}) \to \mathrm{Cl}(\mathbb{D}) \text{ such that } F \text{ is continuous, measure-preserving and invertible}\}$$

and also consider maps with a priori known Lipschitz constant,

$$\Omega_{\mathbb{D}}^L = \{F \in \Omega_{\mathbb{D}} : \mathrm{Lip}(F) \leq L\},$$

where $\mathrm{Lip}(F)$ is the (optimal) Lipschitz constant if F is Lipschitz and $+\infty$ otherwise. We take $\Lambda = \{F : x \mapsto F(x) : x \in \mathrm{Cl}(\mathbb{D})\}$, though our upper bounds can be restricted to evaluations of F at rational points (in Cartesian or polar coordinates).

Theorem B.7. *Given the above setup, we have the following classifications for the spectrum*

$$\Delta_2^{\mathbb{P}} \not\ni \{\Xi_{\mathrm{Sp}}, \Omega_{\mathbb{D}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Sigma_2^A, \qquad \Delta_1^{\mathbb{P}} \not\ni \{\Xi_{\mathrm{Sp}}, \Omega_{\mathbb{D}}^L, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Sigma_1^A.$$

Moreover, the same classifications hold for the pseudospectrum Sp_ϵ *(and the approximate point spectrum and approximate point pseudospectrum). In other words, to compute the spectral sets in one limit, we must be able to bound the variability of* F*.*

Let $F_0 \in \Omega_{\mathbb{D}}$ be the map $(r, \theta) \mapsto (r, \theta + \pi)$, where we use polar coordinates. Using the eigenfunctions of the Dirichlet Laplacian, we see that $\mathrm{Sp}(\mathcal{K}_{F_0}) = \{\pm 1\}$. To prove Theorem B.7, we will need the following technical lemma regarding F_0.

Lemma B.8. *Let* $\mathcal{A} = \{x \in \mathbb{R}^2 : 0 < R \leq |x| \leq r < 1\}$ *be an annulus and* $X = \{x_1, \ldots, x_N\}$ *and* $Y = \{y_1, \ldots, y_N\}$ *two sets of* N *distinct points in* $\mathcal{A}$ *with* $X \cap Y = \emptyset$*. Then there exists a measure-preserving homeomorphism* H *that acts as the identity on* $\mathrm{Cl}(\mathbb{D})\backslash\mathcal{A}$ *such that* $H^{-1} \circ F_0 \circ H(x_j) = y_j$ *for* $j = 1, \ldots, N$*.*

Proof. It suffices to show that there exists a measure-preserving homeomorphism H that acts as the identity on $\mathrm{Cl}(\mathbb{D})\backslash\mathcal{A}$ such that

$$H(x_j) = (r_j, 0), \quad H(y_j) = (r_j, \pi), \quad j = 1, \ldots, N,$$

where we employ polar coordinates and $r > r_1 > \cdots > r_N > R$. Consider a family $\mathcal{G}$ of nested, smooth, closed simple curves filling $\mathrm{Cl}(\mathbb{D})$ such that $\mathcal{G}$ consists of concentric circles on $\mathrm{Cl}(\mathbb{D})\backslash\mathcal{A}$ and otherwise all curves enclose $R \cdot \overline{\mathbb{D}}$. We can construct such a family $\mathcal{G}$ so that x_j and y_j lie on the same curve in $\mathcal{G}$ for all j. Let $\mathfrak{c}$ be a Jordan arc from $\partial\mathbb{D}$ to the origin, such that $\mathfrak{c}$ meets each curve in $\mathcal{G}$ at a single point. Assume that $\mathfrak{c}$ passes through x_j for all j, starting with x_1 then passing through x_2, ..., x_N consecutively on the way to the origin (after reordering the pairs (x_j, y_j) if necessary). Without loss of generality, $\mathcal{G}$ may be represented implicitly by a smooth positive function on $\mathrm{Cl}(\mathbb{D})$ via the formula

$$f(x) = a, \qquad 0 < a < 1$$

with $f(x_1) > \cdots > f(x_N)$. By [109, Theorem 1], there exists a measure-preserving homeomorphism H, acting as the identity on $\mathrm{Cl}(\mathbb{D})\backslash\mathcal{A}$, which maps each the family of curves $\mathcal{G}$ to a family of nested concentric circle in $\mathrm{Cl}(\mathbb{D})$ about the origin. Furthermore, the theorem ensures that H maps the chord $\mathfrak{c}$ onto the radius of $\mathrm{Cl}(\mathbb{D})$ corresponding to zero angular coordinates. It follows that $H(x_j) = (r_j, 0)$ and $|H(y_j)| = r_j$ for all j.

It remains to prove that we can choose f so that the angular coordinate of $H(y_j)$ is π. We employ the formula for H provided in [109, Equations 3 and 4],

$$|H(x)| = f(x), \qquad \theta = \frac{1}{f(x)} \int_{T(x)}^{x} \frac{\mathrm{d}s}{f_n} = \frac{1}{f(x)} \int_{T(x)}^{x} \frac{\mathrm{d}s}{\sqrt{[\partial_1 f(x)]^2 + [\partial_2 f(x)]^2}}, \tag{15}$$

where $T(x)$ is the point of intersection with $\mathfrak{c}$ of the curve of $\mathcal{G}$ through x, f_n is the directional derivative in the outer normal direction, and the integral is evaluated along the curve of $\mathcal{G}$ with s measured in the sense for which the interior is on the left. Let θ_j denote the angular coordinate of $H(y_j)$. By adjusting f only in a small neighborhood of the curve passing through x_j and y_j (avoiding the curves passing through the other points) and making f_n small or large along the path of integration from $T(y_j) = x_j$ to y_j, we can ensure that $\theta_j = \pi$. □

Proof of Theorem B.7. We split the proof into three parts: upper bounds, and proving lower bounds for $\Omega_{\mathbb{D}}$ and $\Omega_{\mathbb{D}}^L$.

Step 1: Upper bounds. To prove the upper bounds, we must alter the proof of Theorem B.5 to arithmetic algorithms. We let $\{g_n\}_{n=1}^{\infty}$ be an orthonormal basis of $L^2(\mathbb{D})$ made up of the eigenfunctions of the Dirichlet Laplacian. These functions can be explicitly expressed in terms of Fourier basis functions in the angle coordinate and Bessel functions of the first kind in the radial coordinate. Let $F \in \Omega_{\mathbb{D}}^L$ and consider the two types of integrals

$$A_{i,j} = \langle \mathcal{K}_F g_j, g_i \rangle = \int_{\mathbb{D}} g_j(F(x)) \overline{g_i(x)}\, \mathrm{d}x, \quad L_{i,j} = \langle \mathcal{K}_F^* \mathcal{K}_F g_j, g_i \rangle = \langle \mathcal{K}_F g_j, \mathcal{K}_F g_i \rangle = \int_{\mathbb{D}} g_j(F(x)) \overline{g_i(F(x))}\, \mathrm{d}x, \quad i, j \in \mathbb{N}.$$

We can bound the modulus of continuity and modulus of the products $g_j(F(x))\overline{g_i(x)}$ and $g_j(F(x))\overline{g_i(F(x))}$. It follows that the integrals can be computed to any desired accuracy using quadrature, for example, Riemann sums. (In practice, other choices may be better.) Similarly, if $F \in \Omega_{\mathbb{D}}$, we can construct arithmetic algorithms that converge to these inner products without error control. The rest of the proofs of upper bounds follow those of the upper bounds in Theorem B.5.

Step 2: $\{\Xi_{\mathrm{Sp}}, \Omega_{\mathbb{D}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \notin \Delta_2^{\mathbb{P}}$. Suppose for a contradiction that $\{\Gamma_n\}$ is an SPGA for $\{\Xi_{\mathrm{Sp}}, \Omega_{\mathbb{D}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1}$ such that

$$\inf_{F\in\Omega_{\mathbb{D}}} \mathbb{P}\left(\lim_{n\to\infty} \Gamma_n(F) = \mathrm{Sp}(\mathcal{K}_F)\right) = \epsilon > 1/2. \tag{16}$$

We will obtain a contradiction by constructing an $F \in \Omega_{\mathbb{D}}$ so that this condition cannot hold. In the following, we work in polar coordinates in $\mathrm{Cl}(\mathbb{D})$. We will also need a smooth surjective bump function $\phi : [0,1] \to [0,1]$ with $\mathrm{supp}(\phi) \in [1/4, 3/4]$. All of the elements of $\Omega_{\mathbb{D}}$ we consider act as the identity on $\{0\} \cup \partial\mathbb{D}$, and hence we may remove these points from the evaluation set Λ without loss of generality. We construct F from a sequence of functions $\{F_k\}_{k=1}^{\infty} \subset \Omega_{\mathbb{D}}$ that are defined inductively as follows.

For the base case $k = 1$, define the function

$$\tilde{F}_1(r,\theta) = (r, \theta + \pi + \phi(r)).$$

A simple argument using Jacobians shows that $\tilde{F}_1$ is measure-preserving. It is also smooth and invertible. We claim that $\mathrm{Sp}(\mathcal{K}_{\tilde{F}_1}) = \mathbb{T}$. To see this, let r_0 be any point in $[0,1]$ for which $\phi(r_0)/\pi \notin \mathbb{Q}$. Let $f_n : [0,1] \to [0,1]$ be non-vanishing continuous functions that are supported on $I_n = [\max\{r_0 - 1/n, 0\}, \min\{r_0 + 1/n, 1\}]$ and for $j \in \mathbb{Z}$, set $f_{n,j}(r,\theta) = c_n f_n(r) e^{ij\theta}$, where c_n is a normalization constant so that $\|f_{n,j}\|_{L^2} = 1$. Then

$$\left\|\mathcal{K}_{\tilde{F}_1} f_{n,j} - e^{ij(\pi+\phi(r_0))} f_{n,j}\right\|_{L^2}^2 = |c_n|^2 \int_0^1 \int_0^{2\pi} f_n(r)^2 |e^{ij\phi(r)} - e^{ij\phi(r_0)}|^2 r \,\mathrm{d}\theta \,\mathrm{d}r \le \sup_{r\in I_n} |e^{ij\phi(r)} - e^{ij\phi(r_0)}|^2.$$

This bound converges to zero and hence $e^{ij(\pi+\phi(r_0))} \in \mathrm{Sp}(\mathcal{K}_{\tilde{F}_1})$. Since $j \in \mathbb{Z}$ was arbitrary, $\mathrm{Sp}(\mathcal{K}_{\tilde{F}_1}) = \mathbb{T}$.

From Equation (16) it follows that there exists $n_1 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\mathrm{dist}(i, \Gamma_{n_1}(\tilde{F}_1)) \le 1\right) \ge \frac{\epsilon}{2} + \frac{1}{4} > 1/2.$$

Moreover, there exists a finite set $X_1 = \{x_{1,1}, x_{1,2}, \ldots, x_{1,N_1}\}$ of points in $\mathbb{D}\backslash\{0\}$ and $\delta_1 > 0$ so that

$$\mathbb{P}\left(\mathrm{dist}(i, \Gamma_{n_1}(\tilde{F}_1)) \le 1 \text{ and } C_1(\tilde{F}_1)\right) \ge \frac{\epsilon}{4} + \frac{3}{8} > 1/2,$$

where $C_1(F)$ is the event that Γ_{n_1} samples F to output $\Gamma_{n_1}(F)$ at the fixed finite set of x-points X_1, and $\delta_1 > 0$ is the minimum allowed error of these measurements. Without loss of generality, we may assume that all the points in X_1 are distinct. We may also assume that the Δ_1-information consists of distinct points $Y_1 = \{y_{1,1}, y_{1,2}, \ldots, y_{1,N_1}\}$ with $|y_{1,j} - \tilde{F}_1(x_{1,j})| \le \delta_1$ such that $X_1 \cap Y_1 = \emptyset$. Let $\mathcal{A}_1 = \{R_1 \le r \le r_1\}$ be a closed annulus in $\mathbb{D}\backslash\{0\}$, whose interior contains $X_1 \cup Y_1$. Using Lemma B.8, there is a measure-preserving homeomorphism H_1 that acts as the identity on $\mathrm{Cl}(\mathbb{D})\backslash\mathcal{A}_1$ such that $H_1^{-1} \circ F_0 \circ H_1(x_{1,j}) = y_{1,j}$. We then define $F_1 = H_1^{-1} \circ F_0 \circ H_1$ so that $F_1(x_{1,j}) = y_{1,j}$. We then provide Γ_{n_1} with exact information for input F_1. By Definition A.12 of an SPGA and Definition A.3,

$$\mathbb{P}\left(\mathrm{dist}(i, \Gamma_{n_1}(F_1)) \le 1 \text{ and } C_1(F_1)\right) = \mathbb{P}\left(\mathrm{dist}(i, \Gamma_{n_1}(\tilde{F}_1)) \le 1 \text{ and } C_1(\tilde{F}_1)\right) \ge \frac{\epsilon}{4} + \frac{3}{8} > 1/2.$$

This completes the base case.

We now proceed inductively. Suppose that $F_k \in \Omega_{\mathbb{D}}$, $n_k \in \mathbb{N}$ and $R_k \in (0,1)$ have been defined. Let $r_{k+1} > 0$ be such that $r_{k+1} < \min\{(k+1)^{-1}, R_k\}$ and define the function

$$\tilde{F}_{k+1}(r,\theta) = \begin{cases} (r, \theta + \pi + \phi(r/r_{k+1})), & \text{if } r \le r_{k+1}, \\ F_k(r,\theta), & \text{otherwise.} \end{cases}$$

Clearly $\tilde{F}_{k+1} \in \Omega_{\mathbb{D}}$. The above argument for $\tilde{F}_1$ extends to show that $\mathrm{Sp}(\mathcal{K}_{\tilde{F}_{k+1}}) = \mathbb{T}$. We may also argue as above to see that there exists $n_{k+1} \in \mathbb{N}$ with $n_{k+1} > n_k$, a finite set $X_{k+1} = \{x_{k+1,1}, x_{k+1,2}, \ldots, x_{k+1,N_{k+1}}\}$ of points $x = (r,\theta)$ with $r < r_{k+1}$ and $\delta_{k+1} > 0$ such that

$$\mathbb{P}\left(\mathrm{dist}(i, \Gamma_{n_{k+1}}(\tilde{F}_{k+1})) \le 1 \text{ and } C_{k+1}(\tilde{F}_{k+1})\right) \ge \frac{\epsilon}{4} + \frac{3}{8} > 1/2,$$

where $C_{k+1}(F)$ is the event that $\Gamma_{n_{k+1}}$ samples F to output $\Gamma_{n_{k+1}}(F)$ at the fixed set of x-points $X_{k+1} \cup \{x : |x| > r_{k+1}\}$ and $\delta_{k+1} > 0$ is the minimum allowed error of these measurements. Without loss of generality, we may assume that all of the points in X_{k+1} are distinct. We may also assume that the Δ_1-information for X_{k+1} consists of distinct points $Y_{k+1} = \{y_{k+1,1}, y_{k+1,2}, \ldots, y_{k+1,N_{k+1}}\}$ with $|y_{k+1,j} - \tilde{F}_{k+1}(x_{k+1,j})| \le \delta_{k+1}$ such that $X_{k+1} \cap Y_{k+1} = \emptyset$. We may also assume that $X_{k+1} \cup Y_{k+1}$ are contained in the interior of the annulus $A_{k+1} = \{R_{k+1} \le r \le r_{k+1}\}$ for some $R_{k+1} < r_{k+1}$. Using Lemma B.8, there is a measure-preserving homeomorphism H_{k+1} that acts as the identity on $\mathrm{Cl}(\mathbb{D})\backslash A_{k+1}$ such that $H_{k+1}^{-1} \circ F_0 \circ H_{k+1}(x_{k+1,j}) = y_{k+1,j}$. We then define

$$F_{k+1}(r,\theta) = \begin{cases} H_{k+1}^{-1} \circ F_0 \circ H_{k+1}(r,\theta), & \text{if } r \le r_{k+1}, \\ F_k(r,\theta), & \text{otherwise,} \end{cases}$$

so that $F_{k+1}(x_{k+1,j}) = y_{k+1,j}$. We then provide $\Gamma_{n_{k+1}}$ with exact information for input F_{k+1}. By Definition A.12 of an SPGA and Definition A.3,

$$\mathbb{P}\left(\operatorname{dist}(i, \Gamma_{n_{k+1}}(F_{k+1})) \leq 1 \text{ and } C_{k+1}(F_{k+1})\right) = \mathbb{P}\left(\operatorname{dist}(i, \Gamma_{n_{k+1}}(\tilde{F}_{k+1})) \leq 1 \text{ and } C_{k+1}(\tilde{F}_{k+1})\right) \geq \frac{\epsilon}{4} + \frac{3}{8} > 1/2.$$

This completes the inductive step.

We now let $F = \lim_{k\to\infty} F_k$ and $H = \lim_{k\to\infty} H_k \circ H_{k-1} \circ \cdots \circ H_1$. We provide each Γ_{n_k} with exact input for F. Since F agrees with F_k for all x with $|x| > r_{k+1}$, this implies that for all $k \in \mathbb{N}$,

$$\mathbb{P}\left(\operatorname{dist}(i, \Gamma_{n_k}(F)) \leq 1\right) \geq \mathbb{P}\left(\operatorname{dist}(i, \Gamma_{n_k}(F)) \leq 1 \text{ and } C_k(F)\right) = \mathbb{P}\left(\operatorname{dist}(i, \Gamma_{n_k}(F_k)) \leq 1 \text{ and } C_k(F_k)\right) > 1/2.$$

The function H is a measure-preserving homeomorphism and $F = H^{-1} \circ F_0 \circ H$. Hence, $\mathcal{K}_H$ is unitary and $\mathcal{K}_F = \mathcal{K}_H^* \mathcal{K}_{F_0} \mathcal{K}_H$ so that $\operatorname{dist}(i, \operatorname{Sp}(\mathcal{K}_F)) > 1$. We can now argue as in step 3 of the proof of Theorem B.1 to see that this contradicts Equation (16).

Step 3: $\{\Xi_{\mathrm{Sp}}, \Omega_{\mathbb{D}}^L, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \notin \Delta_1^{\mathbb{P}}$. Suppose for a contradiction that $\{\Gamma_n\}$ is an SPGA for $\{\Xi_{\mathrm{Sp}}, \Omega_{\mathbb{D}}^L, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1}$ such that

$$\inf_{F \in \Omega_{\mathbb{D}}^L} \mathbb{P}\left(d_{\mathrm{H}}(\Gamma_n(F), \operatorname{Sp}(\mathcal{K}_F)) \leq 2^{-n} \text{ for all } n \in \mathbb{N}\right) = \epsilon > 1/2.$$

Let F_1 be the identity map. We may argue as above to show that there exists a finite set $X \subset \mathcal{X}$ and $\delta > 0$ such that

$$\mathbb{P}\left(d_{\mathrm{H}}(\Gamma_1(F_1), \{1\}) \leq 2^{-1} \text{ and } C(F_1)\right) > 1/2,$$

where $C(F)$ is the event that Γ_1 samples F to output $\Gamma_1(F)$ at the fixed finite set of x-points X, and $\delta > 0$ is the minimum allowed error of these measurements. We follow the argument in step 1 of the proof of Theorem B.1. There exists $r_0 \in [0, 1]$ and $\delta_0 > 0$ such that with

$$F_2(r, \theta) = (r, \theta + \delta_0 \phi(r/r_0)),$$

we have $F_2 \in \Omega_{\mathbb{D}}^L$ and

$$\mathbb{P}\left(d_{\mathrm{H}}(\Gamma_1(F_2), \{1\}) \leq 2^{-1} \text{ and } C(F_2)\right) = \mathbb{P}\left(d_{\mathrm{H}}(\Gamma_1(F_1), \{1\}) \leq 2^{-1} \text{ and } C(F_1)\right) > 1/2.$$

But this contradicts $\operatorname{Sp}(\mathcal{K}_{F_2}) = \mathbb{T}$. □

Remark B.9. We can alter the argument in step 2 of the above proof to smooth functions F on the punctured disk $\mathrm{Cl}(\mathbb{D})\backslash\{0\}$.

B.4 Maps on the unit interval

We now complement Theorem B.7 and show that, even for smooth, invertible systems, where we can bound the variability of F, the spectrum cannot be computed in one limit if we drop the measure-preserving assumption from Theorem B.5. We consider $\mathcal{X} = [0, 1]$ equipped with the Euclidean metric and Lebesgue measure.

Let $c = \{c_n\}_{n=0}^{\infty}$ be a strictly increasing sequence in $[1/2, 1)$ with $c_0 = 1/2$ and $\lim_{n\to\infty} c_n = 1$. We set $c_{-n} = 1 - c_n$ for $n \in \mathbb{N}$ and define the intervals

$$S_n = [c_n, c_{n+1}), \quad n \in \mathbb{Z}.$$

We consider a continuous bijection $F : [0, 1] \to [0, 1]$ with $F(0) = 0$ and $F(1) = 1$ such that

$$F(c_n) = c_{n+1}, \quad F(S_n) = S_{n+1}.$$

In other words, F acts a bijection between S_n and S_{n+1}. To study the spectrum of $\mathcal{K}_F$, we define the ratios

$$a_n = \frac{|S_{n+1}|}{|S_n|}, \quad n \in \mathbb{Z}.$$

We assume that F and F^{-1} are smooth on each subinterval S_n and for $\delta > 0$, we define

$$\Omega_{[0,1]}^{\delta} = \left\{F : [0, 1] \to [0, 1] \text{ such that the above holds for a sequence } c = \{c_n\} \text{ with } \max\{\|F'\|_{L^\infty}, \|(F^{-1})'\|_{L^\infty}\} \leq 1 + \delta\right\}.$$

(The 1 in the $1 + \delta$ here comes from the fact that F is a bijection on the whole interval $[0, 1]$.) Note that if the ratios a_n are bounded and bounded below away from zero, then the piecewise affine function F_c (uniquely) defined by acting as an affine function on each S_n is a particular case for which

$$\max\{\|F_c'\|_{L^\infty}, \|(F_c^{-1})'\|_{L^\infty}\} \leq \max\left\{\sup_{n\in\mathbb{N}} a_n, \sup_{n\in\mathbb{N}} a_n^{-1}\right\}.$$

The map F_c is an example of an affine interval exchange map [110], which are well-studied in dynamical systems theory.

Theorem B.10. *Given the above setup and $\delta > 0$, we have the following classifications for the approximate point spectrum*

$$\Delta_2^{\mathbb{P}} \not\ni \{\Xi_{\mathrm{Sp}_{\mathrm{ap}}}, \Omega_{[0,1]}^{\delta}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Pi_2^A.$$

Moreover, the same classifications hold when restricting to smooth functions in $\Omega_{[0,1]}^{\delta}$, or when restricting to piecewise affine functions in $\Omega_{[0,1]}^{\delta}$.

Remark B.11. Since $\Omega_{[0,1]}^{\delta} \subset \Omega_{\mathcal{X}}^{\alpha}$ for $\alpha(x) = (1+\delta)x$, Theorem B.5 and Theorem B.10 demonstrate that computing $\mathrm{Sp}_{\mathrm{ap}}$ is strictly harder than computing $\mathrm{Sp}_{\mathrm{ap}}$ for this class of Koopman operators.

We need the following technical lemma to prove Theorem B.10. The first part of this lemma states that any $F \in \Omega_{[0,1]}^{\delta}$ induces a bounded Koopman operator $\mathcal{K}_F$ on $L^2([0,1])$. The second part of the lemma gives conditions for the spectrum to be contained in (or equal to) certain annuli.

Lemma B.12. *If $F \in \Omega_{[0,1]}^{\delta}$, then $\|\mathcal{K}_F\| \leq \sqrt{1+\delta}$. Furthermore, suppose that $\lim_{n\to\infty} a_n = \alpha \in (0,1]$ and that*

$$(1-\delta_F)a_n \leq \frac{\mathrm{d}F}{\mathrm{d}x}(x) \leq (1+\delta_F)a_n \qquad \text{and} \quad (1-\delta_F)/a_{n-1} \leq \frac{\mathrm{d}F^{-1}}{\mathrm{d}x}(x) \leq (1+\delta_F)/a_{n-1} \quad \forall x \in S_n$$

for some $\delta_F \geq 0$. Then

$$\mathrm{Sp}(\mathcal{K}_F) \subset \left\{ z \in \mathbb{C} : \sqrt{\alpha/(1+\delta_F)} \leq |z| \leq \sqrt{(1+\delta_F)/\alpha} \right\}.$$

If $\alpha < 1$, then

$$\left\{ z \in \mathbb{C} : \sqrt{\alpha} \leq |z| \leq 1/\sqrt{\alpha} \right\} \subset \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F).$$

and hence $\mathrm{Sp}(\mathcal{K}_{F_c}) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_{F_c}) = \{z \in \mathbb{C} : \sqrt{\alpha} \leq |z| \leq 1/\sqrt{\alpha}\}$.

Proof. For ease of notation, let $b_n = \sup_{t \in S_n} |(F^{-1})'(t)|$. If $g \in L^2([0,1])$ with $\|g\| = 1$, then

$$\|\mathcal{K}g\|^2 = \sum_{n\in\mathbb{Z}} \int_{S_n} |g(F(x))|^2 \,\mathrm{d}x \leq \sum_{n\in\mathbb{Z}} \int_{S_{n+1}} |g(x)|^2 b_{n+1} \,\mathrm{d}x \leq (1+\delta) \sum_{n\in\mathbb{Z}} \int_{S_{n+1}} |g(x)|^2 \,\mathrm{d}x = (1+\delta)\|g\|^2.$$

This bound proves the first part of the lemma, i.e., that $\|\mathcal{K}_F\| \leq \sqrt{1+\delta}$.

Suppose in addition that $\lim_{n\to\infty} a_n = \alpha \in (0,1]$ and that

$$(1-\delta_F)a_n \leq \frac{\mathrm{d}F}{\mathrm{d}x}(x) \leq (1+\delta_F)a_n \qquad \text{and} \quad (1-\delta_F)/a_{n-1} \leq \frac{\mathrm{d}F^{-1}}{\mathrm{d}x}(x) \leq (1+\delta_F)/a_{n-1} \quad \forall x \in S_n$$

for some $\delta_F \geq 0$. We may apply the above argument $l \in \mathbb{N}$ times to see that

$$\|\mathcal{K}^l g\|^{1/l} \leq \left[\sum_{n\in\mathbb{Z}} \int_{S_{n+l}} |g(x)|^2 \,\mathrm{d}x \prod_{j=1}^{l} b_{n+j} \right]^{1/(2l)} \leq \sqrt{\sup_{n\in\mathbb{Z}} \prod_{j=1}^{l} b_{n+j}^{1/l}} \leq \sqrt{\sup_{n\in\mathbb{Z}} \frac{1}{l} \sum_{j=1}^{l} b_{n+j}},$$

where we have used the generalized AM-GM inequality for the last inequality. Now let $d_1 \geq d_2 \geq d_3 \geq \cdots$ be a reordering of the sequence $\{b_n\}_{n\in\mathbb{Z}}$. Then,

$$\limsup_{l\to\infty} \sup_{n\in\mathbb{Z}} \frac{1}{l} \sum_{j=1}^{l} b_{n+j} \leq \limsup_{l\to\infty} \frac{1}{l} \sum_{j=1}^{l} d_j \leq \limsup_{l\to\infty} d_l \leq \frac{(1+\delta_F)}{\alpha}.$$

It follows that $\lim_{l\to\infty} \|\mathcal{K}^l\|^{1/l} \leq \sqrt{(1+\delta_F)/\alpha}$. Gelfand's formula for the spectral radius implies that if $z \in \mathrm{Sp}(\mathcal{K}_F)$, then $|z| \leq \sqrt{(1+\delta_F)/\alpha}$. We can argue in the same manner, using the fact that $\lim_{n\to-\infty} a_n = 1/\alpha \in [1,\infty)$, to see that $\lim_{l\to\infty} \|\mathcal{K}^{-l}\|^{1/l} \leq \sqrt{(1+\delta_F)/\alpha}$ and hence that

$$\inf_{z\in\mathrm{Sp}(\mathcal{K}_F)} |z| = \frac{1}{\sup_{z\in\mathrm{Sp}(\mathcal{K}_{F^{-1}})} |z|} \geq \frac{1}{\sqrt{(1+\delta_F)/\alpha}} = \sqrt{\alpha/(1+\delta_F)}.$$

It follows that

$$\mathrm{Sp}(\mathcal{K}_F) \subset \left\{ z \in \mathbb{C} : \sqrt{\alpha/(1+\delta_F)} \leq |z| \leq \sqrt{(1+\delta_F)/\alpha} \right\}.$$

To prove the inclusion for the approximate point spectrum, suppose that $\alpha < 1$, let $z \in \mathbb{C}$ with $\sqrt{\alpha} < |z| < 1/\sqrt{\alpha}$ and set

$$g = \sum_{n\in\mathbb{Z}} z^n \chi_{S_n}.$$

Given $\eta > 0$ with $\sqrt{\alpha} < \sqrt{\eta} < |z| < 1/\sqrt{\eta} < 1/\sqrt{\alpha}$, let $\rho = \max\{|z|\sqrt{\eta}, \sqrt{\eta}/|z|\} < 1$. For sufficiently large n we have $a_n \leq \eta$. It follows that there exists a constant $C > 0$ such that if $n \geq 0$, then

$$|z|^{2n}|S_n| = |S_0||z|^{2n} \prod_{j=0}^{n-1} a_j \leq C(|z|\sqrt{\eta})^{2n} \leq C\rho^{2n},$$

and if $n > 1$ then

$$|z|^{-2n}|S_{-n}| = |z|^{-2n}|S_{n-1}| = |S_0||z|^{-2n} \prod_{j=0}^{n-2} a_j \leq C(\sqrt{\eta}/|z|)^{2n} \leq C\rho^{2n}.$$

It follows that

$$\|g\|^2 = \sum_{n\in\mathbb{Z}} |z|^{2n}|S_n| \leq 2C \sum_{n=0}^{\infty} \rho^{2n} < \infty.$$

Since F acts as a bijection between S_n and S_{n+1}, $F^{-1}(S_{n+1}) = S_n$ and hence $\mathcal{K}\chi_{S_n} = \chi_{S_{n-1}}$. It follows that $\mathcal{K}g = zg$ and hence z is an eigenvalue. Since z with $\sqrt{\alpha} < |z| < 1/\sqrt{\alpha}$ was arbitrary, the closed annulus $\{z \in \mathbb{C} : \sqrt{\alpha} \leq |z| \leq 1/\sqrt{\alpha}\}$ lies in $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$. The final statement follows from the fact that we can take $\delta_{F_c} = 0$. □

Proof of Theorem B.10. The upper bound immediately follows from Theorem B.5, so we only need to prove the lower bounds. We prove this for piecewise affine functions with exact input (which implies the result for Δ_1-information), and the proof is almost identical when restricting to smooth functions in $\Omega^{\delta}_{[0,1]}$. Suppose for a contradiction that $\{\Gamma_n\}$ is an SPGA for $\{\Xi_{\mathrm{Sp}_{\mathrm{ap}}}, \Omega^{\delta}_{[0,1]}, \mathcal{M}_{\mathrm{H}}, \Lambda\}$ such that

$$\inf_{F\in\Omega^{\delta}_{[0,1]}} \mathbb{P}\left(\lim_{n\to\infty} \Gamma_n(F) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)\right) = \epsilon > 1/2. \tag{17}$$

We consider $F = F_c$ defined by acting as an affine function on each S_n, and choose the sequence $c = \{c_n\}_{n=0}^{\infty}$ to contradict Equation (17). The probabilistic part of the proof is similar to step 2 of the proof of Theorem B.7, so we only sketch the differences.

Let $r_1, r_2 \in (0, 1)$ be such that

$$\max\left\{\frac{1}{r_1}, \frac{1}{r_2}, \frac{1-r_1}{1-r_2}, \frac{1-r_2}{1-r_1}\right\} \leq 1 + \delta. \tag{18}$$

and

$$(1 + 2\tau)\sqrt{r_1} < (1 - 2\tau)\sqrt{r_2}, \tag{19}$$

for some $\tau > 0$. Throughout the rest of the proof, such a choice of r_1, r_2, and τ are fixed. We use a superscript (j) to denote choices of the sequences $\{c_n\}$ and corresponding objects such as the intervals $S_n^{(j)}$. We begin with the sequence $c^{(1)} = \{c_n^{(1)}\}_{n=0}^{\infty}$ chosen so that

$$|S_n^{(1)}| = \beta^{(1)} r_1^n, \quad n = 0, 1, 2, \ldots.$$

Here, $\beta^{(1)} = (1 - r_1)/2$ is such that

$$\sum_{n=0}^{\infty} |S_n^{(1)}| = \frac{\beta^{(1)}}{1 - r_1} = 1/2.$$

Due to the bound Equation (18), $F_{c^{(1)}} \in \Omega^{\delta}_{[0,1]}$. Lemma B.12 implies that

$$\mathrm{Sp}(\mathcal{K}_{F_{c^{(1)}}}) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_{F_{c^{(1)}}}) = \{z \in \mathbb{C} : \sqrt{r_1} \leq |z| \leq 1/\sqrt{r_1}\}.$$

It follows that there exists $n_1, N_1 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\inf\{|z| : z \in \Gamma_{n_1}(F_{c^{(1)}})\} \leq (1 + \tau)\sqrt{r_1} \text{ and } \Lambda_{\Gamma_{n_1}}(F_{c^{(1)}}) \subset \cup_{j=-N_1}^{N_1-1} S_j^{(1)}\right) > 1/2.$$

We now define $c^{(2)}$ by

$$|S_n^{(2)}| = \begin{cases} |S_n^{(1)}|, & \text{if } 0 \leq n \leq N_1, \\ \beta^{(2)} r_2^{n-(N_1+1)}, & \text{if } n > N_1. \end{cases}$$

Here, $\beta^{(2)}$ is chosen so that

$$\frac{\beta^{(2)}}{1 - r_2} = \sum_{n=N_1+1}^{\infty} |S_n^{(2)}| = \sum_{n=N_1+1}^{\infty} |S_n^{(1)}| = |S_{N_1}^{(1)}| \sum_{n=1}^{\infty} r_1^n = |S_{N_1}^{(1)}| \frac{r_1}{1 - r_1}.$$

In particular,

$$\frac{|S_{N_1+1}^{(2)}|}{|S_{N_1}^{(2)}|} = \frac{\beta^{(2)}}{|S_{N_1}^{(1)}|} = \frac{r_1(1 - r_2)}{1 - r_1} \leq \frac{1 - r_2}{1 - r_1}.$$

It follows from the bound Equation (18), that $F_{c^{(2)}} \in \Omega^{\delta}_{[0,1]}$. Lemma B.12 implies that

$$\mathrm{Sp}(\mathcal{K}_{F_{c^{(2)}}}) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_{F_{c^{(2)}}}) = \{z \in \mathbb{C} : \sqrt{r_2} \leq |z| \leq 1/\sqrt{r_2}\}.$$

It follows that there exists $n_2, N_2 \in \mathbb{N}$ such that $n_2 > n_1$, $N_2 > N_1$, and

$$\mathbb{P}\left(\inf\{|z| : z \in \Gamma_{n_2}(F_{c^{(2)}})\} \geq (1-\tau)\sqrt{r_2} \text{ and } \Lambda_{\Gamma_{n_2}}(F_{c^{(2)}}) \subset \cup_{j=-N_2}^{N_2-1} S_j^{(2)}\right) > 1/2.$$

We now continue this process inductively. Let $\hat{r}_j = r_2$ if j is even and $\hat{r}_j = r_1$ if j is odd. Suppose that the intervals $S_n^{(k-1)}$ have been defined for $n \in \mathbb{Z}$ and that $n_{k-1}, N_{k-1} \in \mathbb{N}$ are such that

$$\mathbb{P}\left(\inf\{|z| : z \in \Gamma_{n_{k-1}}(F_{c^{(k-1)}})\} \begin{cases} \leq (1+\tau)\sqrt{r_1}, & \text{if } k \text{ is even} \\ \geq (1-\tau)\sqrt{r_2}, & \text{if } k \text{ is odd} \end{cases} \text{ and } \Lambda_{\Gamma_{n_{k-1}}}(F_{c^{(k-1)}}) \subset \cup_{j=-N_{k-1}}^{N_{k-1}-1} S_j^{(k-1)}\right) > 1/2.$$

We define $c^{(k)}$ by

$$|S_n^{(k)}| = \begin{cases} |S_n^{(k-1)}|, & \text{if } 0 \leq n \leq N_{k-1}, \\ \beta^{(k)} \hat{r}_k^{n-(N_{k-1}+1)}, & \text{if } n > N_{k-1}, \end{cases}$$

Here, $\beta^{(k)}$ is chosen so that

$$\frac{\beta^{(k)}}{1-\hat{r}_k} = \sum_{n=N_{k-1}+1}^{\infty} |S_n^{(k)}| = \sum_{n=N_{k-1}+1}^{\infty} |S_n^{(k-1)}| = |S_{N_{k-1}}^{(k-1)}| \sum_{n=1}^{\infty} \hat{r}_{k-1}^n = |S_{N_{k-1}}^{(k-1)}| \frac{\hat{r}_{k-1}}{1-\hat{r}_{k-1}}.$$

In particular,

$$\frac{|S_{N_{k-1}+1}^{(k)}|}{|S_{N_{k-1}}^{(k)}|} = \frac{\beta^{(k)}}{|S_{N_{k-1}}^{(k-1)}|} = \frac{\hat{r}_{k-1}(1-\hat{r}_k)}{1-\hat{r}_{k-1}} \leq \frac{1-\hat{r}_k}{1-\hat{r}_{k-1}}.$$

It follows from the bound Equation (18), that $F_{c^{(k)}} \in \Omega^{\delta}_{[0,1]}$. Using Lemma B.12 again, we may select $n_k, N_k \in \mathbb{N}$ is such that

$$\mathbb{P}\left(\inf\{|z| : z \in \Gamma_{n_k}(F_{c^{(k)}})\} \begin{cases} \leq (1+\tau)\sqrt{r_1}, & \text{if } k+1 \text{ is even} \\ \geq (1-\tau)\sqrt{r_2}, & \text{if } k+1 \text{ is odd} \end{cases} \text{ and } \Lambda_{\Gamma_{n_k}}(F_{c^{(k)}}) \subset \cup_{j=-N_k}^{N_k-1} S_j^{(k)}\right) > 1/2.$$

and this completes the inductive step.

We can now define c by $c_n = \lim_{k\to\infty} c_n^{(k)}$. We have $F_c \in \Omega^{\delta}_{[0,1]}$. For each $k \in \mathbb{N}$, let B_k be the event

$$\inf\{|z| : z \in \Gamma_{n_k}(F_c)\} \begin{cases} \leq (1+\tau)\sqrt{r_1}, & \text{if } k+1 \text{ is even,} \\ \geq (1-\tau)\sqrt{r_2}, & \text{if } k+1 \text{ is odd.} \end{cases}$$

Due to Definition A.12 of an SPGA and Definition A.3, $\mathbb{P}(B_k) > 1/2$ for all k. From Equation (19), we see that either $\inf\{|z| : z \in \mathrm{Sp}_{\mathrm{ap}}(F_c)\} > (1+\tau)\sqrt{r_1}$ or $\inf\{|z| : z \in \mathrm{Sp}_{\mathrm{ap}}(F_c)\} < (1-\tau)\sqrt{r_2}$ (or both). Assume without loss of generality that the former holds and let $A_m = \cap_{k=m}^{\infty} B_{2k+1}^c$. Note that $\mathbb{P}(A_m) \leq 1 - \mathbb{P}(B_{2m+1}) < 1/2$. However, from Equation (17),

$$\mathbb{P}\left(\cup_{m=1}^{\infty} A_m\right) \geq \mathbb{P}\left(\lim_{n\to\infty} \Gamma_n(F_c) = \mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_{F_c})\right) \geq \epsilon > 1/2.$$

Since $A_1 \subset A_2 \subset A_3 \subset \cdots$, there must exist some A_M with $\mathbb{P}(A_M) > 1/2$, the required contradiction. □

B.5 Computing the spectral type is impossible in one limit

We now consider the problem of determining whether the Koopman operator has a non-trivial eigenvalue. We use the term "trivial" to refer to the eigenvalue $\lambda = 1$, which is an eigenvalue with constant eigenfunction if $\omega(\mathcal{X}) < \infty$. Hence, we consider the problem of determining if there are no eigenvalues except 1. Our results also show that one cannot determine whether the spectral measure on $\mathrm{span}\{1\}^{\perp}$ is continuous. We consider the torus $\mathcal{X} = [-\pi,\pi]_{\mathrm{per}} \times [-\pi,\pi]_{\mathrm{per}} = [-\pi,\pi]^2_{\mathrm{per}}$, equipped with the standard, normalized, Lebesgue measure. We set

$$\Omega_p = \left\{F : [-\pi,\pi]^2_{\mathrm{per}} \to [-\pi,\pi]^2_{\mathrm{per}} \text{ s.t. } F \text{ is smooth, measure-preserving, invertible, } \mathrm{Lip}(F) \leq 2, \mathrm{Lip}(F^{-1}) \leq 2\right\}.$$

In other words, Ω_p includes both assumptions needed to compute spectra in one limit in Theorem B.5. We consider two problems. The first is computing the pure point spectrum (closure of the set of eigenvalues), captured by the problem function:

$$\Xi_p : \Omega_p \ni F \mapsto \mathrm{Sp}_{\mathrm{pp}}(\mathcal{K}_F) = \mathrm{Cl}\left(\{\lambda \in \mathbb{C} : \lambda \text{ is an eigenvalue of } \mathcal{K}_F\}\right) \in \mathcal{M}_{\mathrm{H}}.$$

The second is the following decision problem:

$$\Xi_p^{\mathrm{dec}} : \Omega_p \ni F \mapsto \begin{cases} 0, & \text{if there are no eigenfunctions in } \mathrm{span}\{1\}^{\perp} \\ 1, & \text{otherwise} \end{cases} \in \mathcal{M}_{\mathrm{dec}}.$$

In other words, Ξ_p^{dec} decides whether there are any non-trivial eigenfunctions, which amounts to determining if the system is weakly mixing. The following theorem precisely classifies these two computational problems.

Theorem B.13. *Given the above setup, we have the following classifications for non-trivial eigenvalues*

$$\Delta_2^{\mathbb{P}} \not\ni \{\Xi_p, \Omega_p, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Sigma_2^A, \qquad \Delta_2^{\mathbb{P}} \not\ni \{\Xi_p^{\mathrm{dec}}, \Omega_p, \mathcal{M}_{\mathrm{dec}}, \Lambda\}^{\Delta_1} \in \Sigma_2^A.$$

The following technical lemma is used in the proof of Theorem B.13.

Lemma B.14. *Let $F \in \Omega_p$ be the skew product map $F(x,y) = (x, y + f(x))$, where $f : [-\pi,\pi]_{\mathrm{per}} \to [0,\pi]$ is smooth, symmetric ($f(-x) = f(x)$), $f(-\pi) = 0$, $f(0) = \pi$ and non-decreasing on $[-\pi, 0]$. Then:*

(a) If f is strictly increasing on $[-\pi,0]$, the spectral measure of $\mathcal{K}_F$ with respect to any $g \in L^2(\mathcal{X},\omega)$ is continuous on $\mathrm{span}\{1\}^\perp$. It follows that there are no eigenfunctions in the subspace $\mathrm{span}\{1\}^\perp$.

(b) If there exists an open interval $(a,b) \subset [-\pi,0]$ upon which $f(x) = c > 0$ is constant, then $\{e^{ijc} : j \in \mathbb{Z}\} \subset \mathrm{Sp}_{\mathrm{pp}}(\mathcal{K}_F)$.

Proof. For (a), note that the map F is C^0 conjugate to $G(x,y) = (x, y+x)$. The steps in the proof now follow the proof for the related continuous time system $\dot{x} = 0, \dot{y} = x$ in [70, Section 11]. For (b), suppose that there exists an open interval $(a,b) \subset [-\pi,0]$ upon which $f(x) = c > 0$ is constant. Let $g_j(x,y) = \chi_{(a,b)}(x)e^{ijy}$, where χ_S denotes the indicator function of a set S. Then $\mathcal{K}_F g_j = e^{ijc} g_j$ and the result follows. □

We will also need the following discrete-time RAGE theorem from [91, Proof of Theorem A.2]. The continuous-time RAGE theorem [111, Theorem 2.6], named after Ruelle [112], Amrein and Georgescu [113], and Enss [114], is a classical dynamical characterization of the continuous spectrum of self-adjoint operators, and is a valuable tool in the study of Schrödinger operators.

Theorem B.15. *Let A be a unitary operator acting on a separable Hilbert space $\mathcal{H}$, and $\{\mathcal{P}_n\}_{n\in\mathbb{N}}$ a sequence of increasing finite-rank orthogonal projections such that $\mathcal{P}_n^* \mathcal{P}_n$ converges strongly to the identity. Let $\mathcal{P}_{\mathrm{c}}$ and $\mathcal{P}_{\mathrm{pp}}$ denote the orthogonal projections onto $\mathcal{H}_{\mathrm{c}}$ (continuous part) and $\mathcal{H}_{\mathrm{pp}}$ (pure point part), respectively. Then for any $g \in \mathcal{H}$,*

$$\|\mathcal{P}_{\mathrm{c}} g\|^2 = \lim_{n\to\infty} \lim_{L\to\infty} \frac{1}{2L+1} \sum_{\ell=-L}^{L} \left\| (I - \mathcal{P}_n^* \mathcal{P}_n) A^\ell g \right\|^2, \quad \|\mathcal{P}_{\mathrm{pp}} g\|^2 = \lim_{n\to\infty} \lim_{L\to\infty} \frac{1}{2L+1} \sum_{\ell=-L}^{L} \left\| \mathcal{P}_n A^\ell g \right\|^2.$$

Proof of Theorem B.13. We first prove the upper bounds and then prove the lower bounds.

Step 1: Upper bounds. Let $\{e_j\}_{j=1}^\infty$ be an orthonormal basis of $L^2(\mathcal{X},\omega)$ constructed using a tensor product of Fourier bases. Using the arguments in Step 1 of the proof of Theorem B.5 and those of [84], for any $z \notin \mathbb{T}$ and any $g \in L^2(\mathcal{X},\omega)$ that is a finite linear combination of the $\{e_j\}_{j=1}^\infty$, we can compute $(\mathcal{K}_F - zI)^{-1} g$ with error control using the given Δ_1-information. The proof follows the arguments in [84, Theorem 2.1]. It follows from the unitary version of [51, Theorem 5.1.1.] that $\{\Xi_p, \Omega_p, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Sigma_2^A$.

To deal with Ξ_p^{dec}, let $\mathcal{P}_n$ be the orthogonal projection onto $\mathrm{span}\{e_1, \ldots, e_n\}$. Let g be a finite linear combination of the $\{e_j\}_{j=1}^\infty$. For $n \in \mathbb{N}$ and $\ell \in \mathbb{Z}$, using the unitary version of [51, Theorem 4.4.1], we may compute $m_{n,\ell}(g)$, which is a $1/n$-accurate approximation of $\|\mathcal{P}_n \mathcal{K}_F^\ell g\|^2$. We may also assume that the $m_{n,\ell}(g)$ are increasing in n. We then set,

$$W_{n_2,n_1}(g) = \frac{1}{2n_1+1} \sum_{k=-n_1}^{n_1} m_{n_2,n_1}(g).$$

It follows from Theorem B.15 that

$$\lim_{n_1\to\infty} W_{n_2,n_1}(g) = W_{n_2}(g), \quad \lim_{n_2\to\infty} W_{n_2}(g) = \|\mathcal{P}_{\mathrm{pp}} g\|^2 = \mu_g^{(\mathrm{pp})}(\mathbb{T}),$$

where $\mu_g^{(\mathrm{pp})}$ is the pure point part of the spectral measure of $\mathcal{K}_F$ with respect to g and the convergence as $n_2 \to \infty$ is from below. Let $\{g_k\}_{k=1}^\infty$ be a set of such g that lie in $\mathrm{span}\{1\}^\perp$ and densely fill this subspace. We then set

$$a_{n_2,n_1}(F) = \max_{1\le k\le n_2} W_{n_2,n_1}(g_k).$$

Define the two separated intervals $I_1 = [0, 1/4]$ and $I_2 = [1/2, \infty)$. As $n_1 \to \infty$, $a_{n_2,n_1}(F)$ converges to $a_{n_2}(F) = \max_{1\le k\le n_2} W_{n_2}(g_k)$ and hence cannot visit both I_1 and I_2 infinitely often. For a given n_1, we set $\Gamma_{n_2,n_1}(F) = 0$ if the largest $l = 1, \ldots, n_1$ with $a_{n_2,l} \in I_1 \cup I_2$ has $a_{n_2,l} \in I_1$. If no such l exists, or $a_{n_2,l} \in I_2$, we set $\Gamma_{n_2,n_1}(F) = 1$. If $\Xi_p^{\mathrm{dec}}(F) = 0$, then $\mu_{g_k}^{(\mathrm{pp})}(\mathbb{T}) = 0$ for all k and hence $a_{n_2}(F) = 0$ for all n_2. It follows that $\lim_{n_1\to\infty} \Gamma_{n_2,n_1}(F) = 0$. If $\Xi_p^{\mathrm{dec}}(F) = 1$, then there exists g_k with $\mu_{g_k}^{(\mathrm{pp})}(\mathbb{T}) > 1/2$ and hence $a_{n_2}(F) > 1/2$ for sufficiently large n_2. It follows that $\lim_{n_1\to\infty} \Gamma_{n_2,n_1}(F) = 1$ for sufficiently large n_2. Moreover, since $a_{n_2,n_1}(F)$ are increasing in n_2, $\Gamma_{n_2}(F)$ is increasing in n_2. It follows that $\{\Xi_p^{\mathrm{dec}}, \Omega_p, \mathcal{M}_{\mathrm{dec}}, \Lambda\}^{\Delta_1} \in \Sigma_2^A$.

Step 2: Lower bounds. Suppose for a contradiction that $\{\Gamma_n\}$ is an SPGA for $\{\Xi_p, \Omega_p, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1}$ such that

$$\inf_{F\in\Omega_p} \mathbb{P}\left(\lim_{n\to\infty} \Gamma_n(F) = \mathrm{Sp}_{\mathrm{pp}}(\mathcal{K}_F) \right) = \epsilon > 1/2. \tag{20}$$

We construct $F_f \in \Omega_p$ of the form $F(x,y) = (x, y + f(x))$ with the same conditions as Lemma B.14 (i.e., $f : [-\pi,\pi]_{\mathrm{per}} \to [0,\pi]$ is smooth, symmetric, $f(-\pi) = 0$, $f(0) = \pi$ and non-decreasing on $[-\pi, 0]$) so that Equation (20) cannot hold. By symmetry of f and $f(0) = \pi$, without loss of generality, we may restrict our argument to Λ that sample f on the interval $(-\pi, 0)$.

We use a superscript (j) to denote the choices of f in the following argument. We begin with the function $f^{(1)} \in \Omega_p$ chosen to be strictly increasing at on $[-\pi, 0]$. From Lemma B.14, we see that $\Xi_p(F_{f^{(1)}}) = \{1\}$. It follows that there exists $n_1 \in \mathbb{N}$, $\delta_1 > 0$, and a finite set $X_1 \subset (-\pi, r_1]$ with $r_1 \in (-\pi, 0)$ such that

$$\mathbb{P}\left(\operatorname{dist}(-1, \Gamma_{n_1}(F_{f^{(1)}})) \geq 1 \text{ and } \Lambda_{\Gamma_{n_1}}(F_{f^{(1)}}) \subset X_1 \text{ and } F \text{ is sampled only to within error } \delta_1\right) > 1/2.$$

We choose $f^{(2)} \in \Omega_p$ so that $f^{(2)}(x) = f^{(1)}(x)$ for $x \in (-\pi, r_1]$ and so that $f^{(2)}$ is constant on an open interval containing $r_1/2$, where it takes a value that is not co-rational with π. It follows from Lemma B.14 that $\Xi_p(F_{f^{(2)}}) = \mathbb{T}$. Hence, there exists $n_2 \in \mathbb{N}$ with $n_2 \geq n_1$, $\delta_2 > 0$, and a finite set $X_2 \subset (-\pi, r_2]$ with $r_2 \in (r_1/2, 0)$ such that

$$\mathbb{P}\left(\operatorname{dist}(-1, \Gamma_{n_2}(F_{f^{(2)}})) \leq 1/2 \text{ and } \Lambda_{\Gamma_{n_2}}(F_{f^{(2)}}) \subset X_2 \text{ and } F \text{ is sampled only to within error } \delta_2\right) > 1/2.$$

We choose $f^{(3)} \in \Omega_p$ strictly increasing so that its Δ_1-information for (X_1, δ_1) and (X_2, δ_2) is consistent with that of $f^{(2)}$ and so that $f^{(3)}(x) = f^{(2)}(x) = f^{(1)}(x)$ for $x \in (-\pi, r_1]$. From Lemma B.14, we see that $\Xi_p(F_{f^{(3)}}) = \{1\}$. It follows that there exists $n_3 \in \mathbb{N}$ with $n_3 \geq n_2$, $\delta_3 > 0$, and a finite set $X_3 \subset (-\pi, r_3]$ with $r_3 \in (r_2/2, 0)$ such that

$$\mathbb{P}\left(\operatorname{dist}(-1, \Gamma_{n_3}(F_{f^{(3)}})) \geq 1 \text{ and } \Lambda_{\Gamma_{n_3}}(F_{f^{(3)}}) \subset X_3 \text{ and } F \text{ is sampled only to within error } \delta_3\right) > 1/2.$$

We continue this process inductively, switching between the condition $\operatorname{dist}(-1, \Gamma_{n_k}(F_{f^{(k)}})) \geq 1$ for k odd, and the condition $\operatorname{dist}(-1, \Gamma_{n_k}(F_{f^{(k)}})) \leq 1/2.$ for k even. Moreover, we can enforce $f^{(2m+1)}(x) = f^{(2m)}(x) = f^{(2m-1)}(x)$ for $x \in (-\pi, r_{2m-1}]$ for all $m \in \mathbb{N}$, where $r_{k+1} \geq r_k/2$. It follows that the limit

$$f(x) = \lim_{m \to \infty} f^{(2m+1)}(x)$$

exists. Moreover, we can perform the construction so that $f \in \Omega_p$. For each $k \in \mathbb{N}$, let B_k be the event

$$\operatorname{dist}(-1, \Gamma_{n_k}(F_f)) \begin{cases} \geq 1, & \text{if } k+1 \text{ is even,} \\ \leq 1/2, & \text{if } k+1 \text{ is odd.} \end{cases}$$

Due to Definition A.12 of an SPGA and Definition A.3, $\mathbb{P}(B_k) > 1/2$ for all k. We now argue as in the proof of Theorem B.10 to obtain the required contradiction.

Almost exactly the same proof works for proving $\{\Xi_p^{\mathrm{dec}}, \Omega_p, \mathcal{M}_{\mathrm{dec}}, \Lambda\}^{\Delta_1} \notin \Delta_2^{\mathbb{P}}$. In the above construction, we switched between $\Xi_p^{\mathrm{dec}}(F_{f^{(k)}}) = 0$ for k odd and $\Xi_p^{\mathrm{dec}}(F_{f^{(k)}}) = 1$ for k even. We only need to replace the conditions $\operatorname{dist}(-1, \Gamma_{n_k}(F_{f^{(k)}})) \geq 1$ and $\operatorname{dist}(-1, \Gamma_{n_k}(F_{f^{(k)}})) \leq 1/2.$ by $\Gamma_{n_k}(F_{f^{(k)}}) = 0$ and $\Gamma_{n_k}(F_{f^{(k)}}) = 1$, respectively. □

B.6 Spectra of discrete-space systems

We now consider the statespace $\mathcal{X} = \mathbb{N}$, equipped with the usual counting measure $\omega = \sum_{j=1}^{\infty} \delta_j$, and consider the dynamical systems governed by a function $F : \mathbb{N} \to \mathbb{N} \cup \{\infty\}$. We define the Koopman operator on sequences $x : \mathbb{N} \to \mathbb{C}$ via

$$[\mathcal{K}_F x](j) = \begin{cases} x(F(j)), & \text{if } F(j) \in \mathbb{N}, \\ 0, & \text{otherwise.} \end{cases}$$

We consider $\mathcal{K}_F$ as an operator on $L^2(\mathbb{N}, \omega) \cong l^2(\mathbb{N})$, and assume that it is bounded. Our primary set is therefore

$$\Omega_{\mathbb{N}} = \{F \text{ s.t. } F : \mathbb{N} \to \mathbb{N} \cup \{\infty\} \text{ and } \mathcal{K}_F \text{ is bounded}\}$$

Our evaluation set is $\Lambda = \{a_{i,j} : a_{i,j}(F) = \delta_{F(j),i}\}$ and we are interested in computing $\operatorname{Sp}(\mathcal{K}_F)$ and $\operatorname{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$.

Theorem B.16. *Given the above setup, we have the following classifications:*

$$\{\Xi_{\mathrm{Sp}}, \Omega_{\mathbb{N}}, \mathcal{M}_{\mathrm{H}}, \Lambda\} \notin \Delta_3^G, \qquad \{\Xi_{\mathrm{Sp}}, \Omega_{\mathbb{N}}, \mathcal{M}_{\mathrm{H}}, \Lambda\}^{\Delta_1} \in \Pi_3^A.$$

In other words, we cannot compute the spectrum in two limits via any algorithm, even with exact input. However, we can obtain an arithmetic Π_3-tower even with inexact input. Furthermore, the same classifications hold for computing the approximate point spectrum $\operatorname{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$.

To prove Theorem B.16, our strategy will be to embed a certain combinatorial problem into the spectral problem of interest. This problem will have a known lower bound complexity, allowing us to prove the lower bound in Theorem B.16. Specifically, let Ω_{Mat} be the collection of all infinite matrices $a = \{a_{m_1,m_2}\}_{m_1,m_2 \in \mathbb{N}}$ with entries $a_{m_1,m_2} \in \{0, 1\}$ and Λ_{Mat} be the set of component-wise evaluation functions. We consider the formula

$$Q(a) = \begin{cases} 1, & \text{if for all but a finite number of } i,\ \forall j\ \exists n > j \text{ s.t. } a_{n,i} = 1, \\ 0, & \text{otherwise.} \end{cases}$$

In other words, Q decides whether the matrix has only finitely many columns with only finitely many 1's. It was proven in [115] that $\{Q, \Omega_{\mathrm{Mat}}, [0, 1], \Lambda_{\mathrm{Mat}}\} \notin \Delta_3^G$. Note that the metric space is not $\mathcal{M}_{\mathrm{dec}}$, but $[0, 1]$ with the usual metric.

Proof of Theorem B.16. **Step 1: Upper bounds.** As an orthonormal basis of $L^2(\mathbb{N},\omega)$, we consider $\{\phi_n\}_{n=1}^{\infty}$, where $\phi_n(j)=\delta_{n,j}$. Given $F\in\Omega_{\mathbb{N}}$, let A_F be the matrix respresenting $\mathcal{K}_F$ with respect to this basis. We can evaluate any entry of A_F using finitely many evaluations from Λ. The upper bounds now follow from the Π_3^A towers of algorithms in [50].

Step 2: Lower bounds. First, given a sequence $\{c_i\}_{i\in\mathbb{Z}}$ with each $c_i\in\{0,1\}$, we define the bounded operator $C(\{c_i\})$ acting on $l^2(\mathbb{Z})$ via its matrix entries:

$$[C(\{c_i\})]_{k,l}=\begin{cases}1, & k<l, c_k=c_l=1, \text{ and } c_n=0 \text{ for all } k<n<l,\\ 0, & \text{otherwise.}\end{cases}$$

Let $\{e_n\}_{n\in\mathbb{Z}}$ be the canonical basis of $l^2(\mathbb{Z})$. Note that $C(\{c_i\})$ is a shift operator on $\{e_i : c_i=1\}$ and zero on the span of the other basis elements. Given $a\in\Omega_{\mathrm{Mat}}$, we consider the family of sequences

$$c_i^{(j)}=c_i^{(j)}(a)=\begin{cases}1, & |i|\leq j,\\ a_{|i|-j,j}, & \text{otherwise.}\end{cases}$$

If $\sum_{i=1}^{\infty}a_{i,j}=+\infty$, the operator $C(\{c_i^{(j)}\})$ is unitarily equivalent to the direct sum of a zero operator and the bilateral shift operator on $l^2(\mathbb{Z})$. Otherwise, it is unitarily equivalent to the direct sum of a zero operator and a shift operator acting on a finite-dimensional invariant subspace of dimension at least $2j+1$. We then consider the following operator acting on $\oplus_{j=1}^{\infty}l^2(\mathbb{Z})$:

$$A(a)=\bigoplus_{j=1}^{\infty}C\left(\left\{c_i^{(j)}\right\}\right).$$

The above discussion shows that if $Q(a)=1$, then $\mathrm{Sp}(A(a))\subset\{0\}\cup\mathbb{T}$. Otherwise, for any $N\in\mathbb{N}$, there is a finite-dimensional invariant subspace of dimension at least N upon which $A(a)$ acts as a shift operator. Hence, $\mathrm{Sp}(A(a))$ is the complex unit disc in this case. In either case, one can show that $\mathrm{Sp}(A(a))=\mathrm{Sp}_{\mathrm{ap}}(A(a))$. Hence, we will prove the lower bound for the computation of $\mathrm{Sp}(\mathcal{K}_F)$, and the same argument works for $\mathrm{Sp}_{\mathrm{ap}}(\mathcal{K}_F)$.

Suppose for a contradiction that $\{\Gamma_{n_2,n_1}\}$ is a height-two tower of general algorithms for $\{\Xi_{\mathrm{Sp}},\Omega_{\mathbb{N}},\mathcal{M}_{\mathrm{H}},\Lambda\}$. Given $a\in\Omega_{\mathrm{Mat}}$, we consider the function $F=F_a$ defined as follows. As an orthonormal basis of $L^2(\mathbb{N},\omega)$, we consider $\{\phi_n\}_{n=1}^{\infty}$, where $\phi_n(j)=\delta_{n,j}$. We can identify $\oplus_{j=1}^{\infty}l^2(\mathbb{Z})$ with $l^2(\mathbb{N})$ via a computable bijection between the relevant index sets, and, hence, consider $A(a)$ as a bounded operator on $l^2(\mathbb{N})$. Moreover, when viewed as an infinite matrix, this matrix has at most one 1 in each row, and every other entry is 0. By considering this matrix acting with respect to the basis $\{\phi_n\}_{n=1}^{\infty}$, this matrix induces F_a and a corresponding Koopman operator $\mathcal{K}_{F_a}$. In particular, if the nth row of $A(a)$ contains no 1s, then $F_a(n)=\infty$. Otherwise, if the jth column of the nth row is 1, $F_a(n)=j$. In particular, any test of the form "is $F_a(j)=k$" can be evaluated in finitely many calls to Λ_{Mat} with input a. It follows that

$$\tilde{\Gamma}_{n_2,n_1}(a)=\min\left\{2\times\mathrm{dist}(1/2,\Gamma_{n_2,n_1}(A(a))),1\right\}$$

defines general algorithms that map into $[0,1]$ and that use the evaluation set Λ_{Mat}. Moreover, the above discussion shows that $\lim_{n_2\to\infty}\lim_{n_1\to\infty}\tilde{\Gamma}_{n_2,n_1}(a)=Q(a)$, the required contradiction. □

C Experimental details

Here, we provide experimental details for the numerical examples in the main text.

C.1 Duffing oscillator: Computing the spectrum

As an example of the non-convergence of EDMD, we consider the Duffing oscillator:

$$\dot{x}=y,\quad \dot{y}=-\alpha y+x(1-x^2).$$

The state (x,y) is two-dimensional, and we consider a discrete-time system with a time-step of $\Delta t=0.3$. We examine two choices of the parameter α: $\alpha=0$, corresponding to a measure-preserving (specifically, Hamiltonian) system; and $\alpha=0.3$, which leads to a dissipative system with two stable spirals at $(\pm1,0)$ and a saddle at the origin. We first sample conditions uniformly at random within the square $[-2,2]^2$ and use trajectories of length 5 to create an initial dataset. We apply k-means clustering on this dataset to establish centers for N Gaussian radial basis functions as our dictionary. Subsequently, we sample F uniformly at random within the square $[-2,2]^2$ to collect M snapshots, which corresponds to Monte Carlo integration.

The pseudospectra figures in the main text were computed using $N=500$ basis functions and $M=50000$ snapshots, and the algorithm described in Appendix B.2. To compute the spectrum, we used the adaptive procedure outlined in the proof of Theorem B.5, which searches for local minimizers of the injection moduli $\sigma_{\mathrm{inf}}(\mathcal{K}_F-zI)$ over a grid of z points that become dense as $N\to\infty$. The errors in the plot represent the maximum injection moduli over the algorithm's outputs, which are guaranteed to converge to the spectrum with the errors converging to zero. As detailed in the proof of Theorem B.5, in the measure-preserving case, this error metric provides an upper bound on the distance to the spectrum and, in fact, converges uniformly to this distance as $N\to\infty$. We have also shown the error over the output of EDMD, which does not converge.

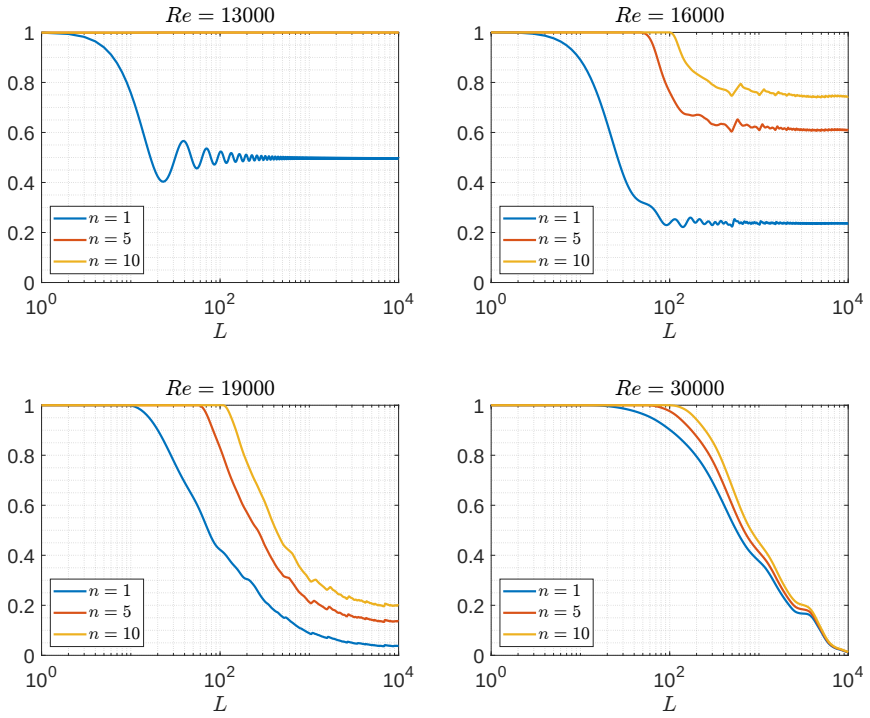

Figure 5: Application of the discrete-time RAGE theorem (Theorem B.15) to the kinetic energy of the lid-driven cavity flow. The plots show $\frac{1}{2L+1}\sum_{\ell=-L}^{L}\|\mathcal{P}_n A^\ell g\|^2$ (for normalized g), which in the double limit $\lim_{n\to\infty}\lim_{L\to\infty}$ converge to the fraction of g that is pure point. This double limit procedure is used in Theorem B.15 to prove upper bounds.

C.2 High Reynolds number fluid flows: Non-trivial eigenfunctions

As an example of detecting non-trivial eigenfunctions using a sharp two-limit procedure, we consider the two-dimensional lid-driven cavity flow of an incompressible viscous fluid confined within a rectangular box with a moving lid and a large Reynolds number, Re. The physical domain is the box (cavity) $B = [-1, 1]^2$ with solid stationary boundaries, except at the top, where the boundary moves with a regularized velocity profile $u_{\text{top}} = (1 - x^2)^2$. This boundary condition is standard in many studies [116, 117, 118] and ensures continuity and incompressibility at the top corners. Using the streamfunction ψ, the Navier–Stokes equations can be expressed as:

$$\frac{\partial}{\partial t}\nabla^2\psi + \frac{\partial\psi}{\partial y}\frac{\partial}{\partial x}\nabla^2\psi - \frac{\partial\psi}{\partial x}\frac{\partial}{\partial y}\nabla^2\psi = \frac{1}{Re}\nabla^4\psi, \qquad \psi|_{\partial B} = 0,\ \frac{\partial\psi}{\partial n}(y=-1) = \frac{\partial\psi}{\partial n}(x=\pm 1) = 0,\ \frac{\partial\psi}{\partial n}(y=1) = u_{\text{top}}.$$

There is a unique solution to these equations, and the flow asymptotically converges to a universal attractor [119]. To solve for ψ, we employ a Chebyshev spectral collocation method [120], using an adaptive discretization size that varies with the Reynolds number and 20000 snapshots.

We apply the discrete-time RAGE theorem (Theorem B.15) to the mean-subtracted total kinetic energy as our observable g. We use time-delay embedding as our projections $\mathcal{P}_n$ corresponding to n time delays. The values $\|\mathcal{P}_n\mathcal{K}^\ell g\|^2$ are computed by applying $\mathcal{K}^\ell$ to the trajectory data of g, which corresponds to a shift of the time series, and then using the Moore–Penrose inverse to apply the projection $\mathcal{P}_n$. The squared norm is computed using the ergodic theorem to approximate inner products. The results are shown in Figure 5 for various choices of Re. We observe the double limit $\lim_{n\to\infty}\lim_{L\to\infty}$ at play. The results indicate that the spectrum is pure point for $Re = 13000$, mixed for $Re = 16000$ and $Re = 19000$, but purely continuous for $Re = 30000$. This is consistent with the findings of [121, 46], where the authors computed spectral measures for this system. The figures in the main text illustrate the extraction of the pure point spectrum for $Re = 19000$. We used the method outlined in the proof of Theorem B.13. Specifically, the top panel shows the smoothed spectral measure with a second-order smoothing kernel (taken from [46]) and a smoothing parameter $\epsilon = 0.001$. The bottom panel displays the extracted pure point spectral measure, computed using the discrete-time RAGE theorem and a partition of the spectrum $\{z : |z| = 1\}$.

References

[1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.

[2] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, July 2021.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.

[4] Yonggang Yao, Qi Dong, Alexandra Brozena, Jian Luo, Jianwei Miao, Miaofang Chi, Chao Wang, Ioannis G. Kevrekidis, Zhiyong Jason Ren, Jeffrey Greeley, Guofeng Wang, Abraham Anapolsky, and Liangbing Hu. High-entropy nanoparticles: Synthesis-structure-property relationships and data-driven discovery. *Science*, 376(6589), April 2022.

[5] Alex Zhavoronkov, Yan A. Ivanenkov, Alex Aliper, Mark S. Veselov, Vladimir A. Aladinskiy, Anastasiya V. Aladinskaya, Victor A. Terentiev, Daniil A. Polykovskiy, Maksim D. Kuznetsov, Arip Asadulaev, Yury Volkov, Artem Zholus, Rim R. Shayakhmetov, Alexander Zhebrak, Lidiya I. Minaeva, Bogdan A. Zagribelnyy, Lennart H. Lee, Richard Soll, David Madge, Li Xing, Tao Guo, and Alán Aspuru-Guzik. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9):1038–1040, September 2019.

[6] C. William Gear, James M. Hyman, Panagiotis G. Kevrekidid, Ioannis G. Kevrekidis, Olof Runborg, and Constantinos Theodoropoulos. Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Communications in Mathematical Sciences*, 1(4):715–762, 2003.

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

[8] Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41(1-3):309–325, August 2005.

[9] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102(21):7426–7431, May 2005.

[10] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, July 2006.

[11] Boaz Nadler, Stéphane Lafon, Ronald R. Coifman, and Ioannis G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, 21(1):113–127, July 2006.

[12] Josh Bongard and Hod Lipson. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 104(24):9943–9948, June 2007.

[13] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, April 2009.

[14] Amit Singer, Radek Erban, Ioannis G. Kevrekidis, and Ronald R. Coifman. Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps. *Proceedings of the National Academy of Sciences*, 106(38):16090–16095, September 2009.

[15] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *Science*, 338(6106):496–500, October 2012.

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

[17] J. Nathan Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. Society for Industrial and Applied Mathematics, November 2016.

[18] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: A review. *Data mining and knowledge discovery*, 33(4):917–963, March 2019.

[19] Steven L. Brunton, Bernd R. Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52(1):477–508, January 2020.

[20] Thomas M. Bury, R. I. Sujith, Induja Pavithran, Marten Scheffer, Timothy M. Lenton, Madhur Anand, and Chris T. Bauch. Deep learning for early warning signals of tipping points. *Proceedings of the National Academy of Sciences*, 118(39), September 2021.

[21] Bernard O. Koopman. Hamiltonian systems and transformation in Hilbert space. *Proceedings of the National Academy of Sciences*, 17(5):315–318, May 1931.

[22] Bernard O. Koopman and John von Neumann. Dynamical systems of continuous spectra. *Proceedings of the National Academy of Sciences*, 18(3):255–263, March 1932.

[23] Marko Budišić, Ryan Mohr, and Igor Mezić. Applied Koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4):047510, December 2012.

[24] Igor Mezić. Analysis of fluid flows via spectral properties of the Koopman operator. *Annual Review of Fluid Mechanics*, 45(1):357–378, January 2013.

[25] Dimitrios Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Applied and Computational Harmonic Analysis*, 47(2):338–396, September 2019.

[26] Samuel E. Otto and Clarence W. Rowley. Koopman operators for estimation and control of dynamical systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):59–87, May 2021.

[27] Steven L. Brunton, Marko Budišić, Eurika Kaiser, and J. Nathan Kutz. Modern Koopman theory for dynamical systems. *SIAM Review*, 64(2):229–340, May 2022.

[28] Matthew J. Colbrook. *The multiverse of dynamic mode decomposition algorithms*, pages 127–230. Elsevier, 2024.

[29] David A. Haggerty, Michael J. Banks, Ervin Kamenar, Alan B. Cao, Patrick C. Curtis, Igor Mezić, and Elliot W. Hawkes. Control of soft robots with inertial dynamics. *Science Robotics*, 8(81), August 2023.

[30] Daniel Bruder, Brent Gillespie, C. David Remy, and Ram Vasudevan. Modeling and control of soft robots using the Koopman operator and model predictive control. In *Robotics: Science and Systems XV*, RSS2019. Robotics: Science and Systems Foundation, June 2019.

[31] Xinyang Wang, Joanna Slawinska, and Dimitrios Giannakis. Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Scientific Reports*, 10(1), February 2020.

[32] Gary Froyland, Dimitrios Giannakis, Benjamin R. Lintner, Maxwell Pike, and Joanna Slawinska. Spectral analysis of climate dynamics with operator-theoretic approaches. *Nature Communications*, 12(1), November 2021.

[33] Antonio Orvieto, Samuel L. Smith, Albert Gu, Anushan Fernando, Caglar Gulcehre, Razvan Pascanu, and Soham De. Resurrecting recurrent neural networks for long sequences. In *International Conference on Machine Learning*, pages 26670–26698. PMLR, 2023.

[34] Joshua L. Proctor and Philip A. Eckhoff. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *International Health*, 7(2):139–145, February 2015.

[35] Bingni W. Brunton, Lise A. Johnson, Jeffrey G. Ojemann, and J. Nathan Kutz. Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *Journal of Neuroscience Methods*, 258:1–15, January 2016.

[36] Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):1–10, November 2018.

[37] Peter J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, July 2010.

[38] Clarence W. Rowley, Igor Mezić, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, November 2009.

[39] Igor Mezić. On numerical approximations of the Koopman operator. *Mathematics*, 10(7):1180, April 2022.

[40] Milan Korda, Mihai Putinar, and Igor Mezić. Data-driven spectral analysis of the Koopman operator. *Applied and Computational Harmonic Analysis*, 48(2):599–629, March 2020.

[41] Dimitrios Giannakis, Amelia Henriksen, Joel A. Tropp, and Rachel Ward. Learning to forecast dynamical systems from streaming data. *SIAM Journal on Applied Dynamical Systems*, 22(2):527–558, May 2023.

[42] Suddhasattwa Das, Dimitrios Giannakis, and Joanna Slawinska. Reproducing kernel Hilbert space compactification of unitary evolution groups. *Applied and Computational Harmonic Analysis*, 54:75–136, September 2021.

[43] Dimitrios Giannakis and Claire Valva. Consistent spectral approximation of Koopman operators using resolvent compactification. *Nonlinearity*, 37(7):075021, June 2024.

[44] Matthew J. Colbrook and Alex Townsend. Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. *Communications on Pure and Applied Mathematics*, 77(1):221–283, July 2023.

[45] Matthew J. Colbrook. The mpEDMD algorithm for data-driven computations of measure-preserving dynamical systems. *SIAM Journal on Numerical Analysis*, 61(3):1585–1608, June 2023.

[46] Matthew J. Colbrook, Catherine Drysdale, and Andrew Horning. Rigged Dynamic Mode Decomposition: Data-driven generalized eigenfunction decompositions for Koopman operators. *arXiv preprint arXiv:2405.00782*, 2024.

[47] Matthew O. Williams, Ioannis G. Kevrekidis, and Clarence W. Rowley. A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science*, 25(6):1307–1346, June 2015.

[48] Milan Korda and Igor Mezić. On convergence of extended dynamic mode decomposition to the Koopman operator. *Journal of Nonlinear Science*, 28(2):687–710, November 2018.

[49] Anders 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, July 2011.

[50] Jonathan Ben-Artzi, Matthew J. Colbrook, Anders C. Hansen, Olavi Nevanlinna, and Markus Seidel. Computing spectra - On the solvability complexity index hierarchy and towers of algorithms. *arXiv*, 2020.

[51] Matthew J. Colbrook. *The foundations of infinite-dimensional spectral computations*. PhD thesis, University of Cambridge, 2020.

[52] Steve Smale. On the efficiency of algorithms of analysis. *Bulletin of the American Mathematical Society*, 13(2):87–122, 1985.

[53] Curt McMullen. Families of rational maps and iterative root-finding algorithms. *Annals of Mathematics*, 125(3):467–493, 1987.

[54] Curt McMullen. Braiding of the attractor and the failure of iterative algorithms. *Inventiones Mathematicae*, 91(2):259–272, 1988.

[55] Peter Doyle and Curt McMullen. Solving the quintic by iteration. *Acta Mathematica*, 163(1):151–180, 1989.

[56] Steve Smale. The work of Curtis T. McMullen. In *Proceedings of the International Congress of Mathematicians 1998*, pages 127–132. EMS Press, 1998.

[57] Gabor Szegő. Beiträge zur Theorie der Toeplitzschen Formen. *Mathematische Zeitschrift*, 6(3-4):167–202, 1920.

[58] Julian Schwinger. The special canonical group. *Proceedings of the National Academy of Sciences*, 46(10):1401–1415, 1960.

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

[60] Ricardo Mañé. *Ergodic Theory and Differentiable Dynamics*. Number v.8. Springer Berlin Heidelberg, Berlin, Heidelberg, 1987. Description based on publisher supplied metadata and other sources.

[61] I. Mezić. Koopman operator, geometry, and learning of dynamical systems. *Notices of the American Mathematical Society*, 68(07):1, August 2021.

[62] Alexandre Mauroy and Igor Mezić. Global stability analysis using the eigenfunctions of the Koopman operator. *IEEE Transactions on Automatic Control*, 61(11):3356–3369, November 2016.

[63] Igor Mezić. On applications of the spectral theory of the Koopman operator in dynamical systems and control theory. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 7034–7041. IEEE, December 2015.

[64] Alexandre Mauroy, Igor Mezić, and Jeff Moehlis. Isostables, isochrons, and Koopman spectrum for the action–angle representation of stable fixed point dynamics. *Physica D: Nonlinear Phenomena*, 261:19–30, October 2013.

[65] Jiajie Chen and Thomas Y. Hou. Stable nearly self-similar blowup of the 2D Boussinesq and 3D Euler equations with smooth data II: Rigorous numerics. *arXiv:2305.05660*, 2023.

[66] J. Cepelewicz. Deep learning poised to "blow up" famed fluid equations, Quanta, April 12, 2022.

[67] Michael I. Gil'. *Operator Functions and Localization of Spectra*. Springer Berlin Heidelberg, 2003.

[68] Peter Henrici. Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices. *Numerische Mathematik*, 4(1):24–40, December 1962.

[69] Zlatko Drmač and Igor Mezić. A data driven Koopman-Schur decomposition for computational analysis of nonlinear dynamics. *arXiv preprint arXiv:2312.15837*, 2023.

[70] Igor Mezić. Spectrum of the Koopman operator, spectral expansions in functional spaces, and state-space geometry. *Journal of Nonlinear Science*, 30(5):2091–2145, December 2020.

[71] Russel E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7:1–49, January 1998.

[72] Alan M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265, 1937.

[73] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale. *Complexity and Real Computation*. Springer eBook Collection. Springer, New York, NY, 1998.

[74] Ricardo Mañé. *Ergodic theory and differentiable dynamics*, volume 8. Springer Science & Business Media, 2012.

[75] Michael Dellnitz and Oliver Junge. On the approximation of complicated dynamical behavior. *SIAM Journal on Numerical Analysis*, 36(2):491–515, January 1999.

[76] Gerhard Keller and Carlangelo Liverani. Stability of the spectrum for transfer operators. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 28(1):141–152, 1999.

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

[78] Gary Froyland, Cecilia González-Tokman, and Anthony Quas. Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools. *Journal of Computational Dynamics*, 1(2):249–278, 2014.

[79] Harry Crimmins and Gary Froyland. Fourier approximation of the statistical properties of Anosov maps on tori. *Nonlinearity*, 33(11):6244–6296, October 2020.

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

[81] Stefan Klus, Feliks Nüske, Péter Koltai, Hao Wu, Ioannis Kevrekidis, Christof Schütte, and Frank Noé. Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, 28(3):985–1010, January 2018.

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

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

[84] Matthew J. Colbrook. Computing spectral measures and spectral types. *Communications in Mathematical Physics*, 384(1):433–501, April 2021.

[85] Hassan Arbabi and Igor Mezić. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator. *SIAM Journal on Applied Dynamical Systems*, 16(4):2096–2126, January 2017.

[86] Nithin Govindarajan, Ryan Mohr, Shivkumar Chandrasekaran, and Igor Mezić. On the approximation of Koopman spectra for measure preserving transformations. *SIAM Journal on Applied Dynamical Systems*, 18(3):1454–1497, January 2019.

[87] Nicolas Boullé and Matthew J. Colbrook. Multiplicative dynamic mode decomposition. *arXiv preprint arXiv:2405.05334*, 2024.

[88] Stefan Klus, Feliks Nüske, Sebastian Peitz, Jan-Hendrik Niemann, Cecilia Clementi, and Christof Schütte. Data-driven approximation of the Koopman generator: Model reduction, system identification, and control. *Physica D: Nonlinear Phenomena*, 406:132416, May 2020.

[89] Nicolas Boullé and Matthew J. Colbrook. On the convergence of Hermitian dynamic mode decomposition. *arXiv preprint arXiv:2401.03192*, 2024.

[90] Tyrus Berry, Dimitrios Giannakis, and John Harlim. Nonparametric forecasting of low-dimensional dynamical systems. *Physical Review E*, 91(3):032915, March 2015.

[91] Jake Fillman and Darren C. Ong. Purely singular continuous spectrum for limit-periodic CMV operators with applications to quantum walks. *Journal of Functional Analysis*, 272(12):5107–5143, June 2017.

[92] Nicolas Boullé, Diana Halikias, and Alex Townsend. Elliptic PDE learning is provably data-efficient. *Proceedings of the National Academy of Sciences*, 120(39), September 2023.

[93] Stefano Luzzatto, Ian Melbourne, and Frederic Paccaut. The Lorenz attractor is mixing. *Communications in Mathematical Physics*, 260(2):393–401, August 2005.

[94] Tanja Eisner, Bálint Farkas, Markus Haase, and Rainer Nagel. *Operator theoretic aspects of ergodic theory*, volume 272. Springer International Publishing, 2015.

[95] Matthew J. Colbrook. Computing semigroups with error control. *SIAM Journal on Numerical Analysis*, 60(1):396–422, 2022.

[96] Simon Becker and Anders Hansen. Computing solutions of schr\" odinger equations on unbounded domains-on the brink of numerical algorithms. *arXiv preprint arXiv:2010.16347*, 2020.

[97] Matthew J. Colbrook, Vegard Antun, and Anders C. Hansen. The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem. *Proceedings of the National Academy of Sciences*, 119(12):e2107151119, 2022.

[98] A. Bastounis, A. C. Hansen, and V. Vlačić. The extended Smale's 9th problem – On computational barriers and paradoxes in estimation, regularisation, computer-assisted proofs and learning. *arXiv preprint arXiv:2110.15734*, 2021.

[99] Matthew J. Colbrook, Bogdan Roman, and Anders C. Hansen. How to compute spectra with error control. *Physical Review Letters*, 122(25):250201, 2019.

[100] Matthew J. Colbrook and Anders C. Hansen. The foundations of spectral computations via the solvability complexity index hierarchy. *Journal of the European Mathematical Society*, 25(12):4639–4728, November 2022.

[101] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[102] Isaac P. Cornfeld, Sergej V. Fomin, and Yakov Grigorevich Sinai. *Ergodic theory*, volume 245. Springer Science & Business Media, 2012.

[103] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate texts in mathematics*. Springer, New York, 1. softcover printing edition, 2000.

[104] Igor Mezić and Stephen Wiggins. A method for visualization of invariant sets of dynamical systems based on the ergodic partition. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 9(1):213–218, 1999.

[105] Donald L. Cohn. *Measure theory*, volume 5. Springer, 2013.

[106] Gerald B. Folland. *Real Analysis: Modern Techniques and their Applications*, volume 40 of *Pure and applied mathematics*. Wiley, New York, 2. ed. edition, 1999. "A Wiley-Interscience publication." - Includes bibliographical references and index.

[107] Béla Sz Nagy, Ciprian Foias, Hari Bercovici, and László Kérchy. *Harmonic analysis of operators on Hilbert space*. Springer New York, 2010.

[108] Matthew J. Colbrook, Lorna J. Ayton, and Máté Szőke. Residual dynamic mode decomposition: Robust and verified Koopmanism. *Journal of Fluid Mechanics*, 955:A21, January 2023.

[109] A. B. Brown and M. Halperin. On certain area-preserving maps. *Annals of Mathematics*, pages 833–837, 1935.

[110] Gilbert Levitt. Feuilletages des surfaces. In *Annales de l'institut Fourier*, volume 32, pages 179–217, 1982.

[111] Michael Aizenman and Simone Warzel. *Random Operators*, volume 168 of *Graduate Studies in Mathematics*. American Mathematical Society, 2015.

[112] David Ruelle. A remark on bound states in potential-scattering theory. *Il Nuovo Cimento A*, 61(4):655–662, 1969.

[113] W. O. Amrein and V. Georgescu. On the characterization of bound states and scattering states in quantum mechanics. *Helvetica Physica Acta. Physica Theoretica. Societatis Physicae Helveticae Commentaria Publica*, 46(5):635–658, 1973.

[114] Volker Enss. Asymptotic completeness for quantum mechanical potential scattering. I. Short range potentials. *Communications in Mathematical Physics*, 61(3):285–291, 1978.

[115] Matthew J. Colbrook. On the computation of geometric features of spectra of linear operators on Hilbert spaces. *Foundations of Computational Mathematics*, 24(3):723–804, December 2022.

[116] Jie Shen. Hopf bifurcation of the unsteady regularized driven cavity flow. *Journal of Computational Physics*, 95(1):228–245, July 1991.

[117] Maciej J. Balajewicz, Earl H. Dowell, and Bernd R. Noack. Low-dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier–Stokes equation. *Journal of Fluid Mechanics*, 729:285–308, July 2013.

[118] Olivier Botella. On the solution of the Navier-Stokes equations using Chebyshev projection schemes with third-order accuracy in time. *Computers & Fluids*, 26(2):107–116, February 1997.

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

[120] Lloyd N. Trefethen. *Spectral methods in MATLAB*. Society for Industrial and Applied Mathematics, January 2000.

[121] Hassan Arbabi and Igor Mezić. Study of dynamics in post-transient flows using Koopman mode decomposition. *Physical Review Fluids*, 2(12):124402, December 2017.