Convergence properties of dynamic mode decomposition for analytic interval maps

Elliz Akindji[1)], Julia Slipantschuk[2)],
Oscar F. Bandtlow[1)], and Wolfram Just[3)†]
[1)]School of Mathematical Sciences,
Queen Mary University of London, London, UK
[2)]Department of Mathematics,
University of Warwick, Coventry, UK
[3)]Institute of Mathematics,
University of Rostock, Rostock, Germany
[†]E-mail: wolfram.just@uni-rostock.de

10th April 2024

Abstract

Extended dynamic mode decomposition (EDMD) is a data-driven algorithm for approximating spectral data of the Koopman operator associated to a dynamical system, combining a Galerkin method of order N and collocation method of order M. Spectral convergence of this method subtly depends on appropriate choice of the space of observables. For chaotic analytic full branch maps of the interval, we derive a constraint between M and N guaranteeing spectral convergence of EDMD.

Keywords: dynamic mode decomposition, transfer operator, Koopman operator

MSC Classification: 37C30, 37E05, 37M10, 37M25, 47A58

1 Context and Results

Extended dynamic mode decomposition (EDMD) is an increasingly popular tool for data analysis in complex systems which is used to identify dynamically relevant modes of a dynamical system based on observations, see for example [28, 27, 30, 31, 6, 38, 26]. The rationale of this approach consists in computing effective modes in dynamical systems, which follows ideas tracing their origins in the context of statistical data analysis [20]. At their core, these methods condense the dynamical observations into a suitably chosen effective linear evolution matrix. The eigenvalues and eigenvectors of this matrix then provide information concerning the relevance and structure of the effective degrees of freedom of the system. While the idea of EDMD has been predominantly pushed by applications in fluid dynamics, the method also aims more broadly to provide useful insight into a variety of real world data analysis problems. Conceptually, it is based on the

observation that general dynamical systems can be described by global evolution operators, known as transfer or Perron-Frobenius operators, or their formal adjoints, known as Koopman operators.

In the last decade considerable progress has been made in the theoretical underpinning of EDMD, for example, by studying data-driven methods to approximate Perron-Frobenius and Koopman operators which feature in the work of Dellnitz and Junge [13] and which have since been extended in many ways, see [12, 15, 23, 22, 16, 11, 40] to name but a few. For broader context, a comparison of various data-driven algorithms can be found in [24], [16] gives a comprehensive overview of the historical context, and [8] provides a recent review of the many variants of dynamic mode decomposition algorithms. In particular, substantial progress has been made in the fundamental understanding of EDMD as a finite-rank approximation scheme of the Koopman operator. For a dynamical system given by a map T, this operator is given by composition with T, and plays a crucial role, in particular for the study of decay of correlations. In most of the literature concerned with data-driven approaches, the operator is considered on the space of square-integrable functions, where for mixing dynamical systems it has a single simple eigenvalue, while other dynamical features are hidden in the continuous spectrum. With data-driven methods it is possible to rigorously approximate the spectral measure in this setting, as for example is done in [25] using Christoffel-Darboux kernels, or in [10] with a residual-based approach to remove spurious eigenvalues, see also [9] for related results.

On the other hand, for certain chaotic dynamical systems of sufficient regularity and with suitably chosen observables, EDMD correctly determines rates of correlation decay. This can be understood by enriching L^2 with generalised functions or distributions, thereby turning the Koopman operator on this extended space into a (quasi)-compact operator with discrete spectrum well-approximated by the EDMD scheme, see [34, 5, 39]. This is explained by the fact that the Koopman operator is adjoint to the transfer operator, which restricted to a dual space (densely and continuously embedded in L^2) enjoys strong spectral properties, see, for example, [21, 2]. This structure can be seen as an instance of a rigged Hilbert space, see [34] for a discussion of this point of view or [36] for a previous use in the dynamical systems context and the very recent work [18] in the data-driven setting.

Nevertheless there remain substantial open questions regarding convergence and quantitative accuracy of the EDMD method. In this article we aim to contribute to this challenge. In order to make quantitative progress, in the spirit of [34, 5] we focus on the simplest kind of complex dynamical systems, chaotic expansive one-dimensional maps, whose statistical long-term behaviour is well understood. We consider a discrete dynamical system given by a map on an interval, say $T\colon [-1,1] \to [-1,1]$, and the associated Koopman operator formally given by $(\mathcal{K}f)(x) = f(T(x))$.

To briefly and informally summarize the EDMD algorithm introduced in [38], one assumes that the dynamics (in our case given by the interval map T) is observed through a set of N observables $\{\psi_0, \dots, \psi_{N-1}\}$. Further one assumes that the dynamics is recorded at M nodes $\{x_0, \dots, x_{M-1}\}$ in the phase space, which might arise from times series data, be sampled from a distribution or be a predefined set. One then solves the generalised eigenvalue problem given by

$$\lambda \sum_{n=0}^{N-1} H_{kn}^{(M)} u_n = \sum_{n=0}^{N-1} G_{kn}^{(M)} u_n \,, \tag{1}$$

where

$$H_{kn}^{(M)} = \frac{1}{M}\sum_{m=0}^{M-1} \psi_k(x_m)\psi_n(x_m)$$
$$G_{kn}^{(M)} = \frac{1}{M}\sum_{m=0}^{M-1} \psi_k(T(x_m))\psi_n(x_m) \tag{2}$$

define the $N \times N$ square matrices $\hat{H}_N^{(M)}$ and $\hat{G}_N^{(M)}$. Ideally, the solutions of the problem (1) are good approximations for the spectral data of the (appropriately defined) Koopman operator. This, however, depends sensitively on the choice and the number N of observables, and the choice and the number M of nodes.

To emphasise the challenge we face, consider the basic textbook example of a piecewise linear full branch map of the interval, the skewed doubling map on $[-1, 1]$ given by

$$T_D(x) = \begin{cases} -1 + 2(x+1)/(1+a) & \text{if} \quad -1 \le x \le a \\ 1 + 2(x-1)/(1-a) & \text{if} \quad a < x \le 1 \end{cases}, \tag{3}$$

where the parameter $a \in (-1, 1)$ determines the skew of the map. For the set of nodes we simply take a lattice of equidistant points, $x_m = -1 + (2m+1)/M$, $m = 0, \ldots, M-1$. For the set of observables, Fourier modes seem to be a sensible choice, that is $\psi_k(x) = \exp(i\pi(k-K)x)$ where $N = 2K+1$ is odd and $k = 0, \ldots, N-1$. If $M > N$, the matrix $\hat{H}_N^{(M)}$ is diagonal[1]. The outcome of the EDMD algorithm, that is, the eigenvalues in (1) for certain values of N and M, are shown in the left panel of Figure 1. These eigenvalues are scattered over the complex plane and the data seem to display a very slow convergence when increasing the number of nodes and observables. Furthermore, these data do not seem to correspond to the rates of correlation decay of the skewed doubling map. For full-branch linear interval maps the (suitably defined) decay rates are in fact well known, see [29] for an elementary account or the appendix of [33] in a slightly more general setting. In particular, for the map (3) these decay rates are given by $\lambda_n = ((1+a)/2)^{n+1} + ((1-a)/2)^{n+1}$, $n \in \mathbb{N}_0$. Since EDMD produces counterintuitive outcomes already in this simple setup there are certainly some issues which need to be clarified.

An alternative setting arises from choosing monomials as the observables, that is, $\psi_k(x) = x^k$ for $k = 0, \ldots, N-1$, while keeping the same set of equidistant points as the nodes. The results of applying EDMD in this case are presented in the right panel of Figure 1. We observe that in this case, the eigenvalues in (1) do approximate λ_n, however the convergence strongly depends on the number of chosen nodes M, and in particular can deteriorate for larger N if M is not increased appropriately.

Our main result (Theorem 2.14) implies the following connection between the eigenvalues of the EDMD matrices and the eigenvalues of the Koopman operator associated to an analytic full branch interval map. The EDMD matrices $\hat{H}_N^{(M)}$ and $\hat{G}_N^{(M)}$ here are defined as in Equation (2), with observables given by monomials, $\psi_k(x) = x^k$ for $k = 0, \ldots, N-1$, and a collection of equidistant nodes $x_m = -1 + \delta + 2m/M \in [-1, 1]$, $m = 0, \ldots, M-1$, for any choice of $0 \le \delta \le 2/M$. Precise definitions and the proof of the result as Corollary 2.17 are given in Section 2; for the moment we simply note that for $\rho > 0$ we write $D_\rho = \{z \in \mathbb{C} : |z| < \rho\}$ for the open disk with radius ρ centred at 0.

Theorem. *Let T be an analytic full branch map on the interval $[-1, 1]$, whose inverse branches $\varphi_\ell, \ell = 1, \ldots, d$, extend analytically to an open disk $D_R \subset \mathbb{C}$ with $\bigcup_\ell \varphi_\ell(D_R) \subseteq D_r$ for some*

[1]Here, for simplicity, we replace the second observable in (2) by its complex conjugate, which amounts to relabelling the matrix indices.

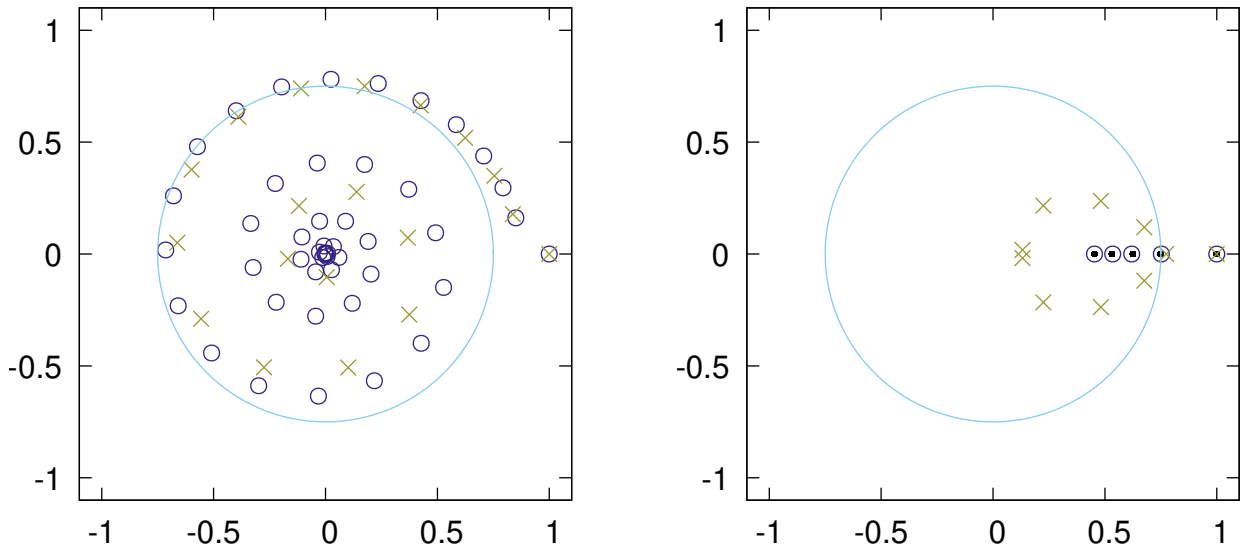

Figure 1: Eigenvalues in the complex plane for the skewed doubling map (Equation (3)) with $a = 1/\sqrt{2}$, obtained via EDMD using Equation (1). Left: The chosen observables are Fourier modes with $N = 50$, $M = 10000$ (blue, circle) and $N = 20$, $M = 5000$ (amber, cross). The circle has a radius determined by the correlation decay rate of smooth observables, $\lambda_1 = (1 + a^2)/2$. Right: The chosen observables are monomials with $N = 5$, $M = 10000$ (blue, circle) and $N = 10$, $M = 10000$ (amber, cross). The dots correspond to $\lambda_n = ((1 + a)/2)^{n+1} + ((1 - a)/2)^{n+1}$ for $n = 0, 1, 2, 3, 4$.

$1 < r < R$, *and assume* $r/R < 1/\gamma$ *with* $\gamma = (1 + \sqrt{2})^2$. *Then there exists a Hilbert space of holomorphic functions* $\mathcal{H}$ *with Banach space dual* $\mathcal{H}'$, *such that the Koopman operator* $\mathcal{K}\colon \mathcal{H}' \to \mathcal{H}'$ *is compact.*

Moreover, for $M \geq N^2 R^N$ *there exist enumerations of the eigenvalues of* $\mathcal{K}$ *and* $(\hat{H}_N^{(M)})^{-1}\hat{G}_N^{(M)}$, $\lambda_k(\mathcal{K})$ *and* $\lambda_k((\hat{H}_N^{(M)})^{-1}\hat{G}_N^{(M)})$ *respectively, such that for each* $k \in \mathbb{N}$, *there are constants* $C > 0$ *and* $b \in (0, 1)$ *so that*

$$|\lambda_k((\hat{H}_N^{(M)})^{-1}\hat{G}_N^{(M)}) - \lambda_k(\mathcal{K})| \leq Cb^N .$$

In short, for analytic full branch interval maps this theorem guarantees exponential convergence of eigendata of the EDMD matrix constructed using N monomials as observables and M equidistant points in the phase space, to those of the associated Koopman operator (defined on a suitable function space), if M satisfies a lower bound in terms of N.

We aim to keep our account self-contained and therefore start with an exposition that can be found elsewhere in the literature and restate some well-known facts. In particular, we introduce in some detail analytic full branch maps and suitable function spaces for the associated transfer operators in Section 2.1, and reiterate some basic features of collocation errors in Section 2.2. The main technical approximation lemmas are presented in Section 2.3 culminating in the proof of our main Theorem 2.14 and its corollaries in Section 2.4. The given estimates are conservative, and a refined approach may allow to relax some of the constraints, such as the constraint on the expansion rate of the map, or the relation between the number of nodes and the number of observables. Therefore we also provide a numerical analysis of EDMD in the context of linear

and nonlinear analytic full branch maps of the interval in Section 3, focusing on maps where the spectrum of the transfer operator is known explicitly, and where we are able to compare EDMD results with the exact expected outcome. These computations indicate that one may indeed be able to derive improved estimates with different tools. Finally, in Section 3.3 we return to the conundrum described above for the choice of Fourier modes presented in the left panel of Figure 1.

2 Error estimates for EDMD

In this section we provide a detailed proof of the convergence properties of EDMD for certain one-dimensional chaotic maps. To keep our presentation self-contained we will include some well-known details and classical facts. Initially, we set up the spectral theory for the dynamical system we consider, and then move on to provide a detailed account of the proof of Theorem 2.14 and of the corollaries.

2.1 Analytic full branch maps and their transfer operators

The following are the standing assumptions we are going to use throughout our exposition. We consider an analytic full branch map $T\colon I \to I$ on the interval $I = [-1, 1]$. By that we mean there exists a collection of closed intervals $\{I_\ell : 1 \leq \ell \leq d\}$ with disjoint interiors, such that $I = \bigcup_\ell I_\ell$, and $T|_{\mathrm{int}(I_\ell)}$ is an analytic diffeomorphism with $\overline{T(I_\ell)} = I$ for each ℓ. We denote by $\varphi_\ell\colon I \to I_\ell$ the analytic inverse branches of the map T, and assume that for some $R > 1$, each φ_ℓ has an analytic extension to an open disk $D_R \subset \mathbb{C}$ with radius R centred at 0. More precisely, we require that $\varphi_\ell \in H^\infty(D_R)$, where

$$H^\infty(D_R) = \{f \in \mathrm{Hol}(D_R) : \sup_{z\in D_R} |f(z)| < \infty\}$$

denotes the usual Banach space of holomorphic functions with the norm

$$\|f\|_{H^\infty(D_R)} = \sup_{z\in D_R} |f(z)|\,.$$

Finally, we assume there exists $1 < r < R$ such that

$$\bigcup_{\ell=1}^{d} \varphi_\ell(D_R) \subseteq D_r\,,$$

which poses a condition on the expansion of the map T.

We associate with the map T the transfer operator $\mathcal{L}\colon L^1(I) \to L^1(I)$ given by

$$(\mathcal{L}f)(z) = \sum_{\ell=1}^{d} \sigma_\ell \varphi_\ell'(z) f(\varphi_\ell(z)), \tag{4}$$

where $\sigma_\ell = \mathrm{sgn}(\varphi_\ell'(0)) = \pm 1$ encodes whether the respective branch of the map is increasing or decreasing. This operator, in this setting also known as the Perron-Frobenius operator, connects to the action of the map via the duality relation

$$\int_{-1}^{1} g(T(x))f(x)\,dx = \int_{-1}^{1} g(x)(\mathcal{L}f)(x)\,dx \quad (f \in L^1(I), g \in L^\infty(I)). \tag{5}$$

We will first show that the expression (4) also gives rise to a well-defined and in fact compact operator when viewed on certain spaces of analytic functions. The main result of this section will be Proposition 2.4, showing that this operator is well-approximated by finite-rank Taylor approximations.

For that purpose we introduce the standard Hardy-Hilbert space

$$H^2(D_R) = \{f \in \mathrm{Hol}(D_R) : \sup_{s<R} \int_0^{2\pi} |f(se^{it})|^2 \, dt < \infty\}$$

with norm

$$\|f\|_{H^2(D_R)} = \sup_{s<R} \sqrt{\frac{1}{2\pi} \int_0^{2\pi} |f(se^{it})|^2 \, dt}\,.$$

Given $f \in \mathrm{Hol}(D_R)$ we can write $f(z) = \sum_{n=0}^{\infty} f_n z^n$, where f_n denotes the n-th Taylor coefficient of f. For fixed $N \in \mathbb{N}_0$ we let

$$(\mathcal{P}_N f)(z) = \sum_{n=0}^{N-1} f_n z^n \tag{6}$$

denote the usual Taylor projection. Later on we will be more specific about the domain and codomain of $\mathcal{P}_N$. For later use we note that $f \in H^2(D_R)$ if and only if $\sum_{n=0}^{\infty} |f_n|^2 R^{-2n} < \infty$, in which case

$$\|f\|^2_{H^2(D_R)} = \sum_{n=0}^{\infty} |f_n|^2 R^{-2n}.$$

We start by recalling the relation between the two classes of spaces introduced above in the following two lemmas.

Lemma 2.1. *Let $\rho \in (r, \infty)$. Then the embedding $\mathcal{J}_1 \colon H^2(D_\rho) \to H^\infty(D_r)$ is compact and $\|\mathcal{J}_1\|_{H^2(D_\rho)\to H^\infty(D_r)} \le \rho/\sqrt{\rho^2 - r^2}$. Moreover,*

$$\|\mathcal{J}_1 - \mathcal{P}_N\|_{H^2(D_\rho)\to H^\infty(D_r)} \le \frac{\rho}{\sqrt{\rho^2 - r^2}} \left(\frac{r}{\rho}\right)^N \qquad \text{for } N \in \mathbb{N}_0\,,$$

where $\mathcal{P}_N \colon H^2(D_\rho) \to H^\infty(D_r)$ denotes the Taylor projection defined in (6).

Proof. Let $f \in H^2(D_\rho)$. Then we can write $f(z) = \sum_{n=0}^{\infty} f_n z^n$ for $z \in D_\rho$. Moreover, for $z \in D_r$ we have

$$\begin{aligned} |((\mathcal{J}_1 - \mathcal{P}_N) f)(z)|^2 = \left|\sum_{n=N}^{\infty} f_n z^n\right|^2 &= \left|\sum_{n=N}^{\infty} f_n \rho^n \left(\frac{z}{\rho}\right)^n\right|^2 \\ &\le \sum_{n=N}^{\infty} |f_n|^2 \rho^{2n} \sum_{n=N}^{\infty} \left|\frac{z}{\rho}\right|^{2n} \le \|f\|^2_{H^2(D_\rho)} \left(\frac{r}{\rho}\right)^{2N} \frac{1}{1 - (r/\rho)^2}\,. \end{aligned}$$

Thus

$$\|(\mathcal{J}_1 - \mathcal{P}_N) f\|_{H^\infty(D_r)} \le \|f\|_{H^2(D_\rho)} \frac{\rho}{\sqrt{\rho^2 - r^2}} \left(\frac{r}{\rho}\right)^N.$$

The estimate for the operator norm follows by taking $N = 0$ in the above. Finally, $\mathcal{J}_1$ is compact, since it is a uniform limit of finite-rank operators. □

Lemma 2.2. *Let $\rho \in (0, R)$. Then the embedding $\mathcal{J}_2 \colon H^\infty(D_R) \to H^2(D_\rho)$ is compact and $\|\mathcal{J}_2\|_{H^\infty(D_R)\to H^2(D_\rho)} \leq 1$. Moreover,*

$$\|\mathcal{J}_2 - \mathcal{P}_N\|_{H^\infty(D_R)\to H^2(D_\rho)} \leq \left(\frac{\rho}{R}\right)^N \qquad \text{for } N \in \mathbb{N}_0\,,$$

where $\mathcal{P}_N \colon H^\infty(D_R) \to H^2(D_\rho)$ denotes the Taylor projection defined in (6).

Proof. Let $f \in H^\infty(D_R)$. Then we can write $f(z) = \sum_{n=0}^\infty f_n z^n$ for $z \in D_R$. For $N \in \mathbb{N}_0$, we have

$$\begin{aligned}\|(\mathcal{J}_2 - \mathcal{P}_N)f\|^2_{H^2(D_\rho)} &= \sup_{s<\rho} \frac{1}{2\pi}\int_0^{2\pi} \left|\sum_{n=N}^\infty f_n s^n e^{int}\right|^2 dt = \sup_{s<\rho} \sum_{n=N}^\infty |f_n|^2 s^{2n} = \sum_{n=N}^\infty |f_n|^2 R^{2n}\left(\frac{\rho}{R}\right)^{2n} \\ &\leq \left(\frac{\rho}{R}\right)^{2N} \sum_{n=R}^\infty |f_n|^2 R^{2n} \leq \left(\frac{\rho}{R}\right)^{2N} \|f\|^2_{H^2(D_R)} \leq \left(\frac{\rho}{R}\right)^{2N} \|f\|^2_{H^\infty(D_R)}\,.\end{aligned}$$

As in the proof of Lemma 2.1, the estimate for the operator norm follows by taking $N = 0$ in the above, and $\mathcal{J}_2$ is compact since it is a uniform limit of finite-rank operators. □

We now turn to the transfer operator defined in (4). Our ultimate aim is to show that it is compact and is well-approximated by operators of finite rank. We start with the following simple observation.

Lemma 2.3. *The transfer operator (4) extends to a bounded operator $\tilde{\mathcal{L}} \colon H^\infty(D_r) \to H^\infty(D_R)$ with*

$$\|\tilde{\mathcal{L}}\|_{H^\infty(D_r)\to H^\infty(D_R)} \leq \sup_{z\in D_R} \sum_{\ell=1}^d |\varphi_\ell'(z)|\,.$$

Proof. Let $f \in H^\infty(D_r)$ and $z \in D_R$. Then

$$|(\tilde{\mathcal{L}}f)(z)| \leq \sum_{\ell=1}^d |\varphi_\ell'(z)||f(\varphi_\ell(z))| \leq \sum_{\ell=1}^d |\varphi_\ell'(z)|\|f\|_{H^\infty(D_r)}$$

so

$$\|\tilde{\mathcal{L}}f\|_{H^\infty(D_R)} \leq \sup_{z\in D_R} \sum_{\ell=1}^d |\varphi_\ell'(z)|\|f\|_{H^\infty(D_r)}$$

and the assertion follows. □

For variants of this result for transfer operators arising from maps with infinitely many branches, see [3].

Now let $\mathcal{J}_1 \colon H^2(D_\rho) \to H^\infty(D_r)$ and $\mathcal{J}_2 \colon H^\infty(D_R) \to H^2(D_\rho)$ denote the canonical embeddings and $\tilde{\mathcal{L}}$ the transfer operator as an operator $\tilde{\mathcal{L}} \colon H^\infty(D_r) \to H^\infty(D_R)$. Then the transfer operator $\mathcal{L}$ viewed as an operator $\mathcal{L} \colon H^2(D_\rho) \to H^2(D_\rho)$ factorises as $\mathcal{L} = \mathcal{J}_2\tilde{\mathcal{L}}\mathcal{J}_1$. Using this factorisation we are now able to estimate the accuracy of particular finite-rank approximations of the transfer operator.

Proposition 2.4. *Let $\rho \in (r, R)$. Then, for $N \in \mathbb{N}_0$ we have*

$$\begin{aligned}\|\mathcal{L} - \mathcal{P}_N\mathcal{L}\|_{H^2(D_\rho)\to H^2(D_\rho)} &\leq C\left(\frac{\rho}{R}\right)^N, \\ \|\mathcal{L} - \mathcal{L}\mathcal{P}_N\|_{H^2(D_\rho)\to H^2(D_\rho)} &\leq C\left(\frac{r}{\rho}\right)^N,\end{aligned}$$

where

$$C = \frac{\rho}{\sqrt{\rho^2 - r^2}} \sup_{z \in D_R} \sum_{\ell=1}^{d} |\varphi_\ell'(z)|$$

and $\mathcal{P}_N \colon H^2(D_\rho) \to H^2(D_\rho)$ *denotes the Taylor projection. In particular*

$$\|\mathcal{L} - \mathcal{P}_N \mathcal{L} \mathcal{P}_N\|_{H^2(D_\rho) \to H^2(D_\rho)} \leq C \left(\left(\frac{\rho}{R}\right)^N + \left(\frac{r}{\rho}\right)^N \right)$$

for $N \in \mathbb{N}_0$. *Choosing* $\rho = \sqrt{Rr}$ *optimises the above convergence rate to*

$$\|\mathcal{L} - \mathcal{P}_N \mathcal{L} \mathcal{P}_N\|_{H^2(D_\rho) \to H^2(D_\rho)} \leq 2C \left(\frac{r}{R}\right)^{N/2} .$$

Proof. Using $\mathcal{L} = \mathcal{J}_2 \tilde{\mathcal{L}} \mathcal{J}_1$, by Lemmas 2.1, 2.2, and 2.3 we have

$$\begin{aligned}
&\|\mathcal{L} - \mathcal{P}_N \mathcal{L}\|_{H^2(D_\rho) \to H^2(D_\rho)} \\
&= \|(\mathcal{J}_2 - \mathcal{P}_N \mathcal{J}_2) \tilde{\mathcal{L}} \mathcal{J}_1\|_{H^2(D_\rho) \to H^2(D_\rho)} \\
&\leq \|\mathcal{J}_2 - \mathcal{P}_N\|_{H^\infty(D_R) \to H^2(D_\rho)} \cdot \|\tilde{\mathcal{L}}\|_{H^\infty(D_r) \to H^\infty(D_R)} \cdot \|\mathcal{J}_1\|_{H^2(D_\rho) \to H^\infty(D_r)} \\
&\leq \left(\frac{\rho}{R}\right)^N \sup_{z \in D_R} \sum_{\ell=1}^{d} |\varphi_\ell'(z)| \frac{\rho}{\sqrt{\rho^2 - r^2}} ,
\end{aligned}$$

and, similarly,

$$\begin{aligned}
&\|\mathcal{L} - \mathcal{L} \mathcal{P}_N\|_{H^2(D_\rho) \to H^2(D_\rho)} \\
&= \|\mathcal{J}_2 \tilde{\mathcal{L}} (\mathcal{J}_1 - \mathcal{J}_1 \mathcal{P}_N)\|_{H^2(D_\rho) \to H^2(D_\rho)} \\
&\leq \|\mathcal{J}_2\|_{H^\infty(D_R) \to H^2(D_\rho)} \cdot \|\tilde{\mathcal{L}}\|_{H^\infty(D_r) \to H^\infty(D_R)} \cdot \|\mathcal{J}_1 - \mathcal{P}_N\|_{H^2(D_\rho) \to H^\infty(D_r)} \\
&\leq \left(\frac{r}{\rho}\right)^N \sup_{z \in D_R} \sum_{\ell=1}^{d} |\varphi_\ell'(z)| \frac{\rho}{\sqrt{\rho^2 - r^2}} .
\end{aligned}$$

The rest follows by observing that

$$\mathcal{L} - \mathcal{P}_N \mathcal{L} \mathcal{P}_N = (\mathcal{L} - \mathcal{P}_N \mathcal{L}) + \mathcal{P}_N (\mathcal{L} - \mathcal{L} \mathcal{P}_N)$$

together with the fact that $\|\mathcal{P}_N\|_{H^2(D_\rho) \to H^2(D_\rho)} = 1$, which follows since $\mathcal{P}_N$ is a self-adjoint projection on $H^2(D_\rho)$. □

2.2 Collocation errors for EDMD

In this section we leverage several classical facts to prove Proposition 2.7, providing basic bounds on the difference between the EDMD matrices in Equation (2) with a finite set of equidistant nodes, and their formal infinite-data limits. The results are stated for expanding analytic full branch maps, but in fact the analyticity assumption can be weakened to piecewise C^1.

Let $M \in \mathbb{N}$. For $m \in \{0, 1, \ldots, M-1\}$ and $0 \leq \delta \leq 2/M$ let $x_m = -1 + \delta + 2m/M$ denote a collection of equidistant nodes in $I = [-1, 1]$. Given $N \in \mathbb{N}$ and $k, \ell \in \{0, \ldots, N-1\}$ let

$$\left(H_N^{(M)}\right)_{k\ell} = \frac{1}{M} \sum_{m=0}^{M-1} x_m^k x_m^\ell, \quad \left(G_N^{(M)}\right)_{k\ell} = \frac{1}{M} \sum_{m=0}^{M-1} (T(x_m))^k \, x_m^\ell, \tag{7}$$

and define the $N \times N$ square matrices $\hat{H}_N^{(M)}$ and $\hat{G}_N^{(M)}$ by $(\hat{H}_N^{(M)})_{k\ell} = (H_N^{(M)})_{k\ell}$ and $(\hat{G}_N^{(M)})_{k\ell} = (G_N^{(M)})_{k\ell}$. These matrices determine EDMD for M equidistant nodes and N observables, where the observables are given by monomials (see Equation (2)).

In order to study the impact of the number of nodes we also introduce for $k, \ell \in \mathbb{N}_0$

$$H_{k\ell} = \frac{1}{2}\int_{-1}^{1} x^k x^\ell \, dx, \quad G_{k\ell} = \frac{1}{2}\int_{-1}^{1} (T(x))^m \, x^\ell \, dx \tag{8}$$

and define the two $N \times N$ square matrices $\hat{H}_N$ and $\hat{G}_N$ by $\left(\hat{H}_N\right)_{k\ell} = H_{k\ell}$ and $\left(\hat{G}_N\right)_{k\ell} = G_{k\ell}$ for $k, \ell \in \{0, \ldots, N-1\}$. These matrices may be viewed as a formal $M \to \infty$ limit of the finite sums occurring in EDMD and we shall sometimes refer to these objects as EDMD matrices with 'infinitely many nodes'. The following lemmas provide the collocation errors of the expressions just introduced.

We begin by recalling the following standard bound for the error in the Riemann integral.

Lemma 2.5. *Let $f\colon [a,b] \to \mathbb{R}$ be piecewise C^1, that is, f is continuously differentiable apart from a finite set of points $S = \{s_1, s_2, \ldots, s_d\} \subset [a,b]$ with $s_1 < s_2 < \ldots < s_d$, and f has an extension to a C^1 function on $[s_{\ell-1}, s_\ell]$ for $\ell \in \{1, \ldots, d+1\}$ where we set $s_0 = a$ and $s_{d+1} = b$. Given $M \in \mathbb{N}$, define $h = (b-a)/M$ and $x_k = a + kh$ for $k \in \{0, \ldots, M\}$. For any $t_k \in [x_{k-1}, x_k]$ and $k \in \{1, \ldots, M\}$ we have*

$$\left| \int_a^b f(x)\, dx - \sum_{k=1}^{M} f(t_k) h \right| \leq \frac{1}{2}\|f'\| \frac{(b-a)^2}{M} + 2d\|f\| \frac{b-a}{M}$$

where

$$\|f'\| = \sup_{x \in [a,b]\setminus S} |f'(x)|, \quad \|f\| = \sup_{x \in [a,b]} |f(x)| \, .$$

Proof. We start by showing that

$$\left| \int_{x_{k-1}}^{x_k} f(x)\, dx - f(t_k) h \right| \leq \frac{1}{2}\|f'\| h^2 + 2\chi_k \|f\| h \tag{9}$$

where

$$\chi_k = \begin{cases} 1 & \text{if} \quad (x_{k-1}, x_k) \cap S \neq \emptyset \\ 0 & \text{if} \quad (x_{k-1}, x_k) \cap S = \emptyset \end{cases} .$$

In order to see this, note that $(x_{k-1}, x_k) \cap S$ is either empty or finite. If it is empty, then

$$\begin{aligned} \left| \int_{x_{k-1}}^{x_k} f(x)\, dx - f(t_k) h \right| &= \left| \int_{x_{k-1}}^{x_k} (f(x) - f(t_k))\, dx \right| \\ &\leq \int_{x_{k-1}}^{x_k} |f(x) - f(t_k)|\, dx \leq \|f'\| \int_{x_{k-1}}^{x_k} |x - t_k|\, dx \leq \frac{1}{2} h^2 \|f'\| \end{aligned}$$

by the Mean Value Theorem, and (9) follows in this case. If $(x_{k-1}, x_k) \cap S \neq \emptyset$, then we use

$$\left| \int_{x_{k-1}}^{x_k} f(x)\, dx - f(t_k) h \right| \leq 2\|f\| h \, ,$$

which proves (9) in this case. Now, using (9) we have

$$\left|\int_a^b f(x)\,dx - \sum_{k=1}^{M} f(t_k)h\right| = \left|\sum_{k=1}^{M}\left(\int_{x_{k-1}}^{x_k} f(x)\,dx - f(t_k)h\right)\right|$$
$$\leq \frac{M}{2}\|f'\|h^2 + \sum_{k=1}^{M}\chi_k\cdot 2\|f\|h \leq \frac{1}{2}\|f'\|\frac{(b-a)^2}{M} + 2d\|f\|\frac{b-a}{M}\,. \quad \square$$

Our subsequent estimates for the matrix norms will rely on a famous lemma of Schur, the short proof of which we recall for the convenience of the reader.

Lemma 2.6. *Let $A \in \mathbb{C}^{N\times N}$ be a complex $N\times N$ matrix and write*

$$R = \sum_{m=1}^{N}\max_{1\leq n\leq N}|A_{mn}|, \quad C = \sum_{n=1}^{N}\max_{1\leq m\leq N}|A_{mn}|\,.$$

Then the spectral norm $\|A\|_2$ of A, that is, the matrix norm induced by the Euclidean norm, satisfies $\|A\|_2 \leq \sqrt{CR}$.

Proof. Let $x \in \mathbb{C}^N$ and $y \in \mathbb{C}^N$. Then by the Cauchy-Schwarz inequality

$$\left|\sum_{k,\ell=1}^{N}\bar{y}_k A_{k\ell}x_\ell\right|^2 \leq \left(\sum_{k,\ell=1}^{N}|y_k||A_{k\ell}|^{1/2}\cdot|A_{k\ell}|^{1/2}|x_\ell|\right)^2$$
$$\leq \sum_{k,\ell=1}^{N}|y_k|^2|A_{k\ell}|\cdot\sum_{k,\ell=1}^{N}|A_{k\ell}||x_\ell|^2 \leq (\|y\|_2)^2 C\cdot R(\|x\|_2)^2\,.$$

Setting $y = Ax$ we obtain $(\|Ax\|_2)^4 \leq CR(\|x\|_2)^2(\|Ax\|_2)^2$, from which the assertion follows. $\square$

We are now able to bound the collocation error caused by taking a finite number of nodes.

Proposition 2.7. *Denote by T an analytic full branch map on $I=[-1,1]$ with d branches, and let S be the set of the $d-1$ 'critical points' of the map T. For $k,\ell \in \{0,\dots,N-1\}$ we have*

$$\left|\left(\hat{H}_N - \hat{H}_N^{(M)}\right)_{k\ell}\right| \leq \frac{k+\ell}{M}\,,$$
$$\left|\left(\hat{G}_N - \hat{G}_N^{(M)}\right)_{k\ell}\right| \leq \max(\|T'\|, 2(d-1))\frac{k+l+1}{M}\,,$$

where $\|T'\| = \sup_{x\in[-1,1]\setminus S}|T'(x)|$.

Moreover

$$\left\|\hat{H}_N - \hat{H}_N^{(M)}\right\|_2 \leq \frac{3}{2}\frac{N^2}{M}\,,$$
$$\left\|\hat{G}_N - \hat{G}_N^{(M)}\right\|_2 \leq \frac{3}{2}\max(\|T'\|, 2(d-1))\frac{N^2}{M}\,.$$

Proof. Using Lemma 2.5 with $f(x) = x^{k+\ell}$ we have

$$\left|\left(\hat{H}_N - \hat{H}_N^{(M)}\right)_{k\ell}\right| = \left|\frac{1}{2}\int_{-1}^{1}f(x)\,dx - \frac{1}{M}\sum_{m=0}^{M-1}f(x_m)\right| \leq \frac{1}{2}\left(\frac{1}{2}\|f'\|\cdot\frac{2^2}{M}\right) = \frac{k+\ell}{M},$$

since $f'(x) = (k+\ell)x^{k+\ell-1}$ and hence $\|f'\| = k+\ell$.

Next we set $f(x) = (T(x))^k x^\ell$ and observe that $f'(x) = k(T(x))^{k-1}T'(x)x^\ell + (T(x))^k \cdot \ell x^{\ell-1}$, and therefore $\|f'\| \le k\|T'\| + \ell \le (k+\ell)\|T'\|$, using that $\|T'\| \ge 1$. Lemma 2.5 now gives

$$\left|\left(\hat{G}_N - \hat{G}_N^{(M)}\right)_{k\ell}\right| = \left|\frac{1}{2}\int_{-1}^{1} f(x)\,dx - \frac{1}{M}\sum_{m=0}^{M-1} f(x_m)\right| \le \frac{1}{2}\left(\frac{1}{2}\|f'\|\frac{2^2}{M} + 2(d-1)\|f\|\frac{2}{M}\right)$$
$$= \frac{1}{2}\left(2\|T'\|\frac{k+\ell}{M} + \frac{2^2(d-1)}{M}\right) \le \max(\|T'\|, 2(d-1))\frac{k+\ell+1}{M}\,.$$

The remaining assertions follow from Lemma 2.6. For instance, if $B \in \mathbb{C}^{N\times N}$ with $|B_{k\ell}| \le k+\ell+1$ for $k,\ell \in \{0,\dots,N-1\}$, then $\max_k |B_{k\ell}| \le N+\ell$ and

$$\sum_{\ell=0}^{N-1}\max_{0\le k\le N-1}|B_{k\ell}| \le \sum_{\ell=0}^{N-1}(N+\ell) = N^2 + \frac{1}{2}N(N-1) \le \frac{3}{2}N^2\,.$$

By symmetry $\sum_{k=0}^{N-1}\max_\ell |B_{k\ell}| \le 3N^2/2$. Thus by Lemma 2.6, we have $\|B\|_2 \le 3N^2/2$. □

2.3 EDMD in an operator setting

This section provides the key technical elements of our paper, preparing the proof of our main results. Proposition 2.11 provides a bound on the Galerkin approximation error, that is, the error incurred from the finite-rank approximation of the transfer operator. Complementarily, in Proposition 2.13 we use the basic bounds from the previous section to establish a bound on the collocation error incurred by the finite-node approximation of the integrals in Equation (8).

The following calculations will be involving matrix representations of operators on $H^2(D_\rho)$ with some fixed $\rho \in (r,R)$ with respect to the standard orthonormal basis $(e_n)_{n\in\mathbb{N}_0}$ given by

$$e_n(z) = \left(\frac{z}{\rho}\right)^n \qquad \text{for } n \in \mathbb{N}_0\,. \tag{10}$$

We will denote by

$$(f,g) = \lim_{s\uparrow\rho}\frac{1}{2\pi}\int_0^{2\pi} f(s\exp(it))\overline{g(s\exp(it))}\,dt$$

the standard inner product in $H^2(D_\rho)$, and by

$$\langle f,g\rangle = \frac{1}{2}\int_{-1}^{1} f(x)\overline{g(x)}\,dx \tag{11}$$

the standard inner product in $L^2([-1,1])$.

With the orthonormal basis above we can write the transfer operator $\mathcal{L}f = \sum_{k\ell}(f,e_\ell)L_{k\ell}e_k$ with

$$L_{k\ell} = (\mathcal{L}e_\ell, e_k)\,. \tag{12}$$

Equation (12) defines a linear operator $L\colon \ell^2(\mathbb{N}_0) \to \ell^2(\mathbb{N}_0)$, where $\|\mathcal{L}\|_{H^2(D_\rho)} = \|L\|_{\ell^2}$. We use Equations (8) to consider G and H as densely defined operators on $\ell^2(\mathbb{N}_0)$. We will also need the following densely defined diagonal operator on $\ell^2(\mathbb{N}_0)$

$$V\colon (x_n)_{n\in\mathbb{N}_0} \mapsto (\rho^n x_n)_{n\in\mathbb{N}_0} \tag{13}$$

and its inverse $V^{-1}\colon \ell^2(\mathbb{N}_0) \to \ell^2(\mathbb{N}_0)$

$$V^{-1}\colon (x_n)_{n\in\mathbb{N}_0} \mapsto (\rho^{-n}x_n)_{n\in\mathbb{N}_0}, \tag{14}$$

which is a compact operator since $\rho > 1$. While V and V^{-1} depend on ρ, we omit to explicitly write out this dependence for notational simplicitly.

We are now going to link the expressions defining EDMD with the transfer operator, and the following lemma is the first in a couple of steps.

Lemma 2.8. *With the above notation, we have $G = HV^{-1}LV$.*

Proof. By Equation (10) we have $x^k = \rho^k e_k(x)$, so that Equations (8) can be written as

$$H_{k\ell} = \rho^{k+\ell}\langle e_\ell, e_k\rangle, \quad G_{k\ell} = \rho^{k+\ell}\langle \mathcal{L}e_\ell, e_k\rangle$$

recalling Equations (5) and (11). Since $\mathcal{L}e_\ell = \sum_n (\mathcal{L}e_\ell, e_n)e_n$, we obtain

$$\begin{aligned} G_{k\ell} = \rho^{k+\ell}\langle \mathcal{L}e_\ell, e_k\rangle &= \sum_{n\in\mathbb{N}_0} \rho^{k+\ell}(\mathcal{L}e_\ell, e_n)\langle e_n, e_k\rangle \\ &= \sum_{n\in\mathbb{N}_0} \rho^{k+\ell}L_{n\ell}\rho^{-n-k}H_{kn} = \sum_{n\in\mathbb{N}_0} H_{kn}\rho^{-n}L_{n\ell}\rho^{\ell}\,. \end{aligned}$$

□

We are now going to introduce operators on $\ell^2(\mathbb{N}_0)$ which capture a finite number of observables. Let $P_N\colon \ell^2(\mathbb{N}_0) \to \ell^2(\mathbb{N}_0)$ denote the canonical orthogonal projection onto the first N coordinates which gives the matrix representation of the Taylor projection $\mathcal{P}_N\colon H^2(D_\rho) \to H^2(D_\rho)$ with respect to the canonical orthogonal basis $(e_n)_{n\in\mathbb{N}_0}$. Let L_N, H_N and G_N denote the orthogonal projections of the operators introduced above, that is,

$$L_N = P_N L P_N, \quad H_N = P_N H P_N, \quad G_N = P_N G P_N \tag{15}$$

defined as operators of rank at most N on $\ell^2(\mathbb{N}_0)$. If we restrict H_N and G_N on the image of P_N, that is, if we consider the $N \times N$ section, we obtain the $N \times N$ square matrices $\hat{H}_N$ and $\hat{G}_N$ mentioned previously. It turns out that $\hat{H}_N$ is invertible (because it is positive definite) and we denote by $H_N^\dagger\colon \ell^2(\mathbb{N}_0) \to \ell^2(\mathbb{N}_0)$ the rank N operator which lifts the matrix $\hat{H}_N^{-1}$ to $\ell^2(\mathbb{N}_0)$. That means $H_N^\dagger$ coincides with $\hat{H}_N^{-1}$ when restricted to the image of P_N, that is, $H_N H_N^\dagger = H_N^\dagger H_N = P_N$ and $H_N^\dagger P_N = P_N H_N^\dagger = H_N^\dagger$. We note that since H_N is symmetric, $H_N^\dagger$ can be seen to be the Moore-Penrose pseudoinverse of H_N.

Lemma 2.9. *For all $N \in \mathbb{N}_0$ we have*

$$\|H_N\|_{\ell^2} \le \|H\|_{\ell^2} \le \pi \quad \text{and} \quad \|H_N^\dagger\|_{\ell^2} \le C(1+\sqrt{2})^{2N}$$

for some $C > 0$.

Proof. Let $F_{k\ell} = 1/(k+\ell+1)$ with $k, \ell \in \mathbb{N}_0$ denote the Hilbert matrix. It is known (see, for example, [7, p.305]) that $\|F\|_{\ell^2} = \pi$. Now

$$H = \frac{1}{2}(F + JFJ)$$

where $J\colon \ell^2(\mathbb{N}_0) \to \ell^2(\mathbb{N}_0)$ is given by $(Jx)_k = (-1)^k x_k$. Since $\|J\|_{\ell^2} = 1$ we have

$$\|H\|_{\ell^2} \le \|F\|_{\ell^2} = \pi\,.$$

Furthermore, with $\|P_N\|_{\ell^2} = 1$ we obtain

$$\|H_N\|_{\ell^2} = \|P_N H P_N\|_{\ell^2} \le \|H\|_{\ell^2}\,.$$

The second inequality follows from [37, Theorem 3.2]. □

We also introduce orthogonal projections of the diagonal operators defined in Equations (13) and (14)

$$V_N = P_N V P_N, \quad V_N^{-1} = P_N V^{-1} P_N \,. \tag{16}$$

Lemma 2.10. *For all $N \in \mathbb{N}_0$ we have*

$$\|V_N H_N^\dagger G_N V_N^{-1} - L_N\|_{\ell^2} \leq C \left(\frac{\gamma\rho}{R}\right)^N$$

for some $C > 0$, where $\gamma = (1+\sqrt{2})^2$.

Proof. By Lemma 2.8 we have

$$P_N G P_N = P_N H V^{-1} L V P_N = P_N H P_N V^{-1} L V P_N + P_N H (1-P_N) V^{-1} L V P_N \,,$$

so using Equations (15) and (16) we have

$$G_N = H_N V_N^{-1} L_N V_N + P_N H (1-P_N) V^{-1} L P_N V_N \,,$$

since $P_N V^{-1} = V_N^{-1} P_N$ and $V P_N = P_N V_N$. Thus using $H_N^\dagger H_N = P_N$ and $V_N V_N^{-1} = P_N$ we obtain

$$\begin{aligned} V_N H_N^\dagger G_N V_N^{-1} &= L_N + V_N H_N^\dagger P_N H (1-P_N) V^{-1} L P_N \\ &= L_N + V_N H_N^\dagger P_N H (1-P_N) V^{-1} (1-P_N) L P_N \end{aligned}$$

and

$$\begin{aligned} &\|V_N H_N^\dagger G_N V_N^{-1} - L_N\|_{\ell^2} \\ &\leq \|V_N\|_{\ell^2} \|H_N^\dagger\|_{\ell^2} \|P_N H (1-P_N)\|_{\ell^2} \|(1-P_N) V^{-1} (1-P_N)\|_{\ell^2} \|(1-P_N) L P_N\|_{\ell^2} \,. \end{aligned}$$

On the other hand, by definition of the diagonal operator we have

$$\|V_N\|_{\ell^2} = \rho^{N-1}, \quad \|(1-P_N) V^{-1} (1-P_N)\|_{\ell^2} = \rho^{-N} \,,$$

by Lemma 2.9 we have

$$\|P_N H (1-P_N)\|_{\ell^2} \leq 2\pi, \quad \|H_N^\dagger\|_{\ell^2} \leq C_1 \gamma^N \,,$$

and Proposition 2.4 yields

$$\|(1-P_N) L P_N\|_{\ell^2} \leq \|(1-P_N) L\|_{\ell^2} \|P_N\|_{\ell^2} = \|(1-\mathcal{P}_N)\mathcal{L}\|_{H^2(D_\rho)\to H^2(D_\rho)} \leq C_2 \left(\frac{\rho}{R}\right)^N$$

with suitable $C_1, C_2 > 0$, and the assertion follows. □

Proposition 2.11. *Writing $\gamma = (1+\sqrt{2})^2$, we have for all $N \in \mathbb{N}_0$*

$$\|V_N H_N^\dagger G_N V_N - L\|_{\ell^2} \leq C \left(\left(\frac{\gamma\rho}{R}\right)^N + \left(\frac{r}{\rho}\right)^N \right)$$

for some $C > 0$. Choosing $\rho = \sqrt{rR/\gamma}$ optimises the convergence speed to $(\gamma r/R)^{N/2}$.

Proof. The assertion follows immediately from Proposition 2.4 and Lemma 2.10, as

$$\begin{aligned}
&\|V_N H_N^\dagger G_N V_N^{-1} - L\|_{\ell^2} \\
&\leq \|V_N H_N^\dagger G_N V_N^{-1} - L_N\|_{\ell^2} + \|L - L_N\|_{\ell^2} \\
&= \|V_N H_N^\dagger G_N V_N^{-1} - L_N\|_{\ell^2} + \|\mathcal{L} - \mathcal{P}_N \mathcal{L} \mathcal{P}_N\|_{H^2(D_\rho)\to H^2(D_\rho)} \\
&\leq C_1 \left(\frac{\gamma\rho}{R}\right)^N + C_2 \left(\left(\frac{\rho}{R}\right)^N + \left(\frac{r}{\rho}\right)^N \right).
\end{aligned}$$

□

Proposition 2.11 is the first main convergence result. It shows that for an infinite number of nodes the EDMD matrix $H_N^\dagger G_N$ of order N, is similar to a matrix that converges uniformly exponentially fast to the matrix representation of the transfer operator, provided the map T is sufficiently expansive, that is, if $r/R < 1/\gamma$.

We will now implement the concept with a finite number of nodes in our setting. For that purpose lift the $N \times N$ square matrix $\hat{H}_N^{(M)}$ (see Equation (7)) to a rank N operator $H_N^{(M)} \colon \ell^2(\mathbb{N}_0) \to \ell^2(\mathbb{N}_0)$ such that the $N \times N$ section of $H_N^{(M)}$ coincides with $\hat{H}_N^{(M)}$ and $H_N^{(M)} = P_N H_N^{(M)} = H_N^{(M)} P_N$. Furthermore we can see that $\hat{H}_N^{(M)}$ is positive definite for $M \geq N$ since

$$\sum_{k,\ell=0}^{N-1} \bar{\alpha}_k (\hat{H}_N^{(M)})_{k\ell} \alpha_\ell = \frac{1}{M} \sum_{m=0}^{M-1} |p(x_m)|^2 , \tag{17}$$

where the polynomial $p(x) = \sum_{k=0}^{N} \alpha_k x^k$ has at most $N-1$ distinct roots as long as $(\alpha_0, \ldots, \alpha_{N-1})$ is nonzero, so that at least one of the terms in the sum of Equation (17) is positive. In the same vein, provided that $M \geq N$, we can lift the inverse matrix $(\hat{H}_N^{(M)})^{-1}$ to a rank N operator $(H_N^{(M)})^\dagger \colon \ell^2(\mathbb{N}_0) \to \ell^2(\mathbb{N}_0)$ which obeys $H_N^{(M)} (H_N^{(M)})^\dagger = (H_N^{(M)})^\dagger H_N^{(M)} = P_N$.

Lemma 2.12. *There exist constants $C_1, C_2 > 0$ such that for $N \in \mathbb{N}_0$ and $M \geq C_1 N^2 \gamma^N$ we have*

$$\|(H_N^{(M)})^\dagger\|_{\ell^2} \leq C_2 \gamma^N,$$

where $\gamma = (1+\sqrt{2})^2$.

Proof. We have $\|H_N^\dagger\| \leq \tilde{C}_1 \gamma^N$ by Lemma 2.9, and

$$\|H_N - H_N^{(M)}\|_{\ell^2} = \|\hat{H}_N - \hat{H}_N^{(M)}\|_2 \leq \frac{3}{2} \frac{N^2}{M}$$

by Proposition 2.7. Define $C_1 = \max(3\tilde{C}_1, 1)$ and suppose that $M \geq C_1 N^2 \gamma^N > N$. Then

$$\|H_N^\dagger (H_N - H_N^{(M)})\|_{\ell^2} \leq \|H_N^\dagger\|_{\ell^2} \|H_N - H_N^{(M)}\|_{\ell^2} \leq \frac{1}{2}$$

and $I - H_N^\dagger (H_N - H_N^{(M)})$ is invertible. From

$$H_N^{(M)} = H_N - (H_N - H_N^{(M)}) = H_N - P_N (H_N - H_N^{(M)}) = H_N (I - H_N^\dagger (H_N - H_N^{(M)}))$$

we conclude (since $M > N$)

$$(H_N^{(M)})^\dagger = (I - H_N^\dagger (H_N - H_N^{(M)}))^{-1} H_N^\dagger$$

and

$$\|(H_N^{(M)})^\dagger\|_{\ell^2} \leq \|(I - H_N^\dagger (H_N - H_N^{(M)}))^{-1}\|_{\ell^2} \|H_N^\dagger\| \leq \frac{1}{1-1/2} \tilde{C}_1 \gamma^N .$$

□

The following proposition essentially estimates the difference between EDMD with a finite and with an infinite number of nodes.

Proposition 2.13. *Let $r/R < 1/\gamma$ with $\gamma = (1+\sqrt{2})^2$, and $\rho \in (r, R)$. There exist $C_1, C_2 > 0$ such that for $N \in \mathbb{N}_0$ and $M \geq C_1 N^2 \gamma^N$ we have*

$$\|V_N (H_N^{(M)})^\dagger G_N^{(M)} V_N^{-1} - V_N H_N^\dagger G_N V_N^{-1}\|_{\ell^2} \leq C_2 (\rho\gamma)^N \frac{N^2}{M} .$$

Proof. Let $C_1 > 0$ be as in Lemma 2.12 and $M \geq C_1 N^2 \gamma^N$. We have

$$\begin{aligned}
&\|V_N (H_N^{(M)})^\dagger G_N^{(M)} V_N^{-1} - V_N H_N^\dagger G_N V_N^{-1}\|_{\ell^2} \\
&= \|V_N (H_N^{(M)})^\dagger (G_N^{(M)} - G_N) V_N^{-1} \\
&\qquad + V_N (H_N^{(M)})^\dagger (H_N^{(M)} - H_N) V_N^{-1} V_N H_N^\dagger G_N V_N^{-1}\|_{\ell^2} \\
&\leq \|V_N\|_{\ell^2} \|(H_N^{(M)})^\dagger\|_{\ell^2} \|G_N^{(M)} - G_N\|_{\ell^2} \|V_N^{-1}\|_{\ell^2} \\
&\qquad + \|V_N\|_{\ell^2} \|(H_N^{(M)})^\dagger\|_{\ell^2} \|H_N^{(M)} - H_N\|_{\ell^2} \|V_N^{-1}\|_{\ell^2} \|V_N H_N^\dagger G_N V_N^{-1}\|_{\ell^2} .
\end{aligned}$$

Using Proposition 2.7, we obtain

$$\begin{aligned}
\|H_N^{(M)} - H_N\|_{\ell^2} &= \|\hat{H}_N^{(M)} - \hat{H}_N\|_2 \leq \frac{3}{2}\frac{N^2}{M} \\
\|G_N^{(M)} - G_N\|_{\ell^2} &= \|\hat{G}_N^{(M)} - \hat{G}_N\|_2 \leq \tilde{C}_1 \frac{N^2}{M} ,
\end{aligned}$$

and by Lemma 2.12,

$$\|(H_N^{(M)})^\dagger\|_{\ell^2} \leq \tilde{C}_2 \gamma^N .$$

Finally, Proposition 2.11 yields

$$\begin{aligned}
\|V_N H_N^\dagger G_N V_N\|_{\ell^2} &= \|V_N H_N^\dagger G_N V_N - L\|_{\ell^2} + \|L\|_{\ell^2} \\
&\leq \tilde{C}_3 \left(\frac{\gamma r}{R}\right)^{N/2} + \|\mathcal{L}\|_{H^2(D_\rho)\to H_2(D_\rho)} \leq \tilde{C}_4 ,
\end{aligned}$$

and

$$\|V_N\|_{\ell^2} \|V_N^{-1}\|_{\ell 2} = \rho^{N-1} ,$$

and the assertion follows. □

2.4 Main results

Our first main result can be stated as follows:

Theorem 2.14. *Let T be an analytic full branch map on the interval $[-1, 1]$ with $r/R < 1/\gamma$, where $\gamma = (1+\sqrt{2})^2$. Then there exist $C_1, C_2 > 0$ such that for all $N \in \mathbb{N}_0$ and all $M \geq C_1 N^2 \gamma^N$ we have*

$$\|V_N (H_N^{(M)})^\dagger G_N^{(M)} V_N^{-1} - L\|_{\ell^2} \leq C_2 \left(\frac{(\gamma r R)^{N/2} N^2}{M} + \left(\frac{\gamma r}{R}\right)^{N/2}\right) .$$

Proof. We note that

$$\begin{aligned}
\|V_N (H_N^{(M)})^\dagger G_N^{(M)} V_N^{-1} - L\|_{\ell^2} \leq &\|V_N (H_N^{(M)})^\dagger G_N^{(M)} V_N^{-1} - V_N H_N^\dagger G_N V_N^{-1}\|_{\ell^2} \\
&+ \|V_N H_N^\dagger G_N V_N^{-1} - L\| .
\end{aligned}$$

Using Proposition 2.11 and Proposition 2.13 with $\rho = \sqrt{rR/\gamma}$ yields the assertion. □

In particular, by imposing a lower-bound condition on the number of nodes in terms of the number of observables, we obtain exponential convergence. More precisely:

Corollary 2.15. *Under the hypothesis of Theorem 2.14, there exists a $C > 0$ such that for all $N \in \mathbb{N}_0$ and for all $M \geq N^2 R^N$ we have*

$$\|V_N (H_N^{(M)})^\dagger G_N^{(M)} V_N^{-1} - L\|_{\ell^2} \leq C \left(\frac{\gamma r}{R}\right)^{N/2} .$$

Proof. We note that by assumption $R^N > r^N \gamma^N$, and so $M \geq N^2 R^N \geq C_1 N^2 \gamma^N$ for all $N \geq \log C_1 / \log r$. Letting

$$C_3 = \max \left\{ \frac{\|V_N (H_N^{(M)})^\dagger G_N^{(M)} V_N^{-1} - L\|_{\ell^2}}{\frac{(\gamma r R)^{N/2} N^2}{M} + \left(\frac{\gamma r}{R}\right)^{N/2}} : N < \frac{\log C_1}{\log r}, M < C_1 N^2 \gamma^N \right\},$$

the assertion follows with $C = \max\{2C_2, C_3\}$. □

It is a classical textbook result (see, for example, [14, p.1091]) that uniform convergence of compact operators, such as in Corollary 2.15, implies uniform convergence of eigenvalues. However, without additional conditions, the speed of convergence does not carry over. In our case, exponential convergence of eigenvalues is a consequence of [1, Theorem 2.18 and ensuing remarks] (see also [34]), which provide the proof of the following simple useful statement.

Lemma 2.16. *Let V be a Banach space and $\mathcal{F}_N \colon V \to V$ a convergent sequence of finite-rank operators with*

$$\|\mathcal{F}_N - \mathcal{F}\|_{V \to V} \leq C e^{-\alpha N}$$

for some $C > 0$ and $\alpha > 0$. Then there exist enumerations (taking algebraic multiplicities into account) of the eigenvalues of $\mathcal{F}_N$ and of the limit $\mathcal{F}$, denoted $\lambda_k(\mathcal{F}_N)$ and $\lambda_k(\mathcal{F})$ respectively, such that for each fixed k there exist constants $C' > 0$ and $\alpha' > 0$ with

$$|\lambda_k(\mathcal{F}_N) - \lambda_k(\mathcal{F})| \leq C' e^{-\alpha' N} \qquad \text{for } N \in \mathbb{N} .$$

With Corollary 2.15 and Lemma 2.16 the eigenvalues of the EDMD matrices $(H_M^{(N)})^\dagger G_N^{(M)}$ converge at an exponential rate towards the eigenvalues of the (compact) transfer operator $\mathcal{L}$ on $H^2(D_\rho)$. If we denote by $(H^2(D_\rho))'$ the Banach dual of $H^2(D_\rho)$, that is, the Banach space of bounded linear functionals $\ell \colon H^2(D_\rho) \to \mathbb{C}$, then the Banach adjoint $\mathcal{K} \colon (H^2(D_\rho))' \to (H^2(D_\rho))'$ of the transfer operator $\mathcal{L} \colon H^2(D_\rho) \to H^2(D_\rho)$ is defined by $(\mathcal{K}\ell)(f) = \ell(\mathcal{L}f)$. The Banach adjoint $\mathcal{K}$ and the transfer operator $\mathcal{L}$ have the same spectra (including multiplicities of eigenvalues), which together with the previous considerations yields the following result:

Corollary 2.17. *Assume the hypothesis of Theorem 2.14, $M \geq N^2 R^N$, and $\rho = \sqrt{rR/\gamma}$. Denote by $\mathcal{K} \colon (H^2(D_\rho))' \to (H^2(D_\rho))'$ the (compact) Koopman operator extended to the Banach dual of $H^2(D_\rho)$. Then there exist enumerations of the eigenvalues of $\mathcal{K}$ and $(H_N^{(M)})^\dagger G_N^{(M)}$, $\lambda_k(\mathcal{K})$ and $\lambda_k((H_N^{(M)})^\dagger G_N^{(M)})$ respectively, such that for each fixed k there exist constants $C > 0$ and $0 < b < 1$ such that*

$$|\lambda_k((H_N^{(M)})^\dagger G_N^{(M)}) - \lambda_k(\mathcal{K})| \leq C b^N \qquad \text{for } N \in \mathbb{N} .$$

Remark 2.18. *The Banach adjoint $\mathcal{K}$ is indeed an extension of the 'usual' Koopman operator to the dual space $(H^2(D_\rho))'$. To see this, take a function $g \in L^2([-1,1])$ and introduce the functional $\ell_g \colon H^2(D_\rho) \to \mathbb{C}$ by*

$$\ell_g(f) = \int_{-1}^{1} f(x) g(x)\, dx . \tag{18}$$

Using the continuity of point evaluations on $H^2(D_\rho)$, that is, $|f(x)| \leq \rho/\sqrt{\rho^2 - x^2}\|f\|_{H^2(D_\rho)}$ (see the proof of Lemma 2.1), one can easily show that ℓ_g is bounded. Hence functionals of the type defined in Equation (18) constitute a subspace of $(H^2(D_\rho))'$. With Equation (5) we obtain

$$(\mathcal{K}\ell_g)(f) = \ell_g(\mathcal{L}f) = \int_{-1}^{1} (\mathcal{L}f)(x)g(x)dx = \int_{-1}^{1} f(x)g(T(x))\,dx = \ell_{g\circ T}(f)\,.$$

Hence on the subspace of linear functionals of the form (18), $\mathcal{K}$ acts in the 'usual' way $g \mapsto g \circ T$.

3 Numerical illustration

We compare our previous analytic estimates by applying EDMD to chaotic one-dimensional expanding maps. For the set of observables we take monomials $\psi_k(x) = x^k$, $k = 0, 1, \ldots, N-1$.

3.1 Piecewise linear maps

As our first model we take a piecewise linear full branch map, the skewed doubling map in (3) defined on $[-1, 1]$, repeated for convenience below

$$T_D(x) = \begin{cases} -1 + 2(x+1)/(1+a) & \text{if} \quad -1 \leq x \leq a \\ 1 + 2(x-1)/(1-a) & \text{if} \quad a < x \leq 1 \end{cases},$$

where the parameter $a \in (-1, 1)$ determines the skew of the map. The eigenvalues of the transfer operator (4), defined on a space of analytic functions, say $H^2(D_R)$ with $R > 1$, are simply given by $\lambda_n = ((1+a)/2)^{n+1} + ((1-a)/2)^{n+1}$ with $n = 0, 1, \ldots$, see for example [29] for an elementary account. The finite-rank approximation given by the matrix L_N (see Equation (15)) becomes exact in this piecewise linear case and the first N eigenvalues are given by the diagonal matrix elements $(L_N)_{kk}$. Furthermore, if we perform EDMD with an infinite number of nodes along the lines of Equation (8), EDMD is equivalent to the eigenvalue problem of the matrix L_N (see the proof of Lemma 2.10, since $(1 - P_N)LP_N = 0$ in the case of piecewise linear maps). This is also reflected in the numerical evaluation of EDMD where results for the leading N eigenvalues coincide with the exact values to the numerical accuracy used, when the matrices H_N and G_N are computed via Equation (8). We note that computing the eigenvalues via EDMD requires solving the generalised eigenvalue problem (1), which in itself poses a considerable numerical challenge. For our numerics we have relied on standard commercial numerical solvers contained, for example, in the Maple or Mathematica software packages. Alternatively, one can use standard numerical eigenvalue solvers by employing the ε-pseudoinverse of the matrix H_N, which, in practice, yields the same numerical result for the eigenvalues.

The picture differs if we focus on EDMD for a finite number M of equidistant nodes $x_m = -1 + (2m-1)/M$, $m = 1, \ldots, M$, see the right panel of Figure 1. Then the collocation error in computing the integrals, see Proposition 2.7, causes corresponding errors for the eigenvalues themselves. If we denote by $\Delta_n = |\lambda_n - \lambda_n^{(EDMD)}|$ the absolute difference between the exact eigenvalue and the corresponding result produced by EDMD, then this error shows indeed the characteristic $1/M$ dependence if the leading eigenvalues are considered, see Figure 2. In addition, the error shows large fluctuations which may be the result of the constant expansivity in the piecewise linear chaotic map. The number of observables had no impact on the error when infinite number of nodes were considered. For a finite number of nodes a fairly small number of observables is sufficient to obtain small errors and the error may in fact increase (as seen in the right panel of Figure 1) if the number of observables is increased, see Theorem 2.14. We will elaborate further on this observation in the next section.

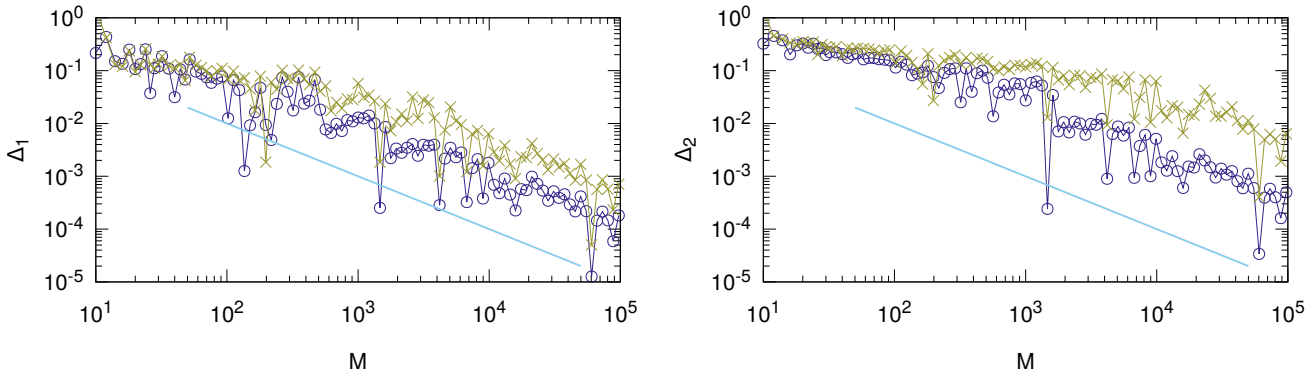

Figure 2: Difference Δ_n between the exact eigenvalue of the transfer operator and the approximate eigenvalue obtained by EDMD using monomials as observables, for the skewed doubling map, Equation (3), with $a = 1/\sqrt{2}$, in dependence on the number of equidistant nodes M on a double logarithmic scale. Left: Error of the subleading eigenvalue, Δ_1, for $N = 5$ (blue, circles) and $N = 6$ (amber, cross). The line (cyan) shows an algebraic decay of order $1/M$. Left: Error of the second nontrivial eigenvalue, Δ_2, for $N = 5$ (blue, circles) and $N = 6$ (amber, cross). The line (cyan) shows an algebraic decay of order $1/M$.

3.2 Blaschke maps

To obtain further insight let us consider a piecewise analytic map with nonlinear branches. For a reliable error estimate we focus on the class of so-called interval Blaschke maps, for which the exact spectrum of the transfer operator on $H^2(D_R)$ can been computed explicitly [32, 4]. We consider here the simplest case which is given by a symmetric nonlinear deformation of the doubling map

$$T_B(x) = \begin{cases} 2x + 1 + \frac{2}{\pi}\arctan\left(\frac{\mu\sin(\pi x)}{1-\mu\cos(\pi x)}\right) & \text{if} \quad -1 \le x \le 0 \\ 2x - 1 + \frac{2}{\pi}\arctan\left(\frac{\mu\sin(\pi x)}{1-\mu\cos(\pi x)}\right) & \text{if} \quad 0 < x \le 1 \end{cases} \tag{19}$$

where $\mu \in (-1, 1)$ denotes the parameter of the map. The exact spectrum consists of the trivial eigenvalue $\lambda_0 = 1$ and two nontrivial families of eigenvalues given by $\{\mu^n : n = 1, 2, \ldots\}$ and $\{(\mu/2 + 1/2)^n : n = 1, 2, \ldots\}$. Eigenvalues in the first family have geometric multiplicity two, while eigenvalues in the second family are simple. The two inverse branches can be computed in a straightforward way and they are given by $\varphi_\ell(x) = x/2 + (-1)^\ell \arccos(\mu\cos(\pi x/2))/\pi$ with $\ell = 1, 2$. If $|\mu| \le 0.3$ then there exists some $R > 1$ so that these branches are analytic on a disk D_R in the complex plane and thus satisfy the conditions required for the transfer operator to be compact, see Section 2.1 for details.

We first focus on EDMD with an infinite number of nodes, see Equation (8). In this case (see Proposition 2.4) the finite-rank approximation by the matrix L_N, Equation (15), produces errors which are exponentially small in N. EDMD yields eigenvalue estimates which are exponentially close to the exact values (Proposition 2.11), and the errors Δ_n decay exponentially with the number of observables used, as can also be observed numerically in Figure 3. We note that the two different sets of eigenvalues which are contained in the spectrum of the transfer operator have different exponential decay rates, which might be caused by the multiplicities of the two sets of eigenvalues.

As a next step we investigate the impact of a finite number of nodes, when EDMD is applied to the Blaschke map (19) with a fixed number N of observables. We again choose M centralised

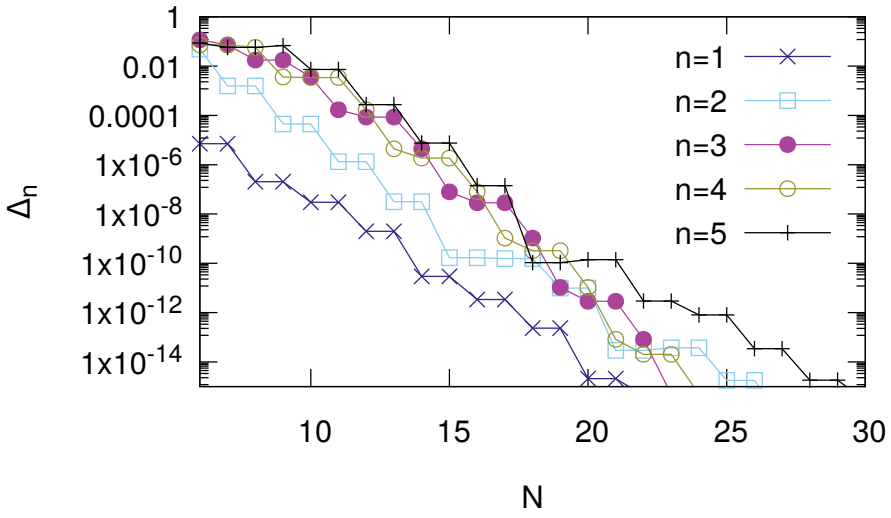

Figure 3: Difference Δ_n between the exact eigenvalue of the transfer operator and the approximate eigenvalue obtained by EDMD using monomials as observables and an infinite number of nodes, for a Blaschke interval map, Equation (19), with $\mu = 0.3$ in dependence on the number of observables N on a semi-logarithmic scale. Results for the first five nontrivial eigenvalues are shown. EDMD has been evaluated for an infinite number of nodes, see Equation (8).

nodes $x_m = -1 + (2m-1)/M$, $m = 1, \ldots, M$ to evaluate the sums in Equations (7). While the finite number of observables causes only an exponentially small error, the total error is now dominated by the collocation errors of the matrix elements (Proposition 2.7), which are algebraic in the number of nodes used. Since we are in a symmetric case with nodes being interval midpoints, we expect the collocation errors to be of order $\mathcal{O}(M^{-2})$, as indeed observed in Figure 4 for the leading eigenvalues. However, the asymptotic behaviour sets in later for larger number of nodes when higher eigenvalues are considered (see also Theorem 2.14).

We will now demonstrate the joint dependence of the eigenvalue approximation error on both the number of observables N and the number of nodes M, again using the Blaschke map (19), with observables given by monomials, and a set of centralised equidistant nodes. As already observed in Figure 2, we can expect to require a sufficiently large number of nodes to deal with expressions containing monomials of high power, so that convergence of eigenvalues depends on a nontrivial relation between N and M (see Corollary 2.15). The results displayed in Figure 5 confirm this expected lack of uniformity as convergence seems to occur in a triangular shaped region when N and M jointly tend to infinity. Numerically, a constraint of the type $M > cN^2$ seems to be required to ensure convergence of the EDMD eigenvalues towards the correct spectrum of the transfer operator. Hence, the numerical result seems to indicate that the terms γ^N occurring in Theorem 2.14 overestimate the actual error, and a tighter bound avoiding such terms might be obtained. In fact, the statement of Theorem 2.14 is uniform across the spectrum, and the estimate might be improved by using ε-pseudoinverses for operators and focusing solely on the leading part of the spectrum.

In line with Theorem 2.14, the numerical investigation of simple toy models shows exponential convergence in the number of observables, so that a small number of observables turns out to be sufficient. However the required number of nodes M is large and grows rapidly with increasing N, and we observe a nontrivial relation between N and M which needs to be satisfied to obtain convergence of the EDMD algorithm.

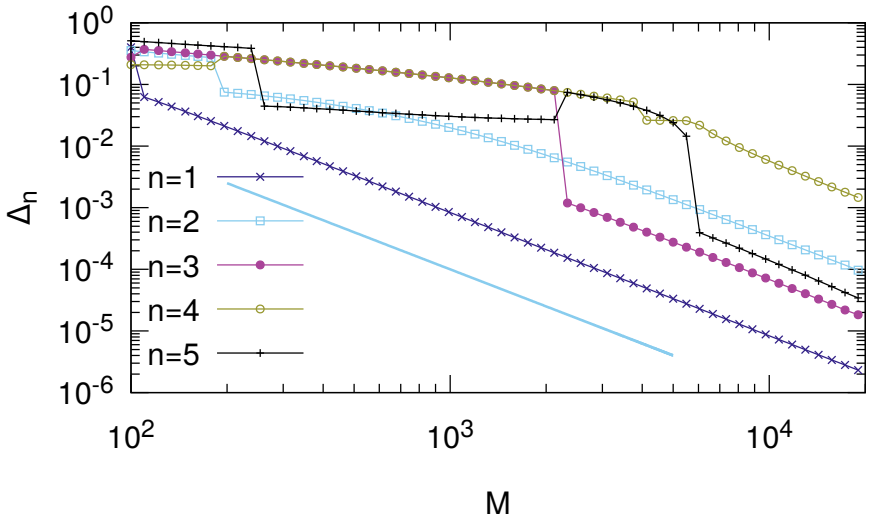

Figure 4: Difference between the exact eigenvalue of the transfer operator and the approximate eigenvalue obtained by EDMD using monomials as observables, for a Blaschke interval map, Equation (19), with $\mu = 0.3$ in dependence on the number of equidistant nodes M on a double logarithmic scale. Results for the first five nontrivial eigenvalues are shown. The jumps are likely to be artefacts caused by sorting eigenvalues by size of modulus. EDMD has been evaluated for $N = 15$ observables. The line (cyan) shows an algebraic decay of order $1/M^2$.

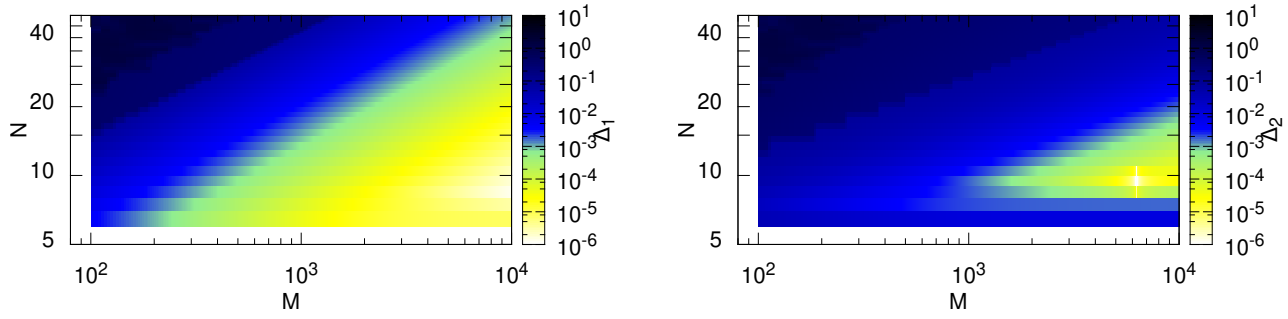

Figure 5: Density plot of the difference between the exact eigenvalue of the transfer operator and the approximate eigenvalue obtained by EDMD using monomials as observables, for a Blaschke interval map with $\mu = 0.3$ in dependence on the number of equidistant nodes M and the number of observables N. The data are shown on a double logarithmic scale. Left: error of the subleading eigenvalue, Δ_1, right: error of the second nontrivial eigenvalue, Δ_2.

3.3 Impact of observables

We have derived our results for simple toy models where a rigorous account for our findings can be provided. In general, however, the quality of the EDMD results is hard to judge, as the outcome depends on the observables in a subtle way. Even when EDMD converges, the meaning of the eigenvalues obtained may not be obvious. We have demonstrated this phenomenon for the skewed doubling map when EDMD is applied with observation through Fourier modes, see the left panel of Figure 1.

Let us look at this case in some more detail. To simplify the analysis consider the case of infinite number of nodes where the relevant matrix elements are simply given by

$$\begin{aligned} H_{k\ell} &= \frac{1}{2}\int_{-1}^{1} \exp(i\pi kx)\exp(-i\pi\ell x)\,dx = \delta_{k\ell} \\ G_{k\ell} &= \frac{1}{2}\int_{-1}^{1} \exp(i\pi kT(x))\exp(-i\pi\ell x)\,dx \\ &= \frac{1+a}{2}e^{i\pi\ell(1-a)/2}\frac{\sin(\pi(k-\ell(1+a)/2))}{\pi(k-\ell(1+a)/2)} \\ &\quad + \frac{1-a}{2}e^{-i\pi\ell(1+a)/2}\frac{\sin(\pi(k-\ell(1-a)/2))}{\pi(k-\ell(1-a)/2)} \end{aligned} \tag{20}$$

when using Equation (3). While not crucial, we have used here the complex conjugate for the second entry as that conveniently results for H in the identity matrix. Figure 6 shows the result for the subleading eigenvalue λ_1 computed from the generalised eigenvalue problem, that is, the subleading eigenvalue of G_N, in dependence on the skew of the map. We obtain convergence when N increases, but convergence slows down considerably for maps with a small skew a. Furthermore, the modulus of the subleading eigenvalue seems to converge towards $(1+|a|)/2$, an expression which coincides with the essential spectral radius of the transfer operator considered on the space of functions of bounded variation, see for example [17]. With hindsight we can also provide an analytic argument supporting this observation. We may consider the skewed doubling map as a piecewise analytic map on the complex unit circle, following the construction which is used for analytic maps, see for example [32]. Fourier modes become complex monomials, and smooth functions on the interval become discontinuous on the unit circle, as smooth functions on an interval normally do not have a continuous periodic extension beyond the endpoints of the interval. If one applies the transfer operator to such discontinuous functions the number of discontinuities may increase, but the variation of the function remains bounded because of the expansivity of the map. While one would naively expect that Fourier modes imply studying transfer operators on a space of square integrable functions we encounter here an unexpected subtle mechanism where Fourier modes result in studying the transfer operator on the space of functions with bounded variation. This observations parallels the approximation of smooth maps by piecewise linear Markov maps, see for example [35, 19], which effectively also amounts to studying transfer operators on the space of functions with bounded variation [33]. In summary, it is not surprising that the corresponding essential spectral radius shows up in the numerical outcome of EDMD. It is also worth mentioning that the essential spectral radius does not have to be an eigenvalue of the transfer operator, so that some care has to be taken when interpreting eigenmodes obtained by EDMD.

Let us also comment on the slow convergence which is visible in Figure 6 close to $a = 0$. The transfer operator of the doubling map, $a = 0$, effectively halves wavenumbers of the Fourier modes, that is, the matrix G_N in Equation (20) has a single eigenvalue 1 and a degenerate eigenvalue zero. The matrix G_N for $a = 0$ consists of Jordan blocks, and the largest block has size b where $2^b \sim N$. For $a \neq 0$, standard perturbation theory tells us that nonvanishing eigenvalues of the size $|\lambda_1| \sim |a|^{1/b} \sim |a|^{\ln 2/\ln N}$ are generated. This simple reasoning is consistent with the numerical data, see the inset in Figure 6. It also supports the assertion that EDMD in this example yields the essential spectral radius for the transfer operator being defined on the space of functions with bounded variation.

In summary, while EDMD is a powerful data analysis tool, results have to be taken with a grain of salt, as the interpretation of the eigenvalues and eigenmodes is far from obvious. In

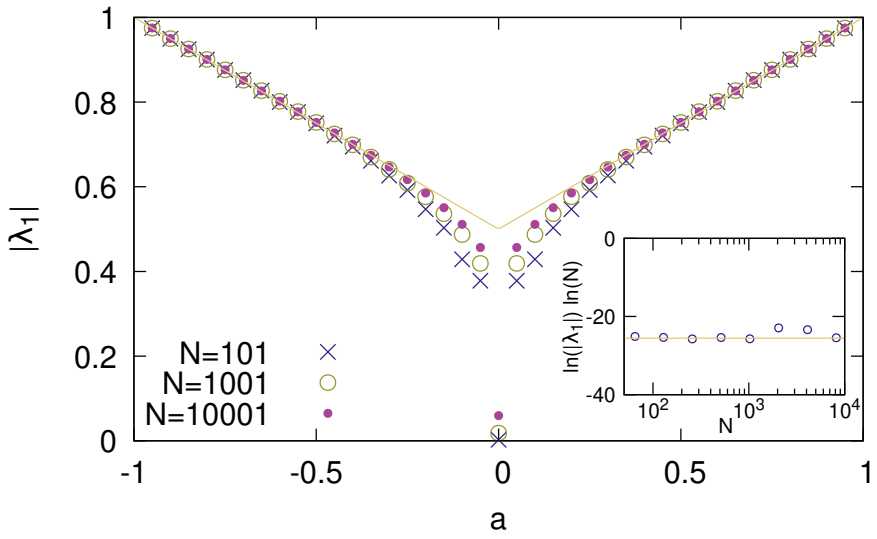

Figure 6: Absolute value of the subleading eigenvalue λ_1 obtained for the skewed doubling map, Equation (3), in dependence on the skew a, using EDMD with N Fourier modes in the limit of infinitely many nodes, see Equations (20) (symbols). The line (amber) shows the essential spectral radius of the transfer operator defined on the space of functions with bounded variation. The inset shows the product $\ln|\lambda_1| \ln N$ as a function of N for a doubling map with tiny skew $a = 10^{-16}$ (symbols). The horizontal line (amber) indicates a crude analytic estimate (see the text for details).

particular, it is often not clear which transfer or Koopman operator is studied by EDMD, as the underlying operator (together with a function space) is determined in a subtle way by the observables used, if it exists at all. Nevertheless, at the phenomenological level EDMD is a very appealing tool, and further rigorous studies are required to complement its empirical success with provable convergence guarantees and rates.

Acknowledgement

O.F.B. gratefully acknowledges support by EPSRC though grant EP/R012008/1, J.S. acknowledges support by ERC-Advanced Grant 833802-Resonances, and W.J. acknowledges funding by German Research Foundation SFB 1270/2 - 299150580.

References

[1] M. Ahues, A. Largillier, and B. Limaye. *Spectral Computations for Bounded Operators*. Chapman and Hall/CRC, Boca Raton, 2001.

[2] V. Baladi. *Positive Transfer Operators and Decay of Correlations*. World Scientific, Singapore, 2000.

[3] O.F. Bandtlow and O. Jenkinson. Invariant measures for real analytic expanding maps. *J. London Math. Soc.*, 75:343, 2007.

[4] O.F. Bandtlow, W. Just and J. Slipantschuk. Spectral structure of transfer operators for expanding circle maps, *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 34:31 , 2017.

[5] O.F. Bandtlow, W. Just, and J. Slipantschuk. EDMD for expanding circle maps and their complex perturbations. *arXiv preprint*, arXiv:2308.01467, 2023.

[6] M. Budišić, R. Mohr, and I. Mezić. Applied Koopmanism. *Chaos*, 22:047510, 2012.

[7] M.-D. Choi. Tricks or treats with the Hilbert matrix. *Am. Math. Month.*, 90:301, 1983.

[8] M.J. Colbrook. The multiverse of dynamic mode decomposition algorithms. *arXiv preprint*, arXiv:2312.00137, 2023.

[9] M.J. Colbrook. The mpEDMD algorithm for data-driven computations of measure-preserving dynamical systems. *SIAM Journal on Numerical Analysis*, 61:1585, 2023.

[10] M.J. Colbrook and A. Townsend. Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. *Comm. Pure Appl. Math.*, 77:221, 2024.

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

[12] M. Dellnitz, G. Froyland, and S. Sertl. On the isolated spectrum of Perron-Frobenius operator. *Nonlin.*, 13:1171, 2000.

[13] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM J. Numer. Anal.*, 36:491, 1999.

[14] N.J. Dunford and J.T. Schwartz. *Linear Operators*. Wiley, Hoboken, New Jersey, 1988.

[15] G. Froyland, C. González-Tokman, and A. Quas. Detecting isolated spectrum of transfer and Koopman operators with Fourier analytic tools. *J. Comput. Dyn.*, 1:249, 2014.

[16] D. Giannakis. Data-driven spectral decomposition and forecasting of ergodic dynamical systems. *Appl. Comput. Harmon. Anal.*, 47:338, 2019.

[17] F. Hofbauer and G. Keller. Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.*, 180:119, 1982.

[18] I. Ishikawa, Y. Hashimoto, M. Ikeda, and Y. Kawahara. Koopman operators with intrinsic observables in rigged reproducing kernel Hilbert spaces. *arXiv preprint*, arXiv:2403.02524, 2024.

[19] W. Just and H. Fujisaka. Gibbs measures and power spectra for type I intermittent maps. *Physica D*, 64:98, 1993.

[20] K. Karhunen. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae.*, 37:1, 1947.

[21] G. Keller and C. Liverani. Stability of the spectrum for transfer operators. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.*, 28:141, 1999.

[22] S. Klus, P. Gelß, S. Peitz, and C. Schütte. Tensor-based dynamic mode decomposition. *Nonlin.*, 31:3359, 2018.

[23] S. Klus, P. Koltai, and C. Schütte. On the numerical approximation of the Perron-Frobenius and Koopman operator. *J. Comp. Dyn.*, 3:51, 2016.

[24] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, and F. Noé. Data-driven model reduction and transfer operator approximation. *J. Nonl. Sci.*, 28:985, 2018.

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

[26] J.N. Kutz, S.L. Brunton, B.W. Brunton, and J.L. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. Society for Industrial and Applied Mathematics, 2016.

[27] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynam.*, 41:309, 2005.

[28] I. Mezić and A. Banaszuk. Comparison of systems with complex behavior. *Physica D*, 197:101, 2004.

[29] H. Mori, B. S. So, and T. Ose. Time-correlation functions of one-dimensional transformations. *Prog. Theor. Phys.*, 66(4):1266, 1981.

[30] W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D.S. Henningson. Spectral analysis of nonlinear flows. *J. Fluid Mech.*, 641:85, 2009.

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

[32] J. Slipantschuk, O.F. Bandtlow, and W. Just. Analytic expanding circle maps with explicit spectra. *Nonl.*, 26:3231, 2013.

[33] J. Slipantschuk, O.F. Bandtlow, and W. Just. On the relation between Lyapunov exponents and exponential decay of correlations. *J. Phys. A*, 46:075101, 2013.

[34] J. Slipantschuk, O.F. Bandtlow, and W. Just. Dynamic mode decomposition for analytic maps. *Comm. Nonl. Sci. Num. Sim.*, 84:105179, 2020.

[35] B. C. So, N. Yoshitake, H. Okamoto, and H. Mori. Correlations and spectra of an intermittent chaos near its onset point. *J. Stat. Phys.*, 36(3/4):367, 1984.

[36] Z. Suchanecki, I. Antoniou, S. Tasaki, and O.F. Bandtlow. Rigged Hilbert spaces for chaotic dynamical systems. *J. Math. Phys.*, 37:5837, 1996.

[37] H.S. Wilf. *Finite Sections of Some Classical Inequalities*. Springer, New York, 1970.

[38] M.O. Williams, I.G. Kevrekidis, and C.W. Rowley. A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. *J. Nonl. Sci.*, 25:1307, 2015.

[39] C. Wormell. Orthogonal polynomial approximation and extended dynamic mode decomposition in chaos. *arXiv preprint*, arXiv:2305.08074, 2023.

[40] C. Zhang and E. Zuazua. A quantitative analysis of Koopman operator methods for system identification and predictions. *Comptes Rendus. Mécanique*, 351, 2023.

Declaration of interests: None