Orthogonal polynomial approximation and Extended Dynamic Mode Decomposition in chaos

Caroline Wormell

June 20, 2023

Abstract

Extended Dynamic Mode Decomposition (EDMD) is a data-driven tool for forecasting and model reduction of dynamics, which has been extensively taken up in the physical sciences. While the method is conceptually simple, in deterministic chaos it is unclear what its properties are or even what it converges to. In particular, it is not clear how EDMD's least-squares approximation treats the classes of regular functions needed to make sense of chaotic dynamics.

We develop for the first time a general, rigorous theory of EDMD on the simplest examples of chaotic maps: analytic expanding maps of the circle. To do this, we prove a new, basic approximation result in the theory of orthogonal polynomials on the unit circle (OPUC) and apply methods from transfer operator theory. We show that in the infinite-data limit, the least-squares projection error is exponentially small for trigonometric polynomial observable dictionaries. As a result, we show that the forecasts and Koopman spectral data produced using EDMD in this setting converge to the physically meaningful limits, exponentially fast with respect to the size of the dictionary. This demonstrates that with only a relatively small polynomial dictionary, EDMD can be very effective, even when the sampling measure is not uniform. Furthermore, our OPUC result suggests that data-based least-squares projections may be a very effective approximation strategy.

1 Introduction

Nonlinear systems with complex dynamics are found across the physical and human sciences, counting among them climate systems, fluid flows and economic systems [16, 6]. Unlike linear systems, they rarely admit useful analytic solutions or have much obvious mathematical structure that can be used to make sense of them. Furthermore, the information we may have on these systems can often be very limited: for example, we may only have access to some empirical observations of its evolution. General methods of studying nonlinear systems in terms of well-understood mathematical objects, that can be performed using empirical data, are therefore an important tool in scientific endeavour.

One approach to do this is to use a surrogate linear system to study the nonlinear system. For example, one can study functions ("observables") on the phase space, and construct a linear *Koopman operator* which maps observables to their forward expectations under the flow [7]. Computationally, only a finite-dimensional vector space of observables is considered: the Koopman operator may not preserve this space of observables, but can be projected onto the finite-dimensional space, with good results if the observables are well-chosen. This projection can be done by least-squares on the empirical data, making it widely applicable. The computational representation of the Koopman operator can then be studied in terms of its spectrum and eignefunctions, which describe dynamical properties such mixing rates, almost-invariant sets in phase space [15, 11, 10].

Many different numerical methods, usually described as Dynamic Mode Decomposition (DMD) variants, employ this approach: for example, the classic DMD uses linear functions of delay coordinates as its observables. The aim of Extended Dynamic Mode Decomposition (EDMD), however, is to choose a large "dictionary" of observables that can be effectively used to approximate any function on the phase space [28]: commonly, polynomials rather than simply linear functions are used. EDMD and its variants have recently been used to great success in a broad variety of settings, including in airflow, molecular dynamics, industrial processes, and decision networks [22, 9, 10, 27].

Given its applicability, there has been a lot of interest in the mathematical theory behind EDMD, and in particular the interpretation of its spectral data. This theory has largely been explored by considering the Koopman operator $\mathcal{K}\varphi = \varphi \circ f$ of the dynamics f as acting on $L^2(\mu)$, where μ is the (dynamically invariant) sampling measure of the empirical data points: the unitary nature on the Koopman operator on this space is exploited to obtain various spectral convergence results regarding the $L^2(\mu)$ spectrum, which lies on the unit circle. However, the actual spectra of Koopman matrices obtained using EDMD bear much closer resemblance to the spectra of quasi-compact Markov operators [18], whose spectrum is mostly located inside the unit circle (see Figure 1 for an example).

The explanation for this can be found in the fact that the Koopman operator is the dual of the transfer operator [26], a typically quasi-compact Markov operator that tracks the movement of probability densities under the dynamics. The transfer operator has been heavily studied in dynamical systems, and has a strong and well-developed functional-analytic theory, including numerical approximation of transfer operators [19, 13, 29, 4]. Nevertheless, with some exceptions [15, 26], cross-pollination between the separates transfer operator and Koopman operator communities has been regrettably sparse.

Among the transfer operator discretisations, Extended DMD bears close resemblance to Fourier methods for transfer operators [29, 12], which make a standard orthogonal projection onto trigonometric polynomals with respect to Lebesgue measure. Proving the convergence of discretisations with respect to these projections can be done by using the orthogonality of derivatives of the basis functions, a property which has been used to prove convergence of EDMD when the sampling measure μ is Lebesgue [26]. It is very rare, however, for a sampling measure to hold this property: only a few, famous, families of measures have this property [21]. For more general sampling measures μ, new approximation theory must be developed to carry over transfer operator results to DMD variants: this is the achievement of this paper.

As systems become more structurally complex, rigorous study of chaotic systems very quickly becomes impractical [5]. For this reason, we choose as an initial example uniformly expanding maps of the torus $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$: that is, maps $f : \mathbb{T} \to \mathbb{T}$ such that for all $\theta \in \mathbb{T}$, $|f'(\theta)| \geq \gamma > 1$. These are common first examples for understanding phenomena in chaotic dynamics [1, 29]. We note that they are sometimes studied as self-maps of the complex unit circle in coordinates $z = e^{i\theta}$. Some additional smoothness is required to obtain good results: we will assume that f is analytic. Since the physical invariant measures of such maps have analytic densities, we will also assume that μ has analytic density.

The crux of the dynamical systems theory is the following: given observable functions φ in a Banach space of analytic functions $\mathcal{H}^2_t \subset C^\infty(\mathbb{T})$ (defined in Section 2) and $\psi \in L^1(\mathbb{T})$, we can write their lag correlations as a sum of exponentially decaying functions:

$$\int_{\mathbb{T}} \varphi\, \psi \circ f^n \,\mathrm{d}\mu \sim \sum_{j=0}^{\infty} \sum_{m=1}^{M_j} \alpha_j^{(m)}(\varphi)\beta_j^{(m)}(\psi)\, n^m \lambda_j^n \tag{1}$$

where $\beta_j^{(m)} : \mathcal{H}^2_t \to \mathbb{R}$ is bounded and $\alpha_j^{(m)}(\varphi) ;= \int a_j^{(m)} \varphi \,\mathrm{d}\mu$ for some function $a_j^{(m)} \in \mathcal{H}^2_t$. The multiplicities M_j are generically 1, in which case we drop the superscripts.

The complex λ_j, which have modulus no greater than one, are known as Ruelle-Pollicott resonances. These resonances and associated linear operators determine many important properties of the dynamical system: for example, for λ_j close to 1, the sets $\{x : a_j(x) > 0\}$ denote almost-invariant sets with respect to the dynamics, as equivalently do regions where β_j is positive on average [14, 15].

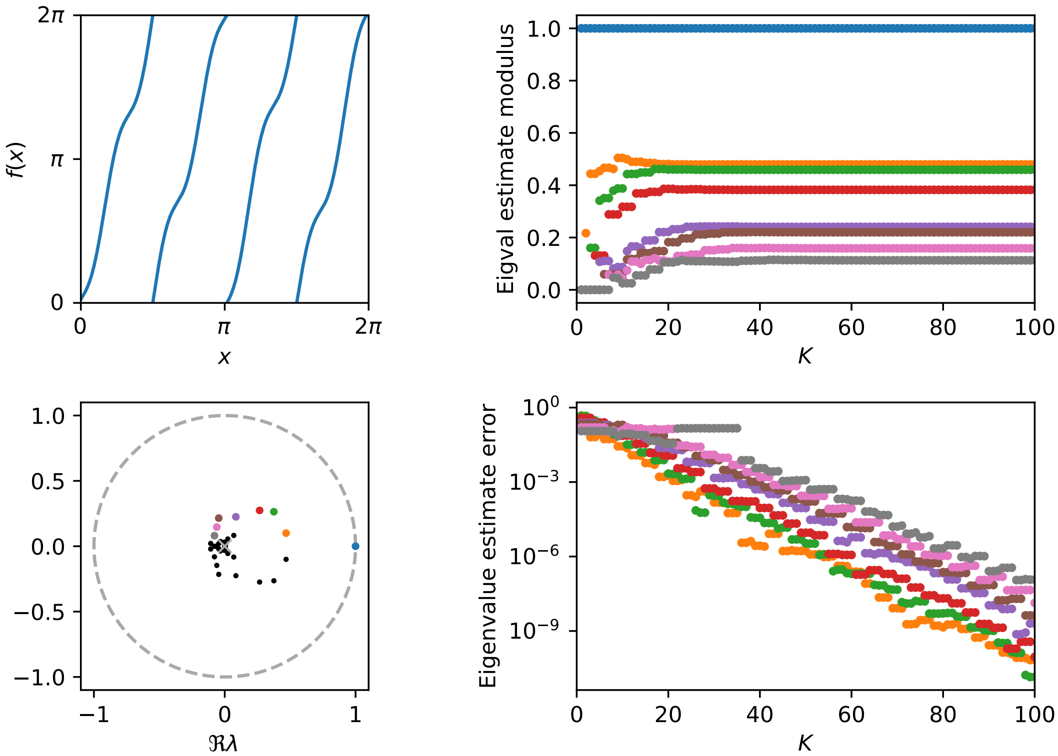

Figure 1: Top left: graph of the circle map $f(x) = 4x - 0.4 \sin 6x + 0.08 \cos 3x \mod 2\pi$. Bottom left: Ruelle-Pollicott resonances of f with certain resonances marked in colour. Top right: modulus of EDMD eigenvalues, with colours corresponding to resonances in the bottom left. Bottom right: exponential convergence with respect to dictionary size of the errors between EDMD eigenvalues and Ruelle-Pollicott responances, with corresponding colours. Ruelle-Pollicott resonances estimated using a Fourier transfer operator discretisation [29].

The aim of EDMD is to make a least-squares best approximation to the Koopman operator $\mathcal{K}$ in $L^2(\mu_N)$ where μ_N is the empirical measure of the data points $\{x_n\}_{n=1,\ldots,N}$ restricted to act on the span of one's function dictionary $\Psi = \{\psi_k\}_{|k|\leq K}$. This can be computed in the Ψ basis as the Koopman matrix

$$((\Psi^{(0)})^*\Psi^{(0)})^{-1}(\Psi^{(0)})^*\Psi^{(1)}$$

where $(\Psi^{(0)})_{nk} = \psi_k(x_n)$ and $(\Psi^{(1)})_{nk} = \mathcal{K}\psi_k(x_n) = \psi_k(f(x_n))$. As the number of points $N \to \infty$ this converges (in any norm) to the continuum limit operator $\mathcal{K}_K$, which is the least-squares best approximation in $L^2(\mu)$, and is therefore still finite-rank [20]. As a result, the Koopman matrix's spectral data also converge to those of the continuum limit operator $\mathcal{K}_K$.

The contribution of this paper is as follows. Let $\lambda_{j,K}$ be the eigenvalues of the EDMD estimate of the Koopman operator $\mathcal{K}_K$ obtained using the function dictionary

$$\Psi_K = \{e^{-i(K-1)x}, e^{-i(K-2)x}, \ldots, e^{-ix}, 1, e^{ix}, \ldots, e^{i(K-1)x}\}, \tag{2}$$

in the limit of infinite data points. Each of these eigenvalues, which will be simple for K large enough if the corresponding λ_j is simple, have left and right eigenvectors $a_{j,K}, b_{j,K} \in C^0$, which we can consider in the observable basis. Let $\beta_{j,K}(w) = \int b_{j,K}(x)w(x)\mathrm{d}\mu(x)$.

Theorem 1.1. *Suppose f is analytic and μ has analytic density. Then for some $c > 0$ $d_{\mathrm{Hausdorff}}(\sigma(\mathcal{K}_K), \{\lambda_j : j \in \mathbb{N}\}) = \mathcal{O}(e^{-c\sqrt{K}})$ as $K \to \infty$.*

Furthermore, if λ_j has multiplicity 1 (see Theorem 2.1 for a more general result), then there exist constants $h, \zeta > 0$ depending on f, μ such that for $t \in [0, \zeta]$,

$$|\lambda_{j,K} - \lambda_j|, \|a_{j,K} - a_j\|_{\mathcal{H}^2_t}, \|b_{j,K} - b_j\|_{\mathcal{H}^2_{-t}} = \mathcal{O}(e^{-htK}).$$

An illustration of this theorem is given in Figure 1, with eigenvalue estimates converging exponentially with the dictionary size K as predicted: note that the leading constants generally increase as the modulus of the eigenvalue decreases.

The consequence of Theorem 1.1 is that, in this setting, EDMD gives very accurate estimates about the system's rates of mixing, encoded through the Ruelle-Pollicott resonances. It also means that the so-called "Koopman modes" obtained as eigenvectors of the Koopman matrix give very accurate information about the long-term dynamical properties of the system: the functions b_j, β_j encode structures that persist under the dynamics, including almost-invariant sets [15, 13]. Notwithstanding that the maps we consider are extremely structurally simple, this may go some way to explaining the success of DMD with small dictionaries.

The overarching proof idea parallels that sketched in [26]. The first component is a study the transfer operator acting on a Hilbert space $\mathcal{H}^2_t$ of analytic functions: we have improved previous results by showing that the norm of the transfer operator in $\mathcal{H}^2_t$ is independent of the analyticity parameter t. The second component, in which we make the most interesting advances on previous work, is a study of the effect of discretisation onto the finite observable space (2) as a projection $\mathcal{P}_K : \mathcal{H}^2_t \circlearrowleft$ in that space.

To understand this operator $\mathcal{P}_K$ we borrow ideas from the distant theory of orthogonal polynomials on the unit circle (OPUC) [25]. This theory formulates a basis of trigonometric polynomials in which $\mathcal{P}_K$ is diagonal, and allows one to relate this basis quite effectively to the usual monomial basis. We obtain the following fundamental result, which suggests that $\mathcal{P}_K$ is as good as the Dirichlet kernel in certain Hardy spaces under some fairly weak stipulations.

Aiming to be as general as possible, we study the Hilbert spaces $W^\sigma := \mathcal{F}^{-1}[\sigma\ell^2(\mathbb{Z})]$ weighted by so-called *Beurling weights*: even functions $\sigma : \mathbb{Z} \to \mathbb{R}^+$ that increase on the natural numbers, and obey $\sigma(j)\sigma(k) \le \sigma(|j| + |k|)$. Among these spaces include the usual fractional Sobolev Hilbert spaces (up to norm equivalence), as well as our Hardy-Hilbert spaces $\mathcal{H}^2_t$.

Theorem 1.2. *Suppose that $\sigma, \tau : \mathbb{Z} \to \mathbb{R}_+$ are Beurling weights with τ/σ decreasing on $\mathbb{N}$.*

Suppose furthermore that $\mu = h\,\mathrm{d}x$ is a positive measure on $\mathbb{T}$ with $M^{-1} \le h(x) \le M$ on $\mathbb{T}$ and $\|\sigma\mathcal{F}[(\log h)']\|_{\ell^q} \le A$ for some $q < \infty$.

Then there is a constant $C_{\mathcal{P}}$ increasing in and dependent only on q, A, M such that

$$\|I - \mathcal{P}_K\|_{W^\sigma \to W^\tau} \le C_{\mathcal{P}} \|I - \mathcal{D}_K\|_{W^\sigma \to W^\tau} = C_{\mathcal{P}} \frac{\tau(K)}{\sigma(K)}.$$

where $\mathcal{P}_K$ is the $L^2(\mu)$-orthogonal projection onto trigonometric polynomials of degree less than K, and $\mathcal{D}_K$ is the Dirichlet kernel.

Note that this implies that the error of $\mathcal{P}_K$ from our Hardy spaces $\mathcal{H}^2_t \to \mathcal{H}^2_s$ is $\mathcal{O}(e^{-(t-s)})$ for small enough t, s when the measure density h is analytic.

This is, as far as we know, a new kind of approximation result even in the theory of OPUC. It suggests that if μ is smooth then approximation of smooth functions in $L^2(\mu)$ is very powerful: we should therefore be eager to make use of this kind of polynomial approximation where it arises, for example in least squares approximation from data. We believe also that the result would generalise to higher dimensions, retaining $\mathcal{O}(e^{-cK})$ convergence for a dictionary of degree $< K$ trigonometric polynomials.

We have taken great pains to optimise our results, but there are places where the results could be meaningfully improved. In particular, we bounded the operator norms of certain operators in Lemma 4.8 via the trace norm, which would scale poorly into higher dimensions, and even in one dimension requires μ to be at least half an order more differentiable than the functions it is approximating. This would make it difficult to apply our methods for less regular densities—which naturally occur even for expanding chaotic systems. For example, the chaotic logistic map $f(x) = ax(1-x)$, which models non-hyperbolic dynamics which are standard in fluid flow, typically has a physical measure with density in $W^{s,p}$ only for certain $p + s < 1/2$ [24]. Further work, perhaps building on OPUC theory in [17], might show a path through, and establish how well DMD works with the kinds of irregular sampling measures typical of most chaotic dynamics.

2 Setup

In this section we will state the precise condition on the map f and sampling density μ that we require, and state our Koopman operator approximation theorems in generality.

The function spaces we will use are extensions of the standard L^p spaces on the torus:

$$\|\varphi\|_{L^p(\mathbb{T})} = \begin{cases} \left(\int_0^{2\pi} |\varphi(x)|^p \, \mathrm{d}x\right)^{1/p}, & p \in [1,\infty) \\ \text{ess-sup}_{x\in\mathbb{T}} |\varphi(x)|, & p = \infty. \end{cases}$$

However, we will need to construct stronger function spaces to obtain effective results. In particular, we will look at on certain open complex strips in phase space around $\mathbb{T}$, which we will parametrise by their half-thickness ζ:

$$\mathbb{T}_\zeta := \{z \in \mathbb{C}/2\pi\mathbb{Z} : |\Im z| < \zeta\}$$

On these sets we will define so-called Hardy spaces $\mathcal{H}^p_\zeta$: the space $\mathcal{H}^p_\zeta$ is the set of holomorphic functions $\varphi : \mathbb{T}_\zeta \to \mathbb{C}$ such that

$$\|\varphi\|_{\mathcal{H}^p_\zeta} := \begin{cases} \sup_{\beta\in[0,\zeta)} \left(\frac{1}{4\pi}\int_{\mathbb{T}} |\varphi(\theta+i\beta)|^p + |\varphi(\theta-i\beta)|^p \, \mathrm{d}\theta\right)^{1/p}, & p \in [1,\infty) \\ \sup_{z\in\mathbb{T}_\zeta} |\varphi|, & p = \infty \end{cases}$$

is finite (in this case the supremum is attained in the limit as $\beta \to \zeta$, and φ is continuous onto $\partial\mathbb{T}_\zeta$ almost everywhere). The $L^p(\mathbb{T})$ norms emerge as a limiting case as $\zeta \to 0$. (For concision, we will define $\|\cdot\|_{\mathcal{H}^p_0} = \|\cdot\|_{L^p(\mathbb{T})}$.

The spaces $\mathcal{H}^2_\zeta$ are Hilbert spaces, and we can characterise these norms in Fourier space. Let $\mathcal{F} : L^2(\mathbb{T}) \to \ell_2(\mathbb{Z})$ be the Fourier series operator:

$$(\mathcal{F}\varphi)(k) = \frac{1}{2\pi}\int_{\mathbb{T}} e^{-ikx}\varphi(x)\, \mathrm{d}x$$

and for σ a Beurling weight let us define the Hilbert spaces $W^\sigma(\mathbb{T}) = \mathcal{F}^{-1}[\sigma\ell^2(\mathbb{Z})]$ with

$$\|\varphi\|_{W^\sigma} = \|\sigma\mathcal{F}\varphi\|_{\ell^2(\mathbb{Z})}.$$

It is a classic result that the standard $L^2(\mathbb{T})$ is such a space with $\sigma(n) \equiv 1$; Sobolev spaces H^r are isometric to those with $\sigma(n) = (1+n^2)^{r/2}$. Similarly, if we define the Beurling weight

$$\sigma_\zeta(k) := \sqrt{\cosh 2k\zeta},$$

then we can write our Hardy space $\mathcal{H}^2_\zeta(\mathbb{T}) = W^{\sigma_\zeta}$.

We can also define adjoints of these spaces: for a Beurling weight σ let $W^{\sigma^{-1}}$ be the completion of $L^2(\mathbb{T})$ with respect to the following norm:

$$\|\varphi\|_{W^{\sigma^{-1}}} = \sup_{\|\psi\|_{W^\sigma}=1} \frac{1}{2\pi}\int_{\mathbb{T}} \varphi\psi \, \mathrm{d}\theta.$$

This Hilbert space $W^{\sigma^{-1}}$ is therefore isometric to $\sigma^{-1}\ell^2$ under the Fourier series operator $\mathcal{F}$. We can of course therefore define $\mathcal{H}^2_{-\zeta} = W^{\sigma_\zeta^{-1}}$ as the dual of $\mathcal{H}^2_\zeta$.

Our basic assumptions will be that map f is uniformly expanding and real-analytic and the sampling measure μ has real-analytic density $h(x)$ with respect to Lebesgue. Note that if μ is the absolutely continuous invariant measure of f, then the first part implies the second.

To better understand rates of convergence, we will make the following quantitative assumptions on f for some $\zeta > 0$:

- The lift of f onto $\mathbb{R}$ has an inverse $v : \mathbb{R} \to \mathbb{R}$ that extends analytically to $\mathbb{R}_\zeta := \mathbb{R}+i[-\zeta,\zeta]$.

- For $z \in \mathbb{R}_\zeta$, $|v'(z)| \leq \gamma^{-1}$ for some $\gamma < 1$.
- $\| \log(v')\|_{C^\alpha(\mathbb{T}_\zeta)} \leq D$ for some constants $\alpha > 0$ and $D \leq \frac{\pi}{3}\zeta^{-\alpha}$.

The last bound can always be achieved by making ζ smaller.

When studying the convergence of extended dynamical mode decomposition, we will also make the following quantitative assumptions on h:

- h extends analytically to $\mathbb{T}_\zeta$.
- For $z \in \mathbb{T}_\zeta$, $M^{-1} \leq |h(z)| \leq M$ for some constant M.
- $\|\sigma \mathcal{F}[(\log h)']\|_{\ell^q} \leq A$ for some constants A and $q < \infty$. (Note this occurs if $(\log h)' \in L^p(\partial \mathbb{T}_\zeta)$ where $1/p + 1/q = 1$.)

The key to our results on spectral convergence is the following strong operator convergence result:

Theorem 2.1. *For all $t \in (0, \zeta]$ the Koopman operator $\mathcal{K} : \mathcal{H}^2_{-t} \circlearrowleft$ is compact, the sequence of operators $\{\mathcal{K}_K : \mathcal{H}^2_{-t} \circlearrowleft\}_{K\in\mathbb{N}}$ is collectively compact, and there exists a constant C depending only on M, A, D, ζ such that*

$$\|\mathcal{K}_K - \mathcal{K}\|_{\mathcal{H}^2_{-t}} \leq C e^{-(1-\gamma^{-1})tK}.$$

This allows us to estimate all the eigenmodes and Ruelle-Pollicott resonances of the system in (1):

Corollary 2.2. *For all $\mathcal{K}$'s eigenvalues λ of multiplicity m, the error in the $\mathcal{K}_K$ estimate as $K \to \infty$ is $\mathcal{O}(e^{-(1-\gamma^{-1})\zeta K/m})$ for the eigenvalues, and $\mathcal{O}(e^{-(1-\gamma^{-1})tK})$ in H^2_{-t} (resp. $H^2_{\gamma^{-1}t}$) for the right (resp. left) generalised eigenspaces.*

3 Orthogonal trigonometric polynomial theory

To understand the μ-orthogonal projection onto our dictionary of complex exponential basis functions $\{e_k\}_{|k|\leq K}$, it will be natural to understand the transformation between the complete, infinite basis $\{e_k\}_{k\in\mathbb{Z}}$ and an orthogonal polynomial basis of $L^2(\mu)$ (in which the $L^2(\mu)$ projection is therefore diagonal). In the correct ordering it is known that the basis change matrices are triangular and asymptotically banded [25]: we will prove some quantitative results on the rate of convergence to bandedness as the order of the polynomial grows.

Recalling that the complex exponential basis on $\mathbb{C}/2\pi\mathbb{Z} \supset \mathbb{T}$ is

$$e_k(z) := e^{ikz},\ k \in \mathbb{Z},$$

let us order the index set $\mathbb{Z}$

$$0 \leq_b -1 <_b 1 <_b -2 <_b 2 <_b \ldots.$$

We will sometimes group the indices into "blocks" $\{0\}, \{-1, 1\}, \{-2, 2\}, \ldots$.

We will always assume that $\mathcal{A}^\dagger$ refers to the adjoint of $\mathcal{A}$ in $L^2(\mathbb{T}, \mathrm{Leb}/2\pi)$; we will call an operator $\mathcal{A} : L^2(\mathbb{T}) \circlearrowleft$ lower-triangular (resp. upper-triangular) if $e_j^\dagger \mathcal{A} e_k = 0$ for all $j <_b k$ (resp. $j >_b k$).

For any positive probability density $h : \mathbb{T} \to \mathbb{R}_+$, define the Fourier space multiplication operator $\mathcal{M} : L^2(\mathbb{T}) \to L^2(\mathbb{T})$:

$$\mathcal{M}_h \varphi = h\varphi. \tag{3}$$

The operator $\mathcal{M}$ is positive-definite and Hermitian in $L^2(\mathbb{T})$. (In fact, in Fourier space it is near-block-Toeplitz with respect to $<_b$.) It turns out we can make a Cholesky decomposition of $\mathcal{M}$ that is very nicely and uniformly bounded in our Hardy or Beurling-weighted Hilbert spaces, under our analyticity assumptions on h:

Theorem 3.1. *Under the assumptions on μ in Theorem 1.2,*

a. There exists a unique lower triangular operator $\mathcal{V}$ acting on $L^2(\mathbb{T})$ such that

$$\mathcal{M} = \mathcal{V}\mathcal{V}^\dagger.$$

Furthermore, $\mathcal{V}$ is invertible. If we let $\mathcal{V}^{-1} =: \mathcal{U}^\dagger$, the operator $\mathcal{U}$ is upper-triangular with

$$\mathcal{M}^{-1} = \mathcal{U}^\dagger\mathcal{U}.$$

b. There exists a constant $C_\triangle$ increasing in q, M, A such that

$$\|\mathcal{U}\|_{W^\sigma}, \|\mathcal{V}\|_{W^\sigma}, \|\mathcal{U}^\dagger\|_{W^\sigma}, \|\mathcal{V}^\dagger\|_{W^\sigma} \leq C_\triangle.$$

Now, define the functions

$$p_k = \mathcal{U}e_k.$$

From the theorem above, it turns out that these are a complete family of (complex) trigonometric polynomials orthonormal with respect to h:

Proposition 3.2. *The p_k are each trigonometric polynomials of order exactly $|k|$, and form a complete orthonormal basis in $L^2(\mu)$.*

The p_k as described here are complex, but they can be transformed via a unitary map into a real orthonormal family (see the proof of Theorem 3.1 part a): we have only chosen a complex exponential basis for ease of presentation.

The corollary of Theorem 3.1(a) is that $L^2(\mu)$-orthogonal projection is conjugate to the usual $L^2(\mathrm{Leb})$ projection (i.e. the Dirichlet kernel, whose properties are very well known):

Proposition 3.3. *Let $\mathcal{P}_K : L^2(\mu) \circlearrowleft$ be the orthogonal projection onto trigonometric polynomials of order $< K$. Then*

$$\mathcal{P}_K = \mathcal{U}^{-1}\mathcal{D}_K\mathcal{U} = \mathcal{V}^\dagger\mathcal{D}_K\mathcal{U}$$

where $\mathcal{D}_K = \mathcal{F}^{-1}\mathbb{1}_{(-K+1,\dots,K-1)}\mathcal{F}$ is convolution by the Dirichlet kernel.

Proof. By Theorem 3.1(b) and the fact that $\{e_k\}_{k\in\mathbb{Z}}$ are a complete basis of W^σ, our polynomials $\{p_k\}_{k\in\mathbb{Z}}$ form a complete basis of W^σ (and therefore also $L^2(\mathrm{Leb}) = L^2(\mu)$ setting $\sigma \equiv 1$).

The action of $\mathcal{P}_K$ on our basis is $\mathcal{P}_K p_k = \mathbb{1}_{|k|<K} p_k$. Using from Proposition 3.2 that $p_k = \mathcal{U}e_k$, we have

$$\mathcal{P}_K\mathcal{U}e_k = \mathbb{1}_{|k|\leq K}\mathcal{U}e_k$$

so

$$\mathcal{U}^{-1}\mathcal{P}_K\mathcal{U}e_k = \mathbb{1}_{|k|\leq K}e_k.$$

The action of $\mathcal{U}^{-1}\mathcal{P}_K\mathcal{U}$ is precisely the action of $\mathcal{D}_K$, so

$$\mathcal{P}_K = \mathcal{U}^{-1}\mathcal{D}_K\mathcal{U} = \mathcal{V}^\dagger\mathcal{D}_K\mathcal{U}.$$

□

Thus motivated to study Cholesky decompositions of multiplication operators, we embark on proving that such things exist and their constituents are bounded in W^σ.

4 Proof of orthogonal polynomial results

We begin by proving the existence of the Cholesky decompositions.

Proof of Theorem 3.1a. Let us consider $\mathcal{M}$ as an infinite matrix acting on the basis

$$1, \sin z, \cos z, \sin 2z, \cos 2z, \ldots$$

This is a real, positive-definite matrix. By [8, Lemma 3.1], there exists a unique lower-triangular operator $\mathcal{V}_{\mathbb{R}}$ such that $\mathcal{M} = \mathcal{V}_{\mathbb{R}}\mathcal{V}_{\mathbb{R}}^{\dagger}$. Now, for any function $\varphi \in L^2(\mathbb{T})$,

$$\varphi^{\dagger}\mathcal{M}\varphi = \varphi^{\dagger}\mathcal{V}_{\mathbb{R}}\mathcal{V}_{\mathbb{R}}^{\dagger}\varphi = \|\mathcal{V}_{\mathbb{R}}^{\dagger}\varphi\|_{L^2}^2 \tag{4}$$

so since $\mathcal{M}\varphi = h\varphi$,

$$M^{-1}\|\varphi\|_{L^2}^2 \leq \|\mathcal{V}_{\mathbb{R}}^{\dagger}\varphi\|_{L^2}^2 \leq M\|\varphi\|_{L^2}. \tag{5}$$

Since, $\mathcal{V}_{\mathbb{R}}^{\dagger}$ is bounded in $L^2(\mathbb{T})$, so too is its adjoint $\mathcal{V}_{\mathbb{R}}$ and its inverse $\mathcal{U}_{\mathbb{R}}$. Now, $\mathcal{V}_{\mathbb{R}}^{\dagger}$ is block-lower triangular in the complex exponential basis e_k, the blocks consisting of $\{e_k, e_{-k}\}$. We can perform a LQ decomposition on each block to obtain a block-diagonal unitary matrix $\mathcal{Q}$ such that $\mathcal{V} = \mathcal{V}_{\mathbb{R}}\mathcal{Q}$ is lower-triangular in the complex exponential basis, and $\mathcal{U} = \mathcal{Q}\mathcal{U}_{\mathbb{R}}$ is upper-triangular in this basis with $\mathcal{V}\mathcal{V}^{\dagger} = \mathcal{M}$ and $\mathcal{U}^{\dagger}\mathcal{U} = \mathcal{M}^{-1}$. $\square$

Following the same argument in (4)–(5) yields that $\|\mathcal{V}\|_{L^2(\mathbb{T})}, \|\mathcal{U}\|_{L^2(\mathbb{T})} \leq \sqrt{M}$.

Let θ^+, θ^- be real-analytic functions $\mathbb{T} \to \mathbb{C}$ such that $\theta^-(\bar{z}) = \overline{\theta^+(z)}$, $\theta^+\theta^- = 1/h$ and θ^+ is holomorphic in the upper half-plane. We can specify them explicitly:

$$(\mathcal{F}\log\theta^+)(k) = \left(\tfrac{1}{2}\delta_{0k} + \mathbb{1}(k > 0)\right)\mathcal{F}(-\log h)(k). \tag{6}$$

and similarly for θ^- with $k < 0$ replacing $k > 0$. In the language of orthogonal polynomials on the unit circle (where one studies the variable $z = e^{ix}$), the function $\theta^+(\log z)$ is known as the *Szegő function* of $\mathrm{d}\mu(\log z)$ [25]. We will also define their reciprocals $\eta^{\pm} = 1/\theta^{\pm}$. The following result states some of its basic properties of $\theta^{\pm}, \eta^{\pm}$ (c.f. [25, Theorem 2.4.1]).

Proposition 4.1. *Under our assumptions on h:*

a. For all $x \in \mathbb{T}$ we have

$$|\theta^+(x)| = |\theta^-(x)| = \sqrt{h(x)}.$$

b. Considering functions as multiplication operators,

$$\|\theta^{\pm}\|_{W^{\sigma}\to W^{\sigma}}, \|\eta^{\pm}\|_{W^{\sigma}\to W^{\sigma}} \leq M^{1/2}e^{qA}.$$

c. If $\hat{\theta}_l^+$ are the Fourier coefficients of θ^+, then

$$\sum_{l=0}^{\infty} l\sigma(l)^2|\hat{\theta}_l^+|^2 \leq qA^2Me^{2qA}.$$

We will find it briefly useful to notate some norms associated to the Beurling weight. For $q \in [1, \infty]$ let $\|\varphi\|_{\sigma;q} = \|\sigma\mathcal{F}\varphi\|_{\ell^q}$.

Lemma 4.2. *Suppose σ is a Beurling weight. Then for any $q \in [1, \infty]$,*

$$\|\varphi\psi\|_{\sigma;q} \leq \|\varphi\|_{\sigma;q}\|\psi\|_{\sigma;1}$$

whenever $\|\varphi\|_{\sigma;r}, \|\psi\|_{\sigma;1} < \infty$.

Proof. All we need to prove is that

$$\|\sigma(\hat{\varphi} * \hat{\psi})\|_{\ell^q} \leq \|\sigma\hat{\varphi}\|_{\ell^q}\|\sigma\hat{\psi}\|_{\ell^q}.$$

By the definition of the Beurling weight, $\sigma(k) \leq \sigma(|j| + |k-j|) \leq \sigma(j)\sigma(k-j)$, so

$$\begin{aligned}|\sigma(k)(\hat{\varphi} * \hat{\psi})(k)| &= \sum_{j\in\mathbb{Z}} \sigma(k)|\hat{\varphi}(j)||\hat{\varphi}(k-j)| \\ &\leq \sum_{j\in\mathbb{Z}} |\sigma(j)\hat{\varphi}(j)||\sigma(k-j)\hat{\varphi}(k-j)| \\ &\leq (|\sigma\hat{\varphi}| * |\sigma\hat{\psi}|)(k)\end{aligned}$$

As a result,

$$\|\sigma(\hat{\varphi} * \hat{\psi})\|_{\ell^q} \leq \||\sigma\hat{\varphi}| * |\sigma\hat{\psi}|\|_{\ell^q} \leq \|\sigma\hat{\varphi}\|_{\ell^q}\|\sigma\hat{\psi}\|_{\ell^1}$$

as required. □

Proof of Proposition 4.1. We only need to prove the results for θ^+, as $\log\theta^+(\bar{z}) = \overline{\log\theta^-(z)}$.

The first part directly uses this fact: $|\theta^+(x)|^2 = \theta^-(x)\theta^+(x) = h(x)^{-1}$.

To prove the second part, we separate $k = 0$ and apply Hölder's inequality to get

$$\|\log h\|_{\sigma;1} \leq \sigma(0)\left|\int \log h \, \mathrm{d}x\right| + \|\sigma(k)k\mathcal{F}[\log h](k)\|_{\ell^q}\|\mathbb{1}_{k\neq 0}k^{-1}\|_{\ell^p}$$

where $1/p + 1/q = 1$. Noting that by submultiplicativity $\sigma(0) \leq 1$, this gives us

$$\|\log h\|_{\sigma;1} \leq \log M + \|(\log h)'\|_{\sigma;q}(2\zeta(p))^{1/p} \leq \log M + A(2q)^{1/p} \leq \log M + 2qA.$$

Noting that $\log h$ is real on $\mathbb{T}$, (6) gives us that

$$\|\log\theta^+\|_{\sigma;1} = \tfrac{1}{2}\|\log h\|_{\sigma;1} = \log M^{1/2} + qA.$$

Now, $\theta^+, \eta^+ = 1/\theta^+$ are respectively the $t = 1, -1$ solutions of

$$\frac{\partial}{\partial t}\mathcal{E}_t(x) = \log\theta^+\,\mathcal{E}_t(x),\ \mathcal{E}_0(x) = 1$$

so Gronwall's Lemma combined with Lemma 4.2 gives

$$\|\theta^+\|_{\sigma;1}, \|\eta^+\|_{\sigma;1} \leq e^{\|\log\theta^+\|_{\sigma;1}}\|1\|_{\sigma;1} \leq e^{qA}M^{1/2}.$$

The bounds on θ^+, η^+ considered as multipliers on W^σ (which has the $\|\cdot\|_{\sigma;2}$ norm) follows from Lemma 4.2.

For the third part we proceed in the same vein as the second part. From (6) we have $\mathcal{F}(\log\theta_+)' = \mathbb{1}(k > 0)\mathcal{F}(-\log h)(k)$, so $\|\sigma\mathcal{F}[(\log\theta_+)']\|_{\ell^q} \leq A$. Then,

$$\|(\theta^+)'\|_{\sigma;q} = \|(\log\theta^+)'\|_{\sigma;q}\|\theta^+\|_{\sigma;1} \leq AM^{1/2}e^{qA}.$$

Applying Hölder's inequality gives that

$$\begin{aligned}\|k^{1/2}\sigma\mathcal{F}[\theta^+]\|_{\ell_2}^2 &= \|(\mathbb{1}(k\neq 0)k^{-1/2})\sigma\mathcal{F}[(\theta^+)']\|_{\ell^2}^2 \\ &\leq \|\mathbb{1}(k\neq 0)k^{-1}\|_{\ell^{q/(q-2)}}\|(\theta^+)'\|_{\sigma;q}^2 \\ &\leq qA^2Me^{2qA},\end{aligned}$$

which is what needed to be proven. □

With these properties in hand we can now quantitatively characterise the Cholesky decomposition.

Define the projections

$$\begin{aligned}\mathcal{P}^+ &= \mathcal{F}^{-1}\mathbb{1}_{\mathbb{Z}^+}\mathcal{F} \\ \mathcal{P}^\circ &= \mathcal{F}^{-1}\mathbb{1}_{\{0\}}\mathcal{F} = e_0e_0^\dagger \\ \mathcal{P}^- &= \mathcal{F}^{-1}\mathbb{1}_{\mathbb{Z}^-}\mathcal{F}.\end{aligned}$$

Let's also define operators

$$\mathcal{M}\varphi = h\varphi$$

being multiplication by h, our limiting Cholesky factors

$$\bar{\mathcal{U}} = \bar{c}\mathcal{P}^\circ + \mathcal{P}^+\theta^- + \mathcal{P}^-\theta^+.$$

$$\bar{\mathcal{V}} = \bar{c}^{-1}\mathcal{P}^\circ + \eta^+\mathcal{P}^+ + \eta^-\mathcal{P}^-$$

where $\bar{c} = \sqrt{\frac{1}{2\pi}\int h^{-1}\,\mathrm{d}x}$.

These obey the same relation as their equivalents $\mathcal{V}$,$\mathcal{U}$ do:

Proposition 4.3. $\bar{\mathcal{V}}^{-1} = \bar{\mathcal{U}}^\dagger$.

Proof. If ψ^+ contains only non-negative Fourier modes, then $\psi^+\mathcal{P}^+ = \mathcal{P}^+\psi^+\mathcal{P}^+$, with the corresponding result for functions with non-positive modes.

We have that $\bar{\mathcal{U}}^\dagger = \bar{c}\mathcal{P}^\circ + \theta^+\mathcal{P}^+ + \theta^-\mathcal{P}^-$, and the result follows by expanding out $\bar{\mathcal{U}}^\dagger\bar{\mathcal{V}}$ and $\bar{\mathcal{V}}\bar{\mathcal{U}}^\dagger$. □

These operators are uniformly bounded:

Lemma 4.4. *For all $t \in [0, \zeta]$,*

$$\|\bar{\mathcal{U}}\|_{W^\sigma}, \|\bar{\mathcal{U}}^\dagger\|_{W^\sigma}, \|\bar{\mathcal{V}}\|_{W^\sigma}, \|\bar{\mathcal{V}}^\dagger\|_{W^\sigma} \leq (1 + 2e^{qA})M^{1/2}.$$

Proof. Since the projections $\mathcal{P}^+, \mathcal{P}^-, \mathcal{P}^\circ$ have norm 1 in W^σ,

$$\|\bar{\mathcal{U}}\|_{W^\sigma} \leq \bar{c} + \|\theta^+\|_{W^\sigma\to W^\sigma} + \|\theta^-\|_{W^\sigma\to W^\sigma} \leq (1 + 2e^{qA})M^{1/2}.$$

The same argument goes for the other operators. □

Proposition 4.5. *If $\mathcal{P}^+w = 0$, then $\mathcal{P}^+w - w\mathcal{P}^+ = (\mathcal{P}^- + \mathcal{P}^\circ)w\mathcal{P}^+$. The same statement holds when $\mathcal{P}^+$ and $\mathcal{P}^-$ are swapped.*

Proof. For $k \leq 0$ the operator $\mathcal{P}^+e^{-ik\cdot}\mathcal{P}^+$ reduces to $\mathcal{P}^+e^{ik\cdot}$. Since we have that $w(z) = \sum_{k=0}^\infty \hat{w}_{-k}e^{-ikz}$ so $\mathcal{P}^+w\mathcal{P}^+ = \mathcal{P}^+w$. Then,

$$\mathcal{P}^+w - w\mathcal{P}^+ = (I - \mathcal{P}^+)w\mathcal{P}^+ = \mathcal{P}^\circ w\mathcal{P}^+ + \mathcal{P}^-w\mathcal{P}^+.$$

□

We will want to show that in some sense $\bar{\mathcal{V}}$ is a good approximation of $\mathcal{V}$, the Cholesky factor of $\mathcal{M}$, and similarly $\bar{\mathcal{U}}$ is a good approximation of $\mathcal{U}$. We can understand this by attempting to "strip" $\mathcal{M}$ down to the identity, by considering $\bar{\mathcal{V}}^\dagger\mathcal{M}^{-1}\bar{\mathcal{V}}$ and $\bar{\mathcal{U}}^\dagger\mathcal{M}\bar{\mathcal{U}}$. (It is important to study both at the same time.) It will turn out in Lemma 4.8 that all we need to know are the diagonal entries of these operators. In the following two propositions, we will show that these entries converge to those of the identity (c.f. [25, Proposition 2.4.7, Theorem 7.2.1]).

Proposition 4.6. *For all $k \in \mathbb{Z}$,*

$$s'_k = e_k^\dagger\bar{\mathcal{V}}^\dagger\mathcal{M}^{-1}\bar{\mathcal{V}}e_k - 1 = 0.$$

Proof. We have

$$1 + s'_k = (\bar{\mathcal{V}}e_k)^\dagger\mathcal{M}^{-1}(\bar{\mathcal{V}}e_k).$$

When $k = 0$,

$$\bar{\mathcal{V}}e_0 = \bar{c}^{-1}e_0 = (e_0^\dagger h^{-1})^{-1/2}e_0$$

so

$$1 + s'_0 = (e_0^\dagger h^{-1})^{-1}e_0^\dagger h^{-1}e_0 = 1.$$

On the other hand, when $k > 0$ (the $k < 0$ case follows by the same argument), $\bar{\mathcal{V}}e_k = \eta^+\mathcal{P}^+e_k = \eta^+e_k$ so

$$1 + s'_k = e_k^\dagger\eta_-\eta_-\eta_+\eta^+e_k = e_k^\dagger e_k = 1,$$

as required. □

Proposition 4.7. *Let*

$$s_k = e_k^\dagger \bar{\mathcal{U}}^\dagger \mathcal{M} \bar{\mathcal{U}} e_k - 1.$$

Then $s_k > -1$ and there exists a constant C increasing in M, A, q such that

$$\sum_{k\in\mathbb{Z}} \sigma(k)^2 |s_k| \leq C. \tag{7}$$

Proof. In this case we find that

$$1 + s_k = (\bar{\mathcal{U}} e_k)^\dagger \mathcal{M} (\bar{\mathcal{U}} e_k),$$

but the situation is more complicated because, when compared with the previous proposition, the order of projection and multiplication have reversed.

For $k = 0$ we have that

$$\bar{\mathcal{U}} e_0 = \mathcal{P}^+ \theta_- e_0 + \mathcal{P}^- \theta_+ e_0 + \bar{c} \mathcal{P}^\circ e_0 = \bar{c} e_0$$

so

$$s_0 = 1 - \bar{c}^2 e_0^\dagger h e_0 = 1 - \frac{1}{2\pi} \int_T h^{-1} \,\mathrm{d}x \, \frac{1}{2\pi} \int_{\mathbb{T}} h \,\mathrm{d}x$$

from which one may extract that $|s_0| \leq (M-1)^2$.

For $k > 0$ we have that

$$\bar{\mathcal{U}} e_k = \mathcal{P}^+ \theta_- e_k + \mathcal{P}^- \theta_+ e_k + \bar{c} \mathcal{P}^\circ e_k = \mathcal{P}^+ \theta_- e_k,$$

and since $\mathcal{M} = \eta^+ \eta^- = \overline{\eta^-} \eta^-$,

$$(\bar{\mathcal{U}} e_k)^\dagger \mathcal{M} (\bar{\mathcal{U}} e_k) = (\eta^- \mathcal{P}^+ \theta^- e_k)^\dagger \eta^- \mathcal{P}^+ \theta^- e_k.$$

Now, we have that

$$\eta^- \mathcal{P}^+ \theta^- e_k - e_k = \eta^- (\mathcal{P}^+ \theta^- - \theta^- \mathcal{P}^+) e_k = -\eta^- (\mathcal{P}^\circ + \mathcal{P}^-) \theta^- \mathcal{P}^+ e_k$$

with the last equality by Proposition 4.5. Now,

$$(\eta^- \mathcal{P}^+ \theta^- e_k - e_k)^\dagger e_k = -(\theta^- e_k)^\dagger (\mathcal{P}^\circ + \mathcal{P}^-) \eta^+ \mathcal{P}^+ e_k = 0,$$

so

$$\begin{aligned}
(\bar{\mathcal{U}} e_k)^\dagger \mathcal{M} (\bar{\mathcal{U}} e_k) &= e_k^\dagger e_k + (\eta^- \mathcal{P}^+ \theta^- e_k - e_k)^\dagger (\eta^- \mathcal{P}^+ \theta^- e_k - e_k) \\
&= 1 + \| -\eta^- (\mathcal{P}^\circ + \mathcal{P}^-) \theta^- \mathcal{P}^+ e_k \|_{L^2(\mathbb{T})}^2 \\
&\leq 1 + \|\eta^-\|_{L^\infty(\mathbb{T})}^2 \|(\mathcal{P}^\circ + \mathcal{P}^-) \theta^- e_k\|_{L^2(\mathbb{T})}^2.
\end{aligned}$$

Now,

$$\|(\mathcal{P}^\circ + \mathcal{P}^-) \theta^- e_k\|_{L^2(\mathbb{T})}^2 = \sum_{j=0}^{\infty} |\mathcal{F}(\theta^- e_k)(-j)^2| = \sum_{j=0}^{\infty} |\hat{\theta}^-_{-j-k}|^2 = \sum_{l=k}^{\infty} |\hat{\theta}^+_l|^2.$$

Consequently, and using the first part of Proposition 4.1 to bound $\eta^- = \overline{1/\theta^+}$, we have

$$s_k = |(\bar{\mathcal{U}} e_k)^\dagger \mathcal{M} (\bar{\mathcal{U}} e_k) - 1| \leq M \sum_{l=k}^{\infty} |\hat{\theta}^+_l|^2.$$

Combining the cases $k = 0$, $k < 0$, $k > 0$ we find that for all $k \in \mathbb{Z} \backslash \{0\}$,

$$s_k \leq M \sum_{l=|k|}^{\infty} |\hat{\theta}^+_l|^2,$$

which we can use to bound (7). In particular,

$$\begin{aligned}\sum_{k\in\mathbb{Z}}\sigma(k)^2|s_k| &\le (M-1)^2+2M\sum_{k=1}^{\infty}\sigma(k)^2\sum_{l=k}^{\infty}|\hat{\theta}_l^+|^2\\ &=(M-1)^2+2M\sum_{l=1}^{\infty}\left(\sum_{k=1}^{l}\sigma(k)^2\right)|\hat{\theta}_l^+|^2\\ &\le (M-1)^2+2M\sum_{l=1}^{\infty}l\sigma(l)^2|\hat{\theta}_l^+|^2\le M^2(1+qA^2e^{2qA}),\end{aligned}$$

using the last part of Proposition 4.1 in the last part. □

Lemma 4.8. *Let $\mathcal{V}\bar{\mathcal{V}}^{-1}=I+\tilde{\mathcal{V}}$ and $\mathcal{U}\bar{\mathcal{U}}^{-1}=I+\tilde{\mathcal{U}}$. Then there exists C increasing in M, A, q such that*

$$\|\tilde{\mathcal{V}}\|,\|\tilde{\mathcal{V}}^\dagger\|,\|\tilde{\mathcal{U}}\|,\|\tilde{\mathcal{U}}^\dagger\|\le C$$

in the $L^2(\mathbb{T})\to W^\sigma$ operator norm.

Proof of Lemma 4.8. Firstly, this is well-posed since by Proposition 4.3, $\bar{\mathcal{V}}^{-1}=\bar{\mathcal{U}}^\dagger$ and $\mathcal{U}^{-1}=\bar{\mathcal{V}}^\dagger$.

Let's notate the diagonal entries of $\tilde{\mathcal{V}}$ as $\alpha_k=e_k^\dagger\tilde{\mathcal{V}}e_k$. Because $\mathcal{V}$ and $\bar{\mathcal{V}}$ have positive diagonal entries, so does $I+\tilde{\mathcal{V}}$, and hence the $\alpha_k>-1$.

Then for all $k\in\mathbb{Z}$,

$$e_k^\dagger(I+\tilde{\mathcal{V}})(I+\tilde{\mathcal{V}})^\dagger e_k=\|\tilde{\mathcal{V}}^\dagger e_k\|_{L^2}^2+1+2\alpha_k.$$

But we also know that

$$e_k^\dagger(I+\tilde{\mathcal{V}})(I+\tilde{\mathcal{V}})^\dagger e_k=e_k^\dagger\bar{\mathcal{U}}^\dagger\mathcal{M}\bar{\mathcal{U}}e_k=1+s_k$$

so

$$2\alpha_k\le\|\tilde{\mathcal{V}}^\dagger e_k\|_{L^2}^2+2\alpha_k=s_k. \tag{8}$$

On the other hand let

$$\bar{\mathcal{V}}^\dagger\mathcal{M}^{-1}\bar{\mathcal{V}}=(I+\tilde{\mathcal{U}})(I+\tilde{\mathcal{U}})^\dagger.$$

Because $(\bar{\mathcal{U}}^\dagger\mathcal{M}\bar{\mathcal{U}})^{-1}=\bar{\mathcal{V}}\mathcal{M}^{-1}\bar{\mathcal{V}}$, we have that $I+\tilde{\mathcal{U}}=(I+\tilde{\mathcal{V}})^{-1}$ and so the diagonal elements of $\tilde{\mathcal{U}}$ are $(1+\alpha_k)^{-1}-1=-\frac{\alpha_k}{1+\alpha_k}$. Hence

$$1+s_k'=e_k^\dagger\bar{\mathcal{V}}^\dagger\mathcal{M}^{-1}\bar{\mathcal{V}}e_k=e_k^\dagger(I+\tilde{\mathcal{U}})(I+\tilde{\mathcal{U}})^\dagger e_k=\|\tilde{\mathcal{U}}^\dagger e_k\|_{L^2}^2+1-2\tfrac{\alpha_k}{1+\alpha_k},$$

so

$$2\tfrac{\alpha_k}{1+\alpha_k}=\|\tilde{\mathcal{U}}^\dagger e_k\|_{L^2}^2-s_k'\ge -s_k'. \tag{9}$$

As a result,

$$2\alpha_k\ge-\frac{2s_k'}{2+s_k'}\ge -s_k'.$$

Applying this to (8) means we can extract a bound on $\tilde{\mathcal{V}}^\dagger e_k$:

$$\|\tilde{\mathcal{V}}^\dagger e_k\|_{L^2}^2\le s_k+s_k'.$$

Similarly, applying the inequality in (8) to (9) gives us that

$$\|\tilde{\mathcal{U}}^\dagger e_k\|_{L^2}^2=s_k'+2\tfrac{\alpha_k}{1+\alpha_k}\le s_k'+\alpha_k\le s_k'+s_k.$$

This is nice for us because the s_k, s'_k decay very quickly (from earlier propositions). Applying $\tilde{\mathcal{V}}^\dagger$ to some function $\varphi \in L^2$ gives that

$$\begin{aligned}
\|\tilde{\mathcal{V}}^\dagger \varphi\|_{L^2} &\leq \sum_{k\in\mathbb{Z}} \|\tilde{\mathcal{V}}^\dagger e_k\|_{L^2} \mathcal{F}\varphi(k) \\
&\leq \left(\sum_{k\in\mathbb{Z}} \sigma(k)^2 \|\tilde{\mathcal{V}}^\dagger e_k\|_{L^2}^2\right)^{1/2} \left(\sum_{k\in\mathbb{Z}} \sigma(k)^{-2} |\mathcal{F}\varphi(k)|^2\right)^{1/2} \\
&\leq \left(\sum_{k\in\mathbb{Z}} \sigma(k)^2 (s'_k + s_k)\right)^{1/2} \|\varphi\|_{W^{\sigma^{-1}}} \\
&\leq C\|\varphi\|_{W^{\sigma^{-1}}},
\end{aligned}$$

with the last line coming from Propositions 4.6–4.7.

Similarly,

$$\|\tilde{\mathcal{U}}^\dagger \varphi\|_{L^2} \leq C\|\varphi\|_{W^{\sigma^{-1}}}.$$

By the duality of the $W^{\sigma^{-1}}$ and W^σ norms,

$$\|\tilde{\mathcal{U}}\psi\|_{W^\sigma}, \|\tilde{\mathcal{V}}\psi\|_{W^\sigma} \leq C\|\psi\|_{L^2}$$

for all $\psi \in L^2$, as required. On the other hand,

$$\tilde{\mathcal{V}}^\dagger = (I + \tilde{\mathcal{V}}^\dagger)^{-1} - I = -\tilde{\mathcal{V}}^\dagger (I + \tilde{\mathcal{V}}^\dagger)^{-1} = -(I + \tilde{\mathcal{V}}^\dagger)$$

which gives

$$\|\tilde{\mathcal{V}}^\dagger \psi\|_{W^\sigma} \leq \|I + \tilde{\mathcal{V}}^\dagger\|_{L^2} \|\psi\|_{L^2} \leq \|\bar{\mathcal{U}}^\dagger \mathcal{M} \bar{\mathcal{U}}\|_{L^2}^{1/2} C\|\psi\|_{L^2} \leq C\|\psi\|_{L^2}.$$

and similarly for $\tilde{\mathcal{U}}^\dagger$. □

Proof of Theorem 3.1b. For all four operators this result is a simple application of Lemmas 4.4 and 4.8. For example, let us consider

$$\mathcal{U} = (I + \tilde{\mathcal{U}})\bar{\mathcal{U}}.$$

We have that

$$\begin{aligned}
\|\mathcal{U}\|_{W^\sigma} &\leq (1 + \|\tilde{\mathcal{U}}\|_{W^\sigma})\|\bar{\mathcal{U}}\|_{W^\sigma} \\
&\leq (1 + \|\tilde{\mathcal{U}}\|_{L^2\to W^\sigma})\|\bar{\mathcal{U}}\|_{W^\sigma} \\
&\leq (1 + C)\,(1 + 2e^{qA})M^{1/2},
\end{aligned}$$

as required. □

Proof of Proposition 3.2. Integrating two polynomials against each other,

$$\begin{aligned}
\int_{\mathbb{T}} \overline{p_j} p_k \,\mathrm{d}\mu &= (\mathcal{U}e_j)^\dagger h \mathcal{U} e_k \\
&= e_j^\dagger \mathcal{U}^\dagger ((\mathcal{U}^\dagger)^{-1}\mathcal{U}^{-1})\mathcal{U} e_k \\
&= e_j^\dagger e_k = \delta_{jk},
\end{aligned}$$

as required for orthonormality. Because $\mathcal{U}$ is upper-triangular in the complex exponential basis with non-zero diagonals, $\mathcal{U}e_k$ is a combination of complex exponentials of order no greater than $|k|$, so p_k is a trigonometric polynomial of the right order.

On the other hand, the complex exponentials $\{e_k\}_{k\in\mathbb{Z}}$ form a complete basis of $L^2(\mathrm{Leb})$ and hence of $L^2(\mu)$, as the spaces are equivalent. But each e_k can be written as a linear combination of p_j's:

$$\sum_{j=-|k|}^{|k|} (e_k^\dagger \mathcal{V}^\dagger e_j) p_j = \mathcal{V}^\dagger \mathcal{U} e_k = e_k.$$

Hence, $\{p_k\}_{k\in\mathbb{Z}}$ are a complete orthonormal basis of $\mathcal{L}^2(\mu)$. □

The following result arises from the diagonal structure of the Dirichlet kernel in the (orthogonal) Fourier basis:

Proposition 4.9. *For all $K \in \mathbb{N}$ and τ/σ decreasing, the Dirichlet kernel approximates the identity as*

$$\|I - \mathcal{D}_K\|_{W^\sigma \to W^\tau} = \sup_{|k| \le K} \frac{\tau}{\sigma} \le \frac{\tau(K)}{\sigma(K)}.$$

We can now prove the main theorem on polynomial approximation:

Proof of Theorem 1.2. From Proposition 3.3 we have that

$$I - \mathcal{P}_K = \mathcal{U}^{-1}(I - \mathcal{D}_K)\mathcal{U} = \mathcal{V}^\dagger (I - \mathcal{D}_K)\mathcal{U}$$

and so using Theorem 3.1b,

$$\|I - \mathcal{P}_K\|_{W^\sigma \to W^\tau} \le \|\mathcal{V}^\dagger\|_{W^\tau} \|I - \mathcal{D}_K\|_{W^\sigma \to W^\tau} \|U\|_{W^\tau} \le C_{\mathcal{P}} \|I - \mathcal{D}_K\|_{W^\sigma \to W^\tau},$$

where $C_{\mathcal{P}} = C_{\triangle}^2$. Combining this with Proposition 4.9 gives the required result. as required. □

5 Transfer operator results

Let $v = \hat{f}^{-1}$ be the inverse lift of f. We know f is w-to-one for some $w \ge 2$, so we expect v to be $2\pi w$-periodic.

Let $\pm_f = \operatorname{sign} f'(0)$, and define the operator $\mathcal{L}_\mu : \mathcal{H}^2_\zeta \circlearrowleft$ as follows:

$$(\mathcal{L}_\mu \varphi)(z) = \sum_{j=0}^{w-1} J_\mu(z + 2\pi j) \varphi(v(z + 2\pi j)), \tag{10}$$

where

$$J_\mu(z) = \pm_f \frac{v'(z) h(v(z))}{h(z)}. \tag{11}$$

Note that all $\mathcal{L}_\mu$ are conjugate to $\mathcal{L}_1$ as $\mathcal{L}_\mu \varphi = h^{-1}\mathcal{L}_1(h\varphi)$. We will prove a uniform bound on $\mathcal{L}_\mu$ in $\mathcal{H}^2$ spaces:

Theorem 5.1. *Suppose $|\log f'|_{C^\alpha(\mathbb{T}_\zeta)} \le D$. Then there exists $C_{\mathcal{H}^2}$ depending only on D, ζ, α such that for $t \in [0, \zeta]$ and $u > \gamma^{-1} t$,*

$$\|\mathcal{L}_\mu\|_{\mathcal{H}^2_u \to \mathcal{H}^2_t} \le C_{\mathcal{H}^2} M^2.$$

This tightens the bounds in the one-dimensional case of [3, Lemma 5.3].

As part of proving this, we will prove a standard uniform bound in $\mathcal{H}^\infty$ spaces à la [26], and a uniform bound in $\mathcal{H}^1$ spaces:

Proposition 5.2. *There exists C depending only on D, ζ, α such that for $t \in [0, \zeta]$ and $u \ge \gamma^{-1} t$,*

$$\|\mathcal{L}_\mu\|_{\mathcal{H}^\infty_u \to \mathcal{H}^\infty_t} \le C M^2.$$

Proposition 5.3. *There exists C depending only on D, ζ, α such that for $t \in [0, \zeta]$ and $u > \gamma^{-1} t$,*

$$\|\mathcal{L}_\mu\|_{\mathcal{H}^1_u \to \mathcal{H}^1_t} \le C M^2$$

The proofs of these three results are given in the Appendix.

The transfer operator $\mathcal{L}_\mu$ is in fact just the adjoint of the Koopman operator:

Proposition 5.4. *For all $\varphi, \psi \in L^2(\mu)$,*

$$\int_{\mathbb{T}} \overline{\mathcal{K}\psi}\, \varphi \, \mathrm{d}\mu = \int_{\mathbb{T}} \overline{\psi}\, \mathcal{L}_\mu \varphi \, \mathrm{d}\mu$$

Proof. Suppose $\psi, \varphi \in L^2$. Then

$$\int_{\mathbb{T}} \overline{\mathcal{K}\psi}\, \varphi \, \mathrm{d}\mu = \int_{\mathbb{T}} \overline{\psi} \circ f \, \varphi \, \mathrm{d}\mu.$$

With a w-to-one change of variables $f(x) = y$ we find

$$\int_{\mathbb{T}} \overline{\psi} \circ f \, \varphi \, \mathrm{d}\mu = \int_{\mathbb{T}} \sum_{j=0}^{w-1} \overline{\psi(y)} \varphi(v(y + 2\pi j)) \mu(v(y + 2\pi j)) |v'(y + 2\pi j)| \, \mathrm{d}y$$

from which, since $\operatorname{sign} v' = \pm f$, the required identity follows. $\square$

Proposition 5.5. *For all $t \in (0, \zeta]$ and $u \in (\gamma^{-1}t, \zeta]$, $\mathcal{K}$ extends to a bounded operator on $\mathcal{H}^2_{-t} \to \mathcal{H}^2_{-u}$, and for all $\varphi \in \mathcal{H}^2_u, \psi \in \mathcal{H}^2_{-t}$,*

$$\int_{\mathbb{T}} \overline{\mathcal{K}\psi}\, \varphi \, \mathrm{d}\mu = \int_{\mathbb{T}} \overline{\psi}\, \mathcal{L}_\mu \varphi \, \mathrm{d}\mu$$

Proof. Suppose $\psi, \varphi \in L^2(\mu) = L^2(\mathrm{Leb}, \mathbb{T})$. Then the same adjoint relationship is obeyed as in Proposition 5.4, and

$$\begin{aligned} \|\mathcal{K}\psi\|_{H^2_{-u}} &= \sup_{\|\chi\|_{H^2_u} = 1} \left| \int_{\mathbb{T}} \mathcal{K}\psi \, \chi \, \mathrm{d}x \right| \\ &= \sup_{\|\chi\|_{H^2_u} = 1} \left| \int_{\mathbb{T}} \psi \, \mathcal{L}_1 \chi \, \mathrm{d}x \right| \\ &\leq \|\mathcal{L}_1 \varphi\|_{\mathcal{H}^2_t \to \mathcal{H}^2_u} \|\psi\|_{\mathcal{H}^2_{-t}}. \end{aligned}$$

Now from Theorem 5.1, and recalling that the uniform distribution satisfies all the assumptions on μ,

$$\|\mathcal{L}_1 \varphi\|_{\mathcal{H}^2_t \to \mathcal{H}^2_u} \leq C_{\mathcal{H}^2} < \infty$$

so by completion, $\mathcal{K} : \mathcal{H}^2_{-t} \to \mathcal{H}^2_{-u}$ is bounded in norm. Since by considering the Fourier duality, $L^2(\mu) = L^2(\mathrm{Leb}/2\pi)$ is dense in $\mathcal{H}^2_{-t}$, the adjoint relation is preserved. $\square$

Proof of Theorem 2.1. We first show that $\mathcal{K} : \mathcal{H}^2_{-t} \circlearrowleft$ is compact. For all $u \in [0, t)$, inclusion $I : \mathcal{H}^2_t \to \mathcal{H}^2_u$ is compact since from Proposition 4.9, $\{\mathcal{D}_K\}_{K \in \mathbb{N}}$ are a family of finite-rank operators $\mathcal{H}^2_t \to \mathcal{H}^2_u$ with $\mathcal{D}_K \to I$ in operator norm. By duality, inclusion $I : \mathcal{H}^2_{-u} \to \mathcal{H}^2_{-t}$ is also compact. Then for any choice of $u \in (\gamma^{-1}t, t)$, $\mathcal{K} : \mathcal{H}^2_{-t} \to \mathcal{H}^2_{-t}$ is the composition of the bounded operator $\mathcal{K} : \mathcal{H}^2_{-u} \to \mathcal{H}^2_{-t}$ with this compact inclusion, giving compactness of $\mathcal{K}$ in $\mathcal{H}^2_{-t}$.

Similarly, from the above and Theorem 1.2 we know that $\mathcal{K}_K = \mathcal{P}_K \mathcal{K}$ are uniformly bounded $\mathcal{H}^2_{-s} \to \mathcal{H}^2_{-t}$ for any $s \in (u, t)$. Then since $I : \mathcal{H}^2_{-t} \to \mathcal{H}^2_{-s}$ is compact we can compose $\mathcal{K}_K = I \circ \mathcal{K}_K$ to see that they are uniformly compact.

We are given that $\mathcal{K}_K = \mathcal{P}_K \mathcal{K}$. From Theorem 1.2 and the fact that $\sigma_u(K)/\sigma_t(K) \leq 2e^{-K(t-u)}$, we have that

$$\|\mathcal{P}_K - I\|_{\mathcal{H}^2_t \to \mathcal{H}^2_u} \leq 2C_{\mathcal{P}} e^{-K(t-u)},$$

but we need to convert this into the dual norm. Because $\mathcal{P}_K$ is self-adjoint in $L^2(\mu)$, we have for $\psi, \chi \in L^2(\mathrm{Leb}, \mathbb{T})$ (i.e. in $L^2(\mu)$) that

$$\int \mathcal{P}_K \psi \, \chi \, \mathrm{d}x = \int \psi \, h \mathcal{P}_K (h^{-1} \chi) \, \mathrm{d}x$$

and so, by the same argument as in Proposition 5.5, P_K and $I - \mathcal{P}_K$ extend to bounded operators $\mathcal{H}^2_{-u} \to \mathcal{H}^2_{-t}$, with

$$\|I - \mathcal{P}_K\|_{\mathcal{H}^2_{-u} \to \mathcal{H}^2_{-t}} \leq \|h\|_{H^\infty_t} \|I - \mathcal{P}_K\|_{\mathcal{H}^2_t \to \mathcal{H}^2_u} \|h\|_{\mathcal{H}^\infty_u} \leq M^2 \cdot 2C_{\mathcal{P}} e^{-K(t-u)}$$

Consequently, using Proposition 5.5 and Theorem 5.1 we obtain that for all $u > \gamma^{-1}t$,

$$\|\mathcal{K} - \mathcal{K}_K\|_{\mathcal{H}^2_{-t}} \leq \|I - \mathcal{P}_K\|_{\mathcal{H}^2_{-u} \to \mathcal{H}^2_{-t}} \|\mathcal{K}\|_{\mathcal{H}^2_{-t} \to \mathcal{H}^2_{-u}} \tag{12}$$

$$\leq 2M^2 C_{\mathcal{H}^2} C_{\mathcal{P}} e^{-K(t-u)}. \tag{13}$$

Taking the infimum over u we get the required result. $\square$

Proof of Corollary 2.2. The dual of $\mathcal{K} : \mathcal{H}^2_{-t} \circlearrowleft$ is $\mathcal{L}_\mu : \mathcal{H}^2_t \circlearrowleft$. Because the $\mathcal{L}_\mu$ are compact in this space, we have

$$\int \varphi\, \psi \circ f^n \,\mathrm{d}\mu = \int \mathcal{L}_\mu^n \varphi\, \psi \,\mathrm{d}\mu = \sum_{j=1}^{J} \int_{\mathbb{T}} \mathcal{L}_\mu^n \Pi_j \varphi\, \psi \,\mathrm{d}\mu + o(\lambda_J^n)$$

as $n \to \infty$, where Π_j are the spectral projections onto the λ_j-generalised eigenspace, and therefore each $\int_{\mathbb{T}} \mathcal{L}_\mu^n \Pi_j \varphi\, \psi \,\mathrm{d}\mu = \mathcal{O}(n^{m_j} \lambda_j^n)$ where m_j is the multiplicity of λ_j. Note that since $\sigma(\mathcal{L}_\mu) = \sigma(\mathcal{L}_1)$, these are the same eigenvalues we would get if we were considering correlations against Lebesgue measure or the physical measure of f, so they are the Ruelle-Pollicott resonances in the usual sense.

The convergence result follows from [23, Theorems 1,6] applied to $\mathcal{L}_\mu$ and $(\mathcal{L}_\mu)^*$, noting that it doesn't matter which value of t we use to estimate the eigenvalues, and ζ gives the optimal bound. $\square$

Proof of Theorem 1.1. $\{\mathcal{K}_L\}_{L \in \mathbb{N}}$ are rank-$2L-1$ approximations of $\mathcal{K}$, and using (12),

$$\|\mathcal{K}_L - \mathcal{K}\|_{\mathcal{H}^2_{-t}} \leq 4M^2 C_{\mathcal{H}^2} C_{\mathcal{P}} e^{-L(t-u)}.$$

Similarly, $\{\mathcal{K}_L\}_{L \leq K}$ are rank-$2L-1$ approximations of $\mathcal{K}_K$, and using (12),

$$\|\mathcal{K}_L - \mathcal{K}_K\|_{\mathcal{H}^2_{-t}} \leq \|\mathcal{K}_L - \mathcal{K}\|_{\mathcal{H}^2_{-t}} + \|\mathcal{K}_K - \mathcal{K}\|_{\mathcal{H}^2_{-t}} \leq \|\mathcal{K} \leq 4M^2 C_{\mathcal{H}^2} C_{\mathcal{P}} e^{-L(t-u)}$$

for $u \in [\gamma^{-1}t, t]$, so taking infima,

$$\|\mathcal{K}_L - \mathcal{K}_K\|_{\mathcal{H}^2_{-t}}, \|\mathcal{K}_L - \mathcal{K}\|_{\mathcal{H}^2_{-t}} \leq 4M^2 C_{\mathcal{H}^2} C_{\mathcal{P}} e^{-L(1-\gamma^{-1})t}.$$

This means the $2L-1$th and $2L$th approximation numbers (and therefore corresponding singular values) of $\mathcal{K}_K, \mathcal{K}$ are bounded by the above constant. By [2, Corollary 5.3] we have that the Hausdorff distance between $\mathcal{K}_K, \mathcal{K}$ are bounded by $Ce^{-c\sqrt{K}}$ for some $C, c > 0$.

To prove the rest of the theorem, we apply Corollary 2.2 when the Π_j are rank-one: i.e. we can decompose $\Pi_j = b_j \beta_j$ with $\beta_j b_j = 1$, and therefore $\mathcal{L}_\mu^n b_j = \lambda_j^n b_j$. Note that the constant $h = 1 - \gamma^{-1}$. $\square$

A Proofs of transfer operator bounds

In this section, we prove various bounds on the transfer operator, specifically in $\mathcal{H}^p_u \to \mathcal{H}^p_t$ for $p \in \{1, 2, \infty\}$. We will actually prove results for an extension of the transfer operator in these spaces to more general spaces of harmonic functions. For $t > 0$ and $p \in [1, \infty]$ us define the spaces

$$\mathcal{A}^p_t = \{\varphi : \overline{\mathbb{T}_\zeta} \to \mathbb{C} : (\tfrac{\partial^2}{\partial x^2} + \tfrac{\partial^2}{\partial y^2})\varphi(x+iy) = 0; \|\varphi|_{\partial \mathbb{T}_\zeta}\|_{L^p} < \infty\}$$

with the norm

$$\|\varphi\|_{\mathcal{A}^p_t} = \|\varphi|_{\partial \mathbb{T}_\zeta}\|_{L^p} = \begin{cases} \frac{1}{4\pi} \int_{\mathbb{T}} |\varphi(x+it)|^p + |\varphi(x-it)|^p \,\mathrm{d}x, & p < \infty \\ \sup_{z \in \partial \mathbb{T}_\zeta} |\varphi(z)|. \end{cases}$$

Harmonicity means that a function in $\mathcal{A}_t^p$ is uniquely determined by its values on the boundary: in particular from any function on $\overline{\mathbb{T}_\zeta}$ with an L^p restriction to $\partial\mathbb{T}_\zeta$, we can construct a harmonic function $\mathcal{I}_t\varphi \in \mathcal{A}_t^p$ that matches φ on the boundary using a kernel operator

$$\mathcal{I}_t\varphi(x+iy) = \int_{\mathbb{T}} (k_t(\theta + it - z)\varphi(\theta+it) + k_t(\theta+it-z)\varphi(\theta - it))\,\mathrm{d}\theta \tag{14}$$

where

$$k_t(x+iy) = \sum_{n\in\mathbb{Z}} \frac{\pi}{4t} \frac{2\cot\frac{\pi y}{4t}\operatorname{sech}^2\frac{\pi(x+2n)}{4t}}{\cot^2\frac{\pi y}{4t} + \tanh^2\frac{\pi(x+2n)}{4t}} \geq 0. \tag{15}$$

Note that functions in $\mathcal{A}_t^p$ are mapped to themselves under $\mathcal{I}_t$.

Since all holomorphic functions are harmonic, the Hardy space $\mathcal{H}_t^p$ is a closed subset of $\mathcal{A}_t^p$ with the same norm. We can therefore study $\mathcal{L}_\mu : \mathcal{H}_u^p \to \mathcal{H}_t^p$ by the following (not entirely natural) extension to general harmonic functions:

$$\tilde{\mathcal{L}}_{t,\mu}\varphi = \mathcal{I}_t\mathcal{L}_\mu\varphi = \mathcal{I}_t\left[\sum_{j=0}^{w-1} J_\mu(\cdot + 2\pi j)\varphi(v(\cdot + 2\pi j))\right],$$

and it is this operator we will prove is bounded on $\mathcal{A}_t^p \to \mathcal{A}_u^p$ for the different p. Note that for any $\varphi \in \mathcal{A}_t^p$,

$$\|\tilde{\mathcal{L}}_{t,\mu}\varphi\|_{\mathcal{A}_t^p} = \|(\mathcal{I}_t\mathcal{L}_\mu\varphi)|_{\partial\mathbb{T}_\zeta}\|_{L^p} = \|(\mathcal{L}_\mu\varphi)|_{\partial\mathbb{T}_\zeta}\|_{L^p}, \tag{16}$$

even though $\mathcal{L}\mu_\varphi$ may not even be harmonic, simply because $\mathcal{I}_t$ matches functions on the boundary.

Proof of Proposition 5.2. From (16) we have for $\varphi \in \mathcal{A}_u^p$ that

$$\begin{aligned}\|\tilde{\mathcal{L}}_{t,\mu}\varphi\|_{\mathcal{A}_t^\infty} &= \sup_{z\in\mathbb{T}_t} |(\mathcal{L}_\mu\varphi)(z)| \\ &= \sup_{z\in\mathbb{T}_t} \sum_{j=0}^{w-1} J_\mu(z+2\pi j)\varphi(v(z+2\pi j).\end{aligned}$$

Now, from its definition in (11), $|J_\mu(z+2\pi j)| \leq M^2|v'(z+2\pi j)|$. Our Hölder distortion assumption implies that

$$\begin{aligned}|v'(z+2\pi j)| &= \frac{1}{2\pi}\int_{\Re z-\pi}^{\Re z+\pi} |v'(z+2\pi j)|\,\mathrm{d}x \\ &\leq \frac{1}{2\pi}\int_{\Re z-\pi}^{\Re z+\pi} e^{D|z-x|^\alpha}|v'(x+2\pi j)|\,\mathrm{d}x \\ &\leq e^{D(\pi^2+\zeta^2)^{\alpha/2}} \frac{\operatorname{Leb} v([\Re z - \pi + 2\pi j, \Re z + \pi + 2\pi j])}{2\pi}\end{aligned}$$

so

$$\sum_{j=0}^{w-1} |J_\mu(z+2\pi j)| \leq e^{D(\pi^2+\zeta^2)^{\alpha/2}}.$$

Furthermore, $\Im v(z+2\pi j) \leq |v(z+2\pi j) - v(\Re z + 2\pi j)| \leq \gamma \Im z \leq \gamma t$, so $v(z+2\pi j) \in \mathbb{T}_{\gamma^{-1}t} \subset \mathbb{T}_u$, so by the maximum principle

$$\|\tilde{\mathcal{L}}_{t,\mu}\varphi\|_{\mathcal{A}_t^\infty} \leq e^{D(\pi+\zeta)^\alpha} 2\pi M^2 \sup_{z\in\mathbb{T}_u} |\varphi(z)| \leq e^{D(\pi^2+\zeta^2)^{\alpha/2}} 2\pi M^2 \|\varphi\|_{\mathcal{A}_u^\infty}$$

as required. $\square$

Proof of Proposition 5.3. Note that we only need consider u smaller than ζ: larger u follows by inclusion.

Functions in $\mathcal{A}^1_u$ are harmonic and so obey

$$\varphi(z) = (\mathcal{I}_t\varphi)(z) = \int_{\mathbb{T}} (k_u(\theta + iu - z)\varphi(\theta + iu) + k_u(\theta + iu - z)\varphi(\theta - iu))\,\mathrm{d}\theta$$

for k_u given by (15).

Using (16), we have that

$$\|\tilde{\mathcal{L}}_{t,1}\varphi\|_{\mathcal{A}^1_t} \leq \sum_{\pm_a,\pm_b} \sum_{j=0}^{w-1} \int_{\mathbb{T}}\int_{\mathbb{T}} |J'_\mu(x + 2\pi j \pm_b it)| k_u(\theta \pm_a iu - v(x \pm_b + 2\pi j + it)) |\varphi(\theta + iu)|\,\mathrm{d}\theta\,\mathrm{d}\omega$$

and so by Fubini's theorem,

$$\begin{aligned}
\|\mathcal{L}_\mu\varphi\|_{\mathcal{A}^1_t} &\leq \sup_{u\in\mathbb{T},\pm_a} \sum_{\pm_b}\sum_{j=0}^{w-1} \int_{\mathbb{T}} |J'_\mu(x + 2\pi j \pm_b it)| k_u(\theta \pm_a iu - v(x \pm_b + 2\pi j + it))\,\mathrm{d}\omega \\
&\leq M^2 \sup_{\theta\in\mathbb{T}} \sum_{j=0}^{w-1}\sum_{\pm} \int_{\mathbb{T}} |v'(x + 2\pi j + it)| k_u(\theta \pm iu - v(x + 2\pi j + it))\,\mathrm{d}\omega \\
&\leq M^2 \sup_{\theta\in\mathbb{T}} \sum_{\pm} \int_{v(w\mathbb{T}+it)} k_u(\theta \pm iu - z)\,|\mathrm{d}z| \qquad (17)
\end{aligned}$$

where we used the symmetry of v under conjugation and bounds on h, followed by a change of variables.

On $\mathbb{T}_u$ we can bound

$$k_u(x + iy) \leq C_u + \frac{2y}{x^2 + y^2}. \qquad (18)$$

for some C_u increasing in u. Note that k_u blows up near 0 and is not integrable along certain curves (e.g. $y = |x|$, which maximises k_u for fixed x), but has constant integral along lines of constant y. We will show that when our curve $v(w\mathbb{T} + it)$ is close enough to $x + iy = 0$, it must be locally close to a line of constant y.

If $v' > 0$ on $\mathbb{T}$, then $|\arg v'(x+it)| = \Im \log v'(x+it) \leq D\zeta^\alpha \leq \frac{\pi}{3}$, so $|v'(x+it)| \leq 2|\Re v'(x+it)|$, we have that $v(w\mathbb{T} + it)$ on the complex plane identified with $\mathbb{R}^2$ can be written as a graph of a function $V : \mathbb{T} \to [-\gamma^{-1}t, \gamma^{-1}t]$ with $|V'| \leq \sqrt{3}$. A similar thing holds when $v' < 0$ on $\mathbb{T}$. This means that

$$\int_{v(w\mathbb{T}+it)} k_u(\theta \pm iu - z) = \int_{\mathbb{T}} k_u(\theta \pm iu - (x + iV(x)))\sqrt{1 + V(x)^2}\,\mathrm{d}x \leq 2\int_{\mathbb{T}} k_u(\theta \pm iu - (x + iV(x)))\,\mathrm{d}x \qquad (19)$$

Let us consider separately two parts of our curve: the set $X = \{x \in [\theta - 2u, \theta + 2u] : |V(x)| > u/2\}$, and the remainder $\mathbb{T}\backslash X$. Using (18) on $\mathbb{T}\backslash X$, we can bound

$$k_u(\theta \pm iu - (x + iV(x))) \leq C_u + \begin{cases} \frac{u}{(x-\theta)^2 + (u/2)^2}, & |x - \theta| < u/2 \\ \frac{1}{x-\theta}, & |x - \theta| \in [u/2, 2u] \\ \frac{4u}{(x-\theta)^2 + 4u^2}, & |x - \theta| > 2u. \end{cases}$$

This means that the $\mathbb{T}\backslash X$ contribution to (19) gives

$$\int_{\mathbb{T}\backslash X} k_u(\theta \pm iu - (x + iV(x)))\,\mathrm{d}x \leq |\mathbb{T}\backslash X| C_u + \pi + \log 4.$$

On the other hand for $x \in X$,

$$u \leq \Im v(x + it) \leq \int_0^t |v'(x + i\tau)|\,\mathrm{d}\tau \leq t e^{D\zeta^\alpha} |v'(x + it)|$$

by Hölder distortion on $\mathbb{T}_\zeta$, so for $x \in X$,

$$|v'(x+it)|^{-1} \leq 2te^{D\zeta^\alpha}/u \leq 2\zeta e^{D\zeta^\alpha}/u.$$

As a result, the direction of the curve $v(wX+it)$ varies as an α-Hölder function of arc-length with constant $D(2\gamma e^{D\zeta^\alpha})^\alpha$, and so $V' = \tan\arg v'$ has an α-Hölder constant $4D(2\gamma e^{D\zeta^\alpha}/u)^\alpha =: K^\alpha u^{-\alpha}$.

Now let us consider the curve $x + iV(x)$ for x close to θ. Since V must lie in $[-u,u]$, it must reach a maximum (resp. minimum) in $[-u,u]$, and so

$$|V'(\theta)| \leq K' \left(\tfrac{u-V(\theta)}{u}\right)^{\alpha/(\alpha+1)} \tag{20}$$

with $K' = K\frac{\alpha}{\alpha+1}$. Taylor approximation gives us that

$$|\tfrac{V(x)-V(\theta)}{u} - V'(\theta)\tfrac{x-\theta}{u}| \leq K''|\tfrac{x-\theta}{u}|^{1+\alpha}. \tag{21}$$

with $K'' = \frac{1}{1+\alpha}K^\alpha$. Combining the previous two equations we get that

$$K'\left(\tfrac{u-V(\theta)}{u}\right)^{\frac{\alpha}{\alpha+1}}(x-\theta) - K''|\tfrac{x-\theta}{u}|^{1+\alpha} \leq \tfrac{V(x)-V(\theta)}{u} \leq K'\left(\tfrac{u-V(\theta)}{u}\right)^{\frac{\alpha}{\alpha+1}}(x-\theta) + K''|\tfrac{x-\theta}{u}|^{1+\alpha} \tag{22}$$

Noting the scaling of this equation with u, the fact that the length of X must be less than $2u$, and the bound for k_u in (18), we find there exists a uniform bound

$$\int_{\mathbb{T}\setminus X} k_u(\theta \pm iu - (x+iV(x)))\,\mathrm{d}x \leq |X|C_u + C'$$

for C' independent of u. Substituting this and (A) into (19) we get

$$\int_{v(w\mathbb{T}+it)} k_u(\theta \pm iu - z) \leq 2\pi C_u + \pi + \log 4 + C' \leq 2\pi C_\zeta + \pi + \log 4 = C''.$$

Substituting this into (17) gives the required uniform bound. □

We will need the following result to relate different $\mathcal{A}^p$ spaces with each other, allowing us to prove Theorem 5.1:

Lemma A.1. *For $t \in [0,\zeta]$, $u > \gamma^{-1}t$, and $p, p' \in [1,\infty]$, $\tilde{\mathcal{L}}_{t,\mu} : \mathcal{A}^p_u \to \mathcal{A}^{p'}_t$ is bounded.*

Proof. Proposition 5.2 tells us that $\tilde{\mathcal{L}}_{t,\mu}$ is bounded $\mathcal{A}^\infty_{\gamma^{-1}t} \to \mathcal{A}^\infty_t$. We then need to resolve the integrability parameters.

The boundedness of the kernel $k_u(x+iy)$ for $y > 0$ via (18) means that, by the definition of (14), $\mathcal{I}_u$ is bounded $\mathcal{A}^p_u \to \mathcal{A}^\infty_{\gamma^{-1}t}$, and therefore so is function inclusion. On the other hand, L^p inclusions on the boundary gives a bounded inclusion from $\mathcal{A}^{p'}_t \to \mathcal{A}^\infty_t$. □

Proof of Theorem 5.1. The spaces $\mathcal{A}^2_t$ are Hilbert spaces with the $L^2(\partial\mathbb{T}_\zeta)$ inner product. As a result, for any $t \in [0,\zeta]$, $u > \gamma^{-1}t$ there exists an operator $\mathcal{J}^{u,t}_\mu : \mathcal{A}^2_t \to \mathcal{A}^2_u$ adjoint to $\tilde{\mathcal{L}}_{t,\mu} : \mathcal{A}^2_u \to \mathcal{A}^2_t$. We can then say that

$$\|\tilde{\mathcal{L}}_{t,\mu}\|^2_{\mathcal{A}^2_u \to \mathcal{A}^2_t} = \|\tilde{\mathcal{L}}_{t,\mu}\mathcal{J}^{u,t}_\mu\|_{\mathcal{A}^2_u},$$

and in fact by self-adjointness of $\tilde{\mathcal{L}}_{t,\mu}\mathcal{J}^{u,t}_\mu$ that

$$\|\tilde{\mathcal{L}}_{t,\mu}\|^2_{\mathcal{A}^2_u \to \mathcal{A}^2_t} = \sqrt[n]{\|(\tilde{\mathcal{L}}_{t,\mu}\mathcal{J}^{u,t}_\mu)^n\|_{\mathcal{A}^2_u}}.$$

Now,

$$\|(\tilde{\mathcal{L}}_{t,\mu}\mathcal{J}^{u,t}_\mu)^n\|_{\mathcal{A}^2_u} \leq \|\tilde{\mathcal{L}}_{t,\mu}\|_{\mathcal{A}^1_t \to \mathcal{A}^2_u}\|(\mathcal{J}^{u,t}_\mu\tilde{\mathcal{L}}_{t,\mu})\|^{n-1}_{\mathcal{A}^1_t}\|\mathcal{J}^{u,t}_\mu\|_{\mathcal{A}^2_u \to \mathcal{A}^1_t}.$$

Since $\mathcal{A}_t^2$ has a bounded inclusion into $\mathcal{A}_t^1$, $\mathcal{J}_\mu^{u,t}$ is bounded $\mathcal{A}_u^2 \to \mathcal{A}_t^1$; by the previous lemma, $\tilde{\mathcal{L}}_{t,\mu}$ is bounded $\mathcal{A}_t^1 \to \mathcal{A}_u^2$. All the operators in the above expression are therefore bounded, and in particular for some C,

$$\|\tilde{\mathcal{L}}_{t,\mu}\|^2_{\mathcal{A}_u^2 \to \mathcal{A}_t^2} \leq \inf_{n \geq 1} \sqrt[n]{C \|(\mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu})\|^{n-1}_{\mathcal{A}_t^1}} = \|\mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu}\|_{\mathcal{A}_t^1}. \tag{23}$$

Suppose then that $\varphi \in \mathcal{A}_t^2$, and so $\mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu} \varphi$ is as well. Defining $\omega = \mathcal{I}_t \left[\frac{\mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu}\varphi}{|\mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu}\varphi|} \right] \in \mathcal{A}_\mu^\infty$ we have that almost everywhere on the boundary of $\mathbb{T}_t$,

$$\bar{\omega} \mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu} \varphi = |\mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu} \varphi|.$$

We then have

$$\begin{aligned}
\|\mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu}\varphi\|_{\mathcal{A}_t^1} &= \frac{1}{4\pi} \int_{\mathbb{T}} (\bar{\omega} \mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu}\varphi)(\theta + i\zeta) + (\bar{\omega} \mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu}\varphi)(\theta - i\zeta) \,\mathrm{d}\theta \\
&= \langle \omega, \mathcal{J}_\mu^{u,t} \tilde{\mathcal{L}}_{t,\mu}\varphi \rangle_{\mathcal{A}_t^2} = \langle \tilde{\mathcal{L}}_{t,\mu}\omega, \tilde{\mathcal{L}}_{t,\mu}\varphi \rangle_{\mathcal{A}_t^2} \\
&= \frac{1}{4\pi} \int_{\mathbb{T}} (\mathcal{L}_\mu \omega \, \tilde{\mathcal{L}}_{t,\mu}\varphi)(\theta + i\zeta) + (\mathcal{L}_\mu \omega \, \tilde{\mathcal{L}}_{t,\mu}\varphi)(\theta - i\zeta) \,\mathrm{d}\theta \\
&\leq \|\mathcal{L}_\mu\|_{\mathcal{A}_t^\infty} \frac{1}{4\pi} \int_{\mathbb{T}} (\tilde{\mathcal{L}}_{t,\mu}\varphi)(\theta + i\zeta) + (\tilde{\mathcal{L}}_{t,\mu}\varphi)(\theta - i\zeta) \,\mathrm{d}\theta \\
&\leq \|\mathcal{L}_\mu\|_{\mathcal{A}_t^\infty} \|\mathcal{L}_\mu\|_{\mathcal{A}_t^1} \|\varphi\|_{\mathcal{A}_t^1}
\end{aligned}$$

Applying Propositions 5.2 and 5.3 for $\mathcal{A}$ spaces, and substituting into (23), gives us what we want when $\varphi \in \mathcal{A}_t^2$. Since $\mathcal{A}_t^2$ is dense in $\mathcal{A}_t^1$, we obtain the full result by interpolation.

We can of course then go back and restrict to looking at $\mathcal{L}_\mu$ on $\mathcal{H}_u^2 \to \mathcal{H}_t^2$. □

Acknowledgements

The author thanks Matthew Colbrook for inspiring discussion and his comments on the manuscript.

This research has been supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 787304).

References

[1] Viviane Baladi. *Positive transfer operators and decay of correlations*, volume 16. World scientific, 2000.

[2] Oscar F Bandtlow and Ayşe Güven. Explicit upper bounds for the spectral distance of two trace class operators. *Linear Algebra and its Applications*, 466:329–342, 2015.

[3] Oscar F Bandtlow and Oliver Jenkinson. Explicit eigenvalue estimates for transfer operators acting on spaces of holomorphic functions. *Advances in Mathematics*, 218(3):902–925, 2008.

[4] Oscar F Bandtlow and Julia Slipantschuk. Lagrange approximation of transfer operators associated with holomorphic data. *arXiv preprint arXiv:2004.03534*, 2020.

[5] Alex Blumenthal, Jinxin Xue, and Lai-Sang Young. Lyapunov exponents for random perturbations of some area-preserving maps including the standard map. *Annals of Mathematics*, 185(1):285–310, 2017.

[6] William A Brock. Nonlinearity and complex dynamics in economics and finance. In *The economy as an evolving complex system*, pages 77–97. CRC Press, 2018.

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

[8] Charles K Chui, J D Ward, and P W Smith. Cholesky factorization of positive definite bi-infinite matrices. *Numerical Functional Analysis and Optimization*, 5(1):1–20, 1982.

[9] Matthew J Colbrook. The mpedmd algorithm for data-driven computations of measure-preserving dynamical systems. *arXiv preprint arXiv:2209.02244*, 2022.

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

[11] Matthew J Colbrook and Alex Townsend. Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems. *arXiv preprint arXiv:2111.14889*, 2021.

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

[13] Michael Dellnitz and Oliver Junge. *Set Oriented Numerical Methods for Dynamical Systems*, page 221. Gulf Professional Publishing, 2002.

[14] Gary Froyland and Michael Dellnitz. Detecting and locating near-optimal almost-invariant sets and cycles. *SIAM Journal on Scientific Computing*, 24(6):1839–1863, 2003.

[15] 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, 2014.

[16] Armin Fuchs. *Nonlinear dynamics in complex systems*. Springer, 2014.

[17] Jeffrey S Geronimo and Andrei Martínez-Finkelshtein. On extensions of a theorem of Baxter. *Journal of Approximation Theory*, 139(1-2):214–222, 2006.

[18] Hubert Hennion and Loïc Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*. Springer, 2001.

[19] 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.

[20] Stefan Klus and Christof Schütte. Towards tensor-based methods for the numerical approximation of the Perron–Frobenius and Koopman operator. *Journal of Computational Dynamics*, 3(2):139–161, 2016.

[21] HL Krall. On derivatives of orthogonal polynomials. *Bulletin of the American Mathematical Society*, 42(6):423–428, 1936.

[22] Alexandre Mauroy, Y Susuki, and I Mezić. *Koopman operator in systems and control.* Springer, 2020.

[23] John E Osborn. Spectral approximation for compact operators. *Mathematics of computation*, 29(131):712–725, 1975.

[24] David Ruelle. Structure and f-dependence of the a.c.i.m. for a unimodal map f of Misiurewicz type. *Communications in Mathematical Physics*, 287(3):1039–1070, 2009.

[25] Barry Simon. *Orthogonal polynomials on the unit circle.* American Mathematical Soc., 2005.

[26] Julia Slipantschuk, Oscar F Bandtlow, and Wolfram Just. Dynamic mode decomposition for analytic maps. *Communications in Nonlinear Science and Numerical Simulation*, 84:105179, 2020.

[27] Sang Hwan Son, Hyun-Kyu Choi, Jiyoung Moon, and Joseph Sang-Il Kwon. Hybrid Koopman model predictive control of nonlinear systems using multiple EDMD models: An application to a batch pulp digester with feed fluctuation. *Control Engineering Practice*, 118:104956, 2022.

[28] 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*, 25:1307–1346, 2015.

[29] Caroline Wormell. Spectral Galerkin methods for transfer operators in uniformly expanding dynamics. *Numerische Mathematik*, 142(2):421–463, 2019.