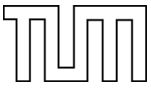

Technische Universität München

School of Computation, Information and Technology

Master's Thesis

Residual-based Methods for Robust Spectral Analysis in Dynamical Systems

April Herwig

april.herwig@tum.de

Supervisor: Prof Oliver Junge

Submission Date:

I assure the single handed composition of this master's thesis only supported by declared resources.

Garching,

Abstract

Koopmania

Contents

1 Introduction

Dynamical systems theory has seen many revolutions. One recent such revolution is to use operator theoretic tools to analyze the global behavior of chaotic systems. Concieved in the 1930's by Koopman and Von Neumann [4], the study of global dynamics under a mapping S is conducted by studying the action of the composition $\psi \circ S$ of S with an observable ψ. Observables which act as eigenvalues for the composition operator and its dual hold information on slowly decaying structures in phase space.

Currently the most popular method by far for analyzing composition operators (named Koopman operators) is dynamic mode decomposition (DMD). The past 2 decades have seen massive growth in this topic, and DMD has recieved many evolutions [1]. One primary concern of nearly all DMD methods is *spectral pollution*. The Koopman operator often has unfavorable spectral qualities such as continuous spectrum which are unstable, and are destroyed by finite approximation. The task presented in the present paper is to identify which candidate eigenvalues (computed e.g. by DMD) are approximations of eigenvalues of the true Koopman operator (or its adjoint), and which arise due to discretization.

2 Background

We consider a discrete dynamical system generated by a map $S : \Omega \to \Omega$ defined on a measure space $(\Omega, \mathcal{A}, dx)$ which is *nonsingular*, that is $\int_{S^{-1}(A)} dx = 0$ for any $A \in \mathcal{A}$ with $\int_A dx = 0$.

2.1 Koopman and Perron-Frobenius Operators

Koopman operator theory shifts the focus from dynamics of points in state space $\mathcal{X}$ to dynamics of observables $g : \Omega \to \mathbb{C}$. A state $x \in \Omega$ evolves by iteratively applying the map S, and similarly observables evolve under the action of the *Koopman operator*

$$\mathcal{K}g = g \circ S. \tag{2.1}$$

The primary benefit of this alternative viewpoint of dynamics is that $\mathcal{K}$ is *linear* and can therefore be analyzed using algebraic and functional analytic tools. However, the Koopman operator acts on function spaces which are typically infinite-dimensional. Even the Krylov space $\{g, \mathcal{K}g, \mathcal{K}^2 g, \ldots\}$ is generically infinite-dimensional: consider e.g. an indicator function $g = \mathbb{1}_{[0,1]}$ and a translation $S(x) = x - 1$.

It should be noted that until now we have not declared a function space to act as a domain for $\mathcal{K}$. This is because there are many potential domains for which $\mathcal{K}$ is not only well-defined, but has interesting properties worth studying. For the current section we consider $\mathcal{K} : L^\infty \to L^\infty$. On this space $\mathcal{K}$ is obviously bounded, in fact it is a contraction.

The Koopman operator tracks how observables evolve with S. There is however a dual viewpoint: that of *densities* and *pushforwards*. From the duality pairing

$$\langle f,\ \mathcal{K}g\rangle = \langle \mathcal{L}f,\ g\rangle \tag{2.2}$$

for $f \in L^1,\ g \in L^\infty$ we may deduce the form of an adjoint operator, known as the *transfer operator* or *Perron-Frobenius operator*. Consider an indicator $g = \mathbb{1}_A$ for an $A \in \mathcal{A}$ and let $f \geq 0$. The left hand side of 2.2 is

$$\int f(x)\, \mathbb{1}_A(S(x))\ dx = \int f(x)\, \mathbb{1}_{S^{-1}(A)}(x)\ dx = \int_{S^{-1}(A)} f(x)\ dx, \tag{2.3}$$

and the right hand side

$$\int \mathcal{L}f(x)\, \mathbb{1}_A(x)\ dx = \int_A \mathcal{L}f(x)\ dx. \tag{2.4}$$

Hence $\mathcal{L}f \in L^1$ should satisfy the equation

$$\int_A \mathcal{L}f(x)\ dx = \int_{S^{-1}(A)} f(x)\ dx. \tag{2.5}$$

A short exercise in measure theory shows that

$$A \mapsto \int_{S^{-1}(A)} f(x)\ dx \tag{2.6}$$

defines a finite absolutely continuous measure, and hence by the Radon-Nikodym theorem the measure has a unique density which by 2.5 is precisely $\mathcal{L}f$. For general $f \in L^1$ decompose $f = f^+ - f^-$ with $f^+, f^- \geq 0$ and set $\mathcal{L}f = \mathcal{L}f^+ - \mathcal{L}f^-$.

2.6 shows that $\mathcal{L}$ is precisely the action of the pushforward $S_\sharp\mu = \mu \circ S^{-1}$ for densities. This provides a useful intuition for the transfer operator: we move away from the viewpoint of single points $x \in \mathcal{X}$, and instead consider *distributions* of point density. If x is chosen from a random distribution with density f, then the distribution of $S(x)$ has density $\mathcal{L}f$.

2.2 Spectral Properties

2.2.1 The Spectrum

We begin with abstract definitions for parts of the spectrum.

Definition 2.1. *Let $M : D(M) \to \mathcal{X}$ be a linear operator on a Banach space $(X, \|\cdot\|)$. The domain $D(M) \subset \mathcal{X}$ of M may be a dense subset of $\mathcal{X}$. M is called* closed *if: whenever $x_k \to x$ is a convergent sequence with Mx_k convergent, then $Mx_k \to Mx$.*

Closedness is weaker than continuity due to the extra assumption that Mx_k is convergent.

Definition 2.2. *The* spectrum *$\sigma(M)$ of a closed operator M is the set of numbers $\lambda \in \mathbb{C}$ for which $M - \lambda I$ does not have a bounded inverse. The complement $\rho(M) := \mathbb{C} \setminus \sigma(M)$ is called* resolvent set *and $(M - \lambda I)^{-1}$ is called the* resolvent*.*

Proposition 2.3. *The spectrum admits a decomposition $\sigma(M) = \sigma_p(M) \oplus \sigma_c(M) \oplus \sigma_r(M)$ into* point-, continuous-, *and* residual-spectrum *with*

$$\sigma_p(M) = \{\lambda \in \mathbb{C} \mid M - \lambda I \text{ is not injective}\}, \tag{2.7}$$

$$\sigma_c(M) = \{\lambda \in \mathbb{C} \mid M - \lambda I \text{ is injective, but its range is a dense subset of } \mathcal{X}\}, \tag{2.8}$$

$$\sigma_r(M) = \{\lambda \in \mathbb{C} \mid M - \lambda I \text{ is injective, but does not have dense range}\}. \tag{2.9}$$

Remark. In some literature, 0 is never part of the spectrum. We do not take this approach.

An element $\lambda \in \sigma_p(M)$ is a *true* eigenvalue in the sense that there exists a vector $x \in \mathcal{X}$ such that $(M - \lambda I)x = 0$. The continuous spectrum has a similar characterization as the set of $\lambda \in \mathbb{C}$ for which $M_\lambda I$ is injective but *not* bounded from below, i.e. there exists a sequence $(x_k)_k$, $\|x_k\| = 1$, for which $(M - \lambda I)x_k \to 0$. Equivalently, $(M - \lambda I)^{-1}$ cannot be extended to a bounded linear operator, but is still a (densely defined) closed linear operator.

One might ask why such detail is required in defining the spectrum. The answer lies in the subtle consequences of infinite-dimensionality. The following example from [3] shows the subtle interplay between residual spectrum and spectrum of the adjoint operator, all of which requires infinite-dimensionality.

Example 2.4 (Shift operators on $\ell^2(\mathbb{N})$). The canonical right-shift operator

$$R : (a_1, a_2, \ldots) \mapsto (0, a_1, a_2, \ldots) \tag{2.10}$$

is adjoint to the left shift

$$L : (a_1, a_2, \ldots) \mapsto (a_2, a_3, \ldots). \tag{2.11}$$

We claim:

$$\sigma_p(L) = \{\lambda \in \mathbb{C} \mid |\lambda| < 1\} \qquad \sigma_p(R) = \emptyset \tag{2.12}$$
$$\sigma_c(L) = \{\lambda \in \mathbb{C} \mid |\lambda| = 1\} \qquad \sigma_c(R) = \{\lambda \in \mathbb{C} \mid |\lambda| = 1\} \tag{2.13}$$
$$\sigma_r(L) = \emptyset \qquad \sigma_r(R) = \{\lambda \in \mathbb{C} \mid |\lambda| < 1\}. \tag{2.14}$$

Indeed, to see 2.12 consider $|\lambda| < 1$, $a^\lambda = (\lambda, \lambda^2, \ldots)$. Then clearly $La^\lambda = \lambda a^\lambda$. On the other hand, suppose $Ra = \lambda a$ for some $a \neq 0$, $\lambda \in \mathbb{C}$. Then letting n be the first index such that $a_n \neq 0$, we must have $0 = a_{n-1} = (Ra)_n = \lambda a_n$ so $\lambda = 0$. But the kernel of R is clearly $\{0\}$.

To see 2.13, take $|\lambda| = 1$ and construct *approximate* (L, λ)-eigenvectors $v^n = (\lambda, \lambda^2, \ldots, \lambda^n, 0, \ldots)$. Then $(L - \lambda I)v^n = (0, \ldots, 0, \lambda^{n+1}, 0, \ldots)$ so $\|(L - \lambda I)v^n\| = 1$ but $\|v^n\| = \sqrt{n}$. It remains to show that λ is of continuous spectrum, as opposed to residual spectrum. Indeed, for any w such that for all $v \in \ell^2$, $0 = \langle (L - \lambda I)v, w\rangle = \langle v, (R - \bar{\lambda} I)w\rangle$ so that w is an eigenvector of R which by 2.12 is a contradiction. An entirely analogous argument shows that the circle is also continuous spectrum for R.

We now reverse the discussion to see that any $|\lambda| < 1$ is in the residual spectrum of R. For $v \in \ell^2$ and w a $\bar{\lambda}$-eigenvector of L we have $0 = \langle v, 0\rangle = \langle v, (L - \bar{\lambda} I)w\rangle = \langle (R - \lambda I)v, w\rangle$ so that the range of $R - \lambda I$ is orthogonal to w.

Finally, $\sigma(M)$ is bounded by $\|M\|$ for any operator M, and clearly $\|L\| = 1$. But $\sigma_p(L) \cup \sigma_c(L) = \{\lambda \in \mathbb{C} \mid |\lambda| \leq 1\}$ so $\sigma_r(L) = \emptyset$. Combined with the previous paragraph, this shows 2.14.

The example already suggests some important relationships regarding the spectrum of adjoint operators.

Theorem 2.5. *Let M be a densely defined closed linear operator on a Hilbert space.*

1. *$\lambda \in \sigma(M)$ iff $\bar{\lambda} \in \sigma(M^*)$.*
2. *If $\lambda \in \sigma_r(M)$, then $\bar{\lambda} \in \sigma_p(M^*)$.*
3. *Conversely, if $\lambda \in \sigma_p(M)$, then $\bar{\lambda} \in \sigma_p(M^*) \cup \sigma_r(M^*)$.*
4. *$\sigma_c(M) = \sigma_c(M^*)$.*

Proof. 1. Suppose $M - \lambda I$ has a bounded inverse B. Then $(M - \lambda I)B = B(M - \lambda I) = I$. Equivalently $B^*(M^* - \bar{\lambda} I) = (M^* - \bar{\lambda} I)B^* = I^* = I$.

2. The range of $M - \lambda I$ is not dense in $\mathcal{X}$. Hence there exists a $v \perp \operatorname{ran}(M - \lambda I)$. But this implies $v \in \ker(M^* - \bar{\lambda} I)$.

3. There exists a $v \in \ker(M - \lambda I)$ which implies $v \perp \operatorname{ran}(M^* - \bar{\lambda} I)$.

4. This follows from 1, 2, and 3.

□

2.2.2 Spectra of Koopman and Perron-Frobenius Operators

The Koopman and Perron-Frobenius operator spectrum holds information about the long-term mixing rates of structures in phase space. We write (with some abuse of notation) $\mathcal{L}|_{\mathcal{X}}$ and $\mathcal{K}|_{\mathcal{X}}$ to denote the Perron-Frobenius / Koopman operators acting on the domain $\mathcal{X}$.

Definition 2.6. *An eigenfunction f for $\mathcal{L}|_{L^1}$ with eigenvalue* 1 *is the density of a* (signed) invariant measure. *We say S* preserves the measure $f\,dx$.

The following are well-known facts about the Koopman and Perron-Frobenius operator. They can be found in many sources e.g. [6].

Theorem 2.7. *1. If S is ergodic*[1]*, then there is at most one invariant density. Conversely, if there is a unique invariant density which is $(dx-)$almost everywhere positive, then S is ergodic.*

2. If S is ergodic, then every eigenvalue of $\mathcal{K}|_{L^1}$ is simple.

3. Suppose S is invertible. Then S is weak-mixing iff 1 *is the only eigenvalue of $\mathcal{K}|_{L^1}$.*

The following theorem is from [2].

Definition 2.8. *A set $A \subset \Omega$ in phase space is called* δ-almost-invariant *if* $\int_{S^{-1}(A)\cap A} dx = \delta \cdot \int_A dx$.

Theorem 2.9. *Let $\mathbb{R} \ni \lambda < 1$ be an eigenvalue corresponding to a real-valued normalized eigenfunction f of $\mathcal{L}|_{L^1}$. Let further $A \subset \Omega$ be such that $\int_A f\,dx = \frac{1}{2}$. Then*

$$\delta + \eta = \lambda + 1 \tag{2.15}$$

if A is δ-almost-invariant and $\Omega \setminus A$ is η-almost-invariant.

[1] Ergodicity and mixing describe how obeservations become decorrelated over time. S is *ergodic* if for all $A, B \in \mathcal{A}$ we have $\frac{1}{n}\sum_{j=0}^{n-1} \int_{S^{-j}(A)\cap B} dx \to \int_A dx \int_B dx$ as $n \to \infty$. S is weak-mixing if we have $\frac{1}{n}\sum_{j=0}^{n-1} \left| \int_{S^{-j}(A)\cap B} dx - \int_A dx \int_B dx \right| \to 0$ as $n \to \infty$.

These are only a few interesting properties of the spectrum of these operators, but there are many more which can be studied.

2.3 Pseudospectra

The reder will have likely noticed the sensitivity required in understanding the spectrum for infinite-dimensional operators. In particular the spectral types can be unstable w.r.t. perturbations of the operator. For example, for any self-adjoint operator M (e.g. the generator for the Koopman semigroup in a continuous-time dynamical system) there exists a compact operator E with arbitrarily small norm such that the perturbation $M + E$ has purely point spectrum.

The situation is even worse when one considers *perturbations of the dynamics* instead of perturbations of the Koopman / Perron-Frobenius operators. Consider a circle rotation $S : \mathbb{T} \to \mathbb{T},\ e^{2\pi i\theta} \mapsto e^{2\pi i(\theta+\alpha)}$. We have $\sigma(\mathcal{L}|_{L^2}) = \sigma_p(\mathcal{L}|_{L^2}) = \alpha^{\mathbb{N}_0}$. If the rotation is rational then the spectrum is discrete, but if the rotation is irrational then the spectrum is dense in the unit circle.

In the present paper we tackle the instability problem for operator perturbations.

2.3.1 Definitions of the Pseudospectrum

Definition 2.10. *Let $M : D(M) \to \mathcal{X}$ be a closed linear operator. The* ϵ-pseudospectrum *of M is the smallest set in $\mathbb{C}$ which contains the spectrum of all perturbations of M with norm less than ϵ:*[2]

$$\sigma_\epsilon(M) = \bigcup_{\|E\|<\epsilon} \sigma(M + E). \tag{2.16}$$

Theorem 2.11. *We have the following equivalent formulation of the pseudospectrum:*

$$\sigma_\epsilon(M) = \left\{ \lambda \in \mathbb{C} \mid \|(M - \lambda I)^{-1}\| > \frac{1}{\epsilon} \right\} \tag{2.17}$$

where we use the convention that $\|(M-\lambda I)^{-1}\| = \infty$ if $M-\lambda I$ is not invertible.

Proof. We prove "$\subset$" by contraposition: assume $\|(M - \lambda I)^{-1}\| > \frac{1}{\epsilon}$ and let $\|E\| < \epsilon$. Then $\|(M - \lambda I)^{-1}M\| < 1$ and hence $I + (M - \lambda I)^{-1}E$ is invertible. This implies $M - \lambda I + E = (M - \lambda I)\,(I + (M - \lambda I)^{-1}E)$ is invertible.

Conversely, we prove "$\supset$" by showing there exists an operator E with $\|E\| < \epsilon$ sich that $M - \lambda I + E$ is not invertible. Since $\|(M - \lambda I)^{-1}\| > \frac{1}{\epsilon}$ there exists a

[2] Some authors will deine the pseudospectrum with "$\leq$" instead of "$<$". This makes $\sigma_\epsilon(M)$ a closed set, but breaks theorem 2.11.

$u \in \mathcal{X}$ with $\|u\| = 1$ and $(M - \lambda I)^{-1}u = v \in D(M)$ with $\|v\| = \frac{1}{\delta} > \frac{1}{\epsilon}$.[3] The Hahn-Banach theorem provides a $v^* \in \mathcal{X}^*$ with $\|v^*\| = 1$ and $v^*v = \|v\| = \frac{1}{\delta}$. Set $E = -\delta uv^*$. Then $\|E\| = \delta < \epsilon$ and

$$Ev = -\delta uv^*v = -u = -(M - \lambda I)(M - \lambda I)^{-1}u = (M - \lambda I)v. \tag{2.18}$$

Rearranging provides $(M + E - \lambda I)v = 0$. □

The following formulation of the pseudospectrum in the Hilbert space case will be the main tool we use in the section on numerical methods.

Definition 2.12. *The function*

$$\sigma_{\inf}(M) = \inf_{\|x\|=1} \|Mx\| \tag{2.19}$$

is known as the injection modulus.

Lemma 2.13. *Let $M : D(M) \to \mathcal{H}$ be a closed linear operator on a Hilbert space. If $\lambda \in \rho(M)$ then*

$$\sigma_{\inf}(M - \lambda I) = \sigma_{\inf}(M^* - \bar{\lambda} I), \tag{2.20}$$

but this is not *necessarily true if $\lambda \in \sigma(M)$.*

Proof. 2.20 for $\lambda \in \rho(M)$ follows from the fact that for bounded operators A we have $\|A\| = \|A^*\|$, applied to $A = (M - \lambda I)^{-1}$. To see that 2.20 does not hold for $\lambda \in \sigma(M)$, consider again the right shift from example 2.4. Clearly $\|Rx\| = \|x\|$ so R is bounded from below but 0 is part of the spectrum. □

Theorem 2.14. *For a Hilbert space operator M,*

$$\frac{1}{\|(M - \lambda I)^{-1}\|} = \min\left\{\sigma_{\inf}(M - \lambda I),\ \sigma_{\inf}(M^* - \bar{\lambda} I)\right\}. \tag{2.21}$$

where we use the convention that $1/\|(M - \lambda I)^{-1}\| = 0$ when $\lambda \in \sigma(M)$.

Proof. When $\lambda \in \sigma_p(M)$, then obviously $\sigma_{\inf}(M - \lambda I) = 0$. Due to theorem 2.5 part 2, $\lambda \in \sigma_r(M)$ has $\sigma_{\inf}(M^* - \bar{\lambda} I) = 0$. Finally, let $\lambda \in \sigma_c(M)$. Assume $\sigma_{\inf}(M - \lambda I) > 0$. Then $(M - \lambda I)^{-1} : \operatorname{ran}(M - \lambda I) \to \mathcal{W} \subset \mathcal{X}$ is bounded. But since $\operatorname{ran}(M - \lambda I) \subset \mathcal{X}$ is dense, $(M - \lambda I)^{-1}$ has a unique extension to $\mathcal{X}$, a contradiction. □

Corollary 2.15. *In the Hilbert space setting, the psudospectrum can be formulated as:*

$$\begin{aligned}\sigma_\epsilon(M) = \{\lambda \in \mathbb{C} \mid \exists\, u : \|u\| = 1 \text{ and either} \\ \|(M - \lambda I)u\| < \epsilon \text{ or } \left\|(M^* - \bar{\lambda} I)u\right\| < \epsilon\}\end{aligned} \tag{2.22}$$

$$= \sigma(M) \cup \{\lambda \in \mathbb{C} \mid \exists\, u : \|u\| = 1,\ \|(M - \lambda I)u\| < \epsilon\}. \tag{2.23}$$

[3] At this point we require the strict inequality. Otherwise, the existence of such a pair u, v is not guaranteed.

References

[1] Matthew J. Colbrook. "Chapter 4 - The multiverse of dynamic mode decomposition algorithms". In: *Numerical Analysis Meets Machine Learning*. Ed. by Siddhartha Mishra and Alex Townsend. Vol. 25. Handbook of Numerical Analysis. Elsevier, 2024, pp. 127–230. DOI: https://doi.org/10.1016/bs.hna.2024.05.004. URL: https://www.sciencedirect.com/science/article/pii/S1570865924000048.

[2] Michael Dellnitz and Oliver Junge. "On the Approximation of Complicated Dynamical Behavior". In: *SIAM Journal on Numerical Analysis* 2.36 (1999).

[3] Paul Garrett. *Exmaples of Operators and Spectra*. Online. 2020. URL: http://www.math.umn.edu/%CB%9Cgarrett/m/fun/notes_2012-13/06b_examples_spectra.pdf.

[4] B. O. Koopman and J. V. Neumann. "Dynamical Systems of Continuous Spectra". In: *Proceedings of the National Academy of Sciences of the United States of America* 18.3 (1932), pp. 255–263. ISSN: 00278424, 10916490. URL: http://www.jstor.org/stable/86259 (visited on 11/13/2024).

[5] Christian Kühn. "Lecture Notes on Dynamical Systems". 2023.

[6] Andrzej Lasota and Michael Mackey. *Chaos, Fractals, and Noise*. Springer New York, NY, 1993. DOI: https://doi.org/10.1007/978-1-4612-4286-4.

5 Appendix

5.1 Transfer Operator for Blaschke Products

We consider the analytic mapping on the unit circle in $\mathbb{C}$,

$$\tau : \mathbb{T} \circlearrowleft, \quad z \mapsto z\frac{z-\mu}{1-\bar{\mu}z}$$

dependent on a parameter $\mu \in \mathbb{D}$ in the open unit disk.

τ is a Lebesgue-measure preserving 2-to-1 map. The spectrum $\sigma\left(\mathcal{K}|_{L^2(d\theta)}\right) = \overline{\mathbb{D}}$ is the whole unit disk, point spectrum $\sigma_p\left(\mathcal{K}|_{L^2(d\theta)}\right) = \{1\}$.

When considering instead the adjoint $\mathcal{L} = \mathcal{K}^*$, the point spectrum has the interesting property that $\left\{\mu^k\right\}_{k=0}^{\infty} \cup \left\{\overline{\mu}^k\right\}_{k=0}^{\infty} \subset \sigma_p\left(\mathcal{L}|_{L^2(d\theta)}\right)$.

More specifically, let $\mathbb{T} \subset A$ be a (suitably chosen) open annulus containing the unit circle, and $H^2(A)$ be the Hardy Hilbert space of holomorphic functions on A which are square integrable on the edge ∂A. Then on this space, $\sigma\left(\mathcal{L}|_{H^2(A)}\right) = \left\{\mu^k\right\}_{k=0}^{\infty} \cup \left\{\overline{\mu}^k\right\}_{k=0}^{\infty}$.

EDMD with monomials $\left\{z \mapsto z^k\right\}_{k=-N}^{N}$ (Fourier basis) results in eigendecompositions which converge exponentially to $\sigma\left(\mathcal{L}|_{H^2(A)}\right)$ as $N \to \infty$.

5.2 Some scribbles about kernels and residuals

EDMD: EDMD is a Petrov-Galerkin method as written in Péter's 2016 paper. Considering a dictionary $\{\psi_i\}_{i=1}^{N} \subset \mathcal{H}$ ($\mathcal{H} = L^2$ mostly) and snapshots $\{x_j\}_{j=1}^{M} \subset X$, EDMD makes the following construction:

$$\Psi(x) = [\psi_1 \dots \psi_N] \tag{5.1}$$

We find a matrix $K \in \mathbb{C}^{N\times N}$ that minimizes

$$\|\mathcal{K}\Psi - \Psi K\|_{\mathcal{H}(X,\mathbb{C}^{1\times N})} \tag{5.2}$$

The $\mathcal{H}$ norm is approximated by a quadrature:

$$\Psi_X = \Psi\begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix}, \quad \Psi_Y = \Psi \circ S\begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix} \tag{5.3}$$

$$\|\mathcal{K}\Psi - \Psi K\|_{\mathcal{H}(X,\mathbb{C}^{1\times N})} \approx \|\Psi_Y - \Psi_X K\|_F. \tag{5.4}$$

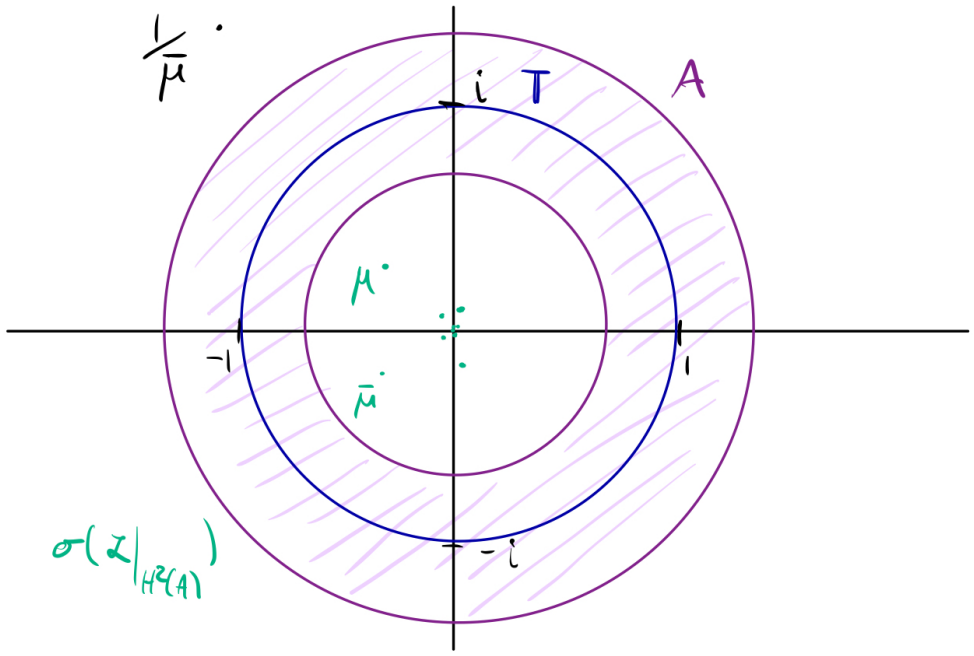

Figure 5.1: Unit circle $\mathbb{T}$ and Annulus A.

5.2 is minimized by

$$K = \Psi_X^\dagger \Psi_Y = G^\dagger A \tag{5.5}$$

where $G = \Psi_X^* \Psi_X$, $A = \Psi_X^* \Psi_Y$.

EDMD for Perron-Frobenius: Since

$$A_{ij} \overset{M\to\infty}{\approx} \langle \psi_i, \mathcal{K}\psi_j \rangle = \langle \mathcal{L}\psi_i, \psi_j \rangle \,, \tag{5.6}$$

an equivalent Galerkin approximation for $\mathcal{L} = \mathcal{K}^*$ is given by

$$P = G^\dagger A^* = \Psi_X^\dagger {\Psi_X^*}^\dagger \Psi_Y^* \Psi_X. \tag{5.7}$$

kEDMD: Let $k : X \times X \to \mathbb{C}$ be a kernel so that

$$k(x_i, x_j) = \langle \Psi(x_i), \Psi(x_j) \rangle_{\ell^2} \;\; (= \Psi(x_i)\Psi(x_j)^* \,) \,. \tag{5.8}$$

(Note that $\Psi(x)$ might be infinite-dimensional, i.e. $N = \infty$, but we think of $N < \infty$ for clarity of the matrix manipulations.)

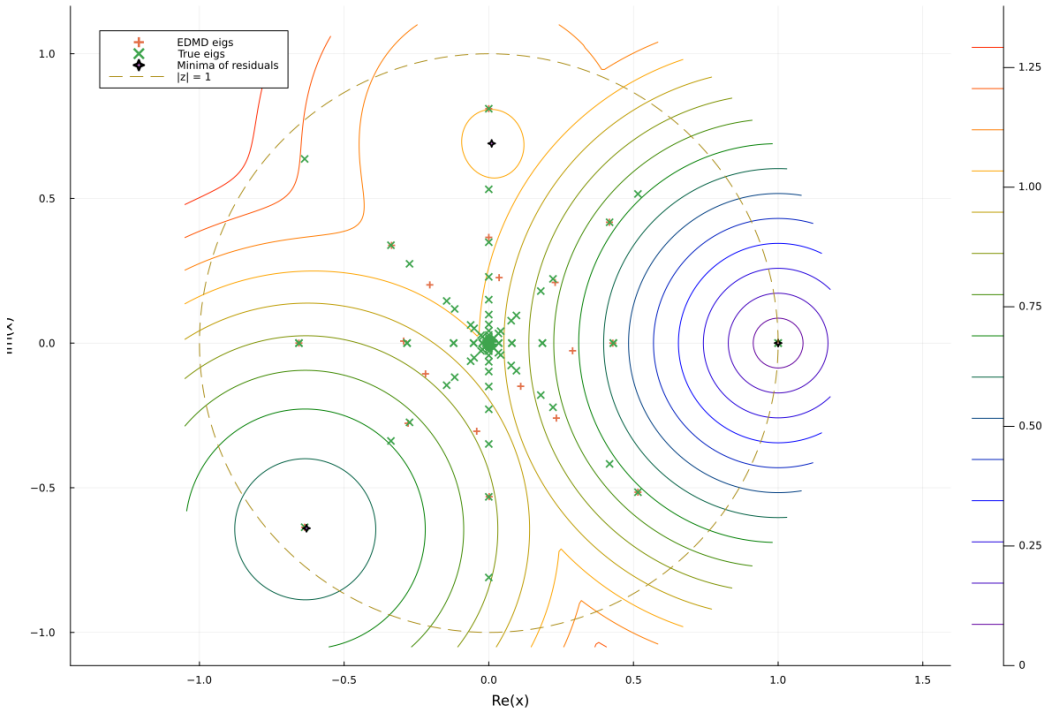

Figure 5.2: ResDMD over the space $L^2(\mathbb{T})$ for a Blaschke map $\tau(z) = z\frac{z-\mu}{1-\bar{\mu}z}$ with $\mu = 0.9e^{\pi i/4}$. True eigenvalues are known in the Hardy space $H^2(A)$ of functions holomorphic on the annulus $A = \{z \in \mathbb{C} \mid r < |z| < R\}$, $R = 11/10 - 1/32$, $r = 1/R$. ResDMD matrices G, A, L are computed w.r.t. this Hilbert space, using a dictionary of $N = 20$ monomials and $M = 200,000$ equally spaced quadrature nodes.

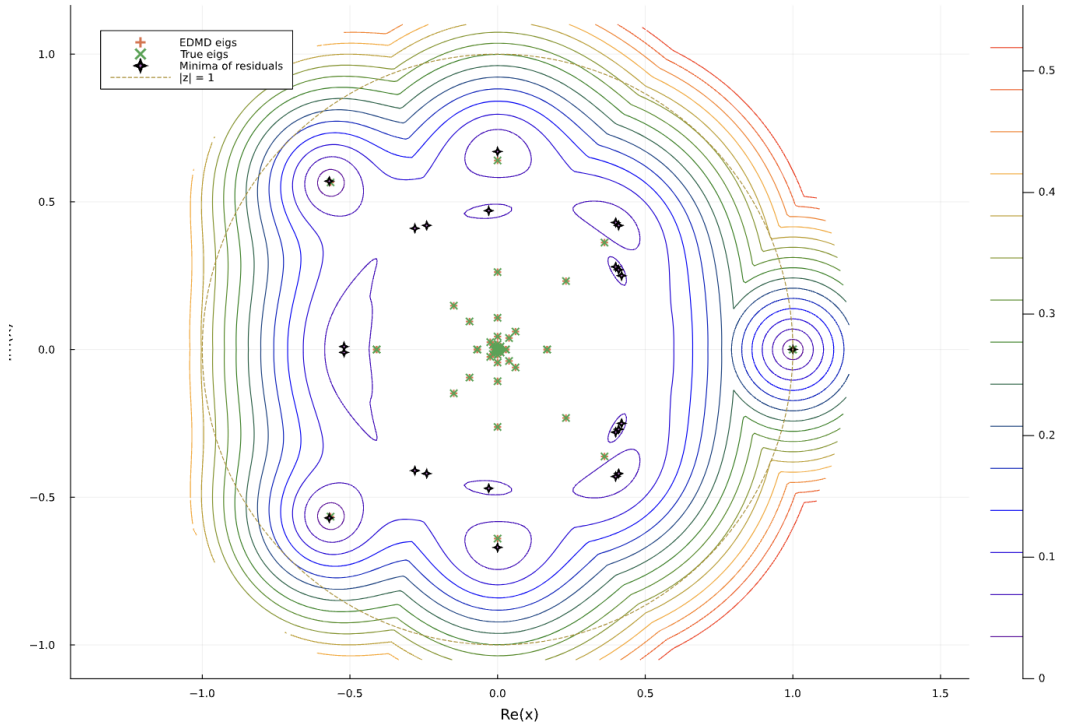

Figure 5.3: ResDMD for $\mu = 0.8e^{\pi i/4}$ over the space $H^2(A)^*$, where $H^2(A) \subset L^2$ is seen as embedded in L^2 and $(\cdot)^*$ the Banach space adjoint. This dual space is isomorphic to the space $H^2(\mathbb{D}_r) \oplus H^2(\mathbb{D}^\infty_{1/r})$ and has a norm $\|f\|^2_{H^2(A)^*} = \sum_{n\in\mathbb{Z}} |\hat{f}|^2 r^{|n|}$.

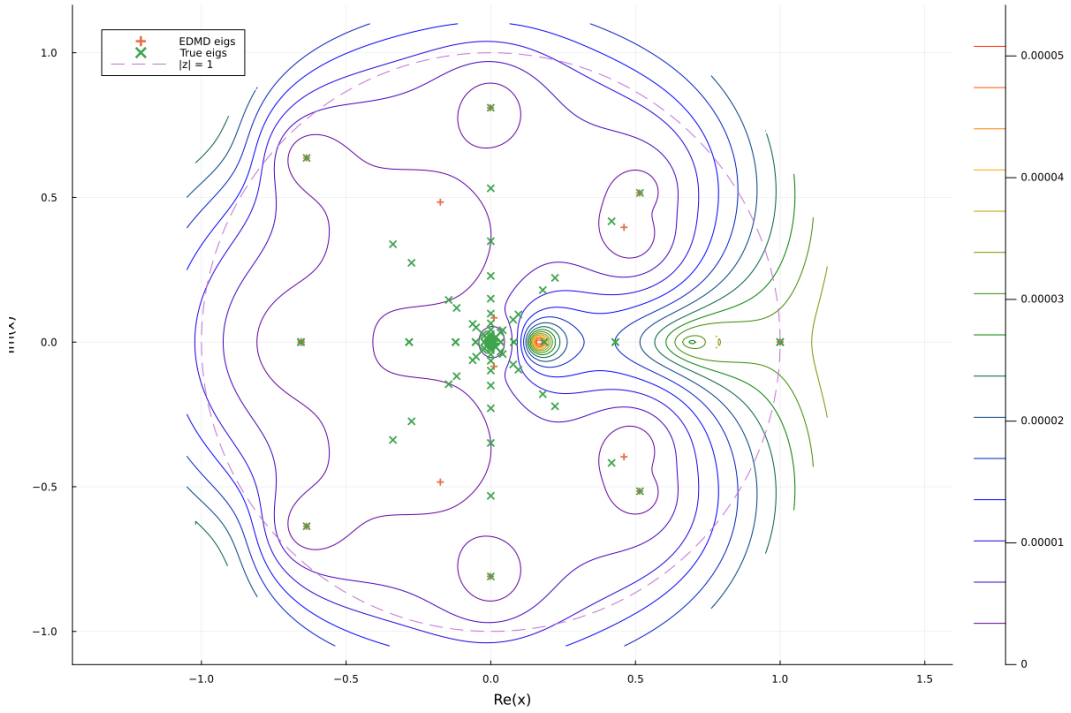

Figure 5.4: Kernelized ResDMD for the Blaschke map using a degree 20 polynomial kernel $k(x, y) = (1 + x'y)^{20}$ and $M = 2,000$ equally spaced quadrature nodes.

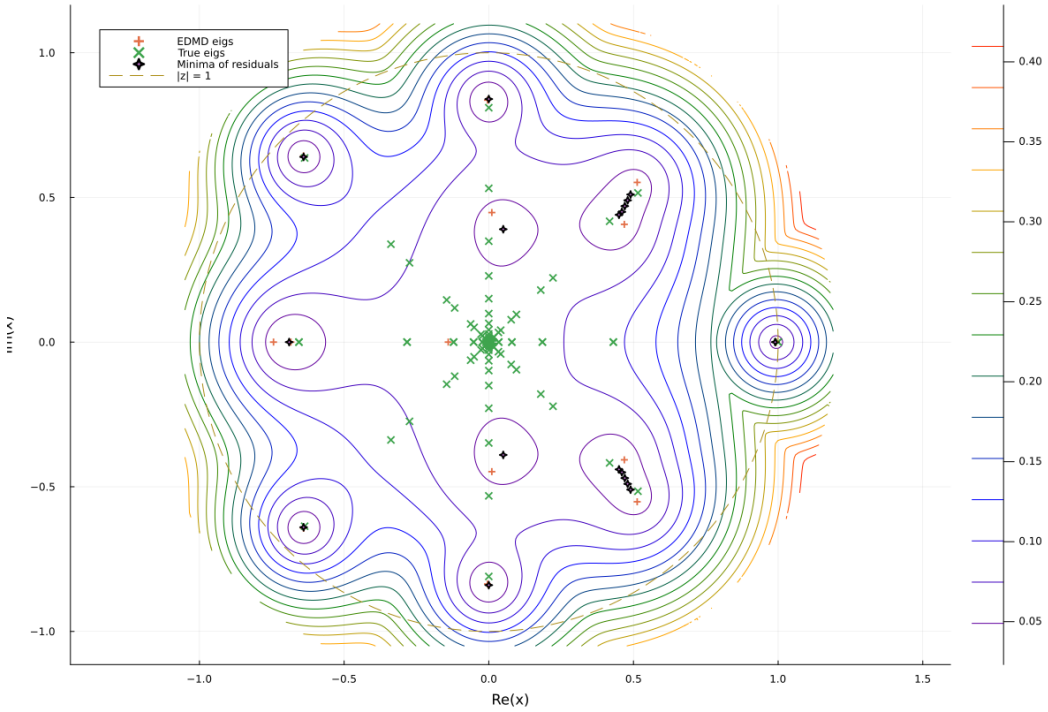

Figure 5.5: Kernelized ResDMD using an RBF kernel $k(x, y) = e^{-|x-y|^2/c}$ with $c = 0.01$ and $M = 2,000$ equally spaced quadrature nodes.

Let further $\Psi_X \overset{\text{SVD}}{=} Q\Sigma Z^*$ and

$$\hat{G} = (k(x_i, x_j))_{ij} = \Psi_X \Psi_X^* \overset{\text{diagonalize}}{=} Q(\Sigma^*\Sigma)Q^*, \tag{5.9}$$

$$\hat{A} = (k(y_i, x_j))_{ij} = \Psi_Y \Psi_X^* \tag{5.10}$$

Then

$$\hat{K} = (\Sigma^\dagger Q^*)\hat{A}(Q\Sigma^\dagger) \in \mathbb{C}^{M\times M} \tag{5.11}$$

has the same eigenvalues as K.

ResDMD: The key idea behind ResDMD is, for an element $z \in \mathbb{C}^N$, $g = \Psi z$:

$$\begin{aligned}
\operatorname{res}(\lambda, z) &:= \|(\mathcal{K} - \lambda)g\|^2_{\mathcal{H}(X,\mathbb{C})} && (5.12)\\
&\overset{M\to\infty}{\approx} \|(\Psi_Y - \lambda\Psi_X)z\|^2_{\ell^2} && (5.13)\\
&= z^* \left(\Psi_Y^*\Psi_Y - \lambda\Psi_Y^*\Psi_X - \overline{\lambda}\Psi_X^*\Psi_Y + |\lambda|^2\Psi_X^*\Psi_X\right) z && (5.14)\\
&= z^*(L - \lambda A^* - \overline{\lambda}A + |\lambda|^2 G)z. && (5.15)
\end{aligned}$$

We interpret the first equation as

$$\textit{regression error} \overset{M\to\infty}{\approx} \textit{operator residual}. \tag{5.16}$$

The function $\lambda \mapsto \operatorname{res}(\lambda) = \min_z \operatorname{res}(\lambda, z)$ is shown in 5.2

kResDMD: We apply idea 5.16 to $\hat{K}$:

$$\hat{K^*} = \left(\Psi_X^* Q\Sigma^\dagger\right)^\dagger \left(\Psi_Y^* Q\Sigma^\dagger\right) =: \hat{\Psi_X}^\dagger \hat{\Psi_Y} \tag{5.17}$$

so that $\hat{K^*}$ solves the regression problem

$$\min_M \|\hat{\Psi_Y} - \hat{\Psi_X} M\|_F. \tag{5.18}$$

The regresion error reduces to

$$\hat{\text{rés}}(\lambda, z) := \|(\hat{\Psi_Y} - \lambda\hat{\Psi_X})z\|_{\ell^2}^2 \tag{5.19}$$
$$= (Q\Sigma^\dagger z)^* \left(\Psi_Y\Psi_Y^* - \lambda\Psi_Y\Psi_X^* - \overline{\lambda}\Psi_X\Psi_Y^* + |\lambda|^2\Psi_X\Psi_X^*\right)(Q\Sigma^\dagger z) \tag{5.20}$$
$$= (Q\Sigma^\dagger z)^* \left(\hat{L} - \lambda\hat{A}^* - \overline{\lambda}\hat{A} + |\lambda|^2\hat{G}\right)(Q\Sigma^\dagger z) \tag{5.21}$$

Note at this point it is unclear whether this regression error in the case of $\hat{K}$ has a physical meaning.

ResDMD for Perron-Frobenius: Consider $\tilde{P} = Z^*PZ$, which has the same eigenvalues as P. Then

$$\tilde{P} = (\Psi_X^* Q\Sigma)^\dagger (\Psi_Y^* Q\Sigma) =: \tilde{\Psi_X}^\dagger \tilde{\Psi_Y} \tag{5.22}$$

so that $\hat{P}$ solves the regresion problem

$$\min_M \|\tilde{\Psi_Y} - \tilde{\Psi_X} M\|_F. \tag{5.23}$$

The interesting thing to note is that

$$\tilde{\text{rés}}(\lambda, z) := \|(\tilde{\Psi_Y} - \lambda\tilde{\Psi_X})z\|_{\ell^2}^2 \tag{5.24}$$
$$= (Q\Sigma z)^* \left(\Psi_Y\Psi_Y^* - \lambda\Psi_Y\Psi_X^* - \overline{\lambda}\Psi_X\Psi_Y^* + |\lambda|^2\Psi_X\Psi_X^*\right)(Q\Sigma z) \tag{5.25}$$
$$= (Q\Sigma z)^* \left(\hat{L} - \lambda\hat{A}^* - \overline{\lambda}\hat{A} + |\lambda|^2\hat{G}\right)(Q\Sigma z) \tag{5.26}$$

and hence whenever Ψ_X has the same number of zero singular values (we assume there are none i.e. $s_{min}(\Psi_X) > 0$),

$$\hat{\text{rés}}(\lambda) := \min_z \ \hat{\text{rés}}(\lambda, z) = \min_\xi \ \tilde{\text{rés}}(\lambda, \xi) =: \tilde{\text{rés}}(\lambda). \tag{5.27}$$

The function rês is shown in 5.4, 5.5.

Interpreting the regression error: Let $y = Q\Sigma z$. For convenience we conjugate λ, i.e. let $\tilde{P}z = \overline{\lambda}z$ be a candidate eigenpair for $\mathcal{L}$.

What follows is a rough, purely formal sketch of the idea that we want to show, from what I calculated this morning (no promises for correctness):

$$\|(\Psi_Y^* - \overline{\lambda}\Psi_X^*)y\|_{\ell^2}^2 \tag{5.28}$$

$$= \sum_{i=1}^{N} \left| \sum_{j=1}^{M} \overline{(\mathcal{K}\psi_i - \lambda\psi_i)(x_j)}\ y_j \right|^2 \tag{5.29}$$

$$\overset{M\to\infty}{\approx} \sum_{i=1}^{N} \left| \int \overline{(\mathcal{K}\psi_i - \lambda\psi_i)(\xi)}\ y(\xi)\ d\xi \right|^2 \tag{5.30}$$

$$= \sum_{i=1}^{N} |\langle(\mathcal{K} - \lambda)\psi_i,\ y\rangle|^2 \tag{5.31}$$

$$= \sum_{i=1}^{N} \left|\left\langle\psi_i,\ (\mathcal{L} - \overline{\lambda})y\right\rangle\right|^2 \tag{5.32}$$

$$\overset{N\to\infty}{\approx} \|(\mathcal{L} - \overline{\lambda})y\|_{\mathcal{H}(X,\mathbb{C})}^2. \tag{5.33}$$

where for 5.30 we say that the vector $y = (y_j)_j$ converges to a function $y : \xi \mapsto y(\xi)$, and for 5.33 we use Plancherel's theorem (hence we need to assume Ψ is an orthonormal family that in the limit has span dense in $\mathcal{H}(X, \mathbb{C})$).

It seems likely that these convergences will work, under the crucial assumption that the singular values stay bounded from above and below as $N \to \infty$, which is given if Ψ is an orthonormal family. There is some literature on kernel QR algorithms, which are essentially Gram-Schmidt for kernel weight matrices. However, I would much prefer if this would work without the orthonormal requirement, as most kernel functions and snapshot datasets will not make orthonormal Gram matrices.

For e.g. Gauß kernels $k(x, y) = \exp(-\|x - y\|^2/c)$, the sharpness parameter c has to be changed as M changes in order for the singular value condition to hold since if c is kept constant and M grows, the Gram matrix slowly becomes the all-ones matrix.

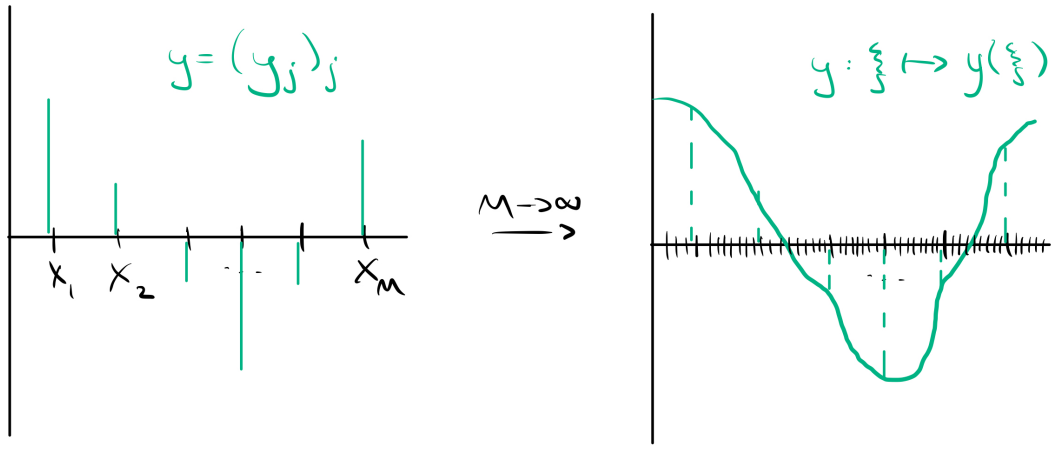

Figure 5.6: Idea behind equation 5.30