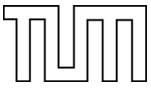

Technische Universität München

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 [2], 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 $S : \mathcal{X} \to \mathcal{X}$ defined on a normed space $(\mathcal{X}, \|\cdot\|)$.

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 $g : \mathcal{X} \to \mathbb{C}$. A state $x \in \mathcal{X}$ 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.

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] 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 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:

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

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

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

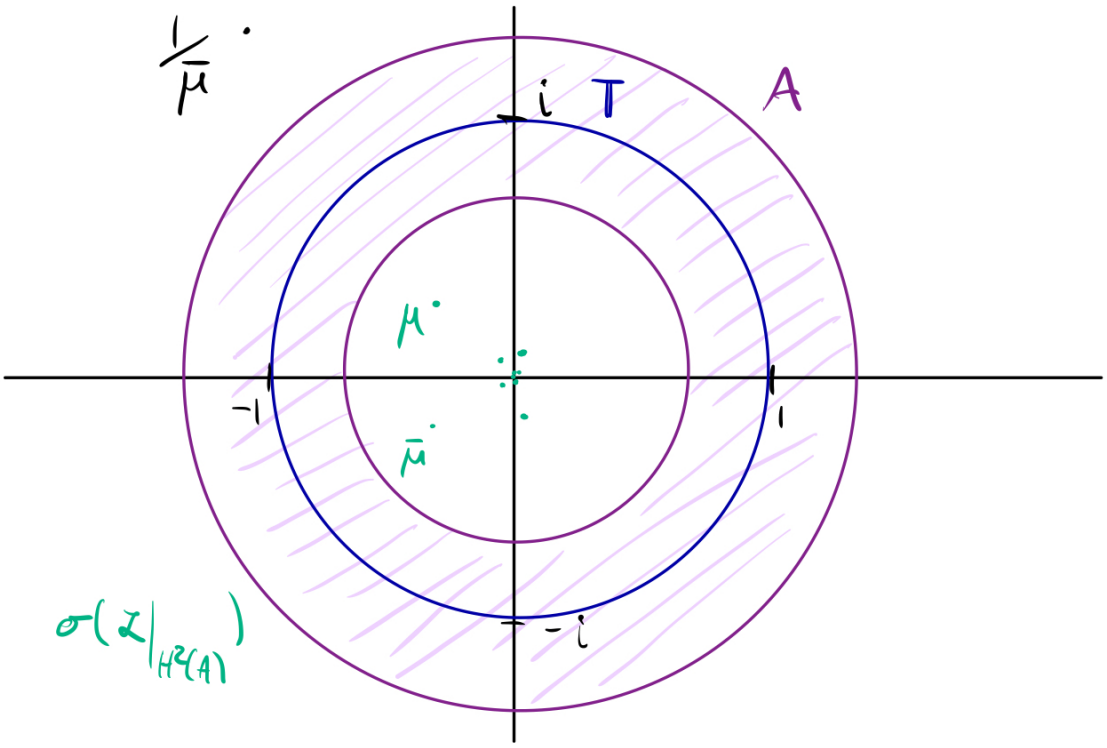

Figure 5.1: Unit circle $\mathbb{T}$ and Annulus A.

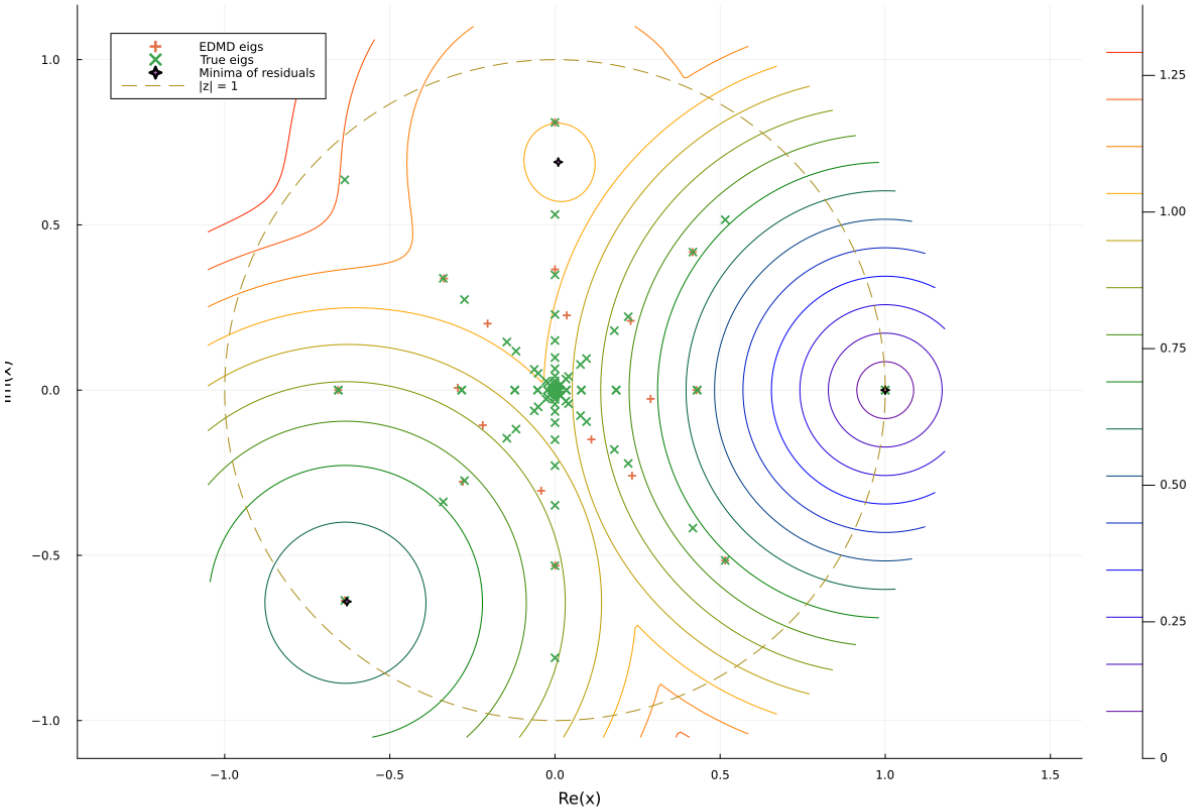

Figure 5.2: ResDMD over the space $L^2(\mathbb{T})$ for a Blaschke map $\tau(z) = z\frac{z-\mu}{1-\bar{\mu}z}$ with $\mu = 0.9e^{\pi i/4}$. True eigenvalues are known in the Hardy space $H^2(A)$ of functions holomorphic on the annulus $A = \{z \in \mathbb{C} \mid r < |z| < R\}$, $R = 11/10 - 1/32$, $r = 1/R$. 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.

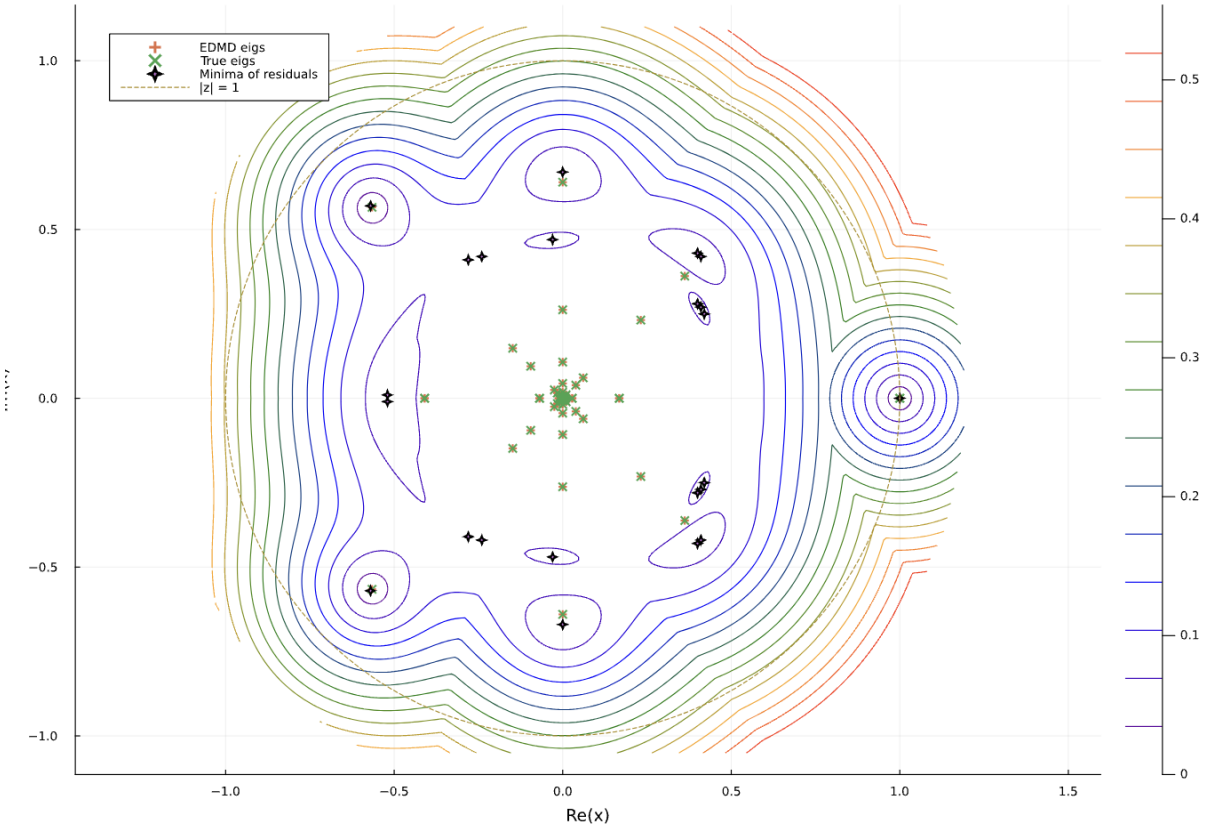

Figure 5.3: ResDMD for $\mu = 0.8e^{\pi i/4}$ over the space $H^2(A)^*$, where $H^2(A) \subset L^2$ is seen as embedded in L^2 and $(\cdot)^*$ the Banach space adjoint. This dual space is isomorphic to the space $H^2(\mathbb{D}_r) \oplus H^2(\mathbb{D}^\infty_{1/r})$ and has a norm $\|f\|^2_{H^2(A)^*}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

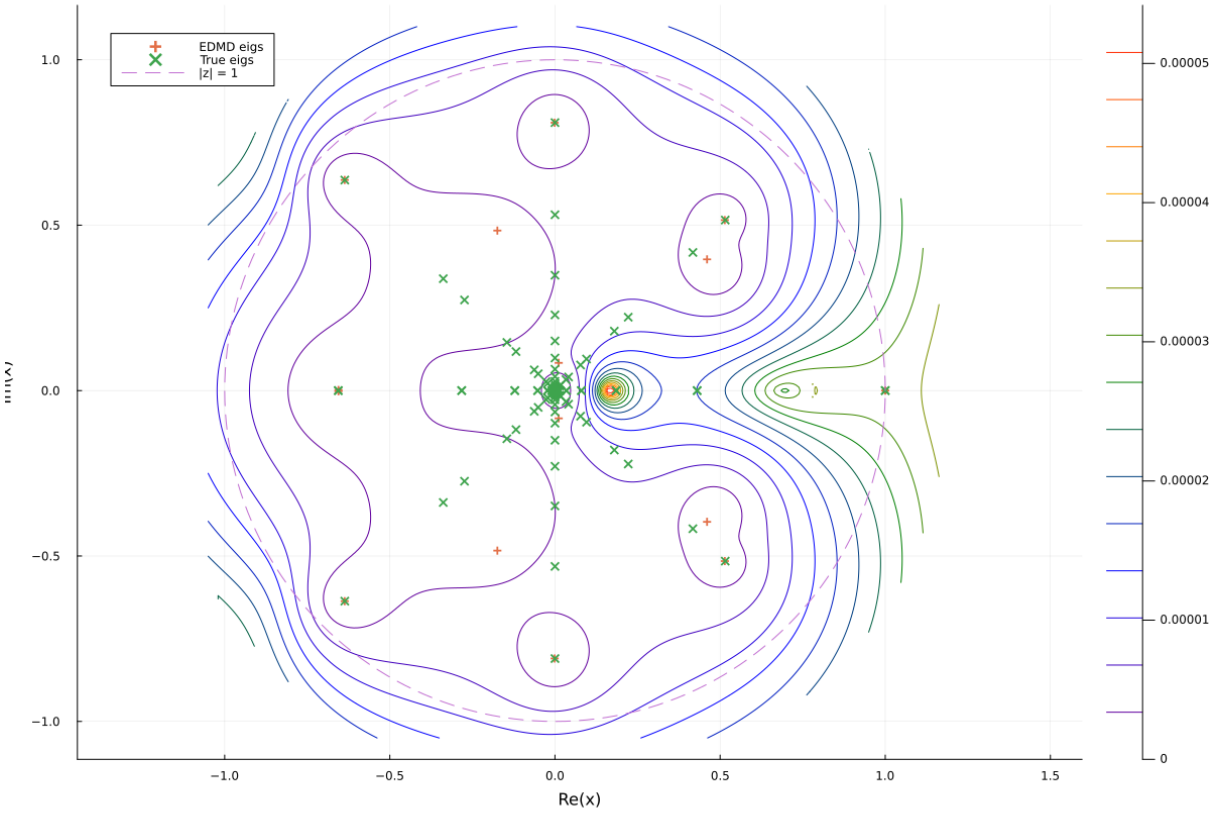

Figure 5.4: Kernelized ResDMD for the Blaschke map using a degree 20 polynomial kernel $k(x, y) = (1 + x'y)^{20}$ and $M = 2,000$ equally spaced quadrature nodes.

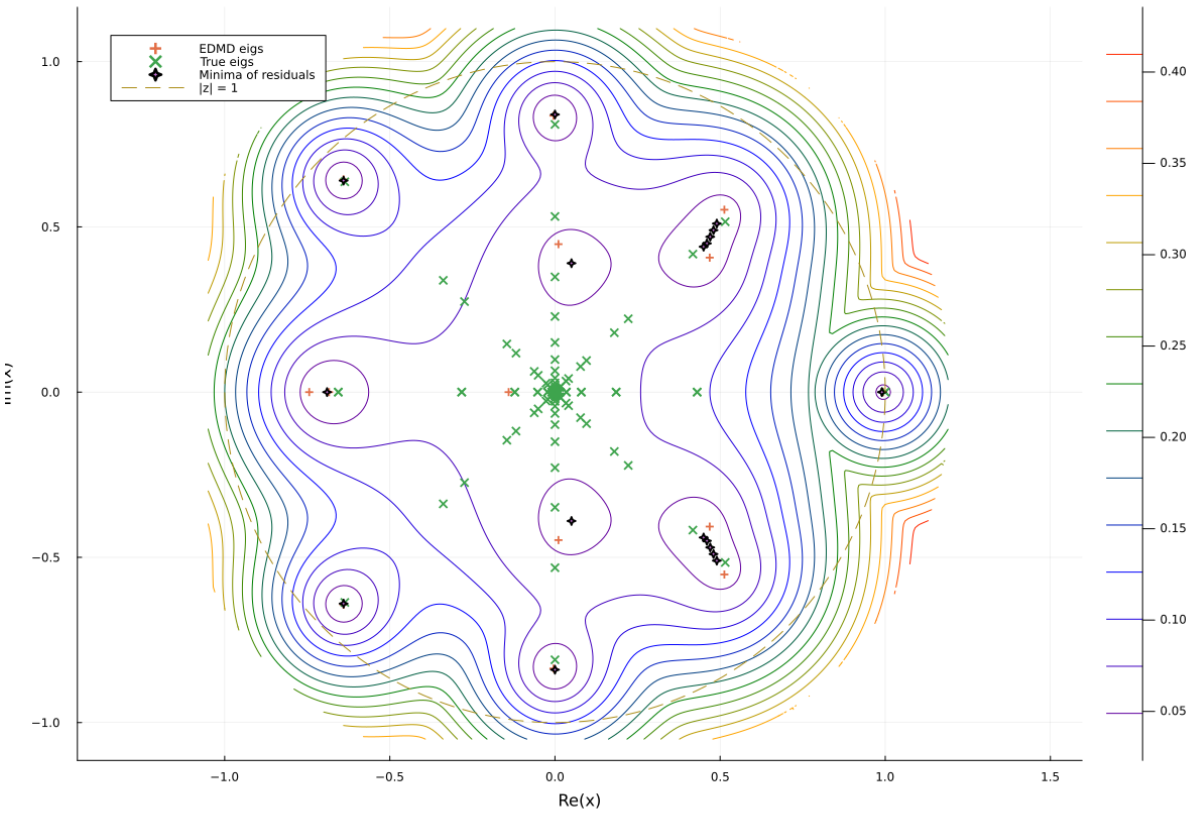

Figure 5.5: Kernelized ResDMD using an RBF kernel $k(x, y) = e^{-|x-y|^2/c}$ with $c = 0.01$ and $M = 2,000$ equally spaced quadrature nodes.

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

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.)
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

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

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

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

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.

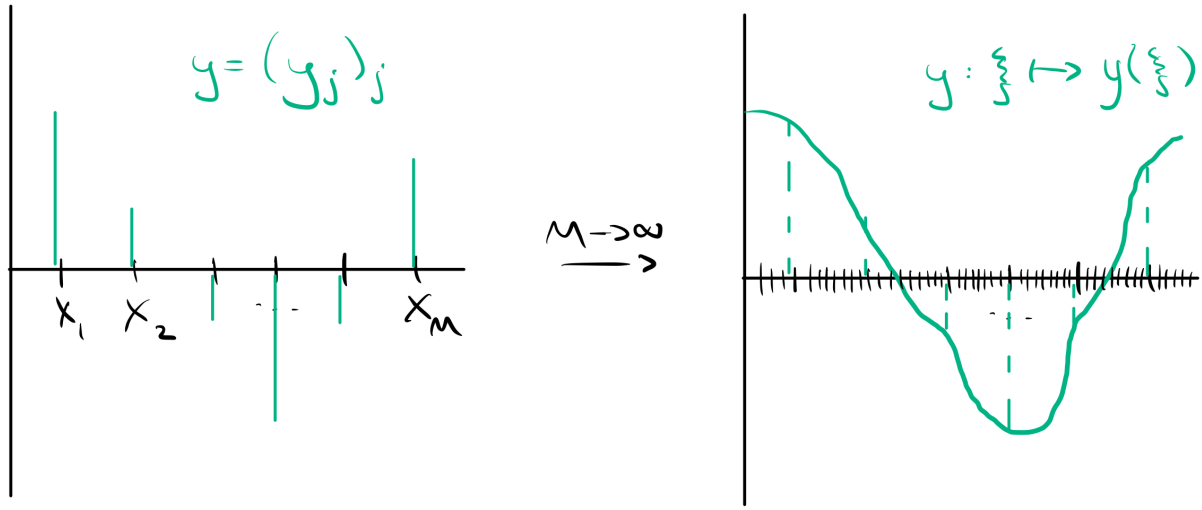

Figure 5.6: Idea behind equation 5.30