# Spectral theory and approximation of Koopman operators in chaos

## Part 3: numerics vs Banach spaces

Caroline Wormell

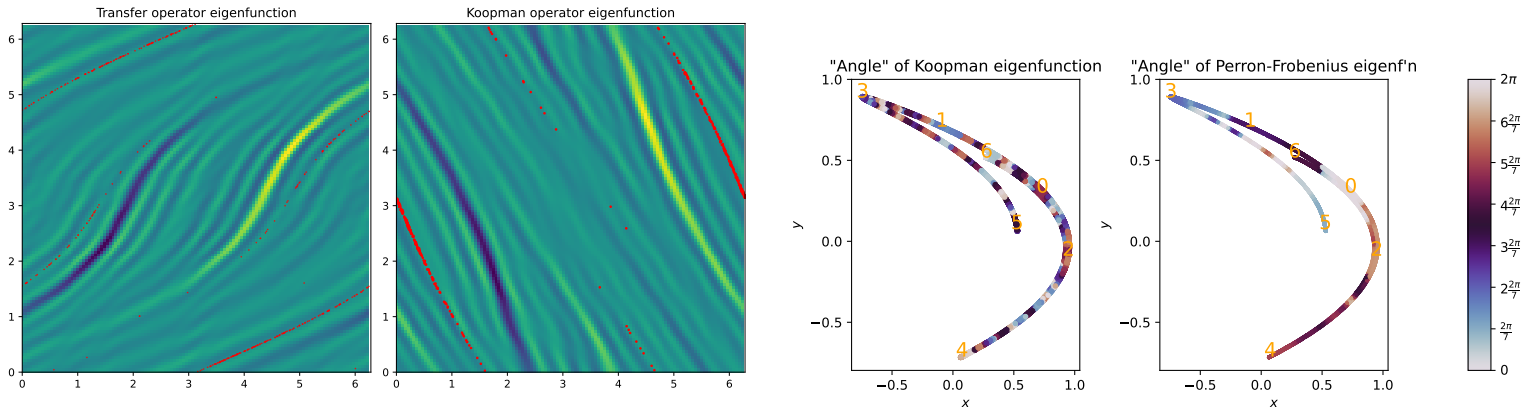
The University of Sydney

July 26, 2024

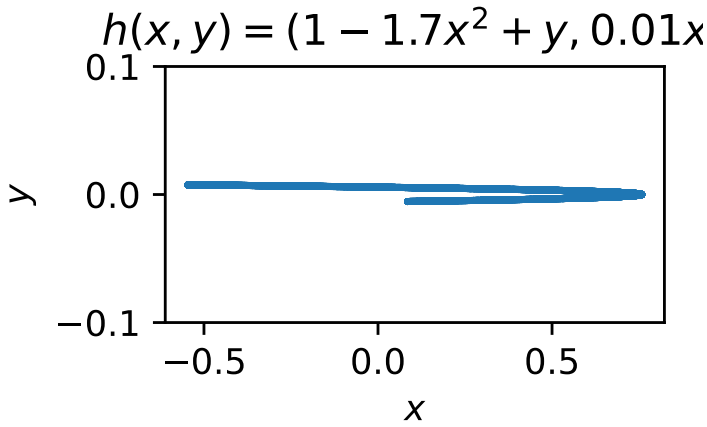
# Today

- Non-removable essential spectrum: the tent map
- A numerical framework for convergence of spectrum
- Data-driven approximations

# Yesterday


Transfer operator eigenfunction
Koopman operator eigenfunction
"Angle" of Koopman eigenfunction
"Angle" of Perron-Frobenius eigenf'n
x
y

# Yesterday



# Yesterday

**Theorem (Butterley, Canestrari, Jain '22)**

*Suppose $f$ is a piecewise monotone map on* $[0,1]$ *with a* non-trivial discontinuity *$c$ and $\mathcal{B}$ is a Banach space satisfying:*

- $C^\infty \subseteq \mathcal{B}$.
- $\mathcal{L}$ *is bounded* $\mathcal{B}$ *to* $\mathcal{B}$.
- $\|\mathcal{L}^n\varphi\|_{\mathcal{B}} \leq \|\mathcal{L}^n\varphi\|_{L^\infty}$ *for all* $\varphi \in C^\infty$.

*Then*

$$\rho_{ess}(\mathcal{L},\mathcal{B}) \geq \frac{1}{\underbrace{\lim_{n\to\infty} |(f^n)'(c)|^{1/n}}_{\text{Lyapunov exponent at } c}}.$$

# Numerics for quasi-compact operators

Remember how we approximated the Koopman operator of a stochastic system. Our ingredients were:

- A weak space $\mathcal{B}_w$ and a strong space $\mathcal{B}_s$, with $\mathcal{B}_s$ compactly embedded in $\mathcal{B}_w$.
- $\mathcal{A}$ is bounded $\mathcal{B}_w \to \mathcal{B}_s$.
- We aimed to approximate $\mathcal{A}$ by $\mathcal{P}_N\mathcal{A}$, where $\mathcal{P}_N$ was a finite rank operator.
  - Since $\mathcal{K}$, $\mathcal{L}$ are fairly easy to approximate pointwise this is a reasonable way to go in general.
- $\mathcal{P}_N - \mathrm{id}$ is small in norm from $\mathcal{B}_s \to \mathcal{B}_w$.
  Say $\|\mathcal{P}_N - \mathrm{id}\|_{\mathcal{B}_s \to \mathcal{B}_w} \leq \epsilon(N)$.
  - This is only possible because $\mathcal{B}_s \Subset \mathcal{B}_w$, i.e. $\mathrm{id} : \mathcal{B}_s \to \mathcal{B}_w$ is compact

# Numerics for quasi-compact operators

Then we had, in operator norm,

$$\mathcal{B}_w \underbrace{\xrightarrow{\mathcal{A}}}_{\text{bounded}} \mathcal{B}_s \underbrace{\xrightarrow{\mathcal{P}_N - \mathrm{id}}}_{\text{small}} \mathcal{B}_w$$

meaning

$$\mathcal{B}_w \underbrace{\xrightarrow{\mathcal{A}_N - \mathcal{A}}}_{\text{small}} \mathcal{B}_w$$

Then, all of the spectrum of $\mathcal{A}$ is discrete, so we get the same convergence of spectrum as we do for finite dimensional operators.

# Numerics for quasi-compact operators

Then we had, in operator norm,

$$\mathcal{B}_w \underbrace{\xrightarrow{\mathcal{A}}}_{\leq C} \mathcal{B}_s \underbrace{\xrightarrow{\mathcal{P}_N - \mathrm{id}}}_{\leq \epsilon(N)} \mathcal{B}_w$$

meaning

$$\mathcal{B}_w \underbrace{\xrightarrow{\mathcal{A}_N - \mathcal{A}}}_{\leq C\epsilon(N)} \mathcal{B}_w$$

Then, all of the spectrum of $\mathcal{A}$ is discrete, so we get the same convergence of spectrum as we do for finite dimensional operators.

# Numerics for quasi-compact operators

On the other hand, suppose we just had a contraction:

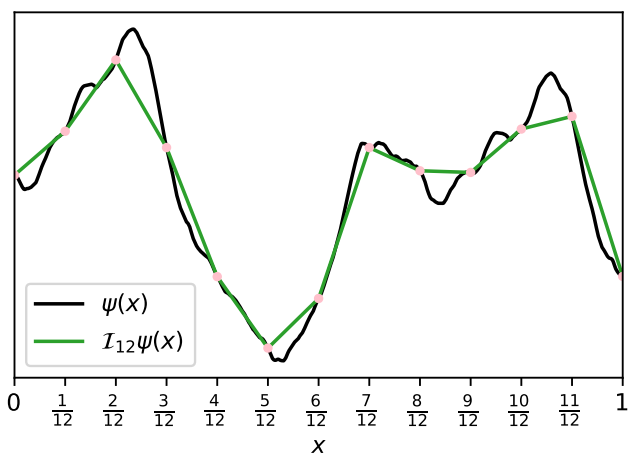
$$\|\mathcal{A}^n\|_{\mathcal{B}_s} \leq Cm^n$$

If $\mathcal{P}_N$ *played nice* in $\mathcal{B}_s$, we could maybe hope that

$$\|(\mathcal{P}_N\mathcal{A})^n\|_{\mathcal{B}_s} \leq C(m+\epsilon(n))^n$$

so that the spectrum of $\mathcal{P}_N\mathcal{A}$ outside $m$ converged to that of $\mathcal{A}$ (i.e. nothing...)

Obviously a basic requirement is that $\mathcal{P}_N$ maps $\mathcal{B}_s$ functions into $\mathcal{B}_s$. This is not always true:



# Linear interpolation again

It turns out you can combine these two ideas to get convergence results for

**Theorem (Keller–Liverani '01)**

*Suppose that*

- $\mathcal{A}$, $\mathcal{P}_N\mathcal{A}$ *all satisfy the same Lasota–Yorke inequalities*
- $\|\mathcal{P}_N - \mathrm{id}\,\|_{\mathcal{B}^s \to \mathcal{B}^w} \to 0$.

*Then the spectrum of* $\mathcal{P}_N\mathcal{A}$ *for* $|\lambda| > m$ *converge to that of* $\mathcal{A}$, *with*

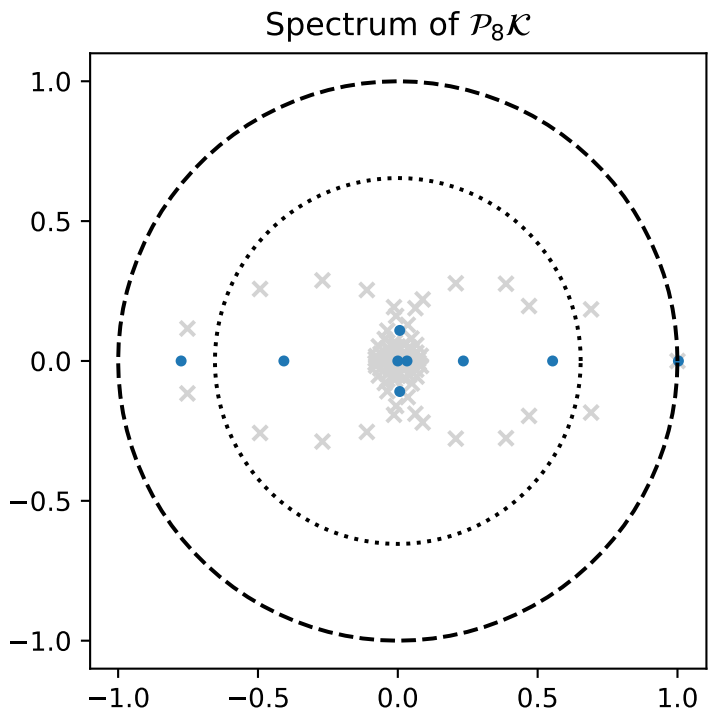
$$|\lambda_N - \lambda|, \|v_N - v\|_{\mathcal{B}^w} \leq \mathcal{O}\left( (\|\mathcal{P}_N - \mathrm{id}\,\|_{\mathcal{B}^s \to \mathcal{B}^w})^{1 - \log\lambda/\log m} \right).$$

In practice (e.g. for EDMD), it is not clear that you can get $\mathcal{P}_N\mathcal{A}$ to satisfy L–Y: maybe there are other ways to prove convergence?

# Linear interpolation again

Let's try and apply our linear interpolation scheme (from way back) to a transfer operator of

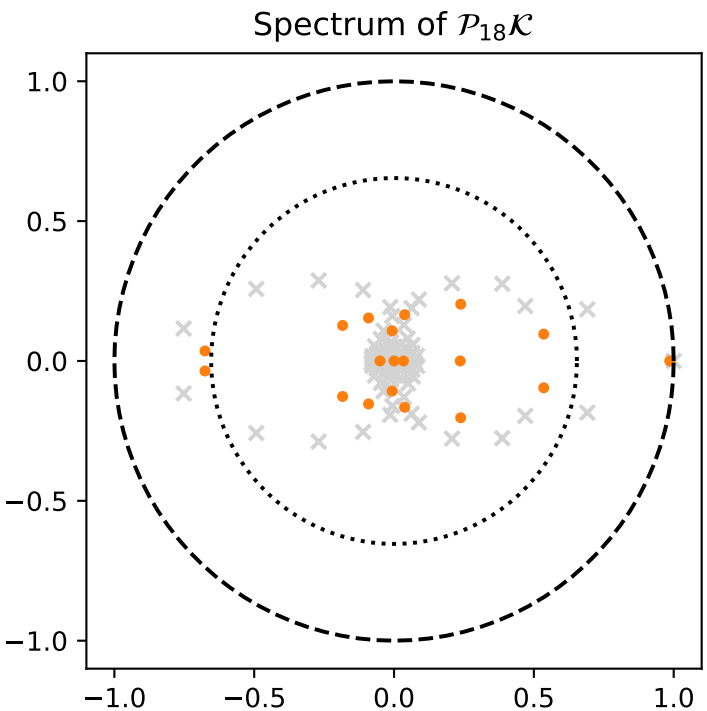
$$f(x) = 2x + 0.2\,(cos(6\pi x) + \sin(6\pi x) - 1)) \mod 1$$



# Linear interpolation again

Let's try and apply our linear interpolation scheme (from way back) to a transfer operator of

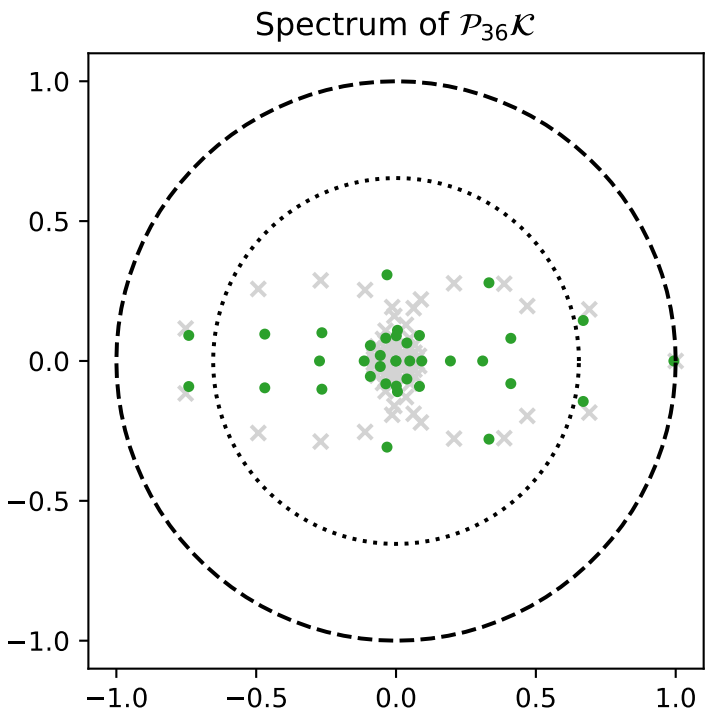
$$f(x) = 2x + 0.2\,(cos(6\pi x) + \sin(6\pi x) - 1)) \mod 1$$



# Linear interpolation again

Let's try and apply our linear interpolation scheme (from way back) to a transfer operator of

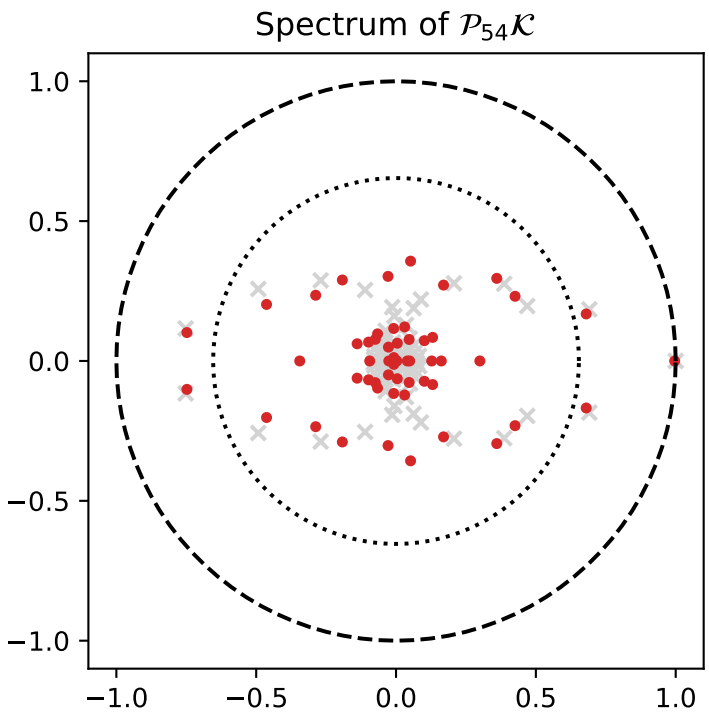
$$f(x) = 2x + 0.2\,(cos(6\pi x) + \sin(6\pi x) - 1)) \mod 1$$



# Linear interpolation again

Let's try and apply our linear interpolation scheme (from way back) to a transfer operator of

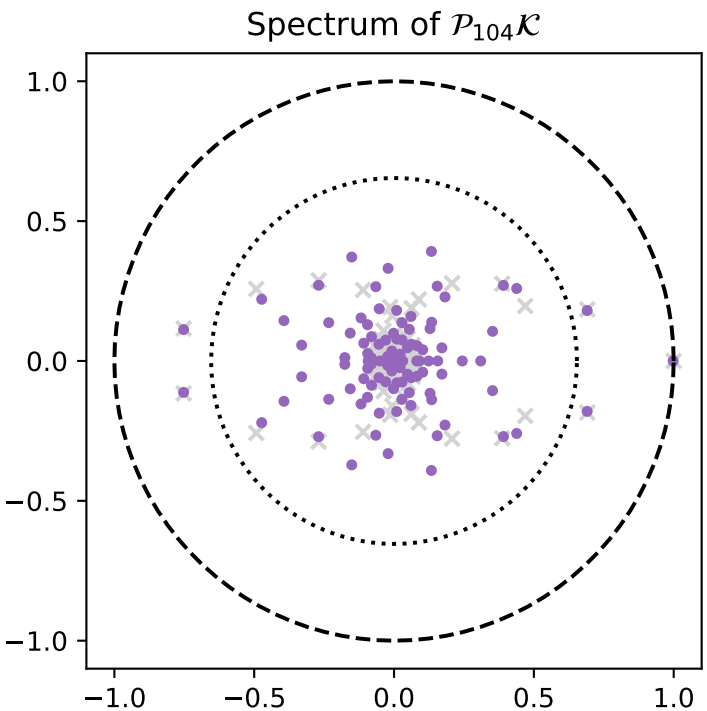
$$f(x) = 2x + 0.2\,(cos(6\pi x) + \sin(6\pi x) - 1)) \mod 1$$



# Linear interpolation again

Let's try and apply our linear interpolation scheme (from way back) to a transfer operator of

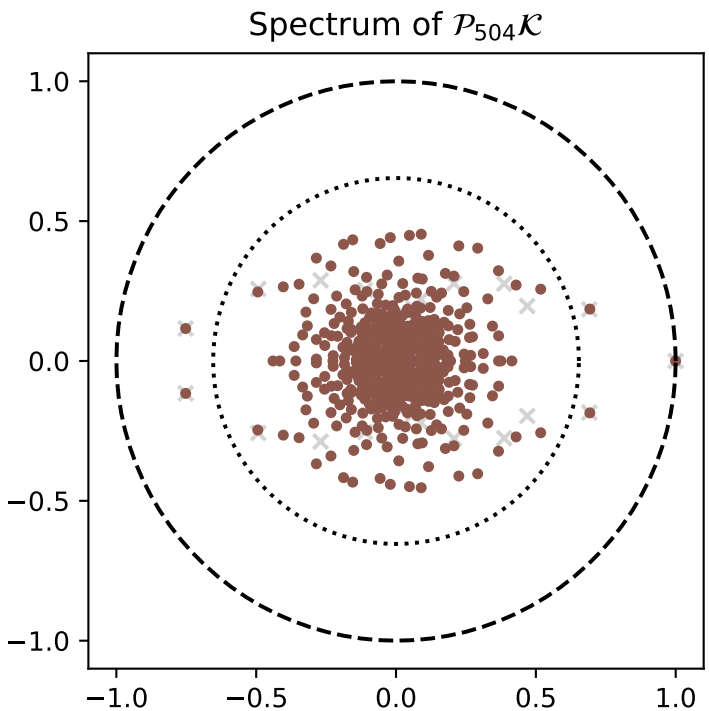
$$f(x) = 2x + 0.2\,(cos(6\pi x) + \sin(6\pi x) - 1)) \mod 1$$



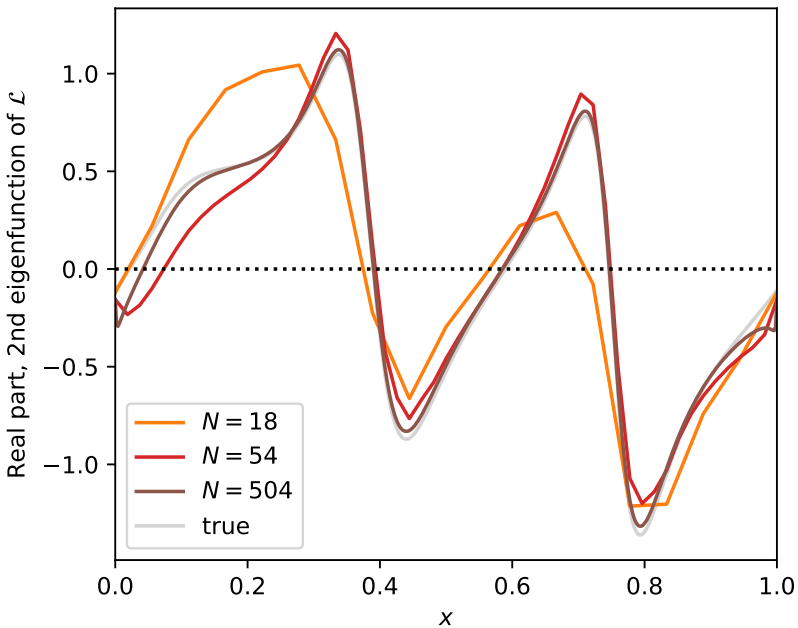
# Linear interpolation again

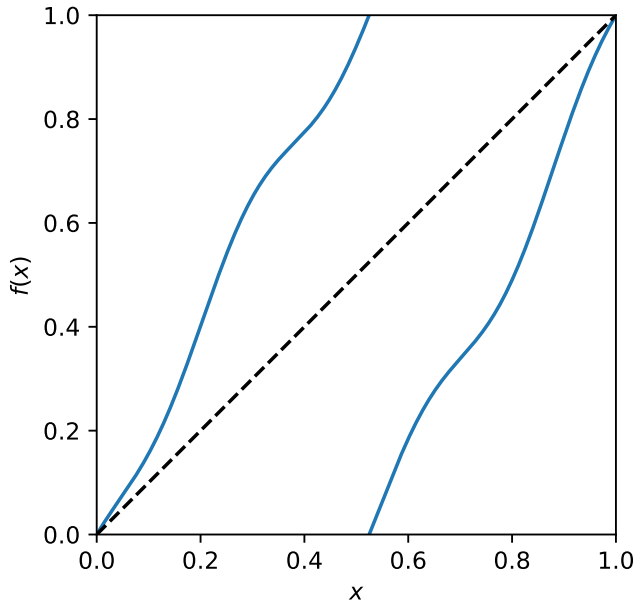
Let's try and apply our linear interpolation scheme (from way back) to a transfer operator of

$$f(x) = 2x + 0.2\,(cos(6\pi x) + \sin(6\pi x) - 1)) \mod 1$$



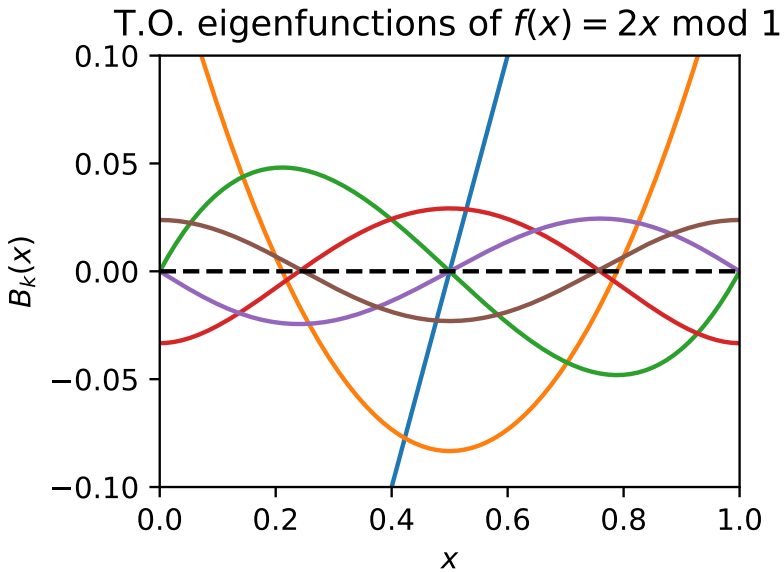
# Computational example



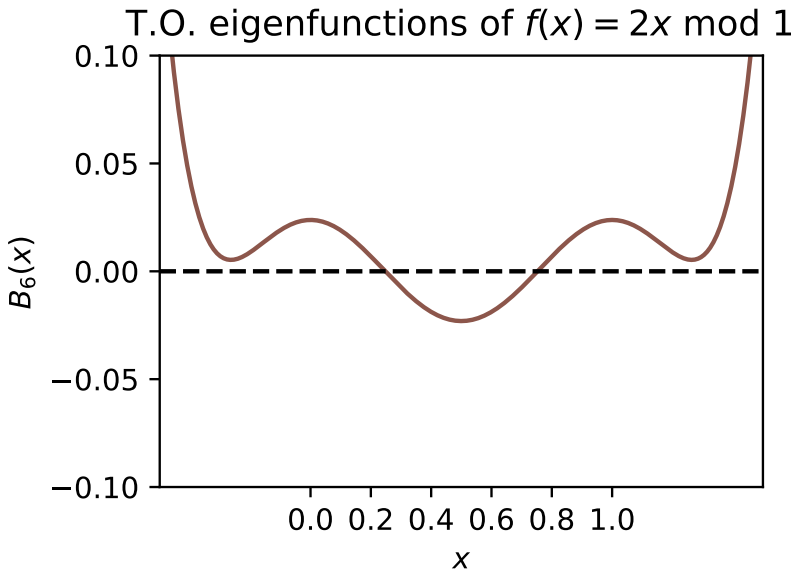
# Constants

Keller-Liverani and similar give asymptotic convergence rates of eigenfunctions, but no explicit constants.
Explicit constants usually linked to explicit bounds on the norm of the resolvent near eigenvalues. Usually this means $\mathcal{B}$ is a Hilbert space with $\mathcal{A}$ normal (i.e. orthogonal eigenfunctions).
However, quasicompact transfer operators in chaos are usually very non-normal.



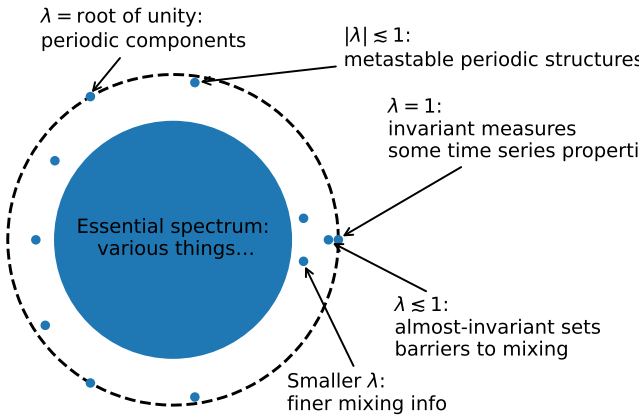
# Constants

Keller-Liverani and similar give asymptotic convergence rates of eigenfunctions, but no explicit constants.
Explicit constants usually linked to explicit bounds on the norm of the resolvent near eigenvalues. Usually this means $\mathcal{B}$ is a Hilbert space with $\mathcal{A}$ normal (i.e. orthogonal eigenfunctions).
However, quasicompact transfer operators in chaos are usually very non-normal.



# Which eigenvalues are geometrically meaningful?

This is an important question! It determines how "differentiable" our function spaces need to be.




Well, it depends on what you mean by "meaningful". Let's say, eigenvalues that describe sets with particularly slow escape rates.

# Which eigenvalues are geometrically meaningful?

This is an interesting question we don't have much of an answer to currently...
For 1D chaos, probably everything with $|\lambda| > 1/\gamma_\infty$, where

$$\gamma_\infty = \inf_{x \in D} \liminf_{n \to \infty} |(f^n)'(x)|^{1/n}$$

i.e. the minimum long-term expansion rate. (c.f. Dellnitz 2000)

# Which eigenvalues are geometrically meaningful?

Why?

- There are an infinite number of periodic orbits $\{p, f(p), f^2(p), \ldots, f^{T-1}(p)\}$ which have $L_{\exp}(p) = |(f^T)'(p)|^{1/T}$ arbitrarily close to $\gamma_\infty$.
- For any small enough $\epsilon$ we can take the set

$$E = \bigcup_{t=0}^{T-1} B\left(f^{t-1}(p), \frac{|(f^{t-1})'(p)|}{L_{\exp}(p)^t}\epsilon\right)$$

  and find that the proportion of measure that stays in $E$ over time $T$ is $\mathcal{O}(L_{\exp}(p)^{-T})\ldots$

This suggests that probably approximation schemes with differentiability order 1 are good enough for the job.
Rather strangely, this includes **Ulam's method** (see worksheet).

# Conceptual framework for actually doing discretisations?

We usually approximate transfer/Koopman operators by a projection: $\mathcal{P}_N\mathcal{K}$ or $\mathcal{P}_N\mathcal{L}$. What kinds of projections are there?

# Conceptual framework for actually doing discretisations?

We usually approximate transfer/Koopman operators by a projection: $\mathcal{P}_N\mathcal{K}$ or $\mathcal{P}_N\mathcal{L}$. What kinds of projections are there?

**Fact**: basically every projection commonly used in practice is an $L^2(\nu)$ least squares projection with respect to some reference measure $\nu$.
This is because:

1. They are very easy to calculate.
2. By and large, you still approximate a quasicompact operator *even though the projection does not necessarily respect the Banach space*.

# Why $L^2$?

Our projection needs to be finite-dimensional. Let's try and determine $\mathcal{P}_N$ by its range:

$$\operatorname{im} \mathcal{P}_N =: E_N = \operatorname{span}\{\psi_0, \psi_1, \ldots, \psi_{N-1}\}.$$

Let's suppose that for any $\omega \in \mathcal{B}$ we want $\mathcal{P}_N\omega$ to be the function in $E_N$ that minimises

$$\mathcal{P}_N\omega = \operatorname{argmin}_{\psi \in E_N} \|\omega - \psi\|_{\mathcal{B}}.$$

Unfortunately, in a general Banach space:

- such a $\mathcal{P}_N\omega$ may be hard or impossible to compute;
- $\mathcal{P}_N : \mathcal{B} \to \mathcal{B}$ is not linear in general.

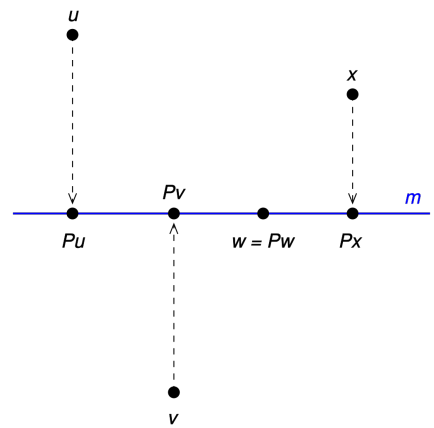
# Projections in a Hilbert space

However, in a Hilbert space (i.e. a Banach space with an inner product), we have no problems in either case.
Let's choose a Hilbert space $\mathcal{H}$ (not necessarily $\mathcal{B}$), with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. Again, we define

$$\mathcal{P}_N \omega = \text{argmin}_{\psi \in E_N} \|\omega - \psi\|_{\mathcal{H}}.$$

With an inner product, we can concieve of "angles". Hence the error $\omega - \mathcal{P}_N \omega$ must be orthogonal to $E_N$.

$$\langle \psi, \mathcal{P}_N \omega - \omega \rangle_{\mathcal{H}} = 0$$



# Projections in a Hilbert space

$$\langle \psi, \mathcal{P}_N \omega - \omega \rangle_{\mathcal{H}} = 0$$

This gives us an easy way to compute $\mathcal{P}_N \omega$. Writing

$$\mathcal{P}_N \omega = \sum_{n=0}^{N-1} a_n \psi_n,$$

we have for all $l = 0, \ldots, N-1$,

$$0 = \langle \psi_l, \mathcal{P}_N \omega - \omega \rangle_{\mathcal{H}} = \sum_{n=0}^{N-1} a_n \langle \psi_n, \psi_l \rangle_{\mathcal{H}} - \langle \psi_l, \omega \rangle_{\mathcal{H}}.$$

If we define the matrix $C_{XX} = (\langle \psi_n, \psi_l \rangle_{\mathcal{H}})_{n,l}$, then this is a vector equation solvable by

$$\mathbf{b} = C_{XX}^{-1} (\langle \psi_l, w \rangle_{\mathcal{H}})_l$$

# Projections in a Hilbert space

$$\mathbf{b} = C_{XX}^{-1}(\langle \psi_l, w \rangle_{\mathcal{H}})_l$$

This is linear! In fact, if we define the *row vector* of functions

$$\mathbf{\Psi}_X(x) = \begin{pmatrix} \psi_0(x) & \psi_1(x) & \cdots & \psi_{N-1}(x) \end{pmatrix},$$

we can write

$$\mathcal{P}_N \omega = \mathbf{\Psi}_X \mathbf{a}$$

where

$$\mathbf{b} = \langle \mathbf{\Psi}_X^\top, \mathbf{\Psi}_X \rangle_{\mathcal{H}}^{-1} \langle \mathbf{\Psi}_X^\top, \omega \rangle_{\mathcal{H}}$$

# Projections in a Hilbert space

Then, our operator $\mathcal{A}_N = \mathcal{P}_N \mathcal{A}$ has the following expression:

$$\mathcal{P}_N \mathcal{A}\psi = \mathbf{\Psi}_X \langle \mathbf{\Psi}_X^\top, \mathbf{\Psi}_X \rangle_{\mathcal{H}}^{-1} \langle \mathbf{\Psi}_X^\top, \mathcal{A}\psi \rangle_{\mathcal{H}}$$

If we restrict the domain of our operator to $E_N$ (!!), $\psi = \mathbf{\Psi}_X \mathbf{a}$ for some vector of coefficients $\mathbf{a}$, and we get the relationship

$$A_N \mathbf{a} := \mathbf{b} = \underbrace{\langle \mathbf{\Psi}_X^\top, \mathbf{\Psi}_X \rangle_{\mathcal{H}}^{-1}}_{C_{XX}^{-1}} \underbrace{\langle \mathbf{\Psi}_X^\top, \mathcal{A}\mathbf{\Psi}_X \rangle_{\mathcal{H}}}_{C_{XY}} \mathbf{a}$$

# Projections in a Hilbert space

This works for any Hilbert space $\mathcal{H}$.
However, computing inner products can be computationally intensive. Suppose we wanted to compute in $\mathcal{H} = H^1$ ($L^2$ functions with $L^2$ derivatives). This has inner product:

$$\langle \psi, \omega \rangle_{H^1} = \int_D \overline{\psi(x)} \omega(x) \, \mathrm{d}x + \int_D \overline{\nabla \psi(x)} \cdot \nabla \omega(x) \, \mathrm{d}x.$$

For the Koopman operator, elements of $C_{XY}$ have the form:

$$\langle \psi_l, \mathcal{K} \psi_n \rangle_{H^1} = \int_D \overline{\psi_l(x)} \, \psi_n(f(x)) \, \mathrm{d}x + \int_D \overline{\psi_l} \cdot Df(x)^* \, \nabla \psi_n(f(x)) \, \mathrm{d}x$$

So, you need to know:

1. The map $f$ (potentially unknown),
2. The Jacobian $Df$ (potentially unknown),
3. Integrals involving the above, over the whole domain (time-consuming).

# Projections in a Hilbert space

We can fix 2. by just looking at $L^2(\nu_M)$ for some reference measure $\nu_M$. Then, elements of $C_{XY}$ have the form:

$$\langle \psi_l, \mathcal{K}\psi_n \rangle_{L^2(\nu_M)} = \int_D \overline{\psi_l(x)}\, \psi_n(f(x))\, \mathrm{d}\nu_M(x)$$

We have

1. The map $f$,
2. ~~The Jacobian $Df$,~~
3. Integrals involving the above, over the support of $\nu$

What are the nicest measures to take integrals against? **Discrete measures.**

# Projections in a Hilbert space

Suppose we can write our measure

$$\nu_M(x) = \sum_{m=1}^{M} w_m \delta(x - x_m).$$

Then

$$\langle \psi_l, \mathcal{K}\psi_n \rangle_{L^2(\nu_M)} = \sum_{m=1}^{M} w_m \overline{\psi_l(x_m)}\, \psi_n(f(x_m)).$$

# Projections in a Hilbert space

In fact, if we set

$$\Psi_X = \begin{pmatrix} \psi_1(x_1) & \psi_1(x_2) & \cdots & \psi_1(x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_M(x_1) & \psi_M(x_2) & \cdots & \psi_M(x_1) \end{pmatrix}$$

$$\Psi_Y = \begin{pmatrix} \psi_1(f(x_1)) & \psi_1(f(x_2)) & \cdots & \psi_1(f(x_N)) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_M(f(x_1)) & \psi_M(f(x_2)) & \cdots & \psi_M(f(x_1)) \end{pmatrix}$$

$$W = \mathrm{diag}(w_m)_m,$$

we get Koopman operator matrix (i.e. a computer representation of $\mathcal{P}_{N,M}\mathcal{K}$ in $\psi_n$ basis)

$$K_{N,M} = \underbrace{(\Psi_X^* W \Psi_X)^{-1}}_{C_{XX}^{-1}} \underbrace{(\Psi_X^* W \Psi_Y)}_{C_{XY}}$$

# Least squares projection

These are the ingredients:

- A finite-dimensional space $E_N \subset \mathcal{B}$ of functions $D \to \mathbb{R}$, with basis $\{\psi_0, \psi_1, \ldots, \psi_N\}$;
- A reference measure $\nu_M$, ideally discrete
- The ability to evaluate your operator $(\mathcal{A}\psi)(x)$ for chosen functions $\psi \in \mathcal{B}$ and points $x \in D$.

The recipe:

$$\Psi_X = (\psi_n(x_m))_{m,n},\ \Psi_Y = (\mathcal{A}\psi_n(x_m))_{m,n},\ W = \mathrm{diag}(w_m)_m,$$

$$A_{N,M} = (\Psi_X^* W \Psi_X)^{-1}(\Psi_X^* W \Psi_Y) = (\sqrt{W}\Psi_X)^+ \sqrt{W}\Psi_Y$$

$\Psi_X^+ =$ pseudoinverse of $\Psi_X$

# Pseudoinverse

(Let's take $W = \mathrm{id}$ for now)
Pseudoinverse is given by

$$\Psi_X^+ = \underbrace{(\Psi_X^* \Psi_X)^{-1}}_{C_{XX}^{-1}} \Psi_X^*$$

... if $C_{XX}$ is invertible.
More generally, $\Psi_X^+ \mathbf{a}$ is the unique vector $\mathbf{v}$ selected by the following process:

- Find the minimisers of $\|\Psi_X \mathbf{v} - \mathbf{a}\|_{\ell^2}$ (i.e. $L^2(\nu_M)$ error of operator approximation)
- Choose the one that minimises $\|\mathbf{v}\|_{\ell^2}$.

# Pseudoinverse

We can compute it by singular value decomposition:

$$\Psi_X = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^*$$

$$\Psi_X^+ = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^*$$

# Pseudoinverse

We can compute it by singular value decomposition:

$$\underbrace{\Psi_X}_{M\times N} = \underbrace{U}_{M\times N} \underbrace{\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}}_{N\times N} \underbrace{V^*}_{N\times N}$$

$$\underbrace{\Psi_X^+}_{N\times M} = \underbrace{V}_{N\times N} \underbrace{\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix}}_{N\times N} \underbrace{U^*}_{N\times M}$$

# Least squares projection

Pretty much every numerical method is a least squares method:

| Name | Op. | $\psi_n$ | $\nu_M$ | Limiting $\nu$ | $C_{XX}$ diag. | Sparse | Examples |
|---|---|---|---|---|---|---|---|
| DMD | $\mathcal{K}$ | linear variables | empirical | ergodic | N | Y | Stefan's talk |
| Extended DMD | $\mathcal{K}$ | polynomials[1] | empirical[1] | ergodic[1] | N | N | Stefan's talk, Williams *et al* '15 |
| Ulam's method | $\mathcal{K}^1$ | $\{\mathbb{1}_E\}_{E\in P}$ | varies | Lebesgue | Y | Y | GAIO (Dellnitz *et al.* '01) |
| Linear interpolation | $\mathcal{K}$ $\mathcal{L}$ |  | dense ($M = N$) | N/A | Y | Y | this course |
| Lagrange–Chebyshev | $\mathcal{L}$ | Chebyshev polynomials | Chebyshev nodes | $\frac{\mathrm{d}x}{(1-x^2)^{1/2}}$ | Y | N | W. '19, Bandtlow–Slipantschuk '20 |
| Lagrange–Fourier | $\mathcal{L}$ | trig. series | Evenly spaced nodes | Lebesgue | Y | N | Froyland *et al.* 14, W. '19 |

[1] usually

# Extended Dynamical Mode Decomposition

As we have seen: you can do EDMD from a time series

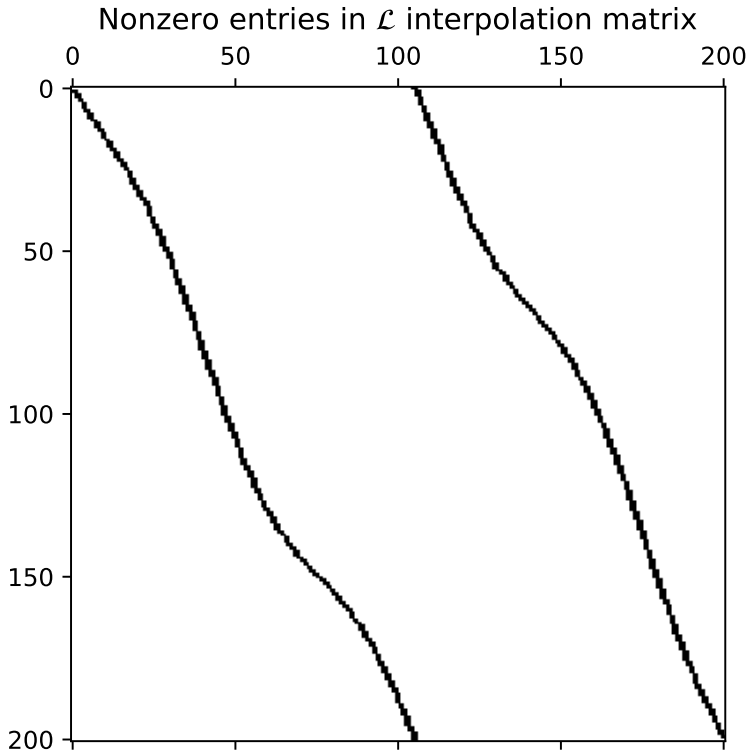
$$\{x_1,\, x_2 = f(x_1),\, x_3 = f(x_2), \ldots, x_{M+1} = f(x_M)\}$$

$$\Psi_X = \begin{pmatrix} \psi_1(x_1) & \psi_1(x_2) & \cdots & \psi_1(x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_M(x_1) & \psi_M(x_2) & \cdots & \psi_M(x_1) \end{pmatrix}$$

$$\Psi_Y = \begin{pmatrix} \psi_1(x_2) & \psi_1(x_3) & \cdots & \psi_1(x_{N+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_M(x_2) & \psi_M(x_3) & \cdots & \psi_M(x_{N+1}) \end{pmatrix}$$

# Sparseness

Sparse basis functions (e.g. Ulam, interpolation hats) tend to lead to sparse $\mathcal{K}, \mathcal{L}$ matrices:




They also tend to be well-conditioned (i.e. singular values of $C_{XX}$ tend to not be small).

Problem: curse of dimensionality.

# Getting transfer operator from Koopman and vice versa

It's very natural to want to get one from the other. If we have

$$\psi = \sum_n a_n \psi_n = \mathbf{\Psi a},\ \varphi = \sum_n b_n \psi_n = \mathbf{\Psi b},$$

we have that the $L^2(\nu_M)$ inner product matrix is given by $C_{XX}$:

$$\begin{aligned}\int_D \overline{\varphi}\psi \,\mathrm{d}\nu_M &= \int_D \mathbf{b}^* \mathbf{\Psi}^* \mathbf{\Psi a} \,\mathrm{d}\nu_M \\ &= \mathbf{b}^* C_{XX} \mathbf{a}\end{aligned}$$

Consequently, if $\nu_M$ approximates Lebesgue measure, the transfer operator (the adjoint of $\mathcal{K}$ wrt $L^2(\mathrm{d}x)$) is going to be represented by matrix

$$L_{M,N} = C_{XX} K_{M,N} C_{XX}^{-1} = C_{XX}^{-1} C_{XY}^*$$

# Getting transfer operator from Koopman and vice versa

What if $\nu$ is more general? Then $L_{M,N} = C_{XX}^{-1} C_{XY}^*$ represents the adjoint of $\mathcal{K}$ in $L^2(\nu)$, which in particular has

$$\int \varphi \mathcal{K} \psi \, \mathrm{d}\nu = \int \psi \mathcal{L}[\varphi\nu] = \int \psi \tfrac{\mathrm{d}}{\mathrm{d}\nu} \mathcal{L}[\varphi\nu] \, \mathrm{d}\nu$$

This is the so-called *Perron–Frobenius operator*.

# Getting transfer operator from Koopman and vice versa

What if $\nu$ is more general? Then $L_{M,N} = C_{XX}^{-1} C_{XY}^*$ represents the adjoint of $\mathcal{K}$ in $L^2(\nu)$, which in particular has

$$\int \varphi \mathcal{K} \psi \, \mathrm{d}\nu = \int \psi \mathcal{L}[\varphi \nu] = \int \psi \tfrac{\mathrm{d}}{\mathrm{d}\nu} \mathcal{L}[\varphi \nu] \, \mathrm{d}\nu$$

This is the so-called *Perron–Frobenius operator*.

# Perron–Frobenius operator

$$\mathbf{P}_\nu = \frac{\mathrm{d}}{\mathrm{d}\nu}\mathcal{L}[\varphi\nu]$$

- Usually called $\mathcal{P}$, but already in use.
- If $\nu$ has a density supported everywhere, then this is just the transfer operator times a function
- If $\nu \in \mathcal{B}$ (for example, $\nu$ is the *physical measure* $\mu$) then this is well-defined at least for nice $\varphi$.
- In fact, let $\mathcal{B}_\nu$ be the completion of $C^\infty(D)$ (etc.) under the norm

  $$\|\varphi\|_{\mathcal{B}_\nu} = \|\varphi\nu\|_{\mathcal{B}}.$$

  So $\nu\mathcal{B}_\nu$ is an invariant subspace of $\mathcal{B}$, and thus $\mathbf{P}_\nu$ has a subset of the same eigenvalues.
- $\mathbf{P}_\nu$ will miss eigenvalues associated to dynamics off $\operatorname{supp}\nu$—eg contraction onto invariant manifolds.

# Constructing these operators is tricky

Beware!
For an invertible map $f$ with invariant measure $\mu$, we can define for

$$\int_D \varphi \, \mathcal{K}\psi \, \mathrm{d}\mu = \int_D \varphi \, \psi \circ f \, \mathrm{d}\mu = \int_D \underbrace{\varphi \circ f^{-1}}_{\mathcal{L}_1 \varphi} \, \psi \, \mathrm{d}\mu$$

whereas the usual transfer operator $\mathcal{L} = \mathcal{L}_{1/|\det Df|}$, from which $\mathbf{P}_\mu$, has form

$$\mathcal{L}\varphi := (\mathcal{L}_{1/|\det Df|}\varphi) = |\det Df^{-1}| \varphi(f^{-1}(x))$$

When both considered on $\mathcal{B}$, these two operators can have different eigenvalues!

# Constructing these operators is tricky

Beware!
For an invertible map $f$ with invariant measure $\mu$, we can define for

$$\int_D \varphi \, \mathcal{K}\psi \, \mathrm{d}\mu = \int_D \varphi \, \psi \circ f \, \mathrm{d}\mu = \int_D \underbrace{\varphi \circ f^{-1}}_{\mathcal{L}_1 \varphi} \, \psi \, \mathrm{d}\mu$$

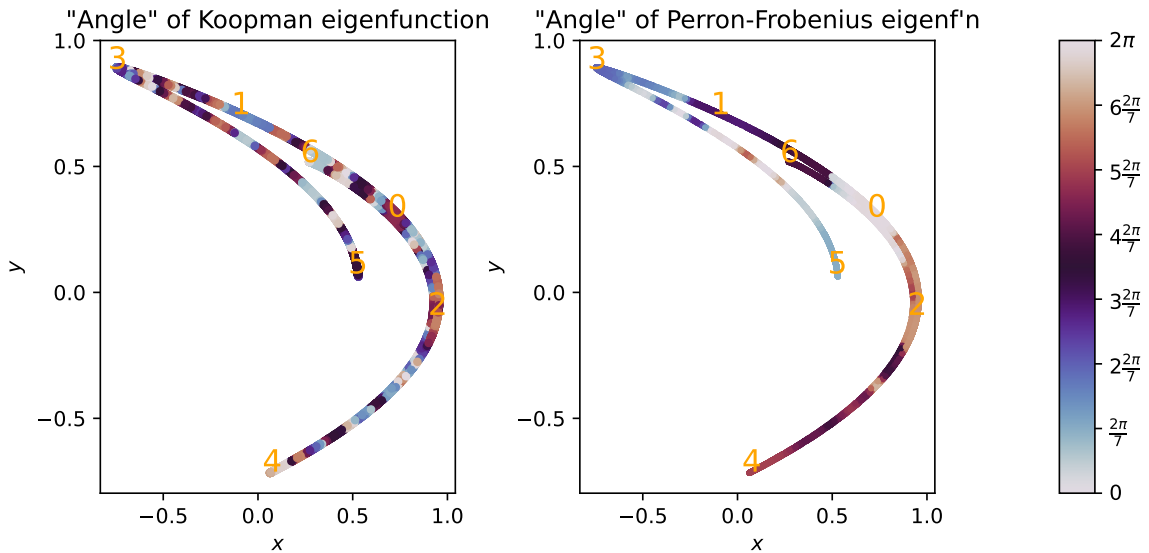
whereas the usual transfer operator $\mathcal{L} = \mathcal{L}_{1/|\det Df|}$, from which $\mathbf{P}_\mu$, has form

$$\mathcal{L}\varphi := (\mathcal{L}_{1/|\det Df|}\varphi) = |\det Df^{-1}| \varphi(f^{-1}(x))$$

When both considered on $\mathcal{B}$, these two operators can have different eigenvalues!
**Moral:** if the reference measure $\nu$ is singular, the weight of your operator may change.

# Getting transfer operator from Koopman and vice versa

- Recall that transfer/Perron-Frobenius operator eigenfunctions are generally smooth in the direction of expansion.
- If the system is dissipative this is along the length of the attractor.
- To some extent, this means that transfer operator eigenfunctions are easier to interpret geometrically. They are often called *Koopman modes* (vs. Koopman eigenfunctions).



# Convergence for EDMD type methods

Least-squares methods are numerically very nice, but can be theoretically harder to pin down.
What is the error of orthogonal projection onto $\{\psi_n\}$ with respect to $\mathcal{L}^2(\nu_M)$ (a discrete measure) from $\mathcal{B}_s \to \mathcal{B}_w$ (some possibly totally unrelated Banach spaces)?
Probably no single answer. Most existing work uses some special properties of the basis functions:

- Ulam's method results use that $\mathcal{P}_N$ is a contraction in $W^{1,1}$
- Fourier/Chebyshev polynomial results use that $\mathcal{P}_{N,M}$ is orthogonal in many different Hilbert spaces
- DMD is time series analysis (geometry independent)

# Convergence rates for different $\nu$

When $\nu$ has a density ($\mathrm{d}\nu = h\mathrm{d}x$) then we have
Suppose we're on $\mathbb{R}/2\pi\mathbb{Z}$. We can define Hilbert spaces of functions with certain Fourier decay rates:

$$\|\varphi\|_{H^\sigma}^2 = \sum_{n\in\mathbb{Z}} |\sigma(n)\hat{\varphi}(n)|^2$$

Examples:

- $\sigma(n) = 1$ gives $L^2$
- $\sigma(n) = (1+|n|^2)^{r/2}$ gives $W^{2,r}$, the space of $L^2$ functions with $r$th derivative in $L^2$

Let's assume $\sigma$ is increasing and $\sigma(n+m) \leq \sigma(n)\sigma(m)$.

# Convergence rates for different $\nu$

We know that truncating Fourier modes at order $N$ has:

$$\| \operatorname{id} - \mathcal{P}_N^{(\mathrm{d}x)} \|_{H^\sigma \to H^\tau} = \frac{\tau(N)}{\sigma(N)}$$

This is an $L^2(\mathrm{d}x)$ least squares projection onto the low-order Fourier modes.

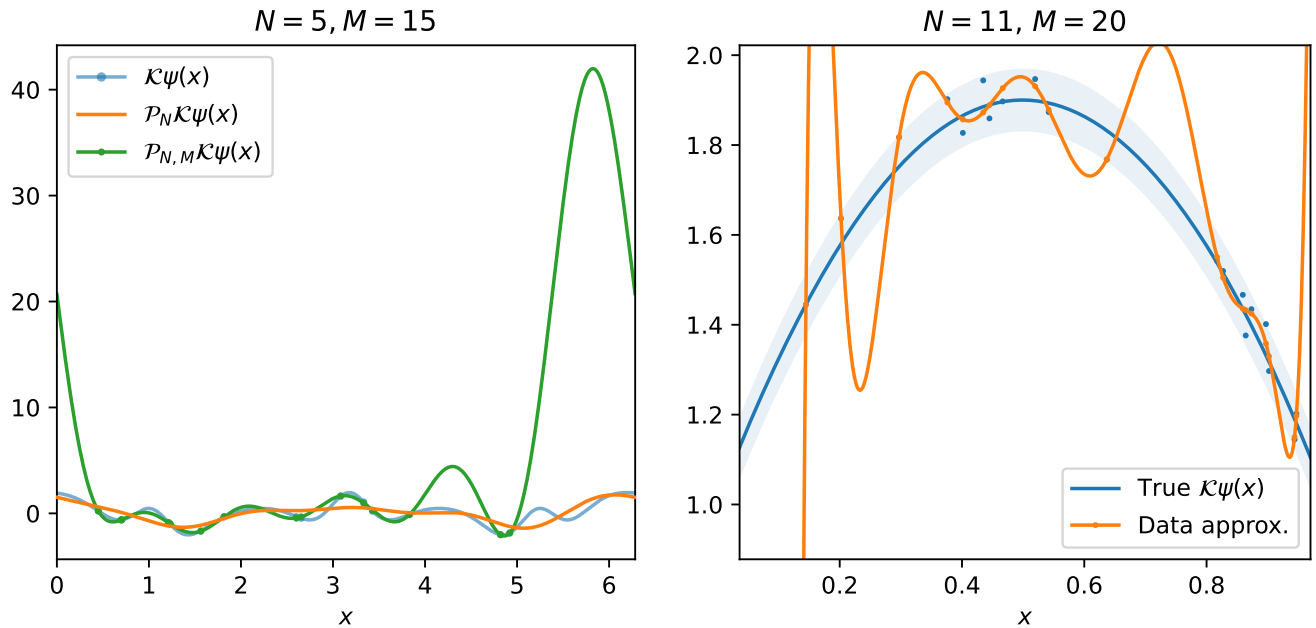
**Theorem (W. '24)**

*Suppose the density of $\nu$ lies in $H^\upsilon$. Then there exists $C$ such that for any $\tau \leq \sigma \leq \upsilon/n$,*

$$\| \operatorname{id} - \mathcal{P}_N^{(\mathrm{d}\nu)} \|_{H^\sigma \to H^\tau} \leq C \| \operatorname{id} - \mathcal{P}_N^{(\mathrm{d}x)} \|_{H^\sigma \to H^\tau}$$

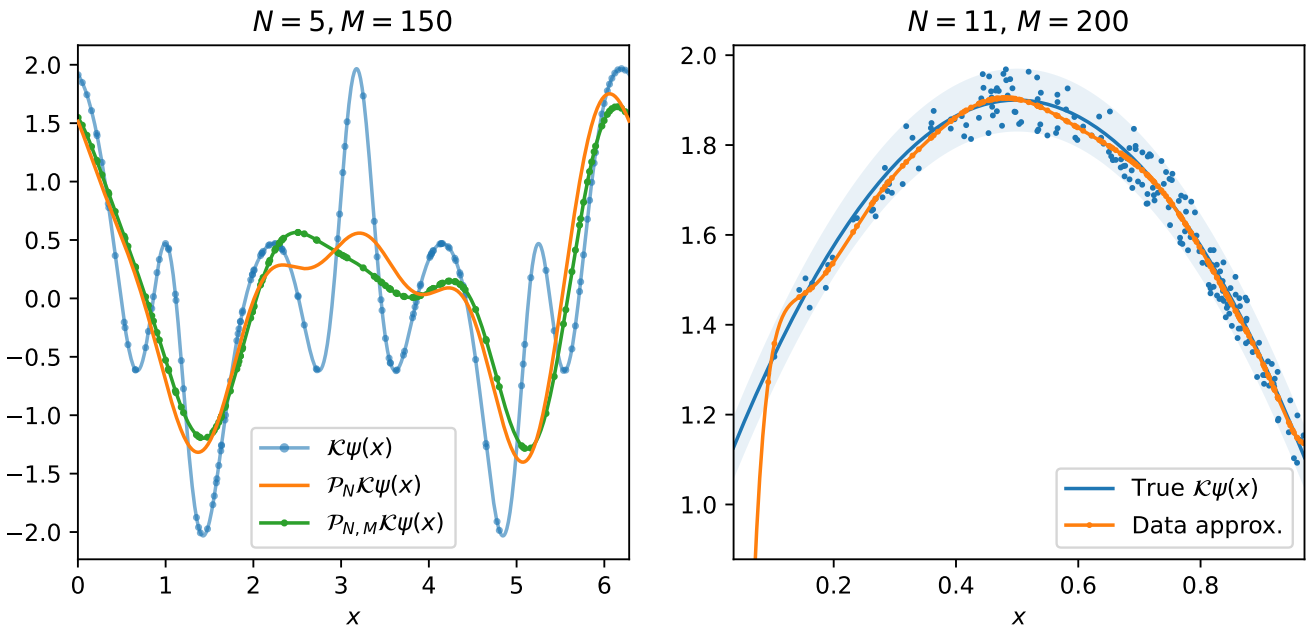
# Convergence rates for data

What happens if we try and fill things from data?
Well, remember that our internal spectrum only appears when we're working with some level of differentiability.
So we need to be able to approximate derivatives (even if they are only fractional):



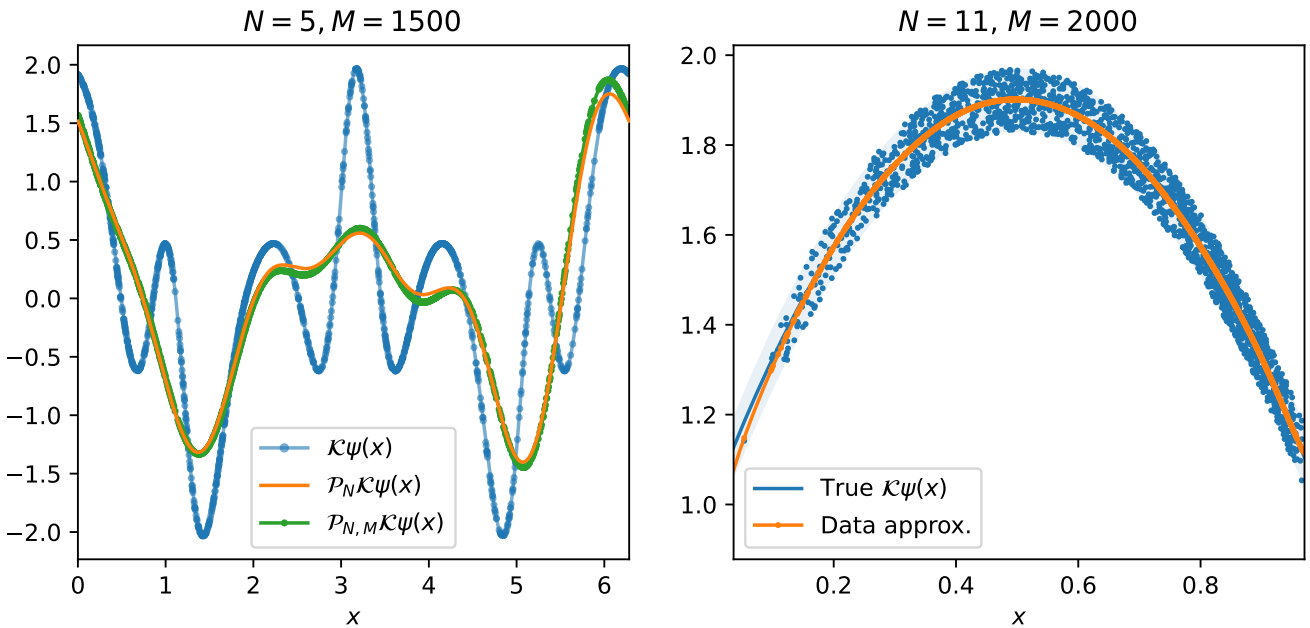
# Convergence rates for data

What happens if we try and fill things from data?
Well, remember that our internal spectrum only appears when we're working with some level of differentiability.
So we need to be able to approximate derivatives (even if they are only fractional):



# Convergence rates for data

What happens if we try and fill things from data?
Well, remember that our internal spectrum only appears when we're working with some level of differentiability.
So we need to be able to approximate derivatives (even if they are only fractional):



# Convergence rates for data

**Modelling assumption (mathematically justified):** data obtained from a chaotic system are randomly distributed $\sim \mu$.

We can think of least squares approximation as something akin to a kernel approximation, which in 1D looks like:

$$\psi(x) \approx \tilde{\psi}_{w,M}(x) := \frac{1}{M} \sum_{m=1}^{M} \frac{1}{w} k(w(x - x_m))\, \psi(x_m)$$

where $w$ is the resolution of the approximation ($\sim 1/N$).
We can estimate the variance as

$$\begin{aligned}
\mathbb{V}[\tilde{\psi}_{w,M}(x)] &\sim \mathbb{E}[\tilde{\psi}_{w,M}(x)^2] \\
&\sim \frac{1}{M^2} \sum_{m=1}^{M} \int w^{-2} |k(w(x - x_m))|^2\, \psi(x_m)|^2 \, \mathrm{d}\nu(x_m) \\
&= \mathcal{O}(w^{-1}/M)
\end{aligned}$$

which should be small if $N \gg M$.

# Convergence rates for data

But, if we want to estimate the derivative, we have

$$\psi'(x) \approx_? \tilde{\psi}'_{w,M}(x) = \frac{1}{M} \sum_{m=1}^{M} w^{-4} k'(w(x - x_m))\psi(x_m)$$

So

$$\begin{aligned}\mathbb{V}[\tilde{\psi}'_{w,M}(x)] &\sim \frac{1}{M^2} \sum_{m=1}^{M} \int w^{-4} |k(w(x - x_m))|^2 \psi(x_m)|^2 \, \mathrm{d}\mu(x_m) \\ &= \mathcal{O}(w^{-3}/M).\end{aligned}$$

which should be small if $N \gg M^3$.

# Convergence rates for data

But, if we want to estimate the derivative, we have

$$\psi^{(r)}(x) \approx_? \tilde{\psi}^{(r)}_{w,M}(x) = \frac{1}{M}\sum_{m=1}^{M} w^{-2+2r} k^{(r)}(w(x-x_m))\psi(x_m)$$

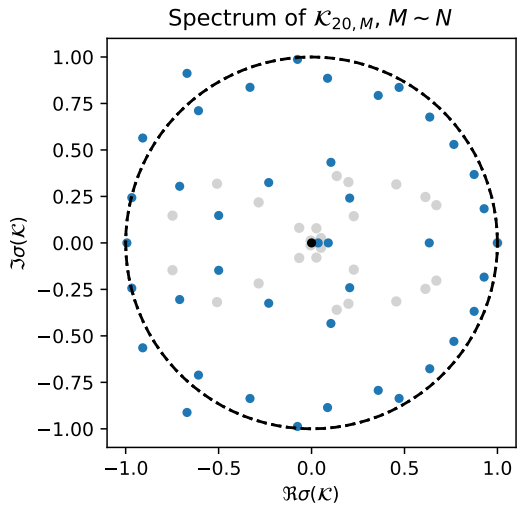
So

$$\begin{aligned}\mathbb{V}[\tilde{\psi}^{(r)}_{w,M}(x)] &\sim \frac{1}{M^2}\sum_{m=1}^{M}\int w^{-2+2r}|k(w(x-x_m))|^2\psi(x_m)|^2\,\mathrm{d}\mu(x_m)\\ &= \mathcal{O}(w^{-1+2r}/M).\end{aligned}$$

which should be small if $N \gg M^{1+2r}$.

# Convergence rates for data

How far we can see the centre of the spectrum depends on how much data we get. . .



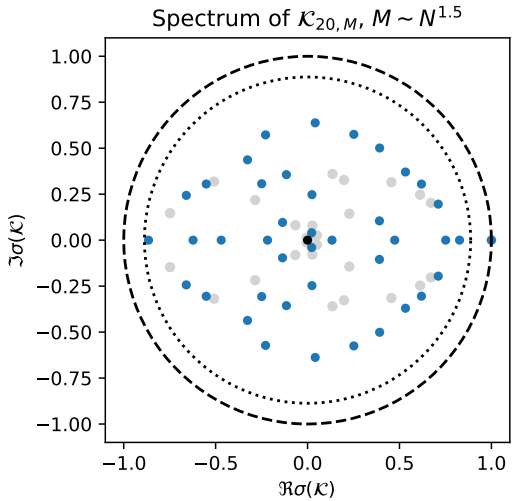

Of course there are practical limits:

- Most systems have "non-removable" essential spectrum;
- The further inside the unit circle, the less physically meaningful.

Perhaps you only need $M \gg N^{1+2\alpha}$ for some $\alpha \in (0, 1/2)$?

# Convergence rates for data

How far we can see the centre of the spectrum depends on how much data we get. . .



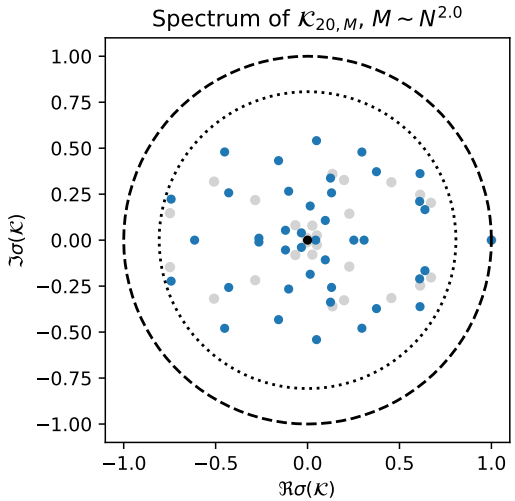

Of course there are practical limits:

- Most systems have "non-removable" essential spectrum;
- The further inside the unit circle, the less physically meaningful.

Perhaps you only need $M \gg N^{1+2\alpha}$ for some $\alpha \in (0, 1/2)$?

# Convergence rates for data

How far we can see the centre of the spectrum depends on how much data we get. . .



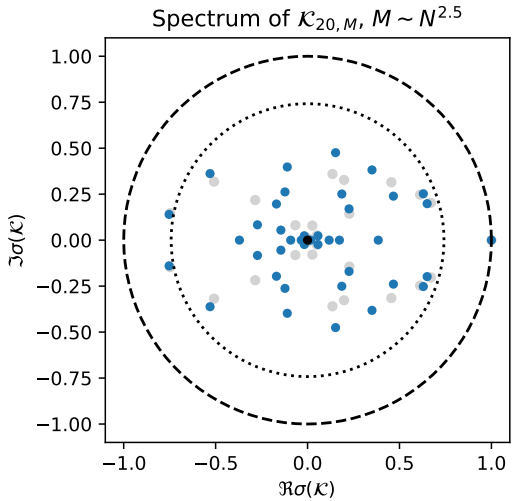

Of course there are practical limits:

- Most systems have "non-removable" essential spectrum;
- The further inside the unit circle, the less physically meaningful.

Perhaps you only need $M \gg N^{1+2\alpha}$ for some $\alpha \in (0, 1/2)$?

# Convergence rates for data

How far we can see the centre of the spectrum depends on how much data we get. . .



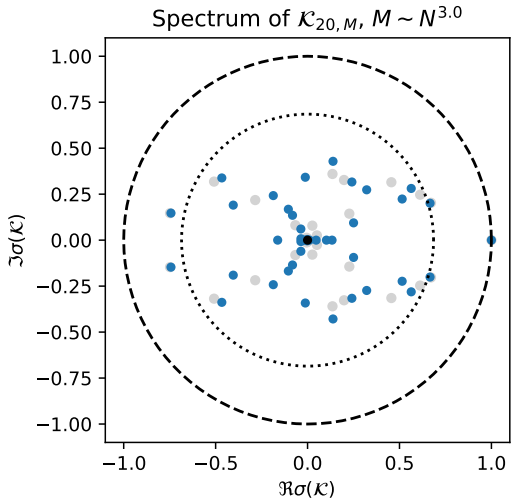

Of course there are practical limits:

- Most systems have "non-removable" essential spectrum;
- The further inside the unit circle, the less physically meaningful.

Perhaps you only need $M \gg N^{1+2\alpha}$ for some $\alpha \in (0, 1/2)$?

# Convergence rates for data

How far we can see the centre of the spectrum depends on how much data we get. . .




Of course there are practical limits:

- Most systems have "non-removable" essential spectrum;
- The further inside the unit circle, the less physically meaningful.

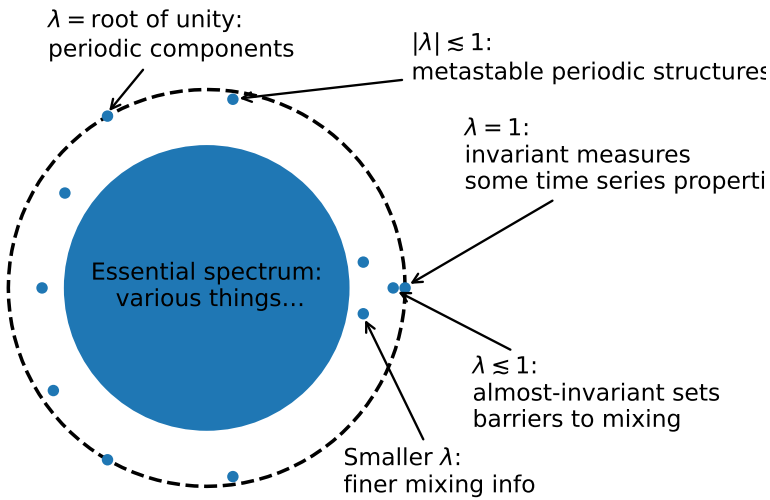
Perhaps you only need $M \gg N^{1+2\alpha}$ for some $\alpha \in (0, 1/2)$?

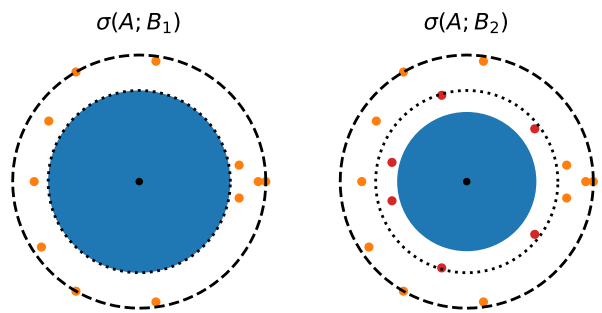
# Convergence rates for data

- This suggests that, in an underfitting regime, you need a lot of data to get a (quantitatively) good answer.
- But if you had that much data to hand, probably you already know the system (or could use SINDy).
- If you do know the system, better to use sample points that are good for approximating derivatives (e.g. Chebyshev nodes on cubes, evenly-spaced on torus. . . )

# Conclusion

- Discrete-time Koopman and transfer operators have quasicompact spectrum (=essential spectrum $\sigma_{\mathrm{ess}}$ in a ball $B(0, \rho_{\mathrm{ess}})$ plus outlying eigenvalues ($\sigma_{\mathrm{d}}$).
- These eigenvalues are more or less independent of the function space, and give information about the system's emergent behaviour.



# Conclusion

- The essential spectrum depends on the Banach space you study $\mathcal{K}$ on.
  - Smaller $\rho_{\text{ess}}$ means you see more eigenvalues.
  - However, in most systems there is some non-removable essential spectrum.
- These Banach spaces depend on the dynamics, and tell you about the structure of the eigenfunctions.
- However, most numerics that are "not too smart" will pick up these spectrum/eigenfunctions, even if they are $L^2(\nu)$ projections—*no knowledge of the Banach space required*.
- However, the discrete spectrum is not very well-conditioned, especially when reconstructing from data.
  - The more physically relevant, the better-conditioned, as a rule of thumb.