Residual Dynamic Mode Decomposition

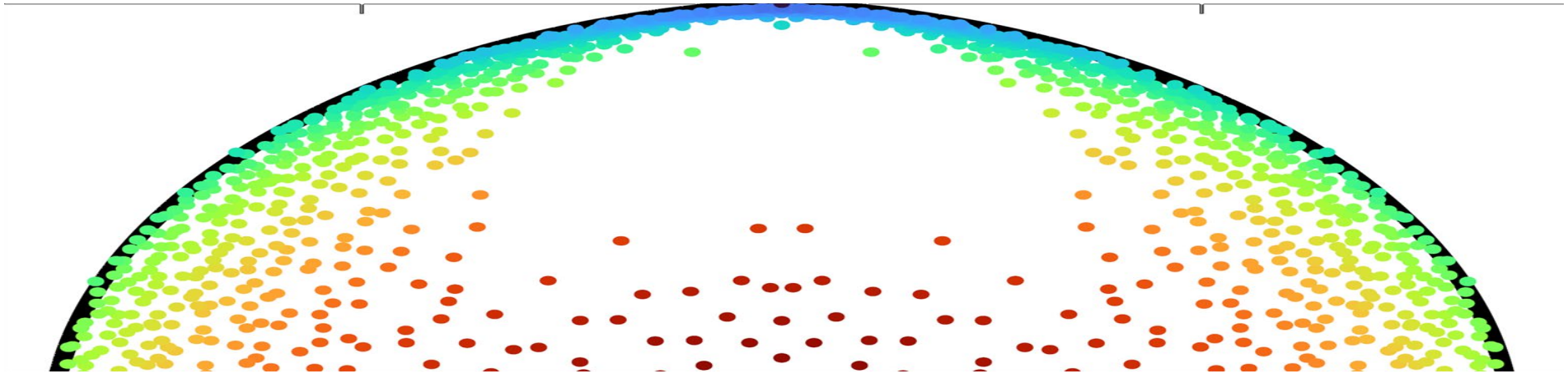
A path towards verified data-driven discovery of dynamics!

Matthew Colbrook (m.colbrook@damtp.cam.ac.uk)

University of Cambridge

Joint work with

Lorna Ayton (Cambridge), **Máté Szőke** (Virginia Tech), **Alex Townsend** (Cornell)



Math

C., Townsend, "*Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems*," preprint.

Structure-preserving method C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition*," **JFM**, 2023. Applications

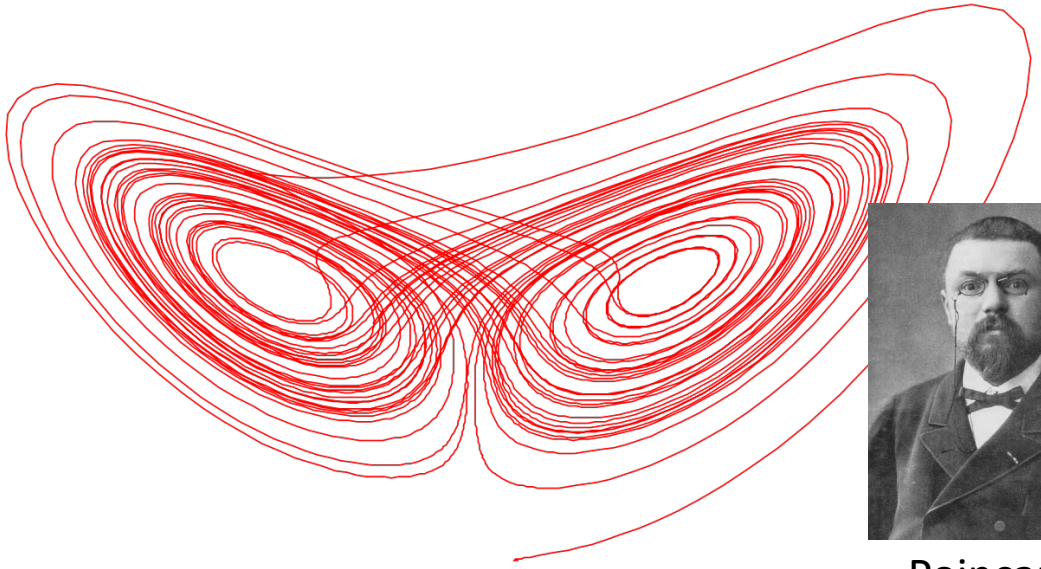
C., "*The mpEDMD Algorithm for Data-Driven Computations of Measure-Preserving Dynamical Systems*," **SINUM**, 2023.

Data-driven dynamical systems

- State $x \in \Omega \subseteq \mathbb{R}^d$, ***unknown*** function $F: \Omega \to \Omega$ governs dynamics

$$x_{n+1} = F(x_n)$$

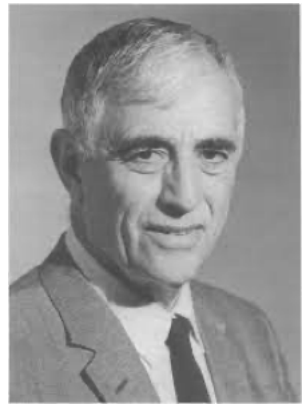
- **Goal:** Learn about system from data $\left\{x^{(m)}, y^{(m)} = F(x^{(m)})\right\}_{m=1}^{M}$
  - **Data:** experimental measurements or numerical simulations
  - E.g., **used for** forecasting, control, design, understanding
- **Applications:** chemistry, climatology, electronics, epidemiology, finance, fluids, molecular dynamics, neuroscience, plasmas, robotics, video processing, etc.



Poincaré

Operator viewpoint

Koopman

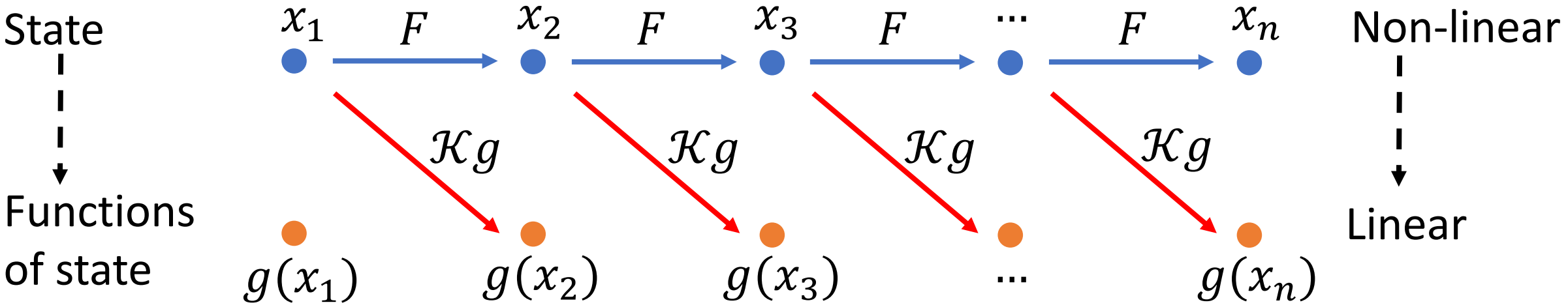


von Neumann

- **Koopman operator** $\mathcal{K}$ acts on <u>functions</u> $g: \Omega \to \mathbb{C}$

$$[\mathcal{K}g](x) = g\big(F(x)\big)$$

- $\mathcal{K}$ is ***linear*** but acts on an ***infinite-dimensional*** space.



- Work in $L^2(\Omega, \omega)$ for positive measure ω, with inner product $\langle \cdot, \cdot \rangle$.

- Koopman, "*Hamiltonian systems and transformation in Hilbert space,*" **Proc. Natl. Acad. Sci. USA**, 1931.
- Koopman, v. Neumann, "*Dynamical systems of continuous spectra,*" **Proc. Natl. Acad. Sci. USA**, 1932.

$x_{n+1} = F(x_n)$

Why is linear (much) easier?

Long-time dynamics become trivial!

- Suppose $F(x) = Ax, A \in \mathbb{R}^{d\times d}, A = V\Lambda V^{-1}$.
- Set $\xi = V^{-1}x$,

$$\xi_n = V^{-1}x_n = V^{-1}A^n x_0 = \Lambda^n V^{-1}x_0 = \Lambda^n \xi_0$$

- Let $w^{\mathrm{T}}A = \lambda w$, set $\varphi(x) = w^{\mathrm{T}}x$,

$$[\mathcal{K}\varphi](x) = w^{\mathrm{T}}Ax = \lambda\varphi(x)$$

Eigenfunction

Much more general (**non-linear** F and even **chaotic** systems).

Koopman mode decomposition

$[\mathcal{K}g](x) = g(F(x))$

eigenfunction of $\mathcal{K}$

generalized eigenfunction of $\mathcal{K}$

$$g(x) = \sum_{\text{eigs } \lambda_j} c_{\lambda_j} \varphi_{\lambda_j}(x) + \underbrace{\int_{[-\pi,\pi]_{\text{per}}} \phi_{\theta,g}(x)\, \mathrm{d}\theta}_{\text{continuous spectra}}$$

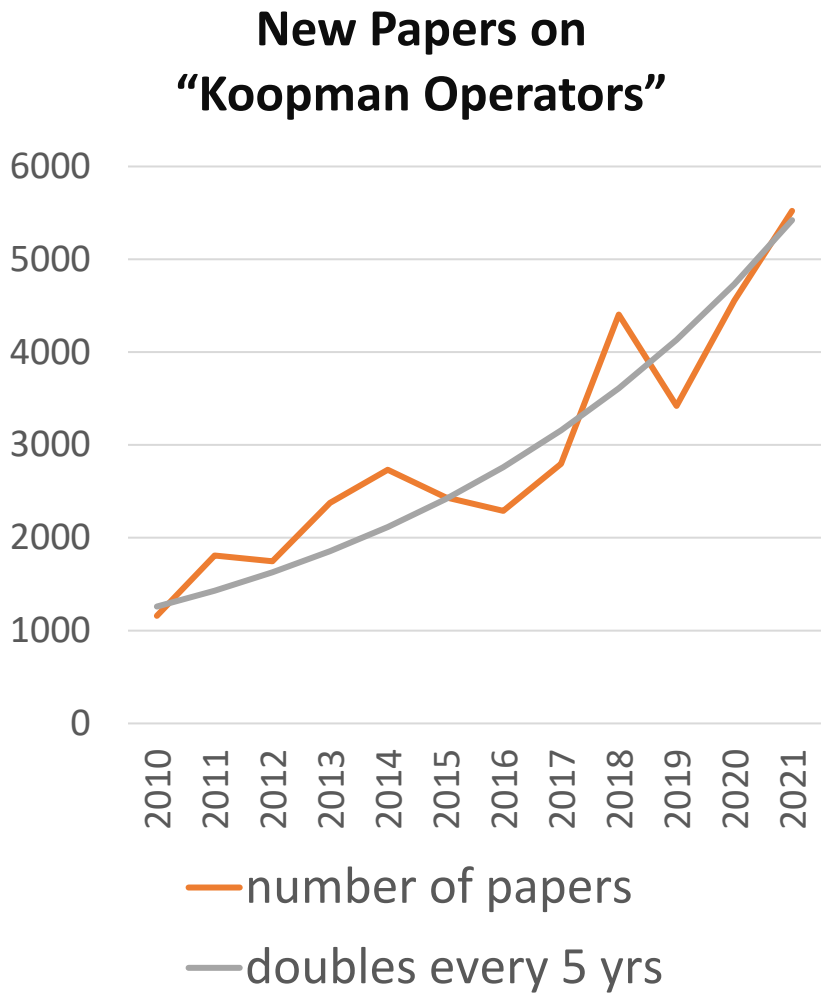
$$g(x_n) = [\mathcal{K}^n g](x_0) = \sum_{\text{eigs } \lambda_j} c_{\lambda_j} {\lambda_j}^n \varphi_{\lambda_j}(x_0) + \int_{[-\pi,\pi]_{\text{per}}} e^{in\theta} \phi_{\theta,g}(x_0)\, \mathrm{d}\theta$$

Encodes: geometric features, invariant measures, transient behavior, long-time behavior, coherent structures, quasiperiodicity, etc.

GOAL: Data-driven approximation of $\mathcal{K}$ and its spectral properties.

- Mezić, "*Spectral properties of dynamical systems, model reduction and decompositions,*" **Nonlinear Dynam.**, 2005.

Koopmania*: A revolution in the big data era?



≈35,000 papers over last decade!

BUT: Very little on verified methods!

Computing spectra in infinite dimensions is notoriously hard!

**Wikipedia*: *"its wild surge in popularity is sometimes jokingly called 'Koopmania'"*

Challenges of computing $\mathrm{Spec}(\mathcal{K}) = \{\lambda \in \mathbb{C}: \mathcal{K} - \lambda I \text{ is not invertible}\}$

Truncate: $\mathcal{K} \longrightarrow \mathbb{K} \in \mathbb{C}^{N_K \times N_K}$

1) **"Too much":** Approximate spurious modes $\lambda \notin \mathrm{Spec}(\mathcal{K})$
2) **"Too little":** Miss parts of $\mathrm{Spec}(\mathcal{K})$
3) **Continuous spectra.**

Verification: Is it right?

Computing spectra

Build the matrix: Dynamic Mode Decomposition (DMD)

Given dictionary $\{\psi_1, \dots, \psi_{N_K}\}$ of functions $\psi_j : \Omega \to \mathbb{C}$,

$$\left\{x^{(m)}, y^{(m)} = F(x^{(m)})\right\}_{m=1}^{M}$$

$$\langle \psi_k, \psi_j \rangle \approx \sum_{m=1}^{M} w_m \overline{\psi_j(x^{(m)})} \psi_k(x^{(m)}) = \left[\underbrace{\begin{pmatrix} \psi_1(x^{(1)}) & \cdots & \psi_{N_K}(x^{(1)}) \\ \vdots & \ddots & \vdots \\ \psi_1(x^{(M)}) & \cdots & \psi_{N_K}(x^{(M)}) \end{pmatrix}}_{\Psi_X}^{*} \underbrace{\begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_M \end{pmatrix}}_{W} \underbrace{\begin{pmatrix} \psi_1(x^{(1)}) & \cdots & \psi_{N_K}(x^{(1)}) \\ \vdots & \ddots & \vdots \\ \psi_1(x^{(M)}) & \cdots & \psi_{N_K}(x^{(M)}) \end{pmatrix}}_{\Psi_X} \right]_{jk}$$

$$\langle \mathcal{K}\psi_k, \psi_j \rangle \approx \sum_{m=1}^{M} w_m \overline{\psi_j(x^{(m)})} \underbrace{\psi_k(y^{(m)})}_{[\mathcal{K}\psi_k](x^{(m)})} = \left[\underbrace{\begin{pmatrix} \psi_1(x^{(1)}) & \cdots & \psi_{N_K}(x^{(1)}) \\ \vdots & \ddots & \vdots \\ \psi_1(x^{(M)}) & \cdots & \psi_{N_K}(x^{(M)}) \end{pmatrix}}_{\Psi_X}^{*} \underbrace{\begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_M \end{pmatrix}}_{W} \underbrace{\begin{pmatrix} \psi_1(y^{(1)}) & \cdots & \psi_{N_K}(y^{(1)}) \\ \vdots & \ddots & \vdots \\ \psi_1(y^{(M)}) & \cdots & \psi_{N_K}(y^{(M)}) \end{pmatrix}}_{\Psi_Y} \right]_{jk}$$

$$\mathcal{K} \longrightarrow \mathbb{K} = (\Psi_X{}^* W \Psi_X)^{-1} \Psi_X{}^* W \Psi_Y \in \mathbb{C}^{N_K \times N_K}$$

Recall open problems: **too much, too little, continuous spectra, verification**

- Schmid, "*Dynamic mode decomposition of numerical and experimental data,*" **J. Fluid Mech.**, 2010.
- Rowley, Mezić, Bagheri, Schlatter, Henningson, *"Spectral analysis of nonlinear flows,"* **J. Fluid Mech.**, 2009.
- Kutz, Brunton, Brunton, Proctor, "*Dynamic mode decomposition: data-driven modeling of complex systems,*" **SIAM**, 2016.
- Williams, Kevrekidis, Rowley "*A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition,*" **J. Nonlinear Sci.**, 2015.

Residual DMD (ResDMD): Approx. $\mathcal{K}$ and $\mathcal{K}^*\mathcal{K}$

$$\langle \psi_k, \psi_j \rangle \approx \sum_{m=1}^{M} w_m \overline{\psi_j(x^{(m)})}\, \psi_k(x^{(m)}) = \left[\underbrace{\Psi_X{}^* W \Psi_X}_{G}\right]_{jk}$$

$$\langle \mathcal{K}\psi_k, \psi_j \rangle \approx \sum_{m=1}^{M} w_m \overline{\psi_j(x^{(m)})}\, \underbrace{\psi_k(y^{(m)})}_{[\mathcal{K}\psi_k](x^{(m)})} = \left[\underbrace{\Psi_X{}^* W \Psi_Y}_{K_1}\right]_{jk}$$

$$\langle \mathcal{K}\psi_k, \mathcal{K}\psi_j \rangle \approx \sum_{m=1}^{M} w_m \overline{\psi_j(y^{(m)})}\, \psi_k(y^{(m)}) = \left[\underbrace{\Psi_Y{}^* W \Psi_Y}_{K_2}\right]_{jk}$$

Residuals: $g = \sum_{j=1}^{N_K} \mathbf{g}_j \psi_j,\ \|\mathcal{K}g - \lambda g\|^2 \approx \mathbf{g}^*[K_2 - \lambda K_1{}^* - \bar{\lambda} K_1 + |\lambda|^2 G]\mathbf{g}$

- C., Townsend, "*Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems,*" preprint.
- C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition,*" **J. Fluid Mech.,** 2023.
- Code*:* https://github.com/MColbrook/Residual-Dynamic-Mode-Decomposition

ResDMD: avoiding "too much"

$$\mathrm{res}(\lambda, \mathbf{g})^2 = \frac{\mathbf{g}^*\left[K_2 - \lambda K_1{}^* - \bar{\lambda} K_1 + |\lambda|^2 G\right]\mathbf{g}}{\mathbf{g}^* G \mathbf{g}}$$

eigenvectors

eigenvalues

Algorithm 1:

1. Compute $G, K_1, K_2 \in \mathbb{C}^{N_K \times N_K}$ and eigendecomposition $K_1 V = G V \Lambda$.
2. For each eigenpair $(\lambda, \mathbf{v})$, compute $\mathrm{res}(\lambda, \mathbf{v})$.
3. **Output:** subset of e-vectors $V_{(\varepsilon)}$ & e-vals $\Lambda_{(\varepsilon)}$ with $\mathrm{res}(\lambda, \mathbf{v}) \leq \varepsilon$ ($\varepsilon =$ input tol).

Theorem (no spectral pollution): Suppose quad. rule converges. Then

$$\limsup_{M \to \infty} \max_{\lambda \in \Lambda^{(\varepsilon)}} \|(\mathcal{K} - \lambda)^{-1}\|^{-1} \leq \varepsilon$$

- C., Townsend, "*Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems*," preprint.

ResDMD: avoiding "too much"

$$\mathrm{res}(\lambda, \mathbf{g})^2 = \frac{\mathbf{g}^*\left[K_2 - \lambda K_1{}^* - \bar{\lambda} K_1 + |\lambda|^2 G\right]\mathbf{g}}{\mathbf{g}^* G \mathbf{g}}$$

eigenvectors

eigenvalues

Algorithm 1:

1. Compute $G, K_1, K_2 \in \mathbb{C}^{N_K \times N_K}$ and eigendecomposition $K_1 V = G V \Lambda$.
2. For each eigenpair $(\lambda, \mathbf{v})$, compute $\mathrm{res}(\lambda, \mathbf{v})$.
3. **Output:** subset of e-vectors $V_{(\varepsilon)}$ & e-vals $\Lambda_{(\varepsilon)}$ with $\mathrm{res}(\lambda, \mathbf{v}) \leq \varepsilon$ $(\varepsilon = \text{input tol})$.

Theorem (no spectral pollution): Suppose quad. rule converges. Then

$$\limsup_{M \to \infty} \max_{\lambda \in \Lambda^{(\varepsilon)}} \|(\mathcal{K} - \lambda)^{-1}\|^{-1} \leq \varepsilon$$

BUT: Typically, does not capture all of spectrum! ("too little")

- C., Townsend, "*Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems*," preprint.

ResDMD: avoiding "too little"

$$\mathrm{Spec}_\varepsilon(\mathcal{K}) = \bigcup_{\|\mathcal{B}\| \le \varepsilon} \mathrm{Spec}(\mathcal{K} + \mathcal{B}), \qquad \lim_{\varepsilon \downarrow 0} \mathrm{Spec}_\varepsilon(\mathcal{K}) = \mathrm{Spec}(\mathcal{K})$$

Algorithm 2:

First convergent method for general $\mathcal{K}$

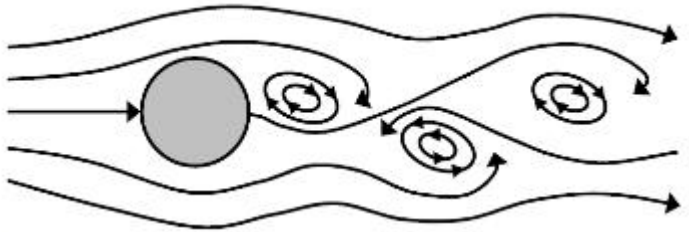
1. Compute $G, K_1, K_2 \in \mathbb{C}^{N_K \times N_K}$.
2. For z_k in comp. grid, compute $\tau_k = \min_{g = \sum_{j=1}^{N_K} \mathbf{g}_j \psi_j} \mathrm{res}(z_k, g)$, corresponding g_k (gen. SVD).
3. **Output:** $\{z_k : \tau_k < \varepsilon\}$ (approx. of $\mathrm{Spec}_\varepsilon(\mathcal{K})$), $\{g_k : \tau_k < \varepsilon\}$ (ε-pseudo-eigenfunctions).

Theorem (full convergence): In the large data limit,

- **Error control:** $\{z_k : \tau_k < \varepsilon\} \subseteq \mathrm{Spec}_\varepsilon(\mathcal{K})$ (as $M \to \infty$)
- **Convergence:** Converges locally uniformly to $\mathrm{Spec}_\varepsilon(\mathcal{K})$ (as $N_K \to \infty$)

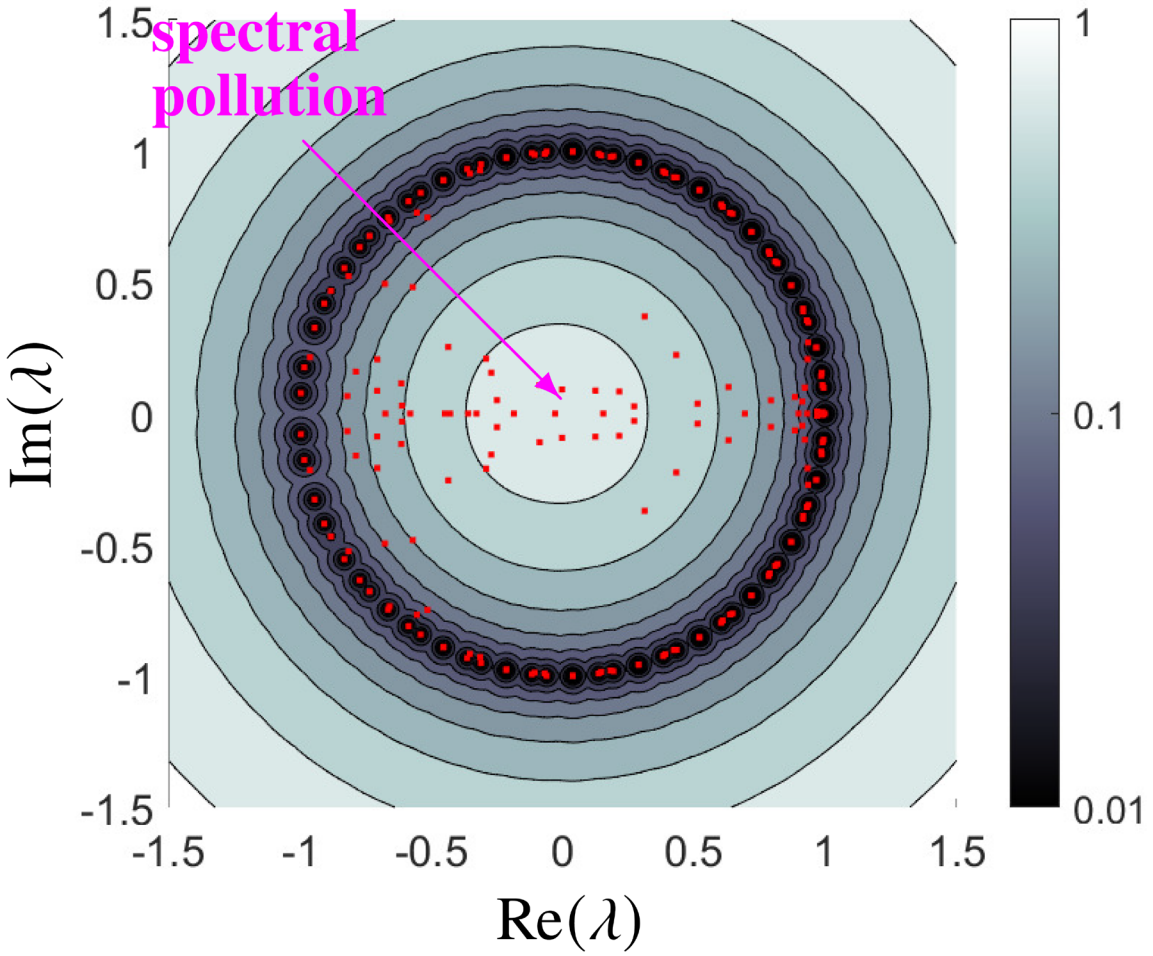
- C., Townsend, "*Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems*," preprint.

Example: Flow past a cylinder wake

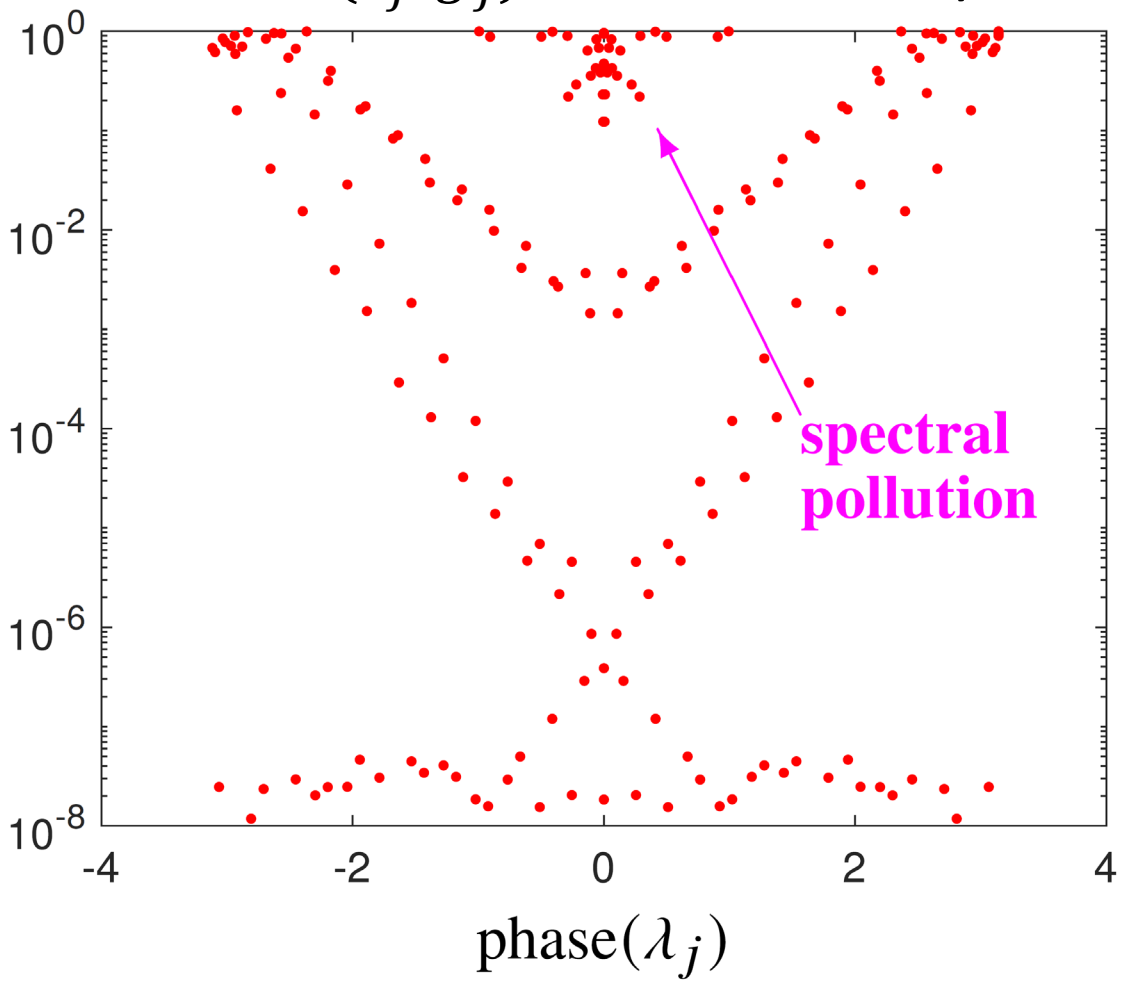


$\mathrm{Re} = 100,$ Dimension $(d) = 80{,}000$ (vel. at grid points)

Pseudospectra, linear dictionary

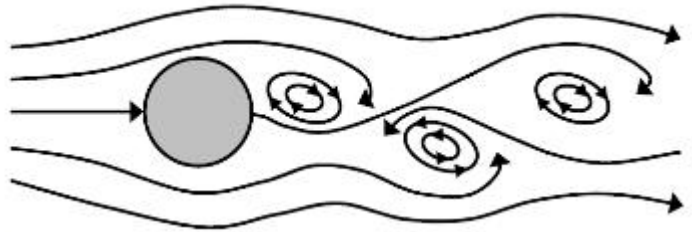


$\mathrm{res}(\lambda_j, g_j)$, linear dictionary



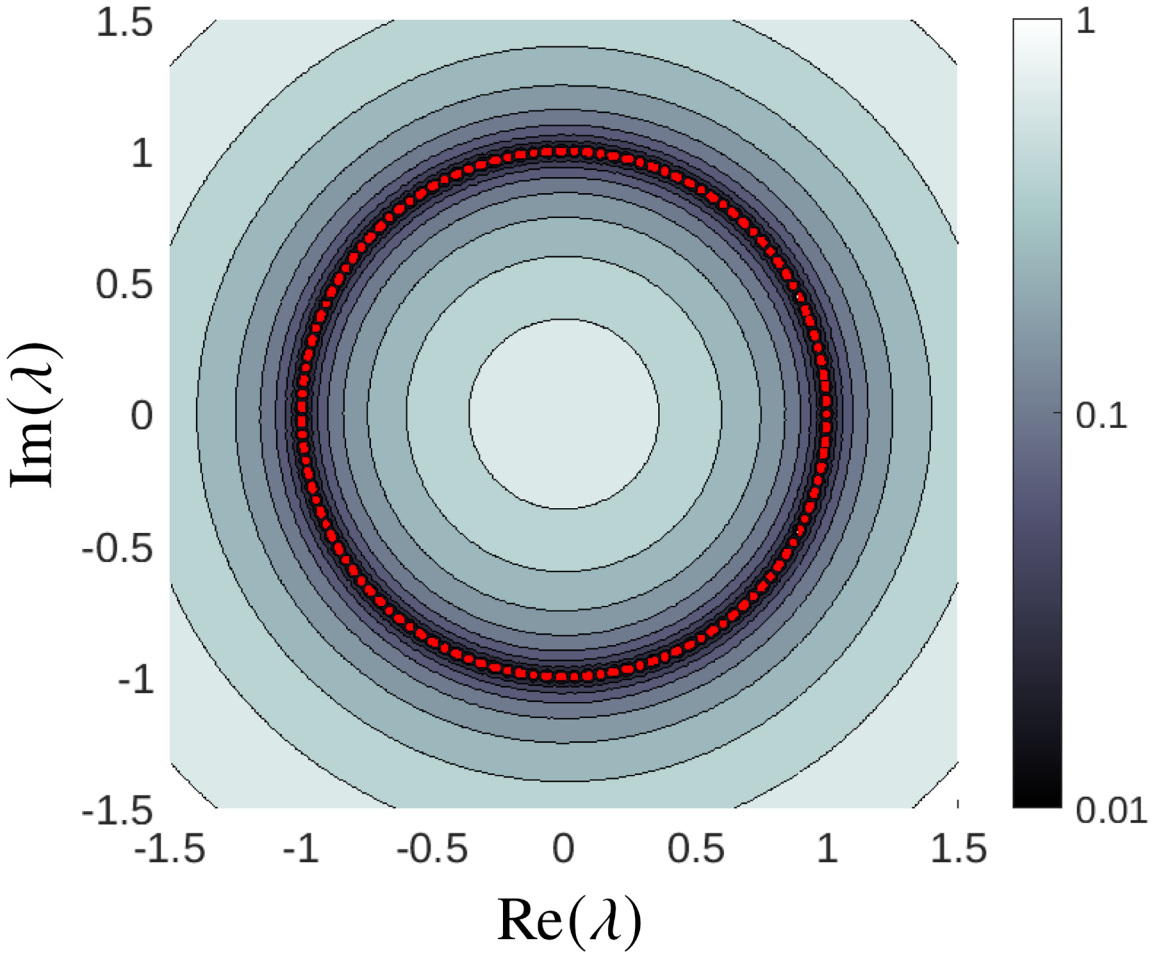
- C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition,*" **J. Fluid Mech.,** 2023.

Example: Flow past a cylinder wake

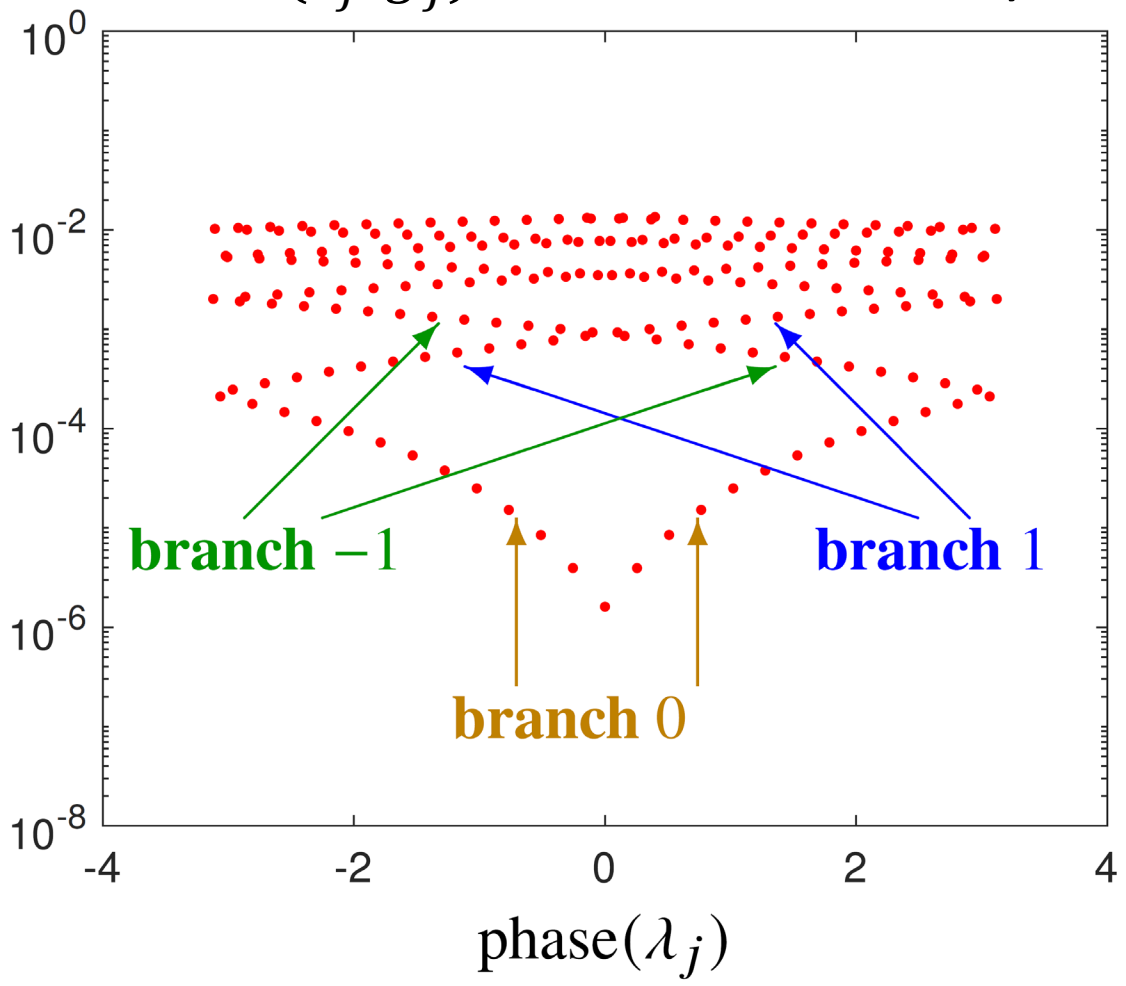


$\mathrm{Re} = 100,$ Dimension $(d) = 80{,}000$ (vel. at grid points)

Pseudospectra, nonlinear dictionary



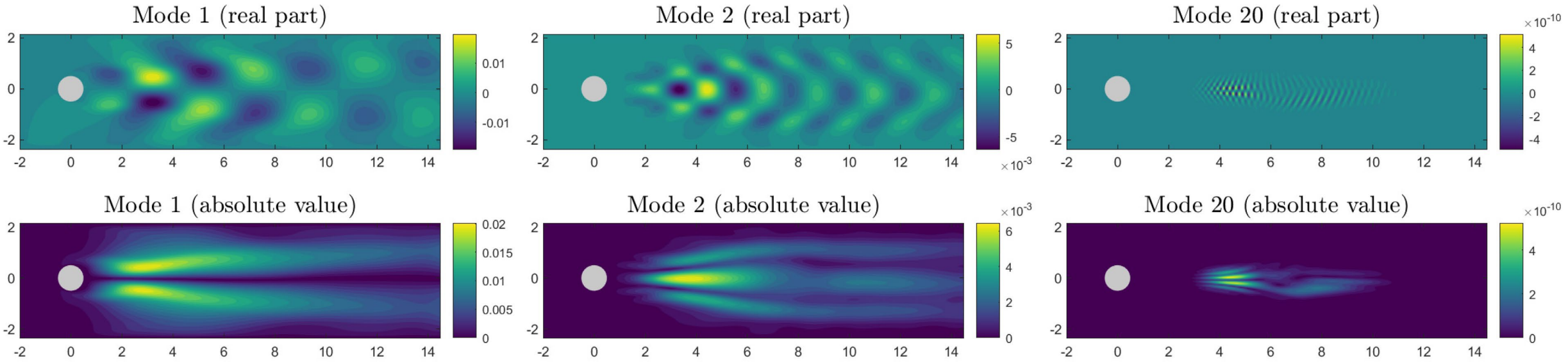
$\mathrm{res}(\lambda_j, g_j)$, nonlinear dictionary



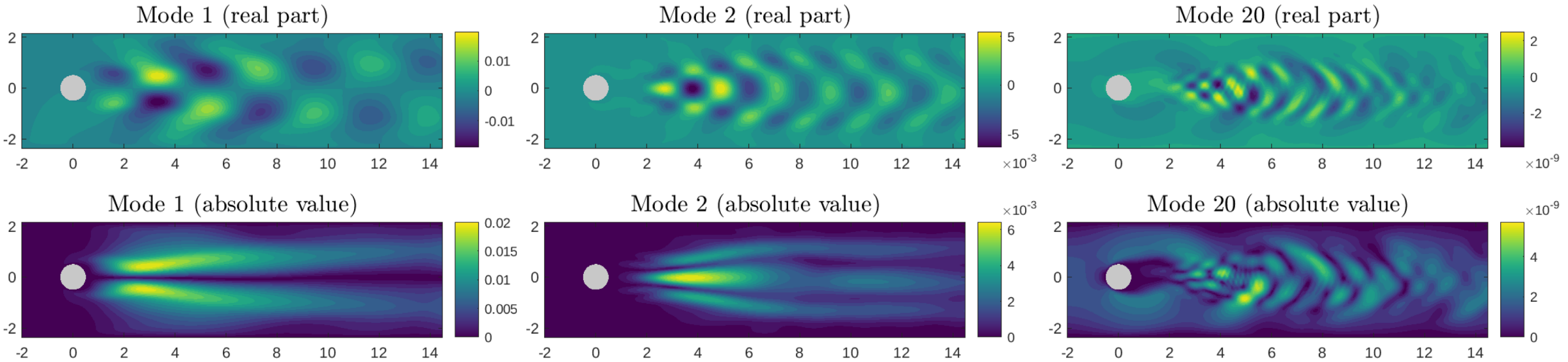
- C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition,*" **J. Fluid Mech.,** 2023.

Koopman Modes

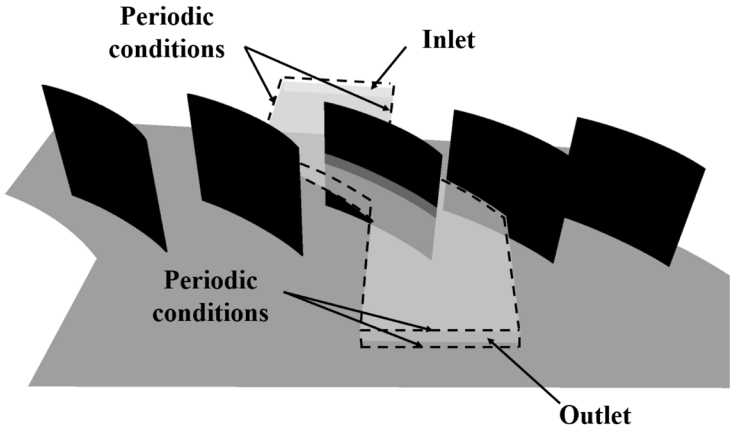
Linear dictionary

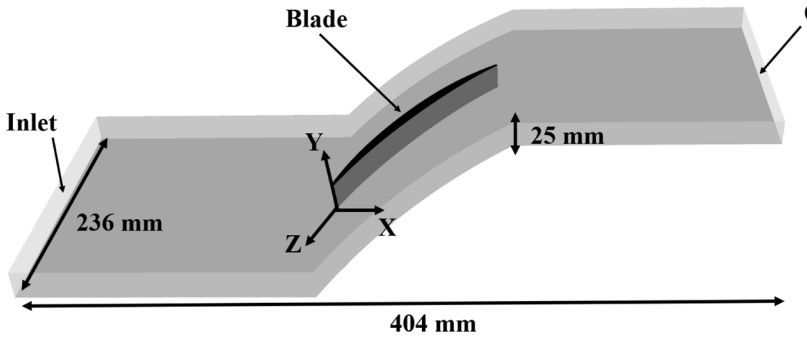


Nonlinear dictionary



Example: Trustworthy computation for large d



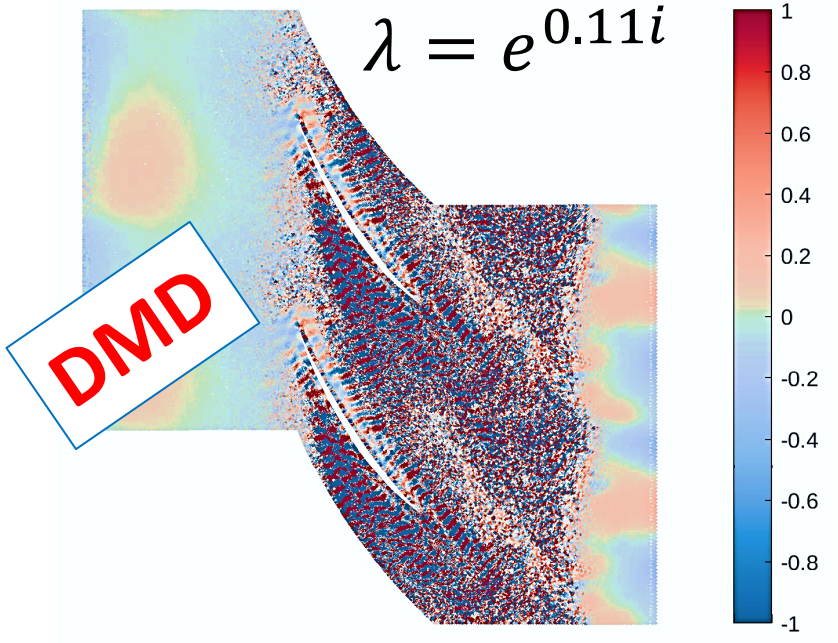


- Reynolds number $\approx 3.9 \times 10^5$
- Ambient dimension (d) $\approx 300{,}000$ (number of measurement points)

*Measurements provided by Stephane Moreau (Sherbrooke)

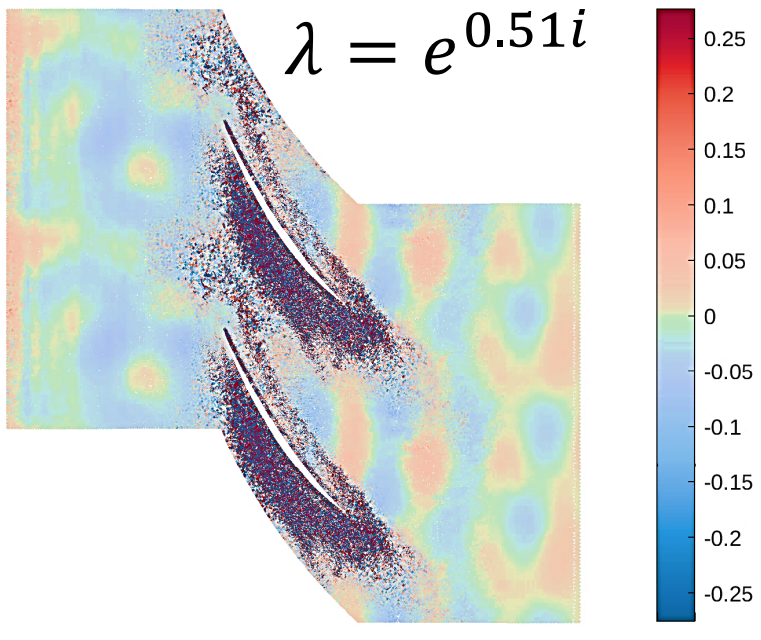
Rel. Error = ?

$\lambda = e^{0.11i}$



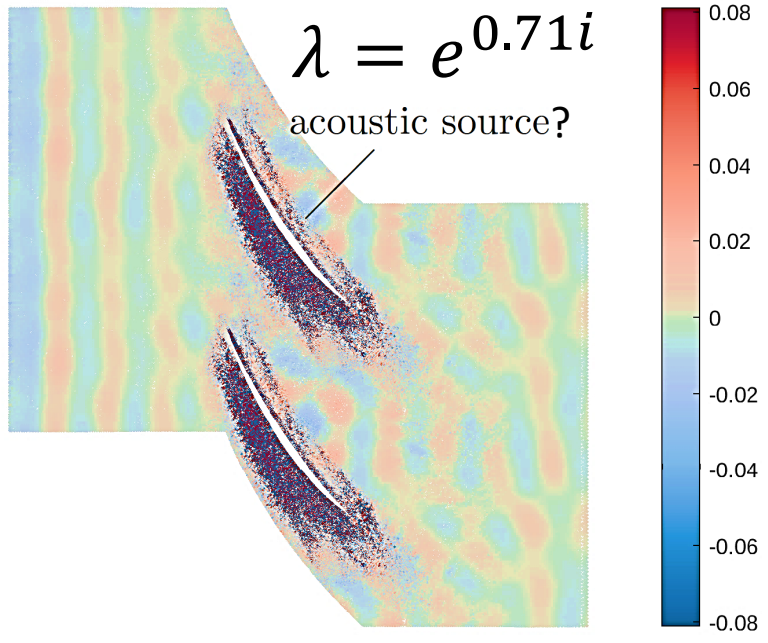
Rel. Error = ?

$\lambda = e^{0.51i}$



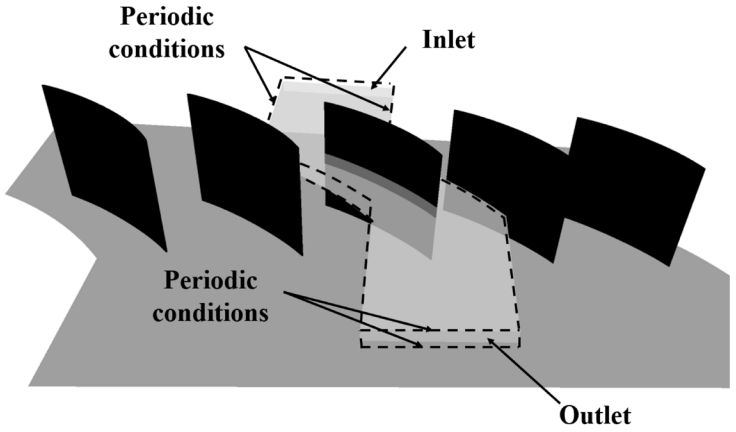
Rel. Error = ?

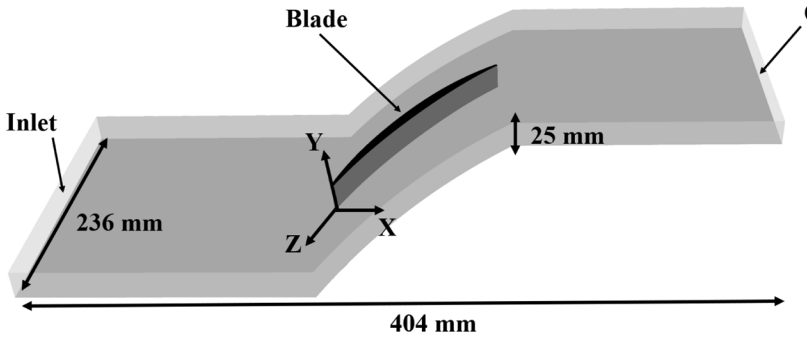
$\lambda = e^{0.71i}$



- C., Townsend, "*Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems*," preprint.

Example: Trustworthy computation for large d



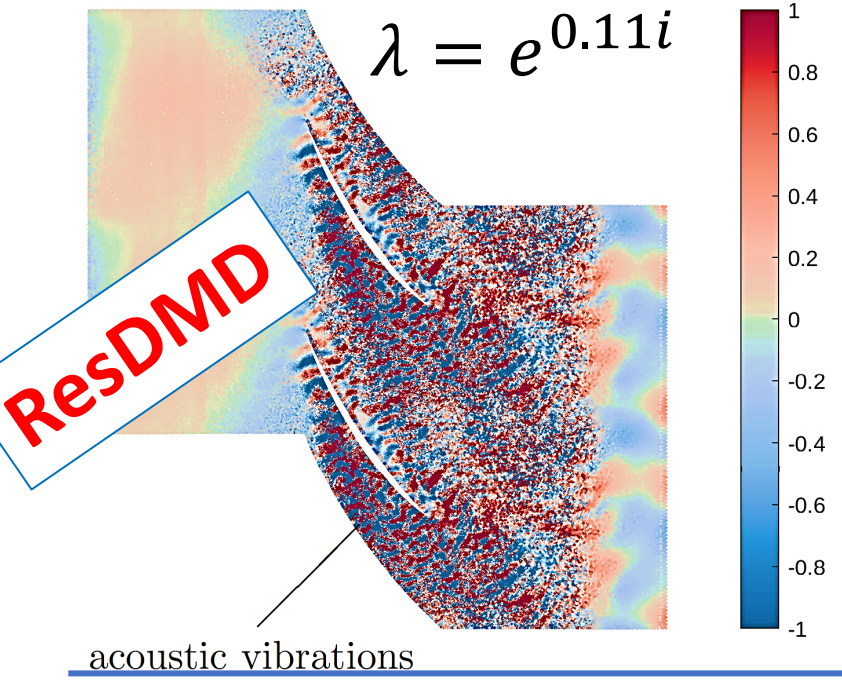


- Reynolds number $\approx 3.9 \times 10^5$
- Ambient dimension (d) $\approx 300{,}000$ (number of measurement points)

*Measurements provided by Stephane Moreau (Sherbrooke)

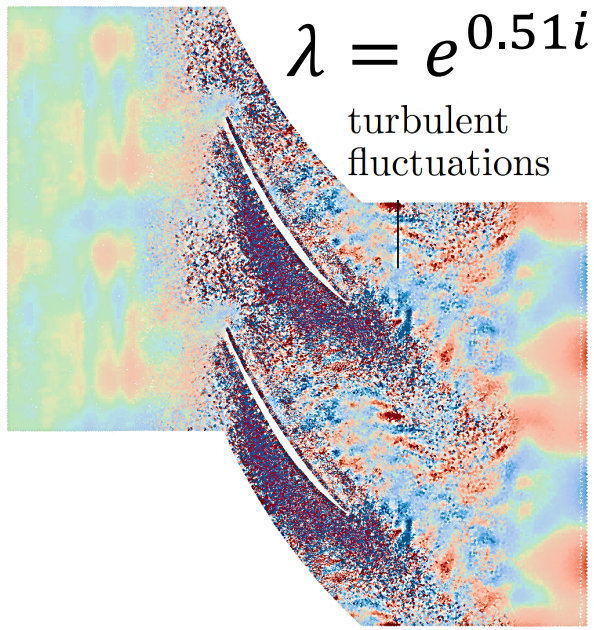
Rel. Error ≤ 0.0054

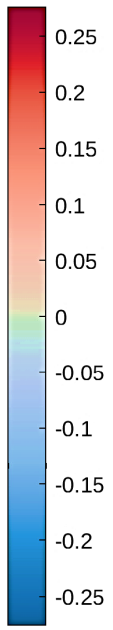
$\lambda = e^{0.11i}$



Rel. Error ≤ 0.0128

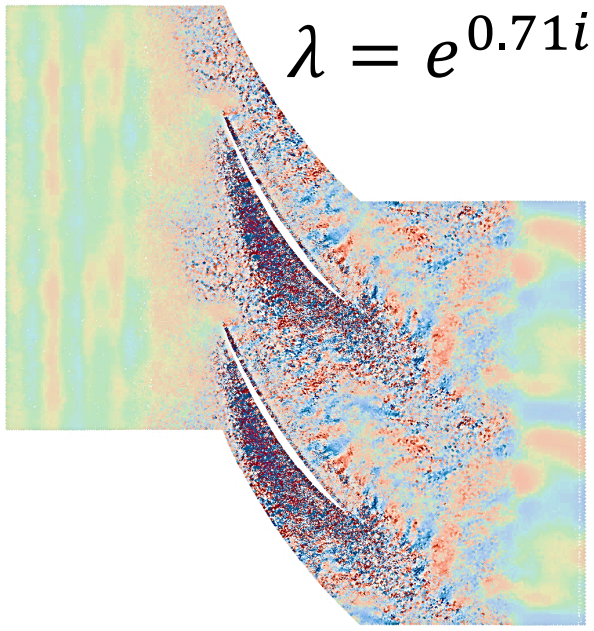
$\lambda = e^{0.51i}$

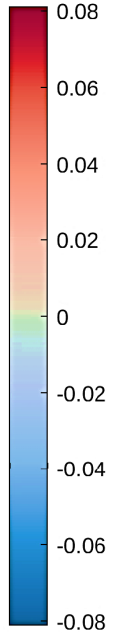




Rel. Error ≤ 0.0196

$\lambda = e^{0.71i}$





- C., Townsend, "*Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems*," preprint.

Dealing with continuous spectra

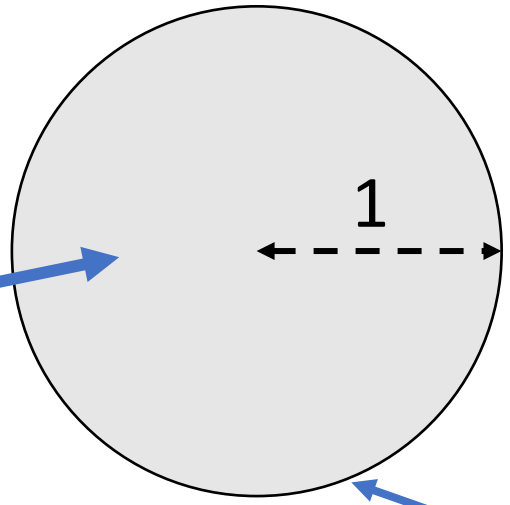
Setup for continuous spectra

No such assumption
was made in first part of talk!

Suppose system is measure-preserving (e.g., Hamiltonian, ergodic, post-transient etc.)

$$\Leftrightarrow \mathcal{K}^*\mathcal{K} = I \text{ (isometry)}$$

$$\Rightarrow \mathrm{Spec}(\mathcal{K}) \subseteq \{z : |z| \leq 1\}$$



spectral
measure
supp. on
boundary

Spectral decomposition of operators

$A \in \mathbb{C}^{n\times n}$ normal $\Longrightarrow$ O.N. basis of eigenvectors $v_1, \dots, v_n$:

$$v = \left(\sum_{k=1}^{n} v_k v_k^*\right) v, \qquad Av = \left(\sum_{k=1}^{n} \lambda_k v_k v_k^*\right) v, \qquad v \in \mathbb{C}^n$$

Projector onto Span(v_k)

eigenvalues

Spectral decomposition of operators

$A \in \mathbb{C}^{n\times n}$ normal $\Longrightarrow$ O.N. basis of eigenvectors $v_1, \dots, v_n$:

$$v = \left(\sum_{k=1}^{n} v_k v_k^*\right) v, \qquad Av = \left(\sum_{k=1}^{n} \lambda_k v_k v_k^*\right) v, \qquad v \in \mathbb{C}^n$$

Projector onto Span(v_k)

eigenvalues

Energy of "v" in each eigenvector: $\mu_v(\lambda_j) = \langle v_j v_j^* v, v\rangle = |v_j^* v|^2$

This is called the spectral measure with respect to a vector v.

Spectral decomposition of operators

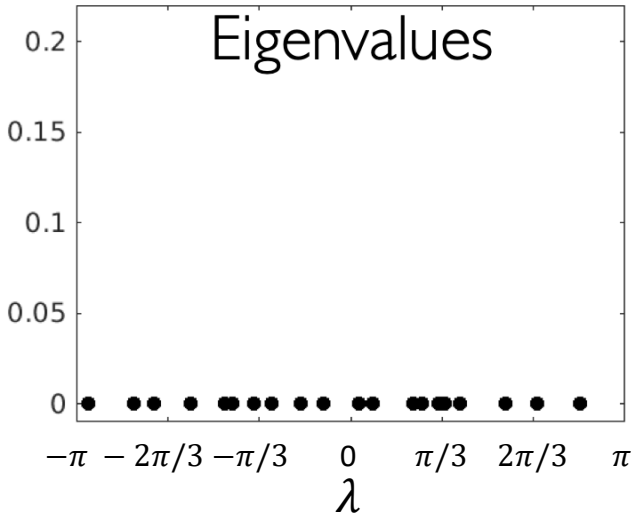
$A \in \mathbb{C}^{n\times n}$ normal $\Longrightarrow$ O.N. basis of eigenvectors $v_1, \dots, v_n$:

$$v = \left(\sum_{k=1}^{n} v_k v_k^*\right) v, \qquad Av = \left(\sum_{k=1}^{n} \lambda_k v_k v_k^*\right) v, \qquad v \in \mathbb{C}^n$$

Projector onto Span(v_k) — eigenvalues

Energy of "v" in each eigenvector: $\mu_v(\lambda_j) = \langle v_j v_j^* v, v\rangle = |v_j^* v|^2$

This is called the spectral measure with respect to a vector v.



Spectral decomposition of operators

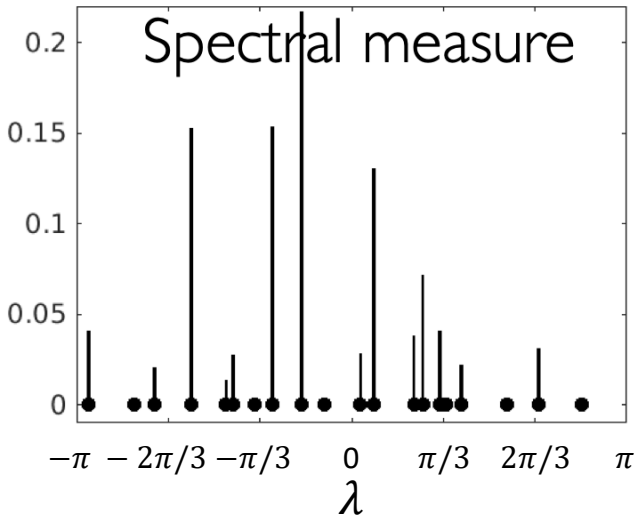
$A \in \mathbb{C}^{n\times n}$ normal $\Longrightarrow$ O.N. basis of eigenvectors $v_1, \dots, v_n$:

$$v = \left(\sum_{k=1}^{n} v_k v_k^*\right) v, \qquad Av = \left(\sum_{k=1}^{n} \lambda_k v_k v_k^*\right) v, \qquad v \in \mathbb{C}^n$$

Projector onto Span(v_k) — eigenvalues

Energy of "v" in each eigenvector: $\mu_v(\lambda_j) = \langle v_j v_j^* v, v\rangle = |v_j^* v|^2$

This is called the spectral measure with respect to a vector v.



Spectral decomposition of operators

$A \in \mathbb{C}^{n\times n}$ normal $\Longrightarrow$ O.N. basis of eigenvectors $v_1, \dots, v_n$:

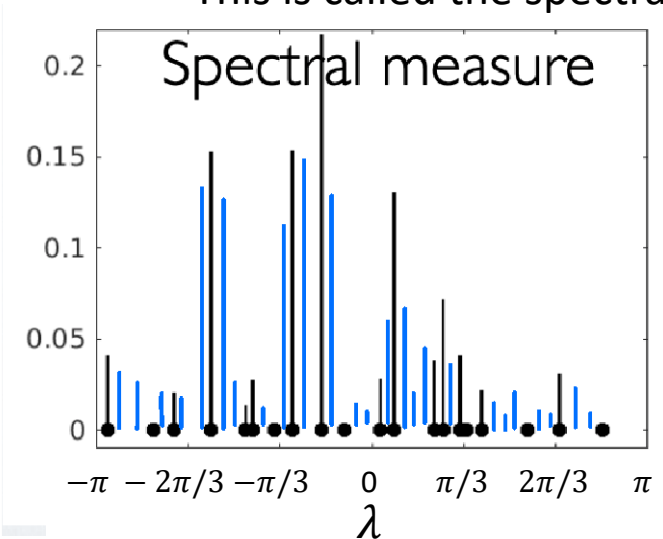
$$v = \left(\sum_{k=1}^{n} v_k v_k^*\right) v, \qquad Av = \left(\sum_{k=1}^{n} \lambda_k v_k v_k^*\right) v, \qquad v \in \mathbb{C}^n$$

Projector onto Span(v_k)

eigenvalues

Energy of "v" in each eigenvector: $\mu_v(\lambda_j) = \langle v_j v_j^* v, v\rangle = |v_j^* v|^2$

This is called the spectral measure with respect to a vector v.



Spectral decomposition of operators

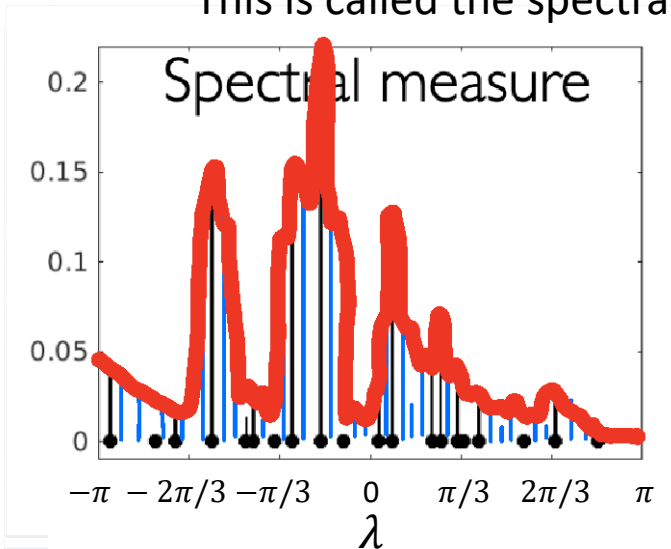
$A \in \mathbb{C}^{n\times n}$ normal $\Longrightarrow$ O.N. basis of eigenvectors $v_1, \dots, v_n$:

$$v = \left(\sum_{k=1}^{n} v_k v_k^*\right) v, \qquad Av = \left(\sum_{k=1}^{n} \lambda_k v_k v_k^*\right) v, \qquad v \in \mathbb{C}^n$$

Projector onto Span(v_k) — eigenvalues

Energy of "v" in each eigenvector: $\mu_v(\lambda_j) = \langle v_j v_j^* v, v\rangle = |v_j^* v|^2$

This is called the spectral measure with respect to a vector v.



Spectral decomposition of operators

$A \in \mathbb{C}^{n\times n}$ normal $\Longrightarrow$ O.N. basis of eigenvectors $v_1, \dots, v_n$:

$$v = \left(\sum_{k=1}^{n} v_k v_k^*\right) v, \qquad Av = \left(\sum_{k=1}^{n} \lambda_k v_k v_k^*\right) v, \qquad v \in \mathbb{C}^n$$

Projector onto Span(v_k) — eigenvalues

Energy of "v" in each eigenvector: $\mu_v(\lambda_j) = \langle v_j v_j^* v, v\rangle = \left|v_j^* v\right|^2$

This is called the spectral measure with respect to a vector v.

$\mathcal{K}$ is unitary $\Longrightarrow$ projection-valued measure ξ

$$g = \left(\int_{\mathbb{T}} d\xi(y)\right) g, \qquad \mathcal{K}g = \left(\int_{\mathbb{T}} y\, d\xi(y)\right) g$$

Spectral measure $\nu_g(B) = \langle \xi(B)g, g\rangle$

Spectral decomposition of operators

$A \in \mathbb{C}^{n\times n}$ normal $\Longrightarrow$ O

$\left(\sum_{k=1}^{n}\right.$

$v,$ A

$\text{an}(v_k)$

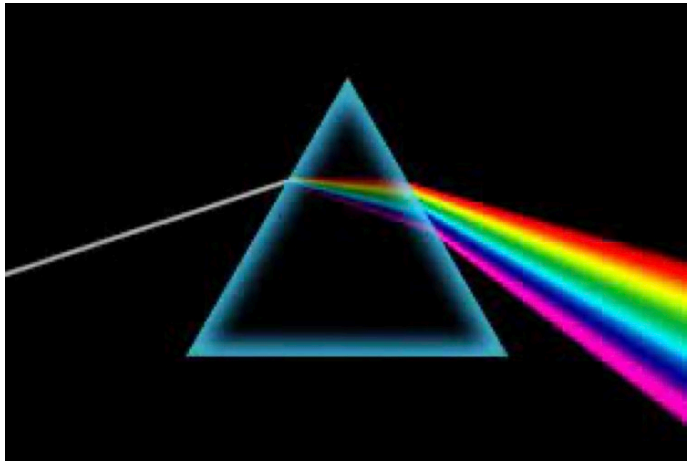
eigenvector:

This is ca

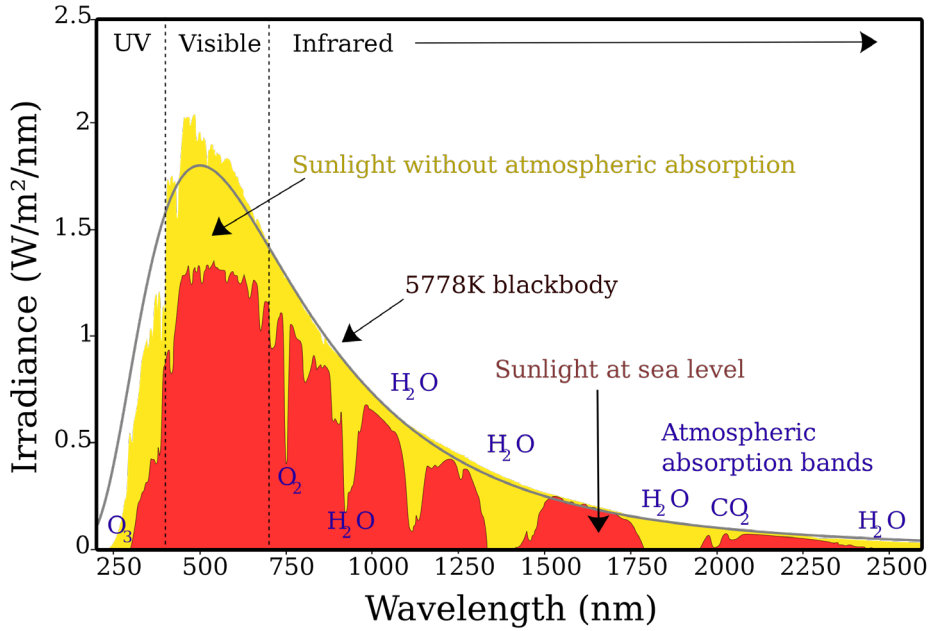
$\Longrightarrow$

$g = \left(\int_{\mathbb{T}} d\xi(y)\right) g,$

White light contains a continuous spectrum



Often interesting to look at
the intensity of each wavelength



Spectral measure $\nu_g(B) = \langle \xi(B)g, g\rangle$

Approximation using autocorrelations

$$\widehat{\nu_g}(n) = \frac{1}{2\pi} \int_{[-\pi,\pi]_{\mathrm{per}}} e^{-in\theta}\, \mathrm{d}\nu_g(\theta) = \frac{1}{2\pi} \begin{cases} \langle \mathcal{K}^{|n|} g, g \rangle, & n < 0 \\ \langle g, \mathcal{K}^{|n|} g \rangle, & n \geq 0 \end{cases}$$

Approximate from trajectory data

$$\nu_{g,N}(\theta) = \sum_{n=-N}^{N} \varphi\left(\frac{n}{N}\right) \widehat{\nu_g}(n) e^{in\theta}$$

Filter function

For $m \in \mathbb{N}$, mth order filter:

- Continuous, even, compactly supported on [-1,1]
- $\in C^{m-1}([-1,1]),\ \in C^{m-1}([0,1])$
- $\varphi(0) = 1, \varphi^j(0) = 0$ for $j = 1, \ldots, m-1$
- $\varphi^j(0) = 0$ for $j = 0, \ldots, m-1$

Approximates ν_g to order $O(N^{-m}\log(N))$ with frequency smoothing scale $O(N^{-m})$

Link with power spectrum

Delay autocorrelation function

$$R_g(n\Delta t) = \langle g, g \circ F_{n\Delta t} \rangle = \begin{cases} \langle \mathcal{K}^{|n|} g, g \rangle, n < 0 \\ \langle g, \mathcal{K}^{|n|} g \rangle, n \geq 0 \end{cases} = 2\pi \widehat{\nu_g}(n)$$

Power spectrum of signal $g(x(t))$

$$S_g(f) = \int_{-T}^{T} R_g(t) e^{2\pi i f t} \, \mathrm{d}t$$

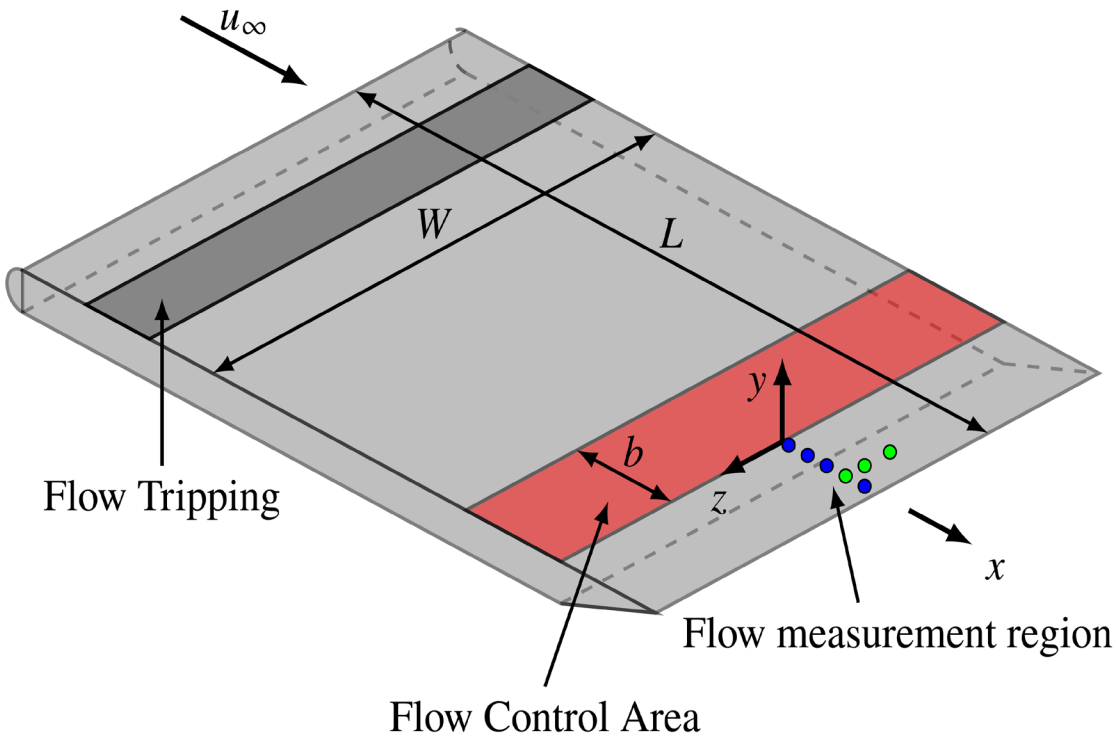
Window (using φ) ***in frequency domain*** and discretize integral:

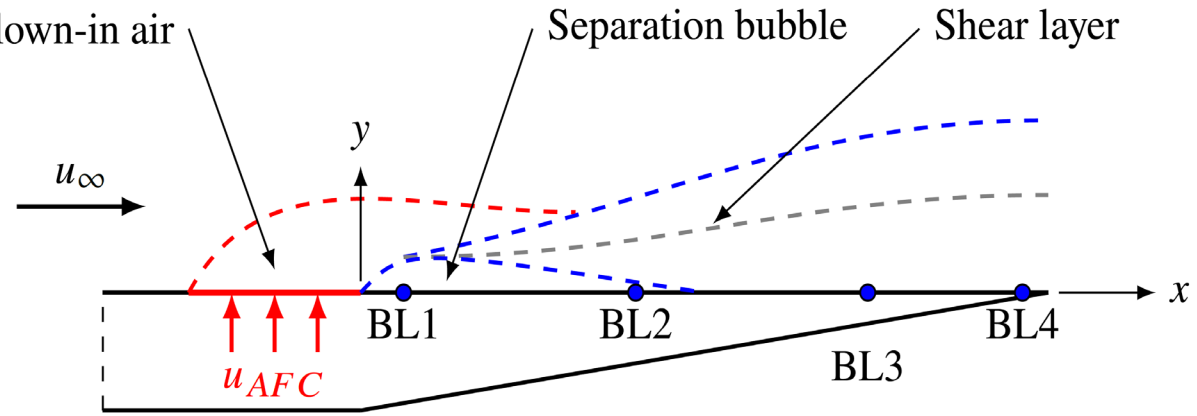
$$\frac{S_g(f)}{2\pi\Delta t} \approx \sum_{n=-N}^{N} \varphi\left(\frac{n}{N}\right) \frac{R_g(n\Delta t)}{2\pi} e^{in(2\pi f \Delta t)} = \nu_{g,N}(2\pi f \Delta t)$$

- Avoid (artificially) periodically extending signal $\Longrightarrow$ avoid broadening.
- Rigorous convergence theory as $N \to \infty$.

Example: Shear layer in turbulent boundary (hotwire experimental data)

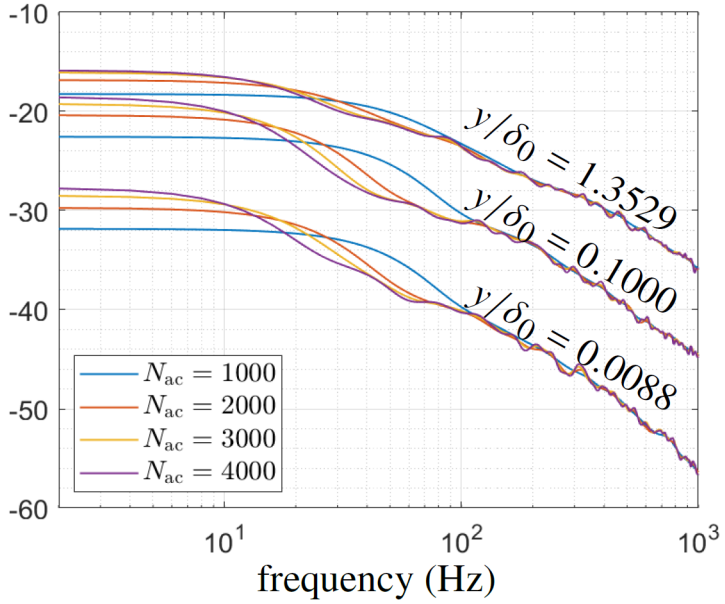
(friction) Re $= 1400$



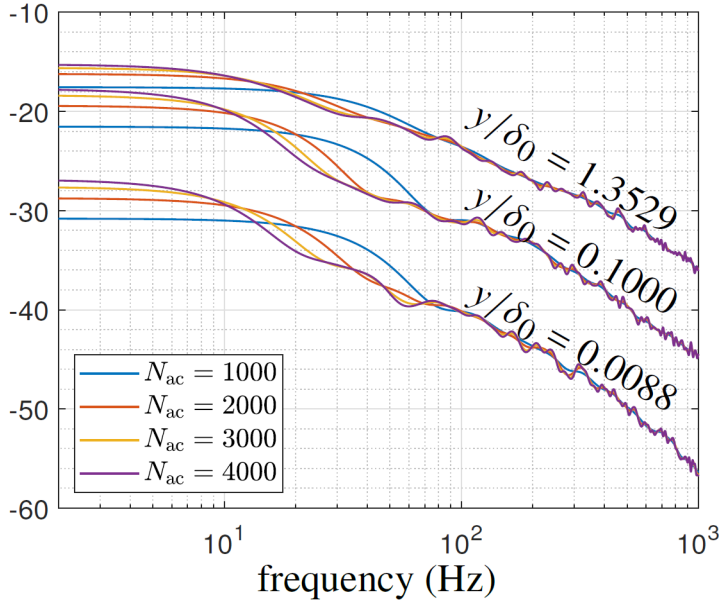


- C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition,*" **J. Fluid Mech.,** 2023.

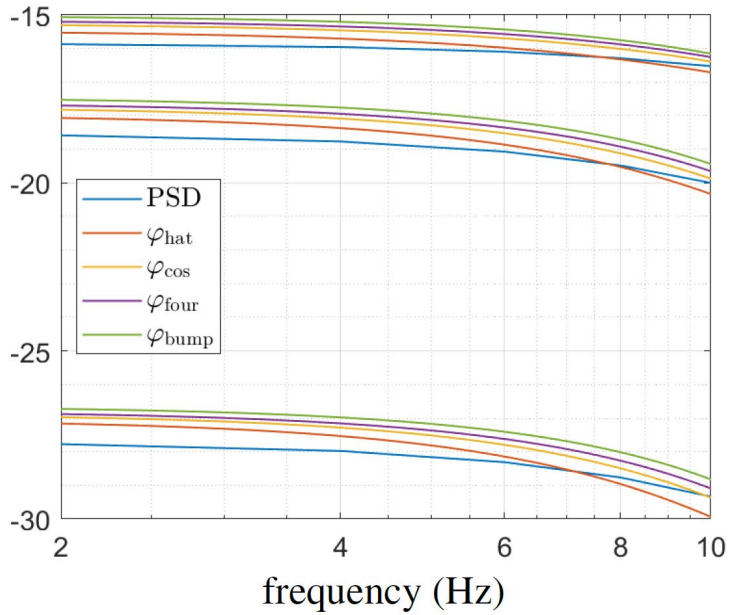
PSD (baseline)



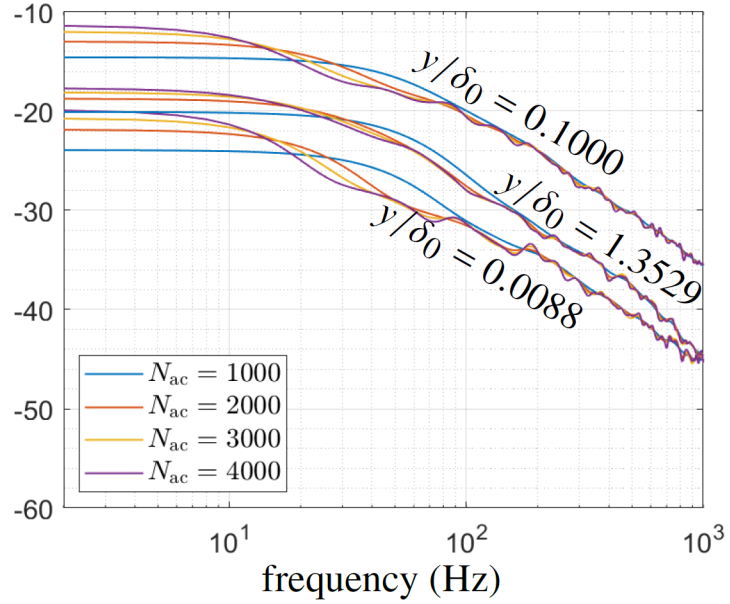
spectral measure (baseline)



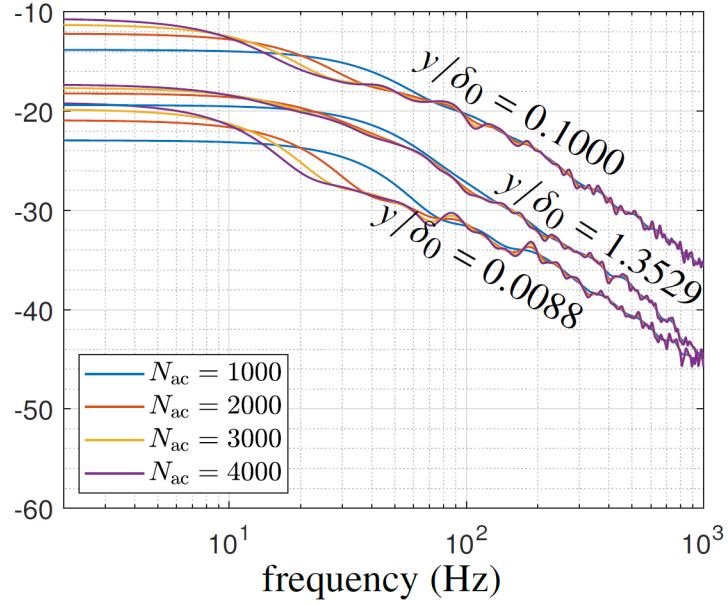
baseline



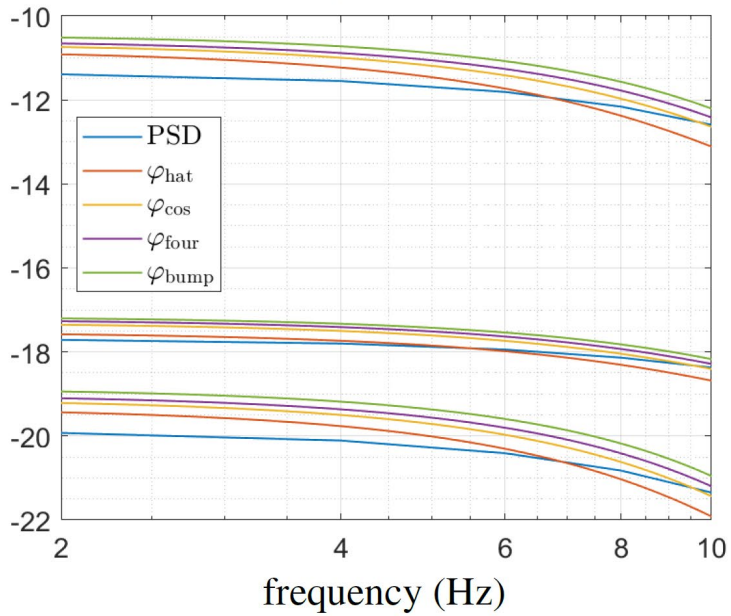
PSD (injection)



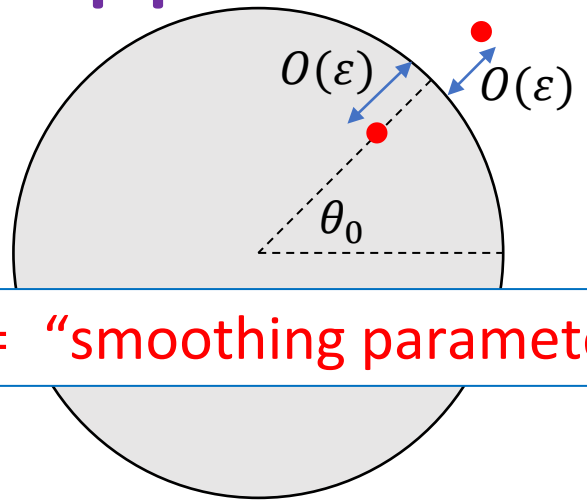
spectral measure (injection)



injection



Approximation using resolvent (Green's function)



$\varepsilon =$ "smoothing parameter"

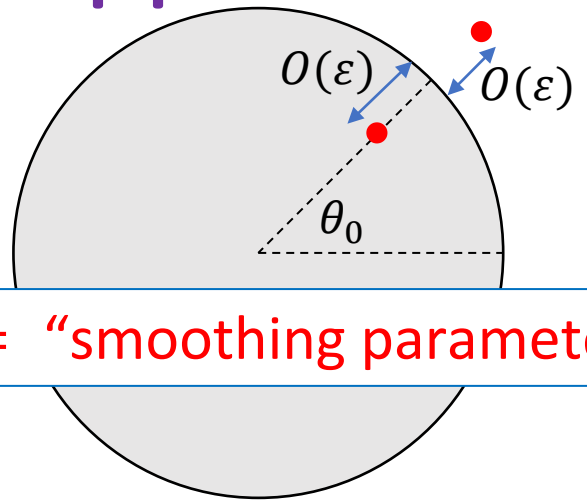
Smoothing convolution

$$[P_\varepsilon * \nu_g](\theta_0) = \int_{[-\pi,\pi]_{\text{per}}} P_\varepsilon(\theta_0 - \theta)\, \mathrm{d}\nu_g(\theta)$$

Poisson kernel for unit disk

$$P_\varepsilon(\theta_0) = \frac{1}{2\pi} \frac{(1+\varepsilon)^2 - 1}{1 + (1+\varepsilon)^2 - 2(1+\varepsilon)\cos(\theta_0)}$$

Approximation using resolvent (Greer



$\varepsilon =$ "smoothing parameter"

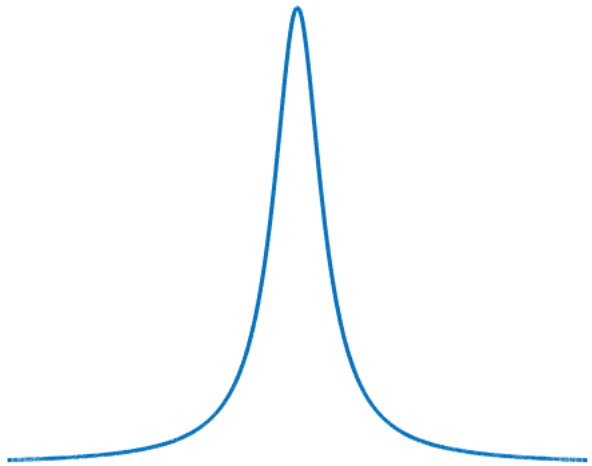
Smoothing cc

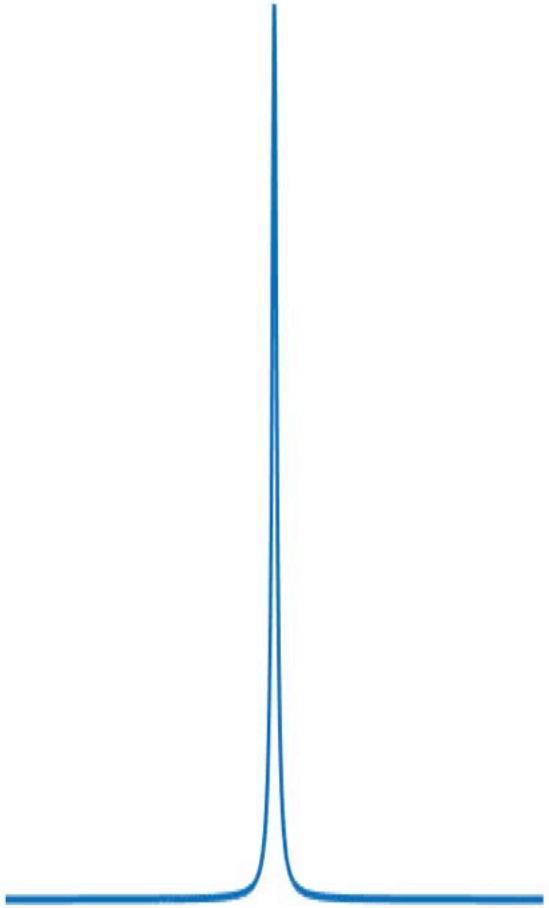
$$[P_\varepsilon * \nu_g](\theta_0) = \int P_\varepsilon(\theta_0 - \theta$$

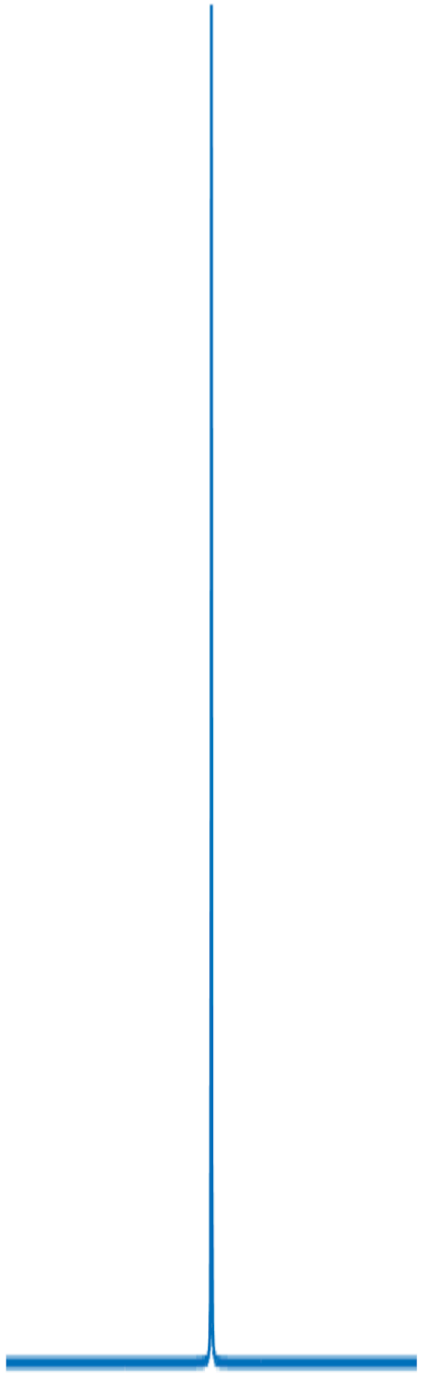
Poisso
unit di

$$\overline{+ (1 + \varepsilon}$$

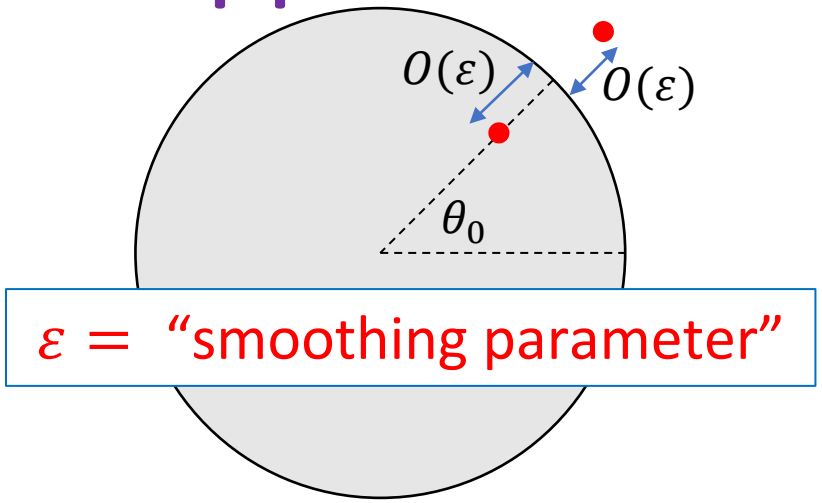
$$\overline{_0)}$$







Approximation using resolvent (Green's function)



$\varepsilon =$ "smoothing parameter"

Smoothing convolution

$$[P_\varepsilon * \nu_g](\theta_0) = \int_{[-\pi,\pi]_{\text{per}}} P_\varepsilon(\theta_0 - \theta)\, \mathrm{d}\nu_g(\theta)$$

Poisson kernel for unit disk

$$P_\varepsilon(\theta_0) = \frac{1}{2\pi} \frac{(1+\varepsilon)^2 - 1}{1 + (1+\varepsilon)^2 - 2(1+\varepsilon)\cos(\theta_0)}$$

$$[P_\varepsilon * \nu_g](\theta_0) = \mathcal{C}_g\left(e^{i\theta_0}(1+\varepsilon)^{-1}\right) - \mathcal{C}_g\left(e^{i\theta_0}(1+\varepsilon)\right)$$

$$\mathcal{C}_g(z) = \int_{[-\pi,\pi]_{\text{per}}} \frac{e^{i\theta}\, \mathrm{d}\nu_g(\theta)}{e^{i\theta} - z} = \begin{cases} \langle (\mathcal{K} - zI)^{-1} g, \mathcal{K}^* g \rangle, & \text{if } |z| > 1 \\ -z^{-1} \langle g, (\mathcal{K} - \bar{z}^{-1} I)^{-1} g \rangle, & \text{if } 0 < |z| < 1 \end{cases}$$

ResDMD computes with error control

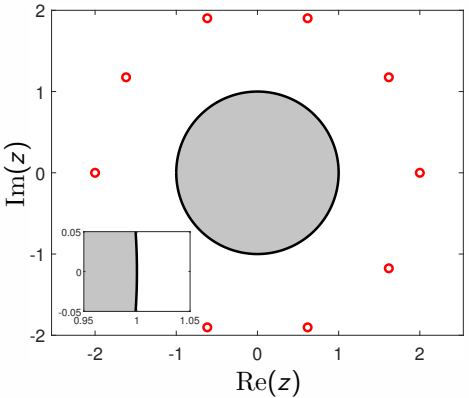
Example

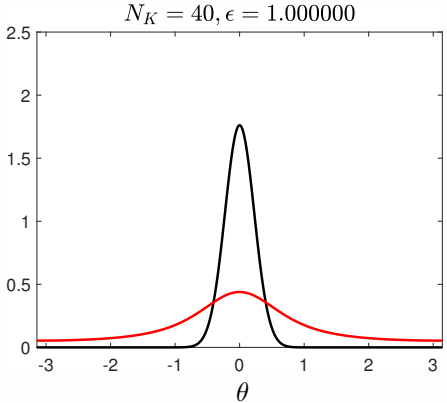
$$\mathcal{K} = \begin{pmatrix} \overline{\alpha_0} & \overline{\alpha_1}\rho_0 & \rho_0\rho_1 & & & \\ \rho_0 & -\overline{\alpha_1}\alpha_0 & -\alpha_0\rho_1 & & & \\ & \overline{\alpha_2}\rho_1 & -\overline{\alpha_2}\alpha_1 & \overline{\alpha_3}\rho_2 & \rho_3\rho_2 & \\ & \rho_2\rho_1 & -\alpha_1\rho_2 & -\overline{\alpha_3}\alpha_2 & -\rho_3\alpha_2 & \ddots \\ & & & \overline{\alpha_4}\rho_3 & -\overline{\alpha_4}\alpha_3 & \ddots \\ & & & \ddots & \ddots & \ddots \end{pmatrix}$$

$$\alpha_j = (-1)^j 0.95^{(j+1)/2}, \qquad \rho_j = \sqrt{1 - |\alpha_j|^2}$$

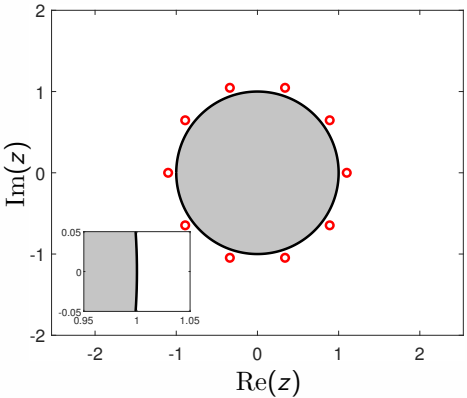
Generalized shift, typical building block of many dynamical systems.

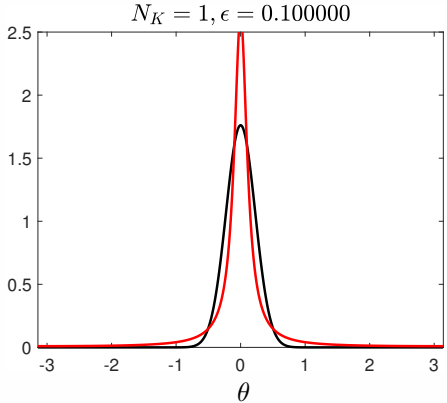
Fix N_K, vary ε: unstable!



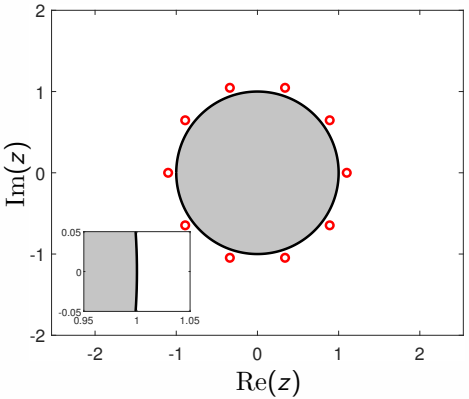


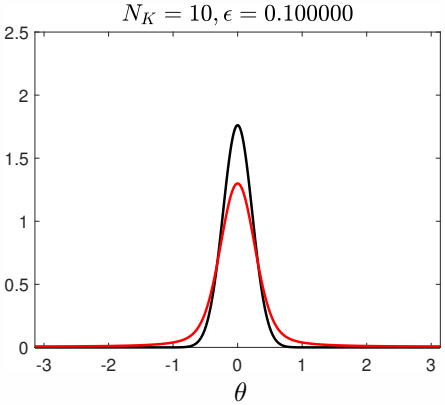
Fix ε, vary N_K: too smooth!





Adaptive: new matrix to compute residuals crucial

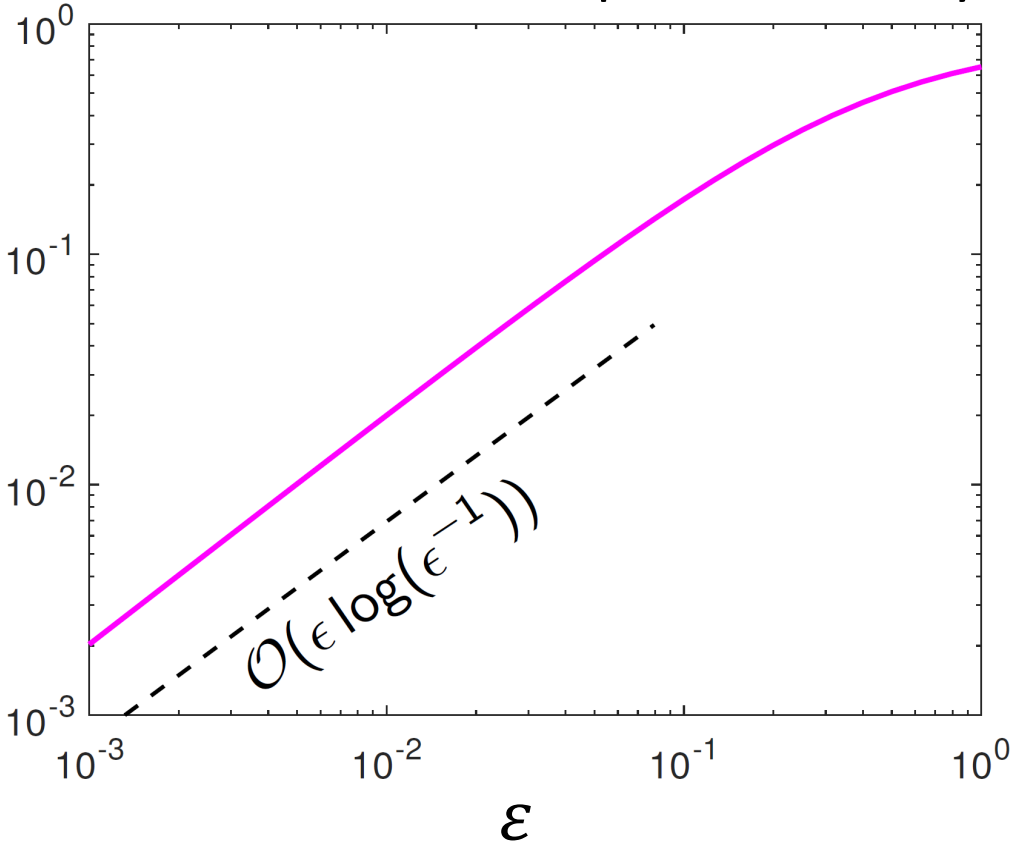




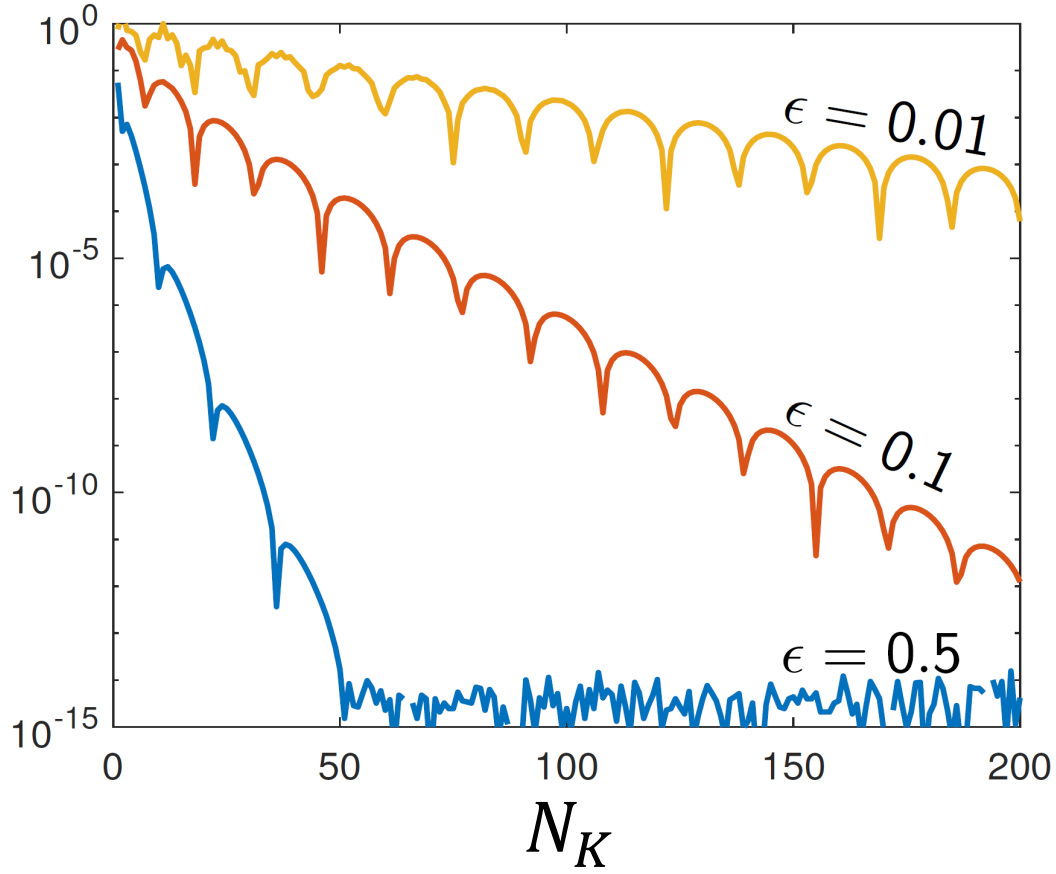
But … slow convergence

Problem: As $\varepsilon \downarrow 0$, error is $O(\varepsilon \log(1/\varepsilon))$ and $N_K(\varepsilon) \to \infty$.

Pointwise error for spectral density



Error due to discretization



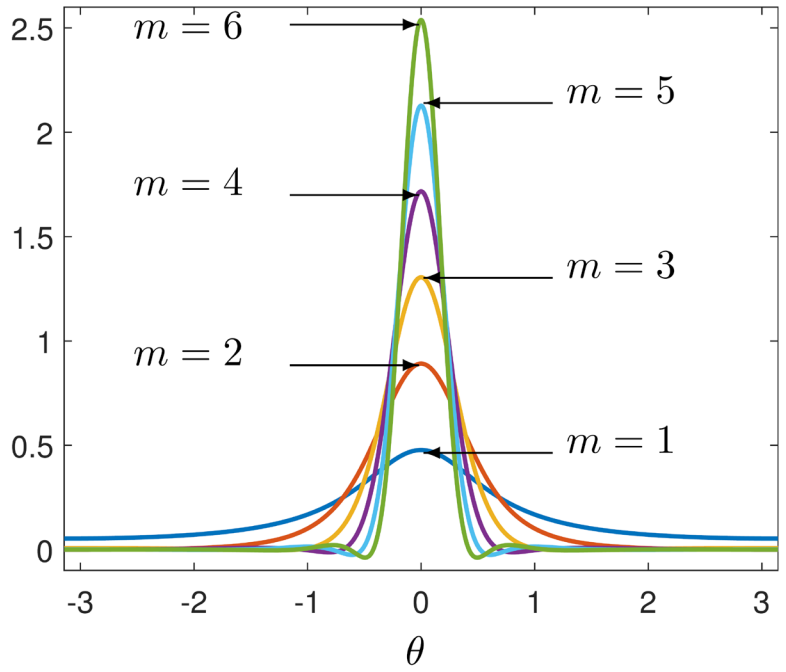
Small N_K critical in <u>data-driven</u> computations. Can we improve convergence rate?

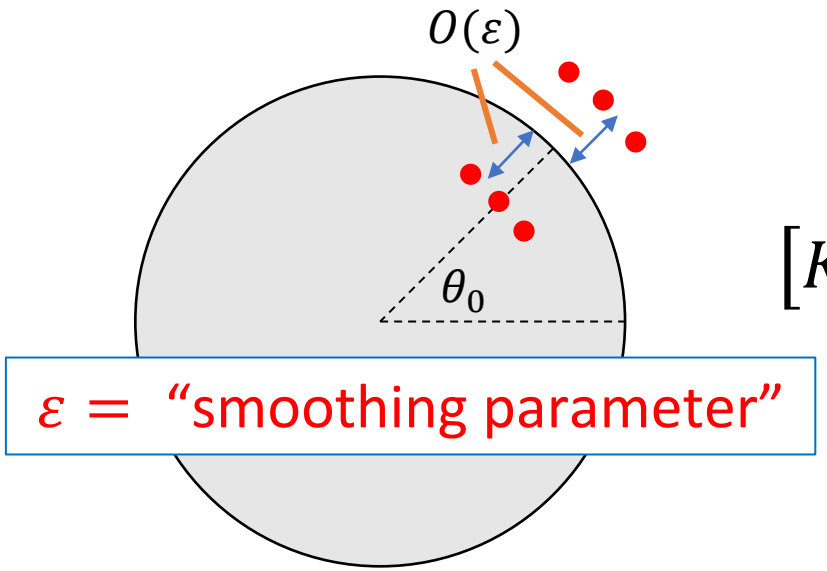
High-order rational kernels

mth order rational kernels:

$$K_\varepsilon(\theta) = \frac{e^{-i\theta}}{2\pi} \sum_{j=1}^{m} \left[\frac{c_j}{e^{-i\theta} - (1+\varepsilon\overline{z_j})^{-1}} - \frac{d_j}{e^{-i\theta} - (1+\varepsilon z_j)} \right]$$

Kernels



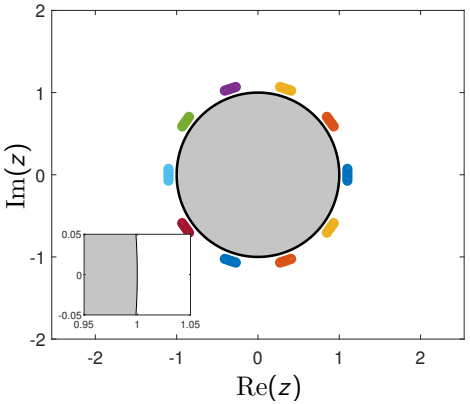


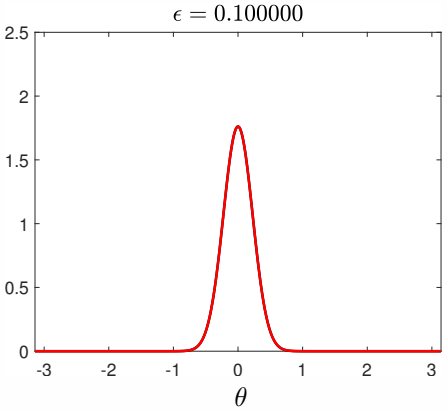
$\varepsilon =$ "smoothing parameter"

ResDMD computes with error control

$$\left[K_\varepsilon * \nu_g\right](\theta_0) = \sum_{j=1}^{m} \left[c_j \mathcal{C}_g\left(e^{i\theta_0}(1+\varepsilon\overline{z_j})^{-1}\right) - d_j \mathcal{C}_g\left(e^{i\theta_0}(1+\varepsilon z_j)\right) \right]$$

Smaller N_K (larger ε)





Convergence

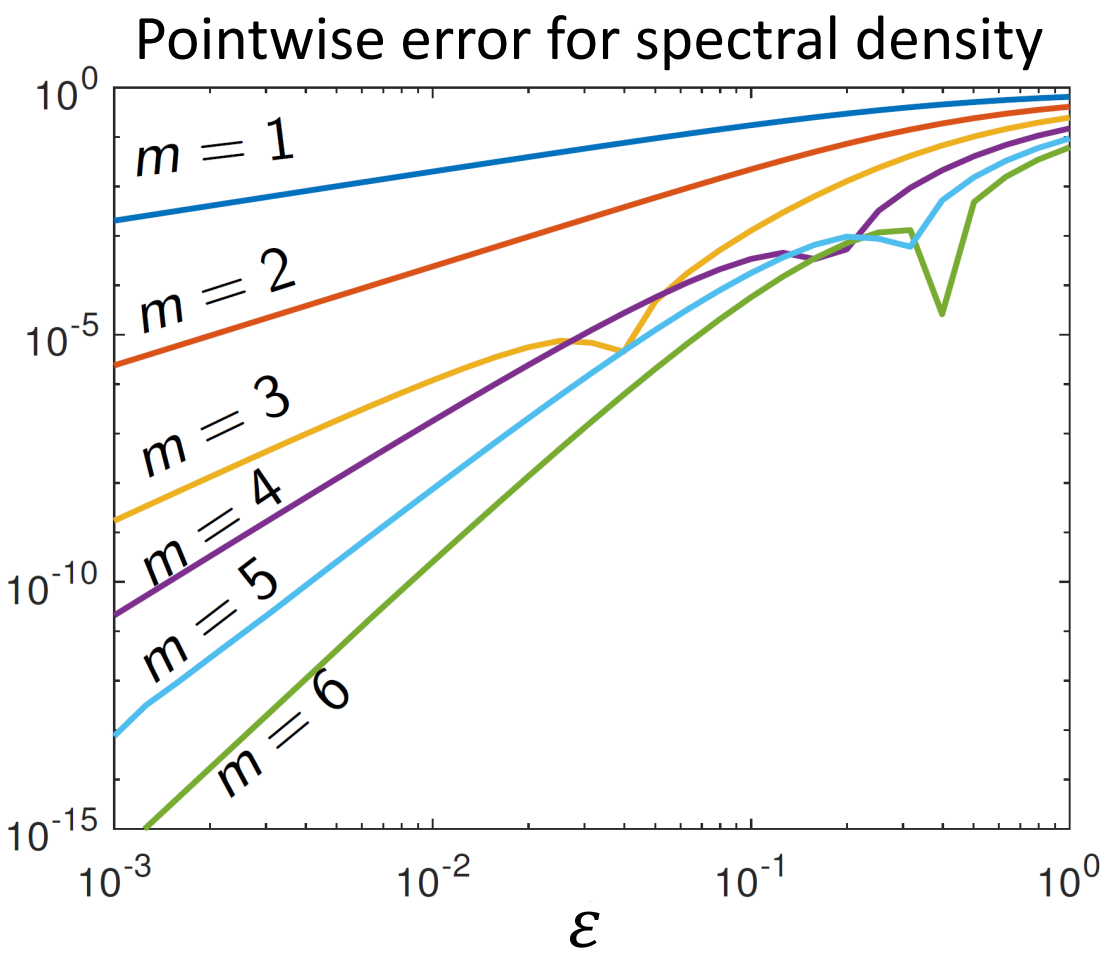
Theorem: *Automatic* selection of $N_K(\varepsilon)$ with $O(\varepsilon^m \log(1/\varepsilon))$ convergence:

- Density of continuous spectrum ρ_g.
  (pointwise and L^p)
- Integration against test functions.
  (weak convergence)

$$\int_{[-\pi,\pi]_{\text{per}}} h(\theta)\left[K_\varepsilon * \nu_g\right](\theta)\ \mathrm{d}\theta$$

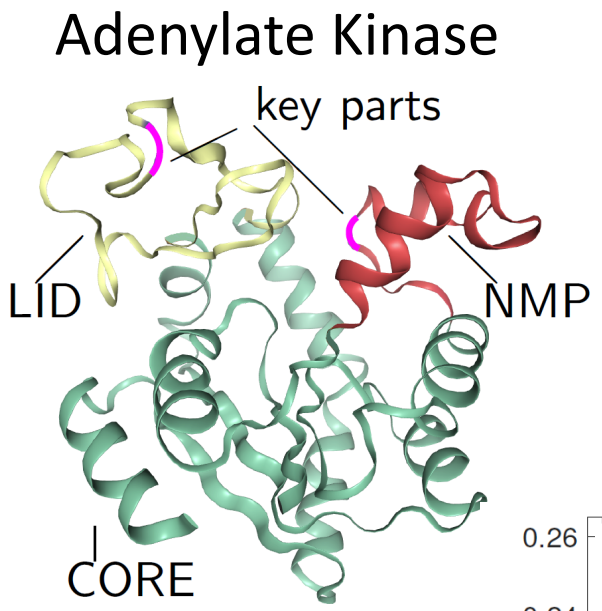
$$= \int_{[-\pi,\pi]_{\text{per}}} h(\theta)\,\mathrm{d}\nu_g(\theta) + O(\varepsilon^m \log(1/\varepsilon))$$

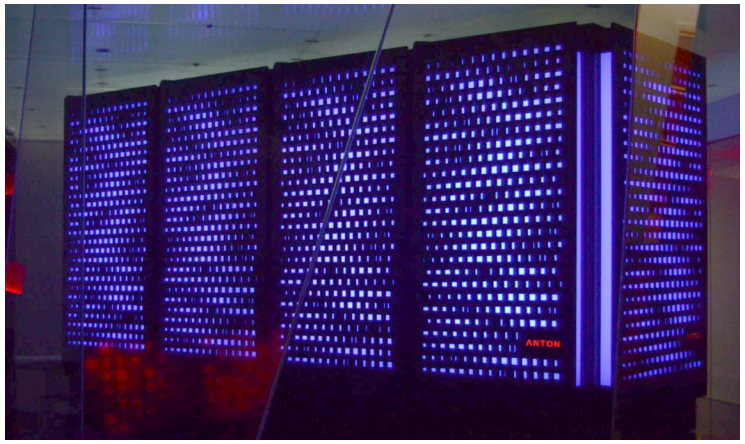
Also recover discrete spectrum.



- C., Townsend, "*Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems*," preprint.

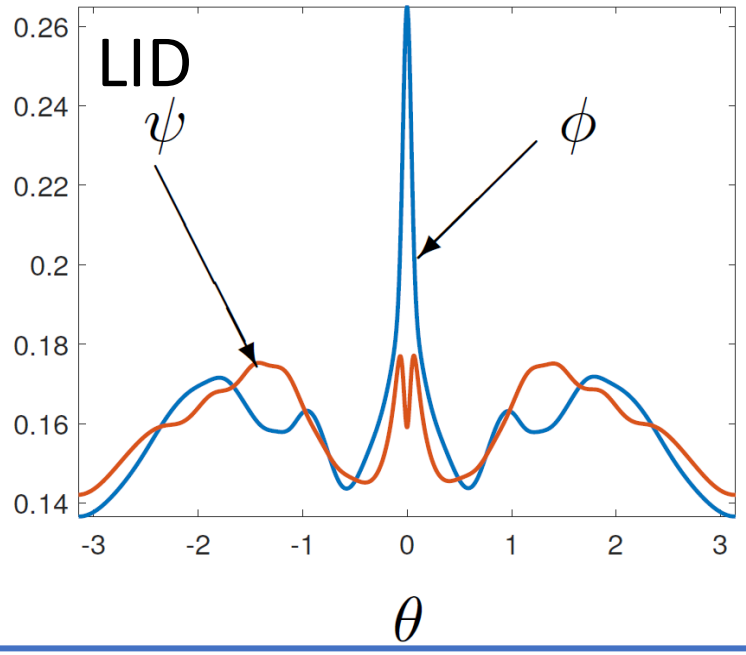
Example: Molecular dynamics (Adenylate Kinase)

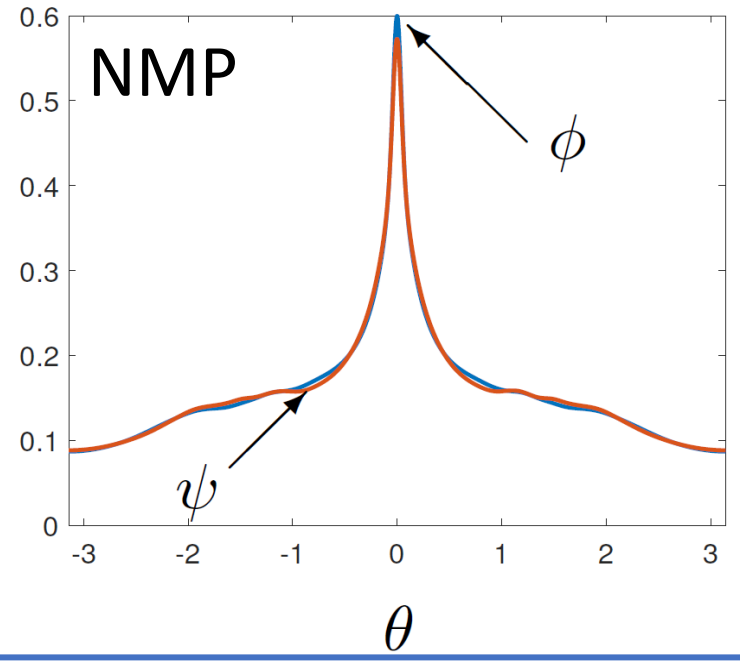




- Ambient dimension (d) $\approx 20{,}000$ (positions and momenta of atoms)
- 6th order kernel (spec res 10^{-6})

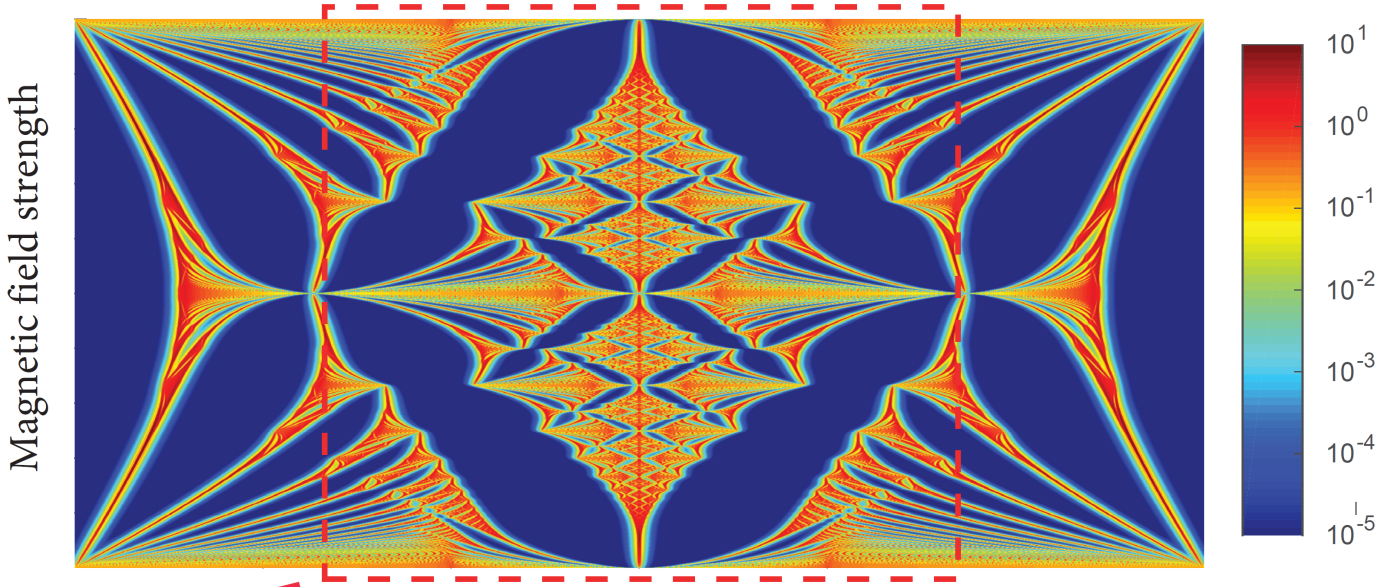
*Dataset: www.mdanalysis.org/MDAnalysisData/adk_equilibrium.html





- C., Townsend, "*Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems*," preprint.

Spectral measures of self-adjoint operators



Horizontal slice = spectral measure at constant magnetic field strength.

Software package

SpecSolve available at https://github.com/SpecSolve
Capabilities: ODEs, PDEs, integral operators, discrete operators.

- C., Horning, Townsend "*Computing spectral measures of self-adjoint operators,*" **SIAM Rev.**, 2021.

Further uses

Large d ($\Omega \subseteq \mathbb{R}^d$): *robust* and *scalable*

Popular to learn dictionary $\{\psi_1, \dots, \psi_{N_K}\}$

E.g., DMD with truncated SVD (linear dictionary, most popular), kernel methods (this talk), neural networks, etc.

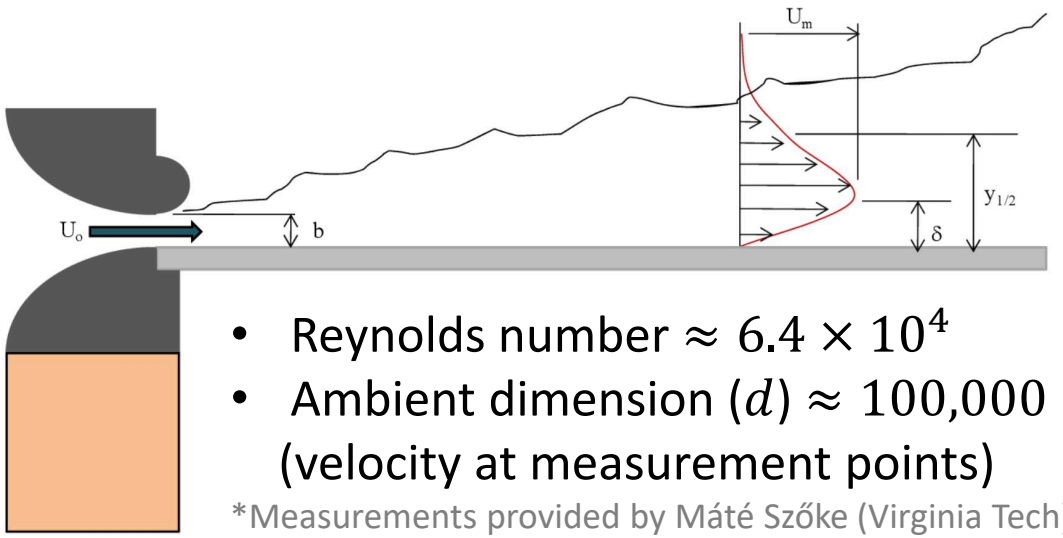
Q: Is discretisation $\mathbf{span}\{\psi_1, \dots, \psi_{N_K}\}$ large/rich enough?

Above algorithms:

- Pseudospectra: $\{z_k : \tau_k < \varepsilon\} \subseteq \mathrm{Spec}_\varepsilon(\mathcal{K})$ **error control**
- Spectral measures: $\mathcal{C}_g(z)$ and smoothed measures **adaptive check**

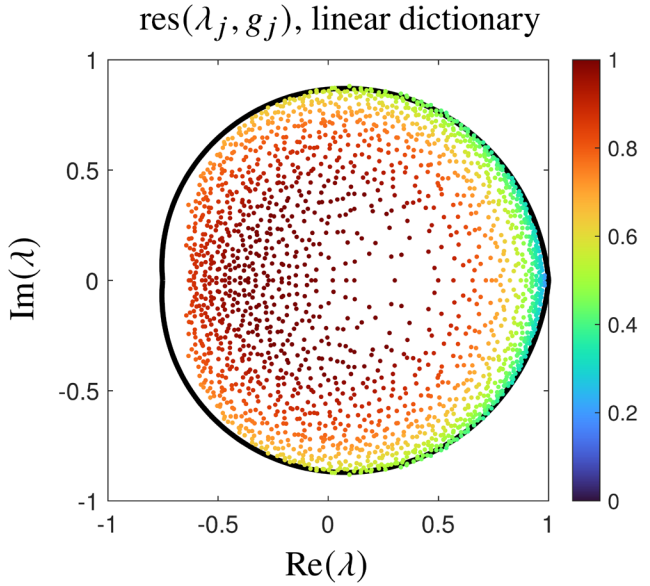
$\Longrightarrow$ Rigorously ***verify*** learnt dictionary $\{\psi_1, \dots, \psi_{N_K}\}$

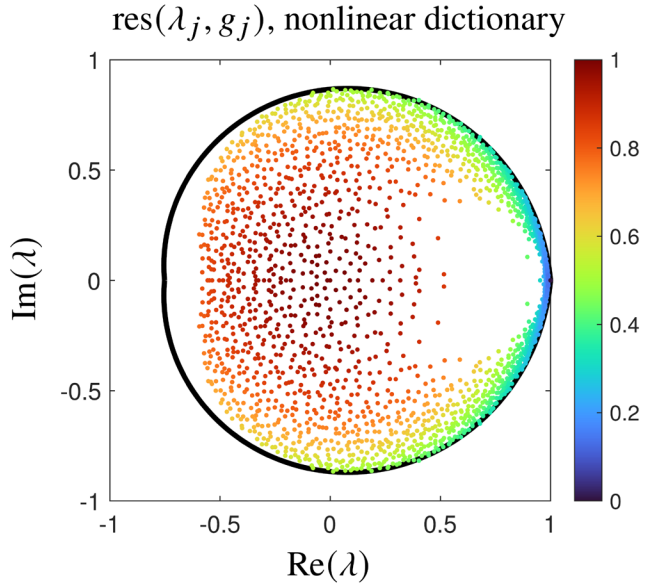
Example: Verify the dictionary



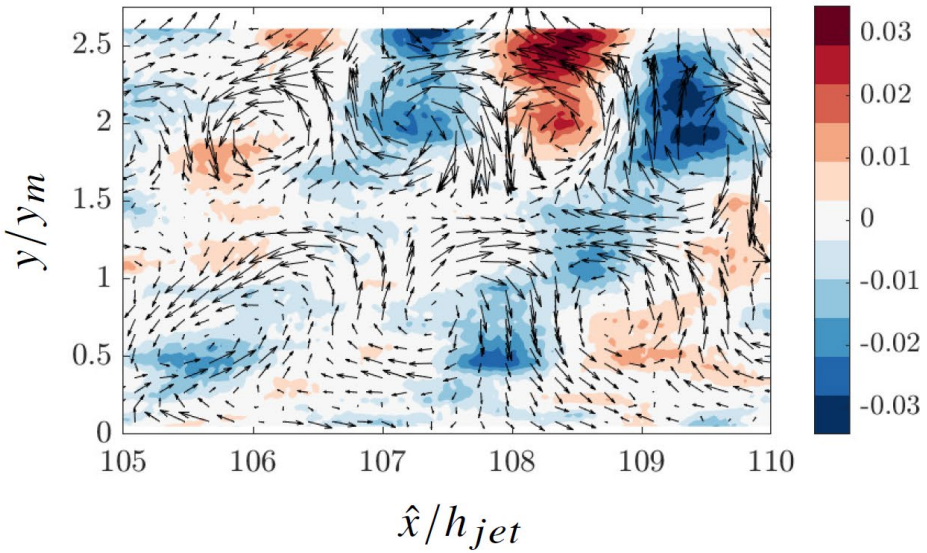
- Reynolds number $\approx 6.4 \times 10^4$
- Ambient dimension $(d) \approx 100{,}000$ (velocity at measurement points)

*Measurements provided by Máté Szőke (Virginia Tech)

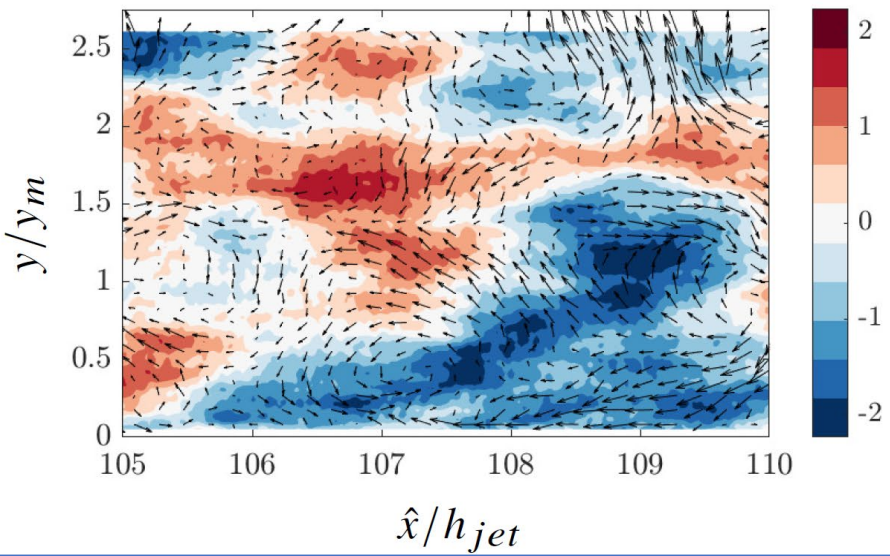




$\lambda = 0.9439 + 0.2458i$, error ≤ 0.0765

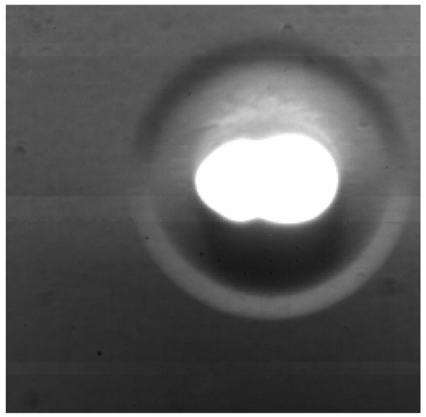


$\lambda = 0.8948 + 0.1065i$, error ≤ 0.1105

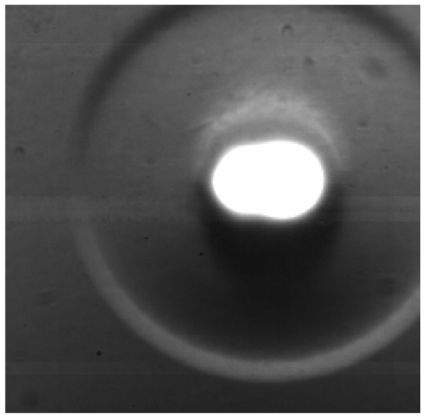


- C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition,*" **J. Fluid Mech.**, 2023.

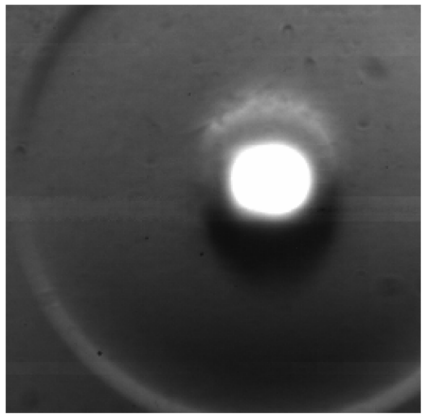
Example: Trustworthy Koopman mode decomposition



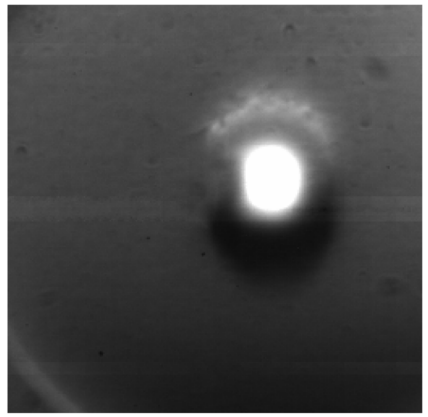
a) $t = 5\ \mu s$



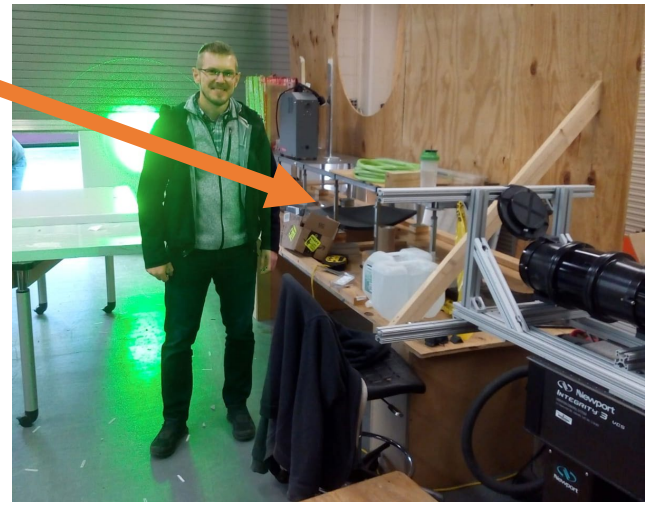
b) $t = 10\ \mu s$



c) $t = 15\ \mu s$

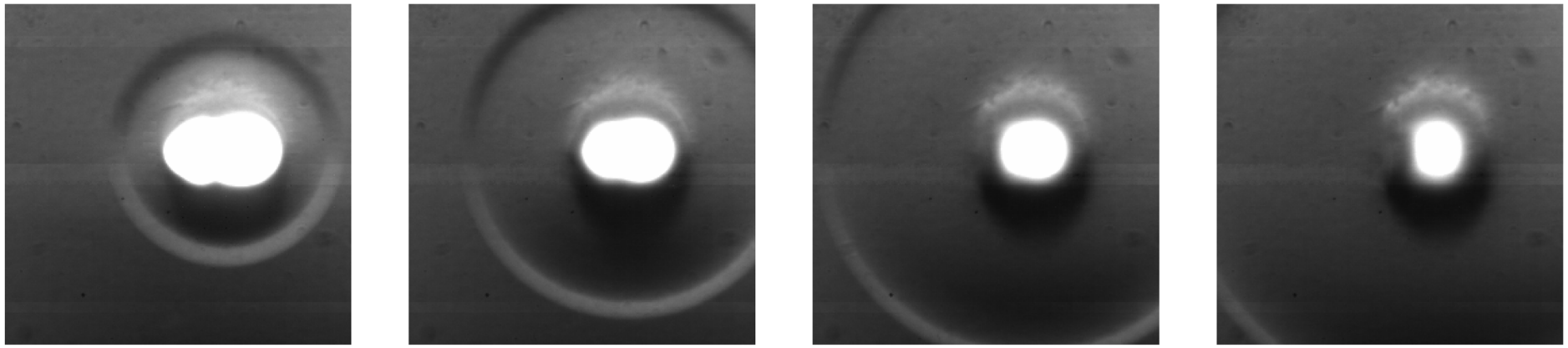


d) $t = 20\ \mu s$

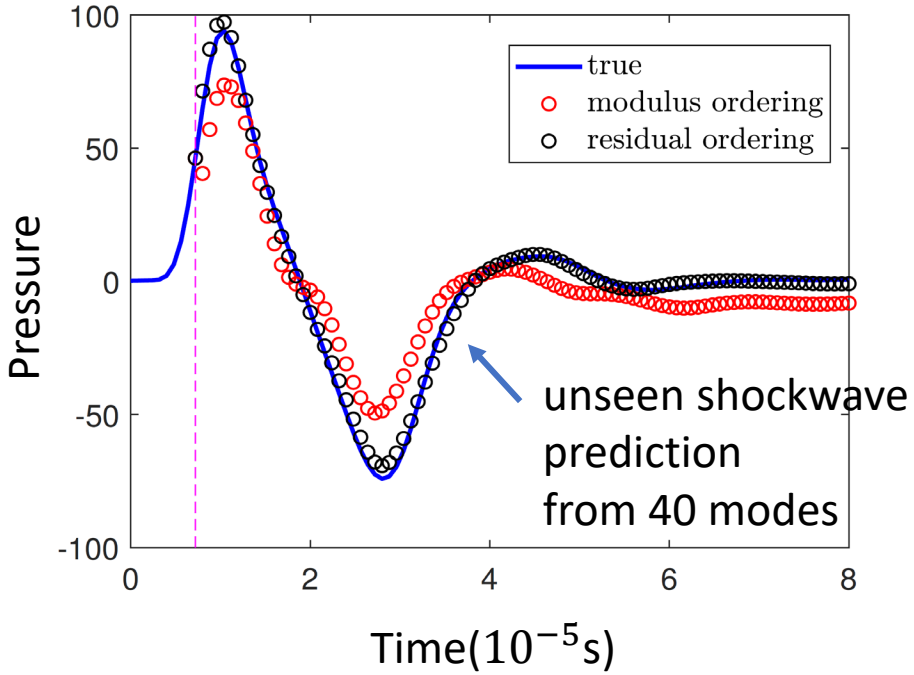


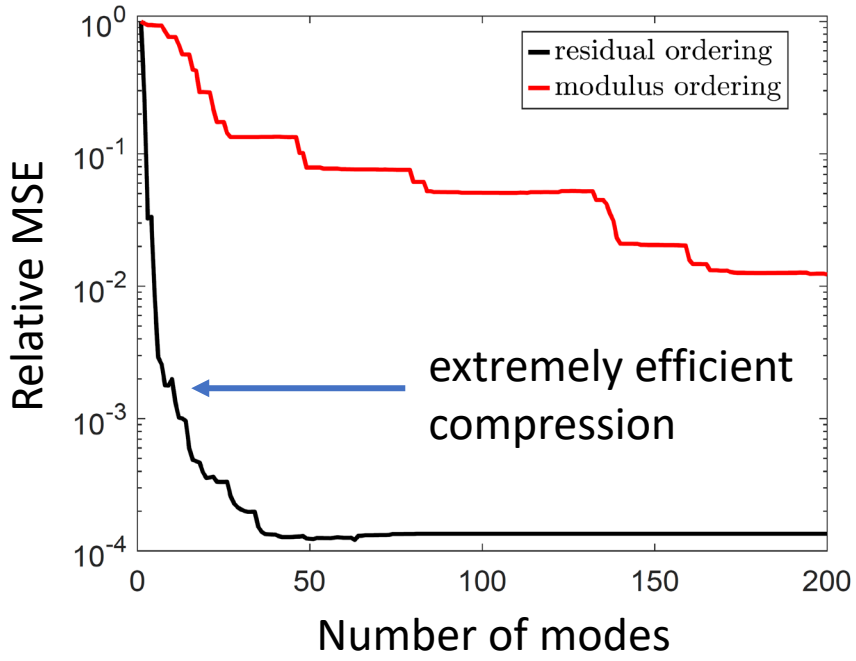
- C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition,*" **J. Fluid Mech.,** 2023.

Example: Trustworthy Koopman mode decomposition



a) $t = 5\ \mu s$ b) $t = 10\ \mu s$ c) $t = 15\ \mu s$ d) $t = 20\ \mu s$





- C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition,*" **J. Fluid Mech.,** 2023.

Wider programme

SCI provides needed assumptions

- Infinite-dimensional computational analysis $\Longrightarrow$ **Practical and rigorous algorithms.**
- Solvability Complexity Index $\Longrightarrow$ **Classify difficulty of problems, prove algorithms are optimal.**
- **Extends to:** Foundations of AI, optimization, computer-assisted proofs, and PDEs etc.

DATA SCIENCE + NUMERICAL ANALYSIS

- C., "*On the computation of geometric features of spectra of linear operators on Hilbert spaces*," **Found. Comput. Math.**, to appear.
- C., "*Computing spectral measures and spectral types,*" **Comm. Math. Phys.,** 2021.
- C., Horning, Townsend "*Computing spectral measures of self-adjoint operators,*" **SIAM Rev.**, 2021.
- C., Roman, Hansen, "*How to compute spectra with error control,*" **Phys. Rev. Lett.**, 2019.
- C., Hansen, "*The foundations of spectral computations via the solvability complexity index hierarchy,*" **J. Eur. Math. Soc.**, 2022.
- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **Proc. Natl. Acad. Sci. USA**, 2022.
- C., *"Computing semigroups with error control,"* **SIAM J. Numer. Anal.**, 2022.
- C., *"The mpEDMD Algorithm for Data-Driven Computations of Measure-Preserving Dynamical Systems,"* preprint.
- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, "*On the solvability complexity index hierarchy and towers of algorithms*," arXiv, 2020.
- Smale, "*The fundamental theorem of algebra and complexity theory*," **Bull. Amer. Math. Soc.**, 1981, *36 pp.*
- McMullen, "*Families of rational maps and iterative root-finding algorithms,*" **Ann. of Math.**, 1987, *27 pp.*

Interested?

Some future directions:

- ResDMD in neuroscience.
- ResDMD + control $\Longrightarrow$ error control?
- Embed & learn symmetries (e.g., check out the algorithm mpEDMD).
- Forecasting with error bounds.
- Koopmanism meets neural nets (and vice versa).
- Foundations results for dynamical systems (i.e., impossibility results)?

Opportunities to collaborate, visit Cambridge, grad students & beyond!

Rigorous data-driven Koopmanism!

"Too much" or "Too little"

Idea: New matrix for residual

⇒ **ResDMD** for computing spectra.

Continuous spectra

Idea: Smoothing via resolvent and **ResDMD**.

Is it right?

ResDMD verifies computations.

E.g., learned dictionaries.

Code: https://github.com/MColbrook/Residual-Dynamic-Mode-Decomposition

Rigorous data-driven Koopmanism!

"Too much" or "Too little"

Idea: New matrix for residual

⇒ **ResDMD** for computing spectra.

Continuous spectra

Idea: Smoothing via resolvent and **ResDMD**.

Is it right?

ResDMD verifies computations.

E.g., learned dictionaries.

Code: https://github.com/MColbrook/Residual-Dynamic-Mode-Decomposition

Newsjournal of the Society for Industrial and Applied Mathematics sinews.siam.org

siam news

Volume 56/ Issue 1
January/February 2023

Optimization and Learning with Zeroth-order Stochastic Oracles

By Stefan M. Wild

Mathematical optimization is a foundational technology for machine learning and the solution of design, decision, and control problems. In most optimization applications, the principal assumption is the availability of at least the

sequence is that material properties are only available via *in situ* and *in operando* characterization. In the context of optimization, this scenario is called a "zeroth-order oracle" — our knowledge about a particular system or property is data driven and limited by the black-box nature of measurement procurement. An additional challenge is

An optimization solver specifies a particular composition of solvents and bases, an operating temperature, and reaction times; this combination is then run through a continuous flow reactor. The material that exits the reactor is then automatically characterized

through an inline nuclear magnetic resonance detector that illuminates properties of the synthesized materials. These stochastic, zeroth-order oracle outputs return to the solver in a closed-loop setting that

See **Optimization** on page 3

search for novel materials for energy storage. In order to create viable new materials, we must move beyond pure theory and account for the actual processes that occur during materials synthesis. A necessary

ers—such as the laboratory environment in Figure 1—doing so is impossible.

Figure 1 displays an instantiation of a data-driven optimization setting in a chem-

Figure 1. *A continuous flow reactor at Argonne National Laboratory's Materials Engineering Research Facility uses the optimization solver ParMOO to perform autonomous discovery [2].*

Read more about these breakthroughs in SIAM News!

Nonprofit Org
U.S. Postage
PAID
Permit No 360
Bellmawr, NJ

Resilient Data-driven Dynamical Systems with Koopman: An Infinite-dimensional Numerical Analysis Perspective

By Steven L. Brunton and Matthew J. Colbrook

Dynamical systems, which describe the evolution of systems in time, are ubiquitous in modern science and engineering. They find use in a wide variety of applications, from mechanics and circuits to climatology, neuroscience, and epidemiology. Consider a discrete-time dynamical system with state $\boldsymbol{x}$ in a state space $\Omega \subset \mathbb{R}^d$ that is governed by an unknown and typically nonlinear function $\boldsymbol{F}: \Omega \to \Omega$:

$$\boldsymbol{x}_{n+1} = \boldsymbol{F}(\boldsymbol{x}_n), \quad n \geq 0. \qquad (1)$$

The classical, geometric way to analyze such systems—which dates back to the seminal work of Henri Poincaré—is based on the local analysis of fixed points, periodic orbits, stable or unstable manifolds, and so forth. Although Poincaré's framework has revolutionized our understanding of dynamical systems, this approach has at least two challenges in many modern applications: (i) Obtaining a global understanding of the nonlinear dynamics and (ii) handling systems that are either too complex to analyze or offer incomplete information about the evolution (i.e., unknown, high-dimensional, and highly nonlinear $\boldsymbol{F}$).

Koopman operator theory, which originated with Bernard Koopman and John von Neumann [6, 7], provides a powerful alternative to the classical geometric view of dynamical systems because it addresses nonlinearity: the fundamental issue that underlies the aforementioned challenges.

We lift the nonlinear system (1) into an infinite-dimensional space of observable functions $g: \Omega \to \mathbb{C}$ via a Koopman operator $\mathcal{K}$:

$$\mathcal{K}g(\boldsymbol{x}_n) = g(\boldsymbol{x}_{n+1}).$$

The evolution dynamics thus become linear, allowing us to utilize generic solution techniques that are based on spectral decompositions. In recent decades, Koopman operators have captivated researchers because of emerging data-driven and numerical implementations that coincide with the rise of machine learning and high-performance computing [2].

One major goal of modern Koopman operator theory is to find a coordinate transformation with which a linear system may approximate even strongly nonlinear dynamics; this coordinate system relates to the spectrum of the Koopman operator. In 2005, Igor Mezić introduced the Koopman mode decomposition [8], which provided a theoretical basis for connecting the dynamic mode decomposition (DMD) with the Koopman operator [9, 10]. DMD quickly became the workhorse algorithm for computational approximations of the Koopman operator due to its simple and highly extensible formulation in terms of linear algebra, and the fact that it applies equally well to data-driven modeling when no governing equations are available. However, researchers soon realized that simply building linear models in terms of the primitive measured variables cannot sufficiently capture nonlinear dynamics beyond periodic and quasi-periodic phenomena. A major breakthrough occurred with the introduction of *extended* DMD (EDMD), which generalizes DMD to a broader class of basis functions in which to expand eigenfunctions of the Koopman operator [11].

See **Dynamical Systems** on page 4

Figure 1. *Koopman modes of a turbulent flow (Reynolds number 3.9×10^5) past a cascade of airfoils that are computed from trajectory data ($d = 300,000$). Koopman modes are projections of the physical field onto eigenfunctions of $\mathcal{K}$; they provide the collective motion of the fluid that occurs at the same spatial frequency, growth, or decay rate according to an approximate eigenvalue λ.* **1a.** *Koopman modes that were computed via existing state-of-the-art techniques. Note the lack of error bounds.* **1b.** *Koopman modes that were computed using residual dynamic mode decomposition (ResDMD). The physical picture in 1b is different from 1a, but we know that it is correct because of the guaranteed relative error bounds (green text). This outcome illustrates the importance of verification. Figure courtesy of Matthew Colbrook.*

Rigorous data-driven Koopmanism!

“Too much” or “Too little”

Idea: New matrix for residual

⇒ **ResDMD** for computing spectra.

Continuous spectra

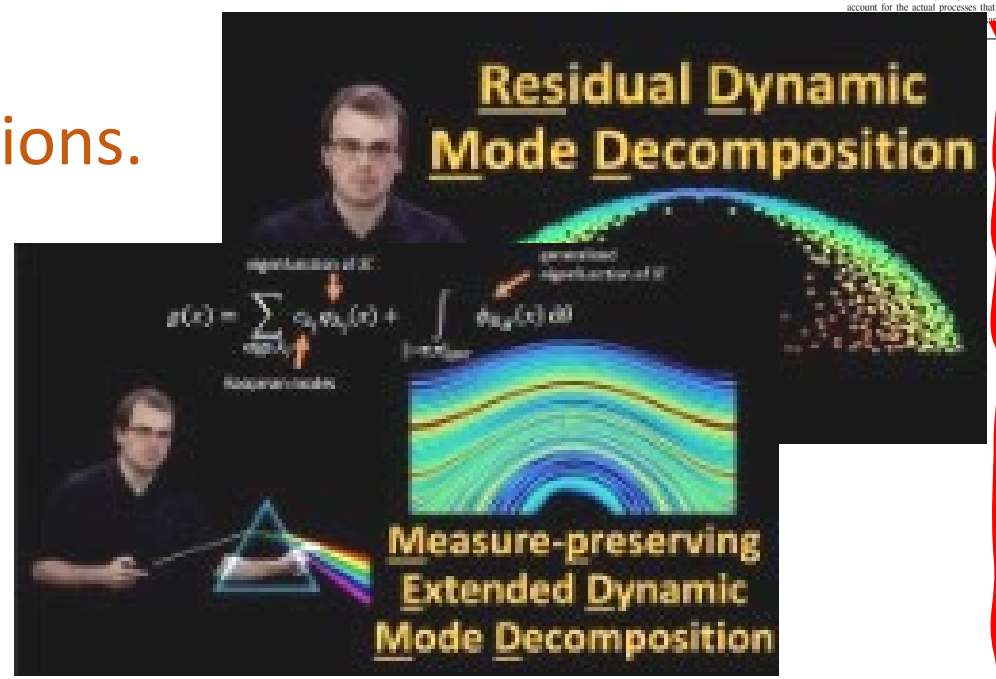
Idea: Smoothing via resolvent and **ResDMD**.

Is it right?

ResDMD verifies computations.

E.g., learned dictionaries.

Short video summaries available on YouTube:
(Thank you Steve Brunton for letting me use your channel!)



Code: https://github.com/MColbrook/Residual-Dynamic-Mode-Decomposition


Newsjournal of the Society for Industrial and Applied Mathematics sinews.siam.org

siam news

Volume 56/ Issue 1
January/February 2023

Optimization and Learning with Zeroth-order Stochastic Oracles

By Stefan M. Wild

Mathematical optimization is a foundational technology for machine learning and the solution of design, decision, and control problems. In most optimization applications, the principal assumption is the availability of at least the

sequence is that material properties are only available via *in situ* and *in operando* characterization. In the context of optimization, this scenario is called a “zeroth-order oracle” — our knowledge about a particular system or property is data driven and limited by the black-box nature of measurement procurement. An additional challenge is

An optimization solver specifies a particular composition of solvents and bases, an operating temperature, and reaction times; this combination is then run through a continuous flow reactor. The material that exits the reactor is then automatically characterized

through an inline nuclear magnetic resonance detector that illuminates properties of the synthesized materials. These stochastic, zeroth-order oracle outputs return to the solver in a closed-loop setting that

See **Optimization** on page 3

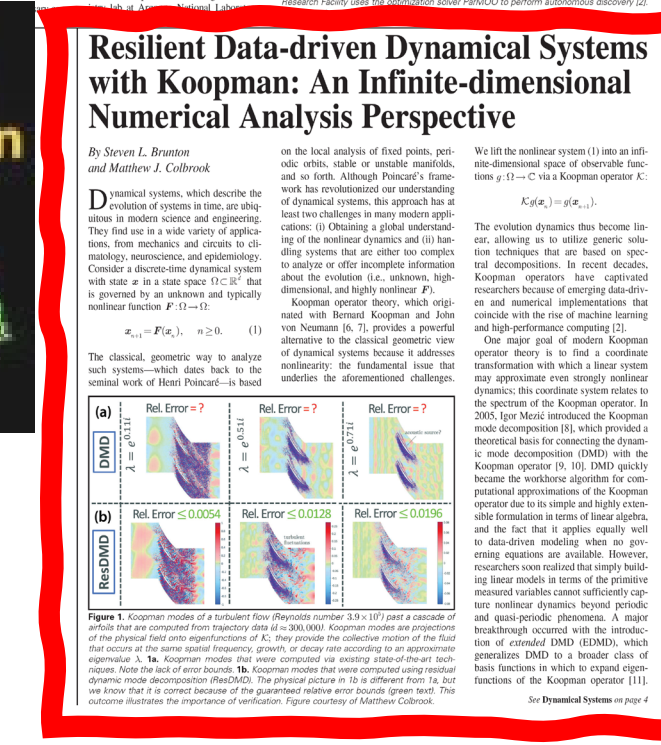
search for novel materials for energy storage. In order to create viable new materials, we must move beyond pure theory and account for the actual processes that occur

ers—such as the laboratory environment in Figure 1—doing so is impossible.

Figure 1 displays an instantiation of a data-driven optimization setting in a chem-

Figure 1. *A continuous flow reactor at Argonne National Laboratory's Materials Engineering Research Facility uses the optimization solver ParMOO to perform autonomous discovery [2].*

Read more about these breakthroughs in SIAM News!


Resilient Data-driven Dynamical Systems with Koopman: An Infinite-dimensional Numerical Analysis Perspective

By Steven L. Brunton and Matthew J. Colbrook

Dynamical systems, which describe the evolution of systems in time, are ubiquitous in modern science and engineering. They find use in a wide variety of applications, from mechanics and circuits to climatology, neuroscience, and epidemiology. Consider a discrete-time dynamical system with state $\boldsymbol{x}$ in a state space $\Omega \subset \mathbb{R}^d$ that is governed by an unknown and typically nonlinear function $\boldsymbol{F}: \Omega \to \Omega$:

$$\boldsymbol{x}_{n+1} = \boldsymbol{F}(\boldsymbol{x}_n), \quad n \geq 0. \qquad (1)$$

The classical, geometric way to analyze such systems—which dates back to the seminal work of Henri Poincaré—is based on the local analysis of fixed points, periodic orbits, stable or unstable manifolds, and so forth. Although Poincaré's framework has revolutionized our understanding of dynamical systems, this approach has at least two challenges in many modern applications: (i) Obtaining a global understanding of the nonlinear dynamics and (ii) handling systems that are either too complex to analyze or offer incomplete information about the evolution (i.e., unknown, high-dimensional, and highly nonlinear $\boldsymbol{F}$).

Koopman operator theory, which originated with Bernard Koopman and John von Neumann [6, 7], provides a powerful alternative to the classical geometric view of dynamical systems because it addresses nonlinearity: the fundamental issue that underlies the aforementioned challenges.

We lift the nonlinear system (1) into an infinite-dimensional space of observable functions $g: \Omega \to \mathbb{C}$ via a Koopman operator $\mathcal{K}$:

$$\mathcal{K}g(\boldsymbol{x}_n) = g(\boldsymbol{x}_{n+1}).$$

The evolution dynamics thus become linear, allowing us to utilize generic solution techniques that are based on spectral decompositions. In recent decades, Koopman operators have captivated researchers because of emerging data-driven and numerical implementations that coincide with the rise of machine learning and high-performance computing [2].

One major goal of modern Koopman operator theory is to find a coordinate transformation with which a linear system may approximate even strongly nonlinear dynamics; this coordinate system relates to the spectrum of the Koopman operator. In 2005, Igor Mezić introduced the Koopman mode decomposition [8], which provided a theoretical basis for connecting the dynamic mode decomposition (DMD) with the Koopman operator [9, 10]. DMD quickly became the workhorse algorithm for computational approximations of the Koopman operator due to its simple and highly extensible formulation in terms of linear algebra, and the fact that it applies equally well to data-driven modeling when no governing equations are available. However, researchers soon realized that simply building linear models in terms of the primitive measured variables cannot sufficiently capture nonlinear dynamics beyond periodic and quasi-periodic phenomena. A major breakthrough occurred with the introduction of *extended* DMD (EDMD), which generalizes DMD to a broader class of basis functions in which to expand eigenfunctions of the Koopman operator [11].

Figure 1. *Koopman modes of a turbulent flow (Reynolds number 3.9×10^5) past a cascade of airfoils that are computed from trajectory data ($d = 300,000$). Koopman modes are projections of the physical field onto eigenfunctions of $\mathcal{K}$; they provide the collective motion of the fluid that occurs at the same spatial frequency, growth, or decay rate according to an approximate eigenvalue λ.* **1a.** *Koopman modes that were computed via existing state-of-the-art techniques. Note the lack of error bounds.* **1b.** *Koopman modes that were computed using residual dynamic mode decomposition (ResDMD). The physical picture in 1b is different from 1a, but we know that it is correct because of the guaranteed relative error bounds (green text). This outcome illustrates the importance of verification. Figure courtesy of Matthew Colbrook.*

See **Dynamical Systems** on page 4

Additional slides…

Quadrature with trajectory data

$$\text{E.g.,} \quad \langle \mathcal{K}\psi_k, \psi_j \rangle = \lim_{M\to\infty} \sum_{m=1}^{M} w_m \overline{\psi_j(x^{(m)})} \underbrace{\psi_k(y^{(m)})}_{[\mathcal{K}\psi_k](x^{(m)})}$$

Three examples:

- **High-order quadrature:** $\{x^{(m)}, w_m\}_{m=1}^{M}$ M-point quadrature rule.
  Rapid convergence. Requires free choice of $\{x^{(m)}\}_{m=1}^{M}$ and small d.
- **Random sampling:** $\{x^{(m)}\}_{m=1}^{M}$ selected at random.
  Large d. Slow Monte Carlo $O(M^{-1/2})$ rate of convergence.
- **Ergodic sampling:** $x^{(m+1)} = F(x^{(m)})$.
  Single trajectory, large d. Requires ergodicity, convergence can be slow.

Most common

Example: Barriers of deep learning

PNAS

RESEARCH ARTICLE | APPLIED MATHEMATICS | FULL ACCESS

The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem

Matthew J. Colbrook, Vegard Antun, and Anders C. Hansen — Authors Info & Affiliations

March 16, 2022 | 119 (12) e2107151119

Significance

Instability is the Achilles' training algorithms findir ones. This foundational i century on the limits of A demonstrate limitations NNs. Despite numerous only in specific cases do t classification theory on w suitable conditions—are number of hidden layers

UNIVERSITY OF CAMBRIDGE | Study at Cambridge | About the University | Research at Cambridge

/ Research / News / Mathematical paradox demonstrates the limits of AI

Research

Research home | News | Our people | Spotlights | About research | Business and enterpr

Mathematical paradox demonstrates the limits of AI

IEEE.ORG | IEEE XPLORE DIGITAL LIBRARY | IEEE STANDARDS | MORE SITES

IEEE Spectrum FOR THE TECHNOLOGY INSIDER — Type to search

NEWS | ARTIFICIAL INTELLIGENCE

Some AI Systems May Be Impossible to Compute

New research suggests there are limitations to what deep neural networks can do

BY CHARLES Q. CHOI | 30 MAR 2022 | 4 MIN READ

…they get things wrong, but …new study, AI generally suffers …ematical paradox.

> *"There are fundamental limits inherent in mathematics and, similarly, AI algorithms can't exist for certain problems"*
>
> — *Matthew Colbrook*

Newsjournal of the Society for Industrial and Applied Mathematics — sinews.siam.org

siam news

Volume 55/ Issue 4
May 2022

Protecting Privacy with Synthetic Data

By Matthew R. Francis

Researchers across every scientific discipline need complete and reliable data sets to draw trustworthy conclusions. However, publishing all data from a given study can be undesirable. For example, medical data in particular include personal information that—if published in full—would violate patients' privacy and potentially expose them to harm. Similarly, many studies in the social sciences include demographic or geographical data that could easily be exploited by unscrupulous parties.

In short, researchers must strike a delicate balance between publishing enough data to verify their conclusions and protecting the privacy of the people involved. Unfortunately, multiple studies have shown that simply anonymizing the data—by removing individuals' names before publication, for instance—is insufficient, as outsiders can use context clues to reconstruct missing information and expose research subjects. "We want to generate synthetic data for public release to replace the original data set," Bei Jiang of the University of Alberta said. "When we design our framework, we have this main goal in mind: we want to produce the same inference results as in the original data set."

In contrast with falsified data, which is one of the deadliest scientific sins, researchers can generate *synthetic* data directly from original data sets. If the construction process is done properly, other scientists can then analyze this synthetic data and trust that their conclusions are no different from what they would have obtained with full access to the original raw data — ideally, at least. "When you [create] synthetic data, what does it mean to be private yet realistic?" Sébastien Gambs of the University of Québec in Montréal asked. "It's still an open research question."

During the 2022 American Association for the Advancement of Science Annual Meeting,[1] which took place virtually in February, Jiang and Gambs each presented formal methods for the generation of synthetic data that ensure privacy. Their models draw from multiple fields to address challenges in the era of big data, where the stakes are higher than ever. "There is always a trade-off between utility and risk," Jiang said. "If you want to protect people [who] are at a higher risk, then you perturb their data. But the utility will be lowered the more you perturb. A better approach is to account for their risks to begin with."

Unfortunately, malicious actors have access to the same algorithmic tools as researchers. Therefore, protection of confidentially also involves testing synthetic data against the types of attacks that such players might utilize. "In practice, this helps one really understand the translation between an abstract privacy parameter and a practical guarantee," Gambs said. In other words, the robustness of a formal mathematical model is irrelevant if the model is not well implemented.

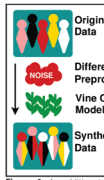


Figure 1. *In addition to adding noise to the data set, Sébastien Gambs' differential privacy-based method processes it with an information theory algorithm to obtain synthetic data that—in principle—shields the privacy of the people involved. Figure courtesy of the author.*

[1] https://aaas.confex.com/aaas/2022/meetingapp.cgi

Differential Privacy Made Simple(r)

Gambs and his collaborators turned to differential privacy: a powerful mathematical formalism that in principle is the best available technique for securing confidentiality. However, the approach is also complex and difficult to implement without a high degree of statistical knowledge. To smooth the

See **Synthetic Data** *on page 3*

Figure 2. *Researchers can protect privacy by performing a full statistical analysis on the original data set, then using a missing-data algorithm called multiple imputation to construct a synthetic data set that has exactly the same statistical characteristics. Figure courtesy of the author.*

Nonprofit Org
U.S. Postage
PAID
Permit No. 360
Bethlehem, NJ

Proving Existence Is Not Enough: Mathematical Paradoxes Unravel the Limits of Neural Networks in Artificial Intelligence

By Vegard Antun, Matthew J. Colbrook, and Anders C. Hansen

The impact of deep learning (DL), neural networks (NNs), and artificial intelligence (AI) over the last decade has been profound. Advances in computer vision and natural language processing have yielded smart speakers in our homes, driving assistance in our cars, and automated diagnoses in medicine. AI has also rapidly entered scientific computing. However, overwhelming amounts of empirical evidence [3, 8] suggest that modern AI is often non-robust (unstable), may generate hallucinations, and can produce nonsensical output with high levels of prediction confidence (see Figure 1). These issues present a serious concern for AI use within legal frameworks. As stated by the European Commission's Joint Research Centre, *"In the light of the recent advances in AI, the serious negative consequences of its use for EU citizens and organisations have led to multiple initiatives [...] Among the identified requirements, the concepts of robustness and explainability of AI systems have emerged as key elements for a future regulation."*[1]

Robustness and trust of algorithms lie at the heart of numerical analysis [9]. The *lack* of robustness and trust in AI is hence the Achilles' heel of DL and has become a serious political issue. Classical approximation theorems show that a continuous function can be approximated arbitrarily well by a NN [5]. Therefore, stable problems that are described by stable functions can be solved stably with a NN. These results inspire the following fundamental question: *Why does DL lead to unstable methods and AI-generated hallucinations, even in scenarios where we can prove that stable and accurate NNs exist?*

Our main result reveals a serious issue for certain problems; while stable and accurate NNs may provably exist, no training algorithm can obtain them (see Figure 2, on page 4). As such, existence theorems on approximation qualities of NNs (e.g., universal approximation) represent only the first step towards a complete understanding of modern AI. Sometimes they even provide overly optimistic estimates of possible NN achievements.

The Limits of AI: Smale's 18th Problem

The strong optimism that surrounds AI is evident in computer scientist Geoffrey Hinton's 2017 quote: *"They should stop training radiologists now."*[2] Such optimism is comparable to the confidence that surrounded mathematics in the early 20th century, as summed up in David Hilbert's sentiment: *"Wir müssen wissen, Wir werden wissen."* ["We must know. We will know"].

Hilbert believed that mathematics could prove or disprove any statement, and that there were no restrictions on which problems algorithms could solve. The seminal contributions of Kurt Gödel [7] and Alan Turing [12] turned Hilbert's idealism upside down by establishing paradoxes that expedited impossibility results about the feasible achievements of mathematics and digital computers.

A similar program on the boundaries of AI is necessary. Stephen Smale already suggested such a program in the 18th problem on his list of mathematical problems for the 21st century: *What are the limits of AI?* [11].

See **Mathematical Paradoxes** *on page 4*

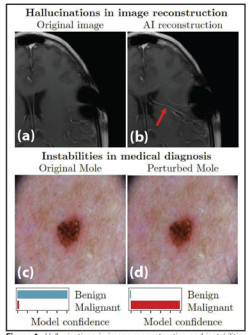


Figure 1. *Hallucinations in image reconstruction and instabilities in medical diagnoses.* **1a.** *The correct, original image from the 2020 fastMRI Challenge.* **1b.** *Reconstruction by an artificial intelligence (AI) method that produces an incorrect detail (AI-generated hallucination).* **1c.** *Dermatoscopic image of a benign melanocytic nevus, along with the diagnostic probability computed by a deep neural network (NN).* **1d.** *Combined image of the nevus with a slight perturbation and the diagnostic probability from the same deep NN. One diagnosis is clearly incorrect, but can an algorithm determine which one? Figures 1a and 1b are courtesy of the 2020 fastMRI Challenge [10], and 1c and 1d are courtesy of [6].*

[1] https://publications.jrc.ec.europa.eu/repository/handle/JRC119336

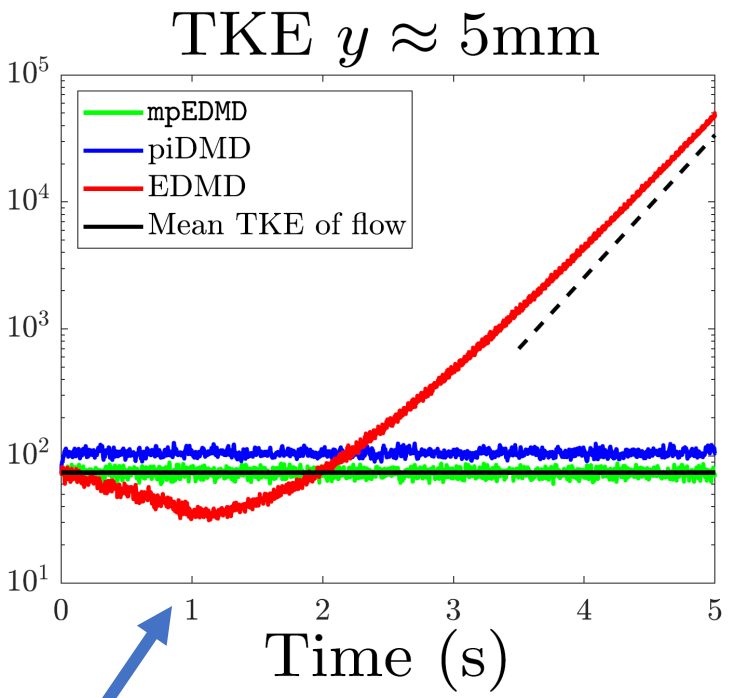
[2] https://www.newyorker.com/magazine/2017/04/03/ai-versus-md

siam. SOCIETY for INDUSTRIAL and APPLIED MATHEMATICS
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA

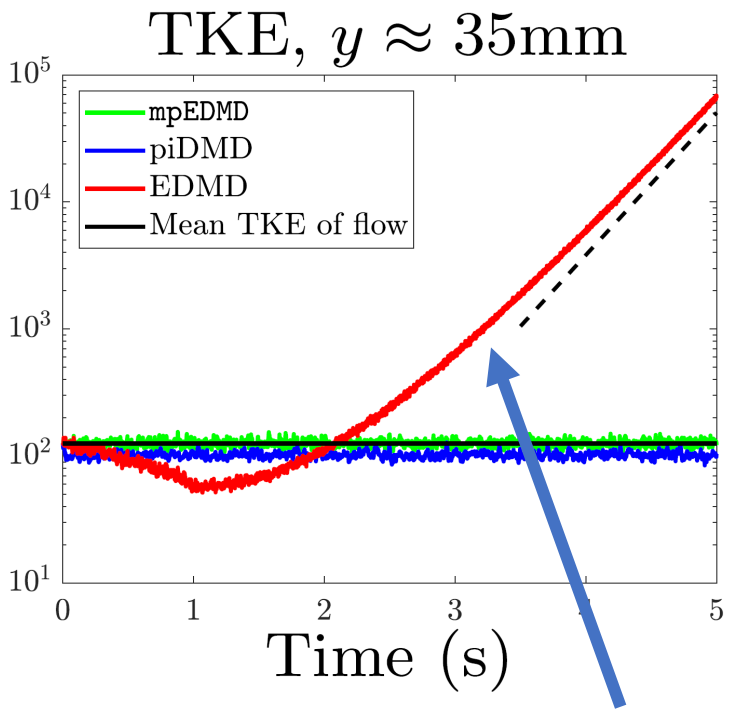
- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **Proc. Natl. Acad. Sci. USA**.

measure-preserving EDMD…

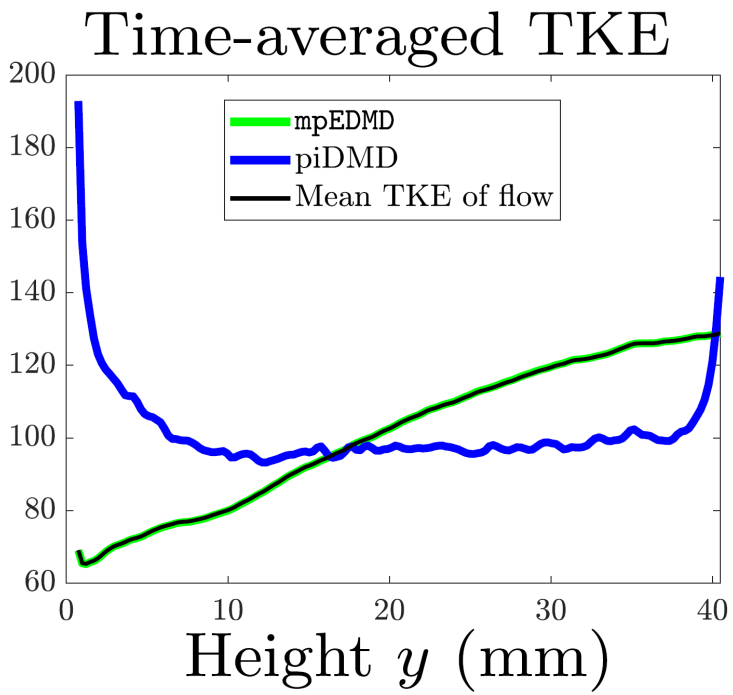
- Polar decomposition of $\mathcal{K}$. Easy to combine with any DMD-type method!
- Converges for spectral measures, spectra, Koopman mode decomposition.
- Measure-preserving discretization for arbitrary measure-preserving systems.



Snapshots collected over 1s



EDMD unstable!



- C., "*The mpEDMD Algorithm for Data-Driven Computations of Measure-Preserving Dynamical Systems*," arXiv 2022.

Solvability Complexity Index Hierarchy

Class $\Omega \ni A$, want to compute $\Xi: \Omega \to (\mathcal{M}, d)$ ← metric space

- Δ_0: Problems solved in finite time (v. rare for cts problems).
- Δ_1: Problems solved in "one limit" with full error control:

$$d(\Gamma_n(A), \Xi(A)) \leq 2^{-n}$$

- Δ_2: Problems solved in "one limit":

$$\lim_{n\to\infty} \Gamma_n(A) = \Xi(A)$$

- Δ_3: Problems solved in "two successive limits":

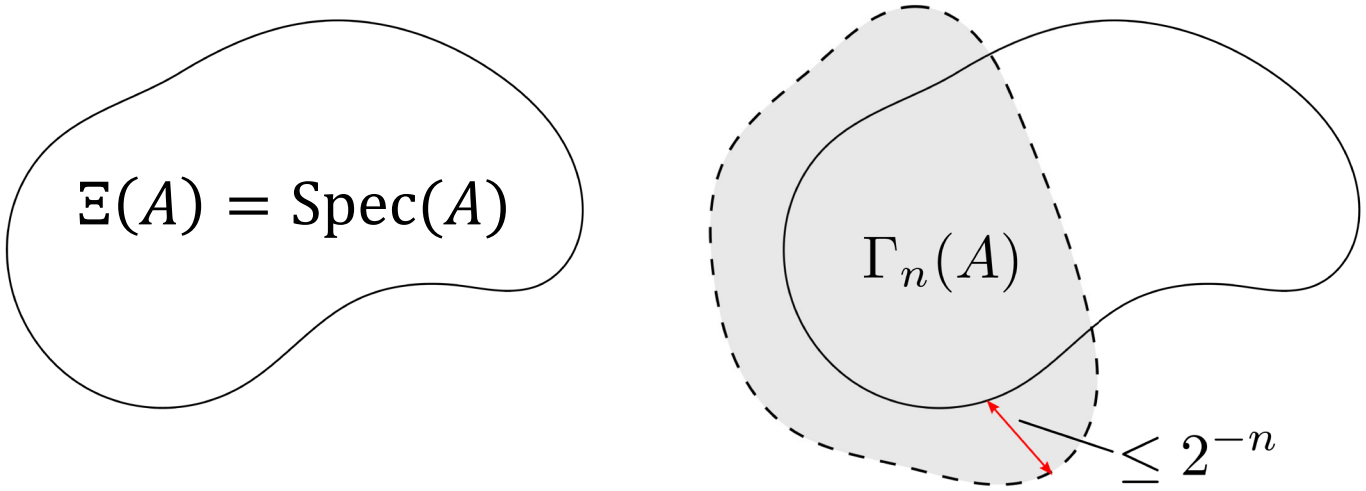
$$\lim_{n\to\infty} \lim_{m\to\infty} \Gamma_{n,m}(A) = \Xi(A)$$

⋮

- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, "*On the solvability complexity index hierarchy and towers of algorithms,*" preprint.
- Hansen, "*On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators,*" **J. Amer. Math. Soc.**, 2011.
- McMullen, "*Families of rational maps and iterative root-finding algorithms,*" **Ann. of Math.**, 1987.
- Doyle, McMullen, "*Solving the quintic by iteration,*" **Acta Math.**, 1989.
- Smale, "*The fundamental theorem of algebra and complexity theory,*" **Bull. Amer. Math. Soc.**, 1981.

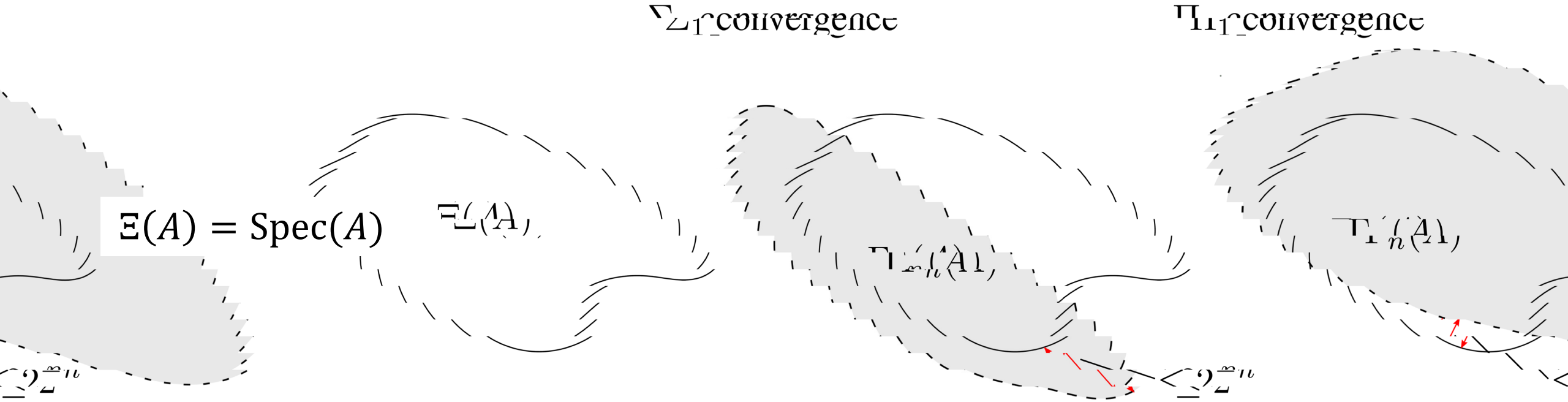
Error control for spectral problems

Σ_1 convergence



- Σ_1: $\exists$ alg. $\{\Gamma_n\}$ s.t. $\lim_{n\to\infty} \Gamma_n(A) = \Xi(A)$, $\max_{z\in\Gamma_n(A)} \mathrm{dist}\big(z, \Xi(A)\big) \leq 2^{-n}$

Error control for spectral problems



- Σ_1: $\exists$ alg. $\{\Gamma_n\}$ s.t. $\lim_{n\to\infty} \Gamma_n(A) = \Xi(A)$, $\max_{z\in\Gamma_n(A)} \mathrm{dist}\big(z, \Xi(A)\big) \le 2^{-n}$
- Π_1: $\exists$ alg. $\{\Gamma_n\}$ s.t. $\lim_{n\to\infty} \Gamma_n(A) = \Xi(A)$, $\max_{z\in\Xi(A)} \mathrm{dist}(z, \Gamma_n(A)) \le 2^{-n}$

Such problems can be used in a proof!

Small sample of classification theorems

Increasing difficulty

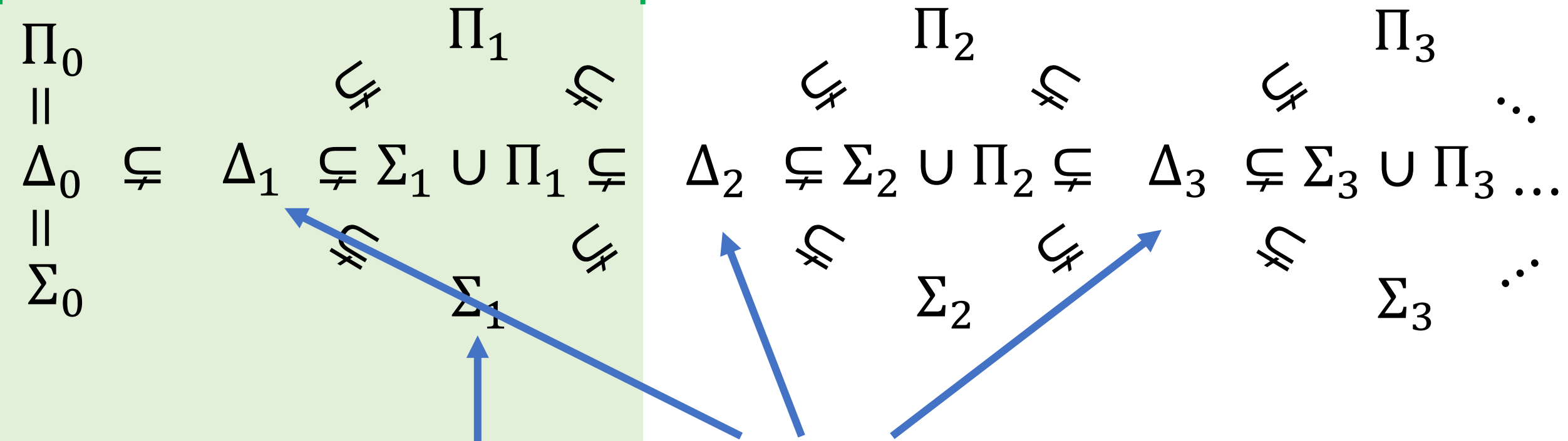
Error control

$$
\begin{array}{ccccccccccc}
\Pi_0 & & & & \Pi_1 & & & & \Pi_2 & & \Pi_3 \\
\| & & & \subsetneq & & \subsetneq & & \subsetneq & & \subsetneq & \ddots \\
\Delta_0 & \subsetneq & \Delta_1 & \subsetneq & \Sigma_1 \cup \Pi_1 & \subsetneq & \Delta_2 & \subsetneq & \Sigma_2 \cup \Pi_2 & \subsetneq & \Delta_3 \quad \subsetneq \Sigma_3 \cup \Pi_3 \ \ldots \\
\| & & & \subsetneq & & \subsetneq & & \subsetneq & & \subsetneq & \vdots \\
\Sigma_0 & & & & \Sigma_1 & & & & \Sigma_2 & & \Sigma_3
\end{array}
$$

Small sample of classification theorems

Increasing difficulty

Error control



Small sample of classification theorems

Increasing difficulty

Error control

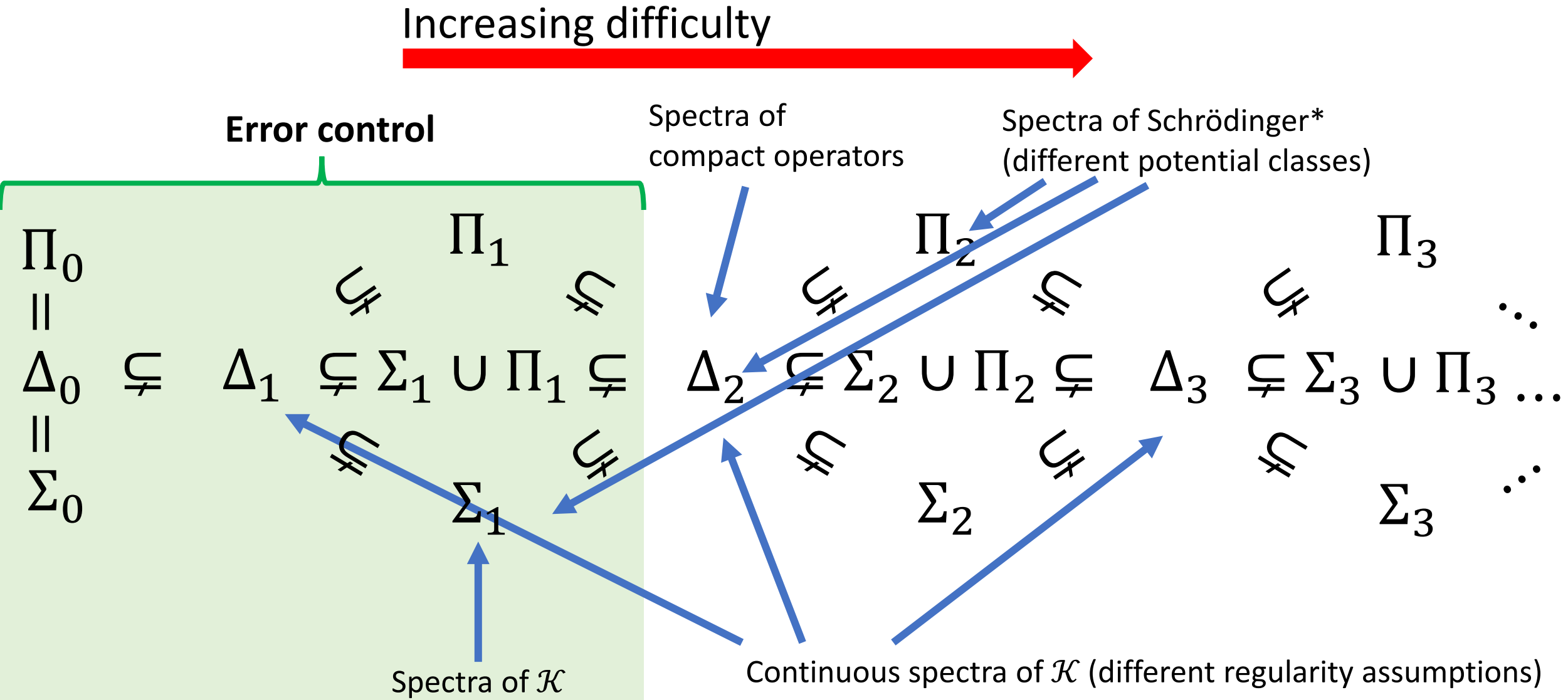
Spectra of
compact operators

$$
\begin{array}{ccccccccccccc}
\Pi_0 & & & & \Pi_1 & & & & \Pi_2 & & & & \Pi_3 \\
\| & & & \cup\!\!\!\!\neq & & \subsetneq & & \cup\!\!\!\!\neq & & \subsetneq & & \cup\!\!\!\!\neq & & \ddots \\
\Delta_0 & \subsetneq & \Delta_1 & \subsetneq & \Sigma_1 \cup \Pi_1 & \subsetneq & \Delta_2 & \subsetneq & \Sigma_2 \cup \Pi_2 & \subsetneq & \Delta_3 & \subsetneq & \Sigma_3 \cup \Pi_3 & \cdots \\
\| & & & \subsetneq\!\!\!\!\cup & & \cup\!\!\!\!\neq & & \subsetneq\!\!\!\!\cup & & \cup\!\!\!\!\neq & & \subsetneq\!\!\!\!\cup & & \iddots \\
\Sigma_0 & & & & \Sigma_1 & & & & \Sigma_2 & & & & \Sigma_3 &
\end{array}
$$

Spectra of $\mathcal{K}$

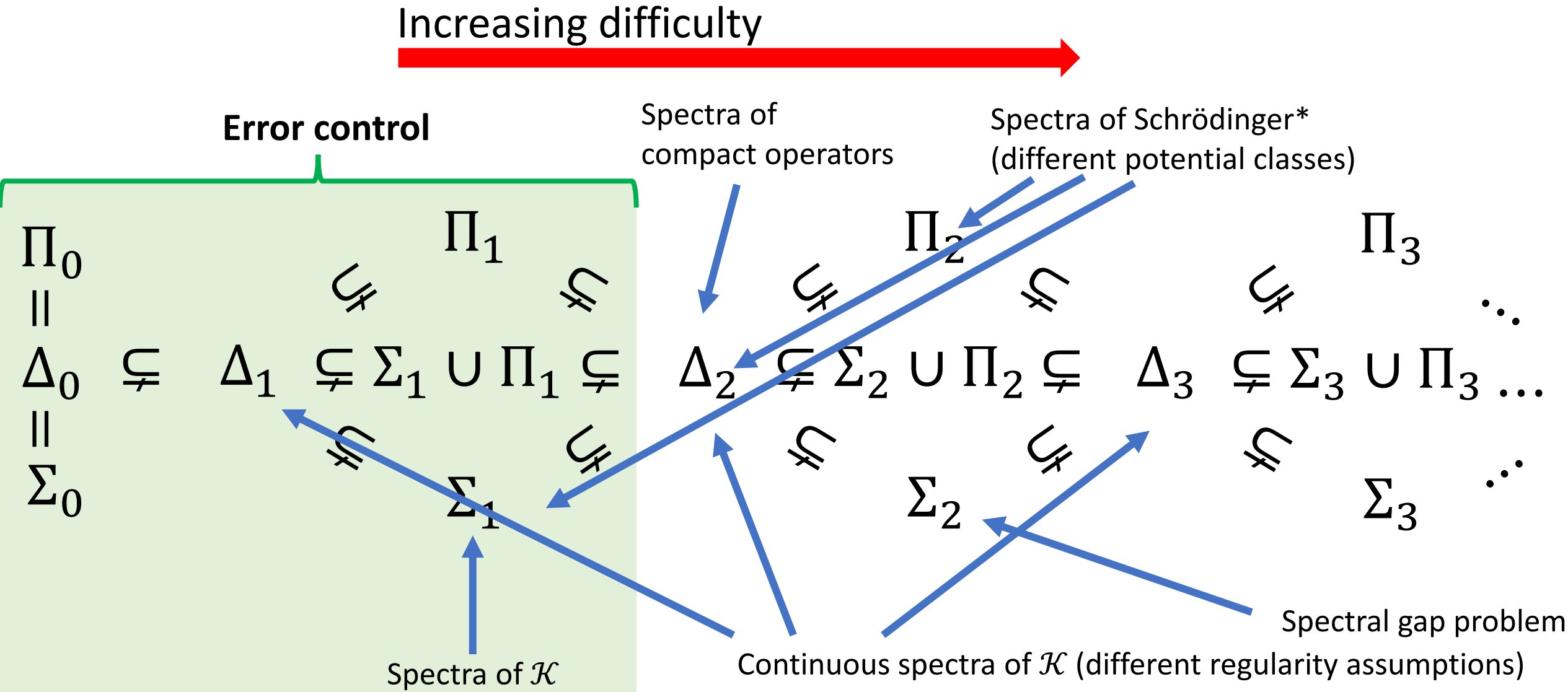
Continuous spectra of $\mathcal{K}$ (different regularity assumptions)

Small sample of classification theorems



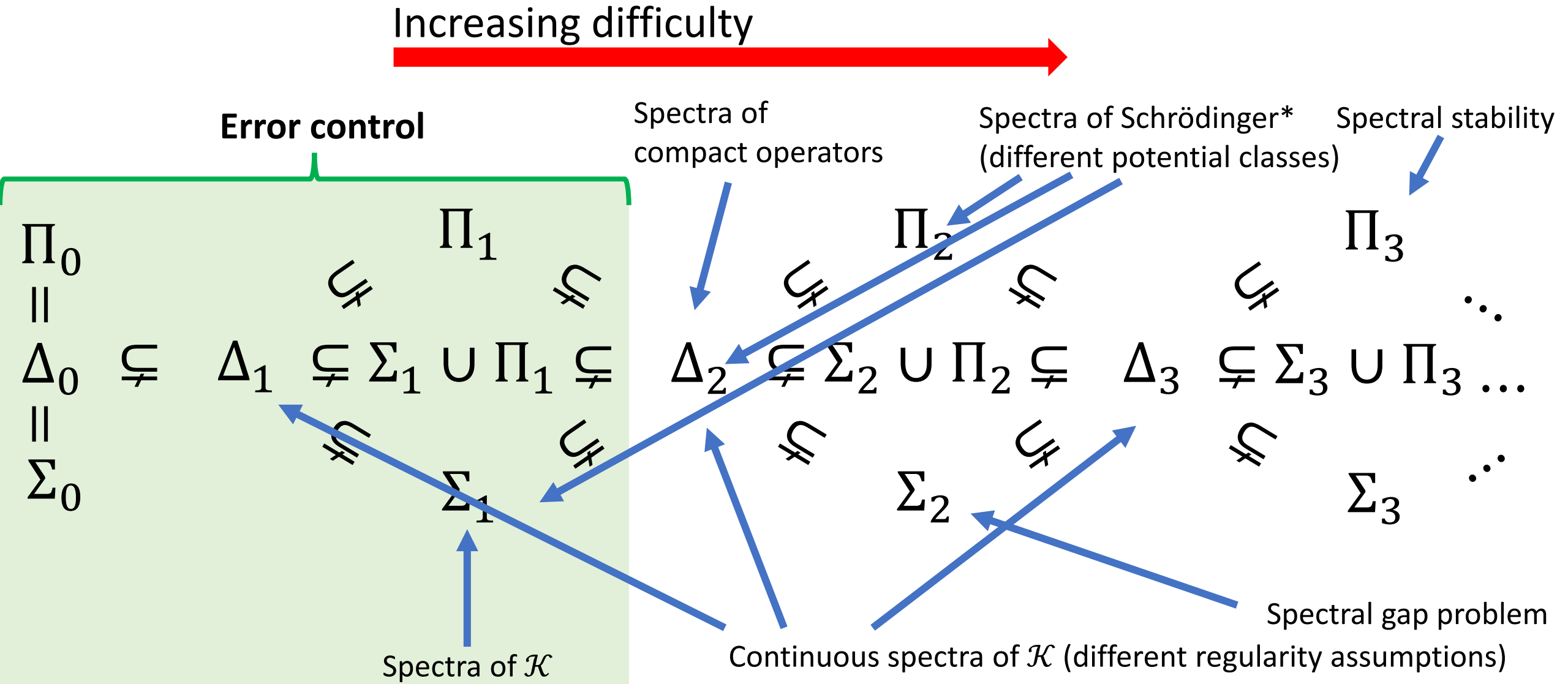
**Open problem of Schwinger*: "The special canonical group," "Unitary operator bases," PNAS, 1960.

Small sample of classification theorems



**Open problem of Schwinger*: "The special canonical group," "Unitary operator bases," PNAS, 1960.

Small sample of classification theorems



**Open problem of Schwinger*: "The special canonical group," "Unitary operator bases," PNAS, 1960.