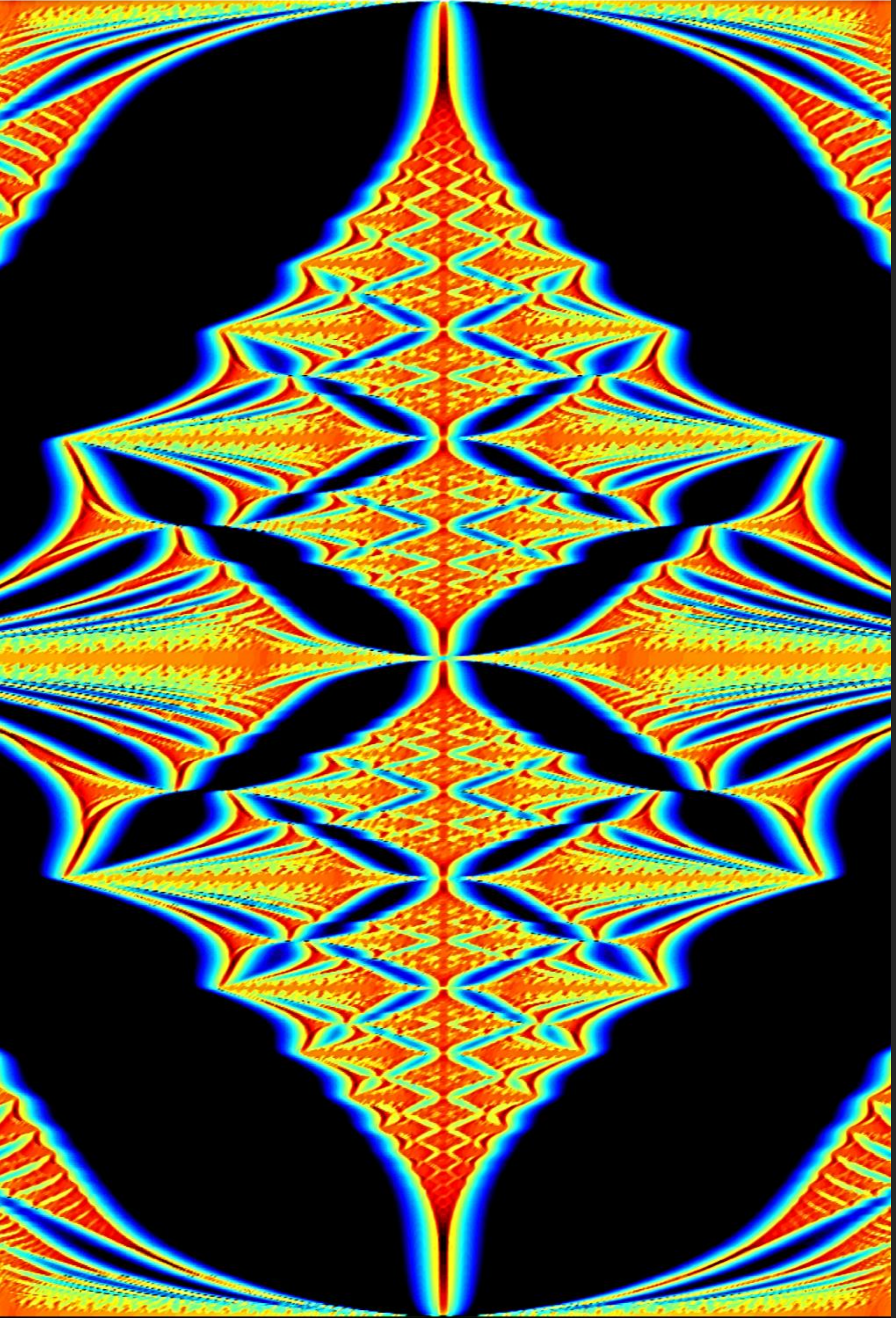


To infinity… and beyond!

The solvability complexity index and the foundations of infinite-dimensional spectral computations

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

The infinite-dimensional spectral problem

$$A \text{"="} \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \qquad A\left(\sum_{k=1}^{\infty} x_k e_k\right) = \sum_{j=1}^{\infty}\left(\sum_{k=1}^{\infty} a_{jk} x_k\right) e_j$$

Canonical basis vectors of $l^2(\mathbb{N})$

Also deal with PDEs, integral operators etc.

Finite-dimensional	$\Longrightarrow$	Infinite-dimensional
Eigenvalues of $B \in \mathbb{C}^{n\times n}$	$\Longrightarrow$	Spectrum, $\mathrm{Spec}(A)$
$\{\lambda_j \in \mathbb{C}: \det(B - \lambda_j I) = 0\}$	$\Longrightarrow$	$\{\lambda \in \mathbb{C}: A - \lambda I \text{ is not invertible}\}$

*"Most operators that arise in practice are not presented in a representation in which they are diagonalized, and it is often very hard to locate even a single point in the spectrum. Thus, one often has to settle for numerical approximations [...] Unfortunately, there is a dearth of literature on this basic problem and, so far as we have been able to tell, **there are no proven [general] techniques**."* W. Arveson, Berkeley (1994)

Why spectra?

Applications: Quantum mechanics, structural mechanics, optics, acoustics, statistical physics, number theory, matter physics, PDEs, data analysis, neural networks and AI, nuclear scattering, optics, computational chemistry, …

Rich history of **computational spectral theory:**

D. Arnold (Minnesota), W. Arveson (Berkeley), A. Böttcher (Chemnitz), W. Dahmen (South Carolina), E. B. Davies (KCL), P. Deift (NYU), L. Demanet (MIT), M. Embree (Virginia Tech), C. Fefferman (Princeton), G. Golub (Stanford), A. Iserles (Cambridge), I. Ipsen (NCS), S. Jitomirskaya (UCI), A. Laptev (Imperial), M. Luskin (Minnesota), S. Mayboroda (Minnesota), W. Schlag (Yale), E. Schrödinger (DIAS), J. Schwinger (Harvard), N. Trefethen (Oxford), V. Varadarajan (UCLA), S. Varadhan (NYU), J. von Neumann (IAS), M. Zworski (Berkeley),…

A motivating problem

In a series of papers in the 1950's and 1960's, J. Schwinger examined the foundations of quantum mechanics. A key problem he considered:

Given a self-adjoint Schrödinger operator $-\Delta + V$ on $\mathbb{R}$, can we approximate its spectrum?

Partial answer: T. Digernes, V. S. Varadarajan and S. R. S. Varadhan (1994) gave a convergent algorithm for a class of V generating compact resolvent.

For which classes of differential operators on unbounded domains do there exist algorithms that converge to the spectrum? Can we guarantee that the output is in the spectrum up to an arbitrarily small tolerance?

- Digernes, Varadarajan, Varadhan, *"Finite approximations to quantum systems,"* **Rev. Math. Phys.**, 1994.

Warm-up: bounded diagonal operators

$$A = \begin{pmatrix} a_1 & & \\ & a_2 & \\ & & \ddots \end{pmatrix}$$

Assumption: Algorithm can query entries of A.

Algorithm: $\Gamma_n(A) = \{a_1, a_2, \ldots, a_n\} \to \mathrm{Spec}(A) = \overline{\{a_1, a_2, \ldots\}}$ in Haus. Metric.

One-sided error control: $\Gamma_n(A) \subset \mathrm{Spec}(A)$

Optimal: Can't obtain $\hat{\Gamma}_n(A) \to \mathrm{Spec}(A)$ with $\mathrm{Spec}(A) \subset \hat{\Gamma}_n(A)$.

Example: compact operators (still easy?)

classic method

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Algorithm: $\Gamma_n(A) = \mathrm{Spec}(P_n A P_n)$ converges to $\mathrm{Spec}(A)$ in Haus. Metric.

Question: Can we verify the output?

i.e., Does there exist $\hat{\Gamma}_n(A) \to \mathrm{Spec}(A)$ with $\hat{\Gamma}_n(A) \subset \mathrm{Spec}(A) + B_{2^{-n}}$?

Example: compact operators (still easy?)

classic method

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Algorithm: $\Gamma_n(A) = \mathrm{Spec}(P_n A P_n)$ converges to $\mathrm{Spec}(A)$ in Haus. Metric.

Question: Can we verify the output?

i.e., Does there exist $\hat{\Gamma}_n(A) \to \mathrm{Spec}(A)$ with $\hat{\Gamma}_n(A) \subset \mathrm{Spec}(A) + B_{2^{-n}}$?

Answer: No!

No alg. can do this on whole class, even for self-adjoint compact operators.

What about Jacobi operators?

$$A = \begin{pmatrix} a_1 & b_1 & & \\ b_1 & a_2 & b_2 & \\ & b_2 & a_3 & \ddots \\ & & \ddots & \ddots \end{pmatrix}, \qquad b_k > 0, \qquad a_k \in \mathbb{R}$$

Non-trivial, e.g., spurious eigenvalues.

What about Jacobi operators?

$$A = \begin{pmatrix} a_1 & b_1 & & \\ b_1 & a_2 & b_2 & \\ & b_2 & a_3 & \ddots \\ & & \ddots & \ddots \end{pmatrix}, \qquad b_k > 0, \qquad a_k \in \mathbb{R}$$

Non-trivial, e.g., spurious eigenvalues.

Enlarge class to **sparse normal operators** - surely now much harder?!

What about Jacobi operators?

$$A = \begin{pmatrix} a_1 & b_1 & & \\ b_1 & a_2 & b_2 & \\ & b_2 & a_3 & \ddots \\ & & \ddots & \ddots \end{pmatrix}, \qquad b_k > 0, \qquad a_k \in \mathbb{R}$$

Non-trivial, e.g., spurious eigenvalues.

Enlarge class to **sparse normal operators** - surely now much harder?!

Answer: $\exists\{\Gamma_n\}$ s.t. $\lim_{n\to\infty} \Gamma_n(A) = \mathrm{Spec}(A)$ and $\Gamma_n(A) \subset \mathrm{Spec}(A) + B_{2^{-n}}$,

for any sparse normal operator A

- C., Roman, Hansen, *"How to compute spectra with error control,"* **Phys. Rev. Lett.**, 2019.
- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.

A curious case of limits

General bounded: $$A = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Algorithm: $\exists\{\Gamma_{n_3,n_2,n_1}\}$ s.t. $\lim_{n_3\to\infty}\lim_{n_2\to\infty}\lim_{n_1\to\infty}\Gamma_{n_3,n_2,n_1}(A) = \mathrm{Spec}(A)$

Question: Can we do better?

- Hansen, *"On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators,"* **J. Amer. Math. Soc.**, 2011.
- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.

A curious case of limits

General bounded: $A = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$

Explains Arveson's lament!

Algorithm: $\exists\{\Gamma_{n_3,n_2,n_1}\}$ s.t. $\lim_{n_3\to\infty}\lim_{n_2\to\infty}\lim_{n_1\to\infty} \Gamma_{n_3,n_2,n_1}(A) = \mathrm{Spec}(A)$

Question: Can we do better?

Answer: No! Canonically embed problems such as:

Given $B \in \{0,1\}^{\mathbb{N}\times\mathbb{N}}$, does B have a column with infinitely many 1's?

$\Longrightarrow$ lower bound on number of "successive limits" needed (ind. of comp. model).

- Hansen, *"On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators,"* **J. Amer. Math. Soc.**, 2011.
- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.
- C., *"On the computation of geometric features of spectra of linear operators on Hilbert spaces,"* **Found. Comput. Math.**, to appear.

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.

Solvability Complexity Index Hierarchy

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

- Δ_0: Problems solved in finite time (v. rare for cts pr…
- Δ_1: Problems solved in "one limit" with f…

$$d(\Gamma_n(A), \Xi \ldots$$

- Δ_2: Problems solved in "…

$$\ldots \Xi(A)$$

- Δ… …o successive limits":

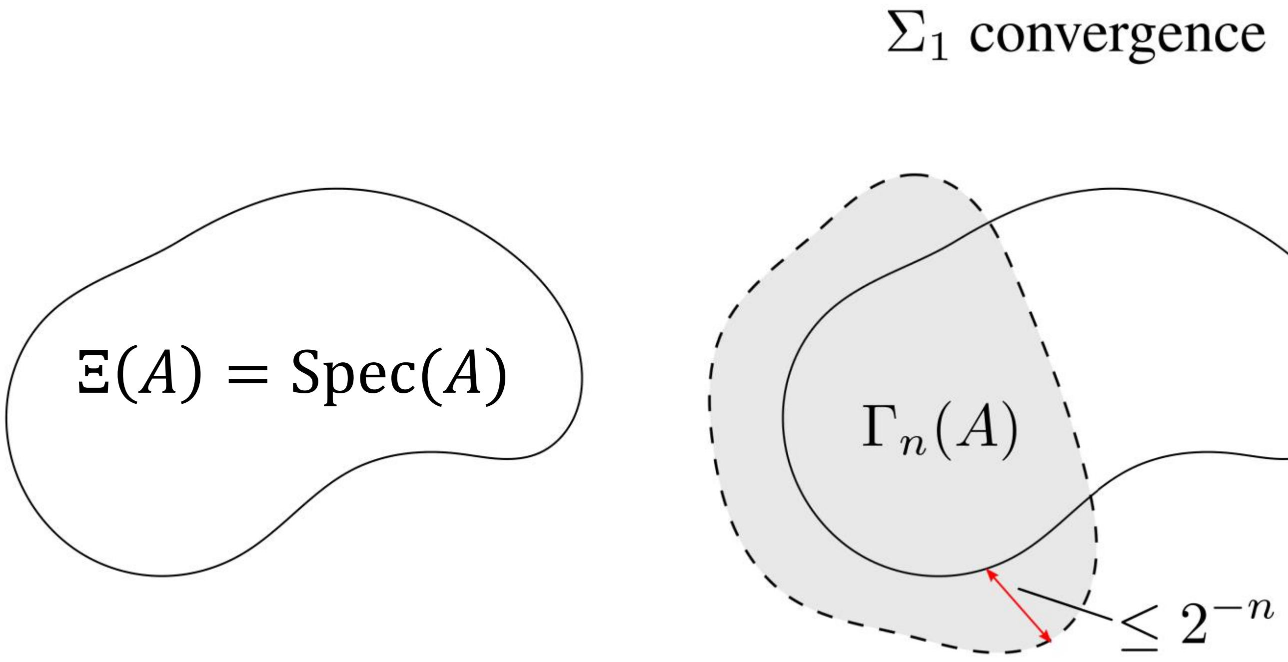
$$\lim_{n\to\infty} \lim_{m\to\infty} \Gamma_{n,m}(A) = \Xi(A)$$

⋮

I will try and convince you that multiple limits is natural and generic!

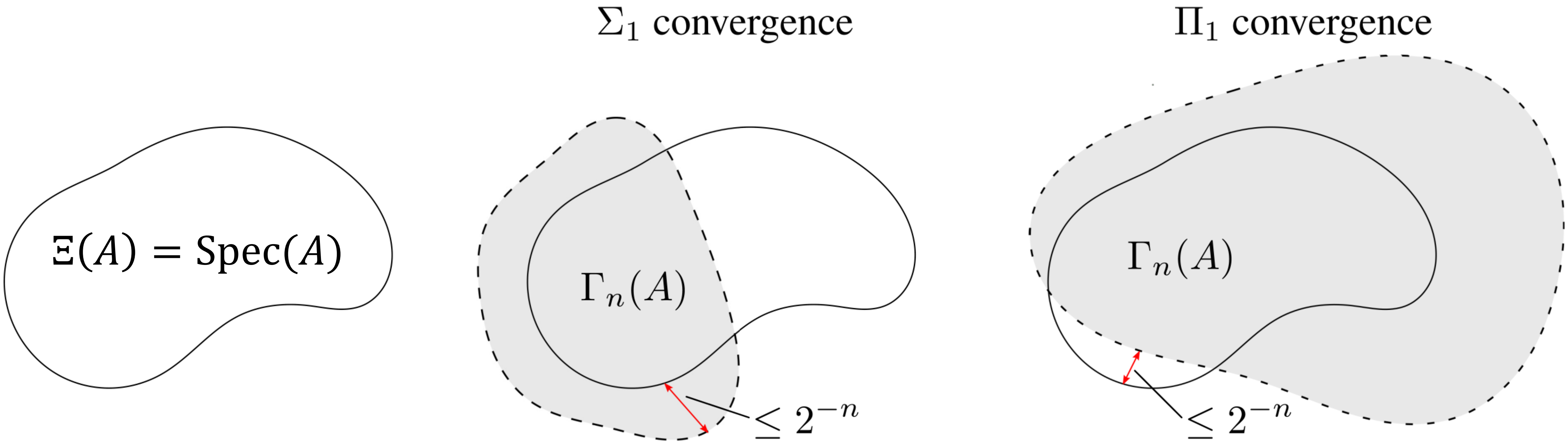
- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.
- Hansen, *"On the solvability complexity index, the n-pseudospectrum and approximations of spectra of operators,"* **J. Amer. Math. Soc.**, 2011.
- McMullen, *"Families 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: $\exists$ 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. $\lim_{n\to\infty} \Gamma_n(A) = \Xi(A)$, $\max_{z\in\Gamma_n(A)} \mathrm{dist}\big(z, \Xi(A)\big) \leq 2^{-n}$
- Π_1: $\exists$ alg. $\{\Gamma_n\}$ s.t. $\lim_{n\to\infty} \Gamma_n(A) = \Xi(A)$, $\max_{z\in\Xi(A)} \mathrm{dist}(z, \Gamma_n(A)) \leq 2^{-n}$

Such problems can be used in a proof!

Sample: some results for bounded op. on $l^2(\mathbb{N})$

Increasing difficulty →

Error control

$$\Pi_0 = \Delta_0 = \Sigma_0$$

$$\Delta_0 \subsetneq \Delta_1 \subsetneq \Sigma_1 \cup \Pi_1 \subsetneq$$

$$\Delta_1 \subsetneq \Pi_1 \subsetneq \Sigma_1 \cup \Pi_1, \qquad \Delta_1 \subsetneq \Sigma_1 \subsetneq \Sigma_1 \cup \Pi_1$$

Sample: some results for bounded op. on $l^2(\mathbb{N})$

Increasing difficulty

Error control

$$\Pi_0 = \Delta_0 = \Sigma_0$$

$$\Delta_1 \subsetneq \Pi_1 \subsetneq \Sigma_1 \cup \Pi_1, \qquad \Delta_1 \subsetneq \Sigma_1 \subsetneq \Sigma_1 \cup \Pi_1$$

$$\Delta_0 \subsetneq \Delta_1 \subsetneq \Sigma_1 \cup \Pi_1 \subsetneq \Delta_2$$

One limit, no error control.

One limit: SCI≤1

Sample: some results for bounded op. on $l^2(\mathbb{N})$

Increasing difficulty →

Error control

$$\Pi_0 = \Delta_0 = \Sigma_0$$

$$\Delta_0 \subsetneq \Delta_1 \subsetneq \Sigma_1 \cup \Pi_1 \subsetneq \Delta_2 \subsetneq \Sigma_2 \cup \Pi_2 \subsetneq \Delta_3$$

$$\Delta_1 \subsetneq \Pi_1 \subsetneq \Delta_2, \quad \Delta_1 \subsetneq \Sigma_1 \subsetneq \Delta_2$$

$$\Delta_2 \subsetneq \Pi_2 \subsetneq \Delta_3, \quad \Delta_2 \subsetneq \Sigma_2 \subsetneq \Delta_3$$

Two limits: SCI ≤ 2

Sample: some results for bounded op. on $l^2(\mathbb{N})$

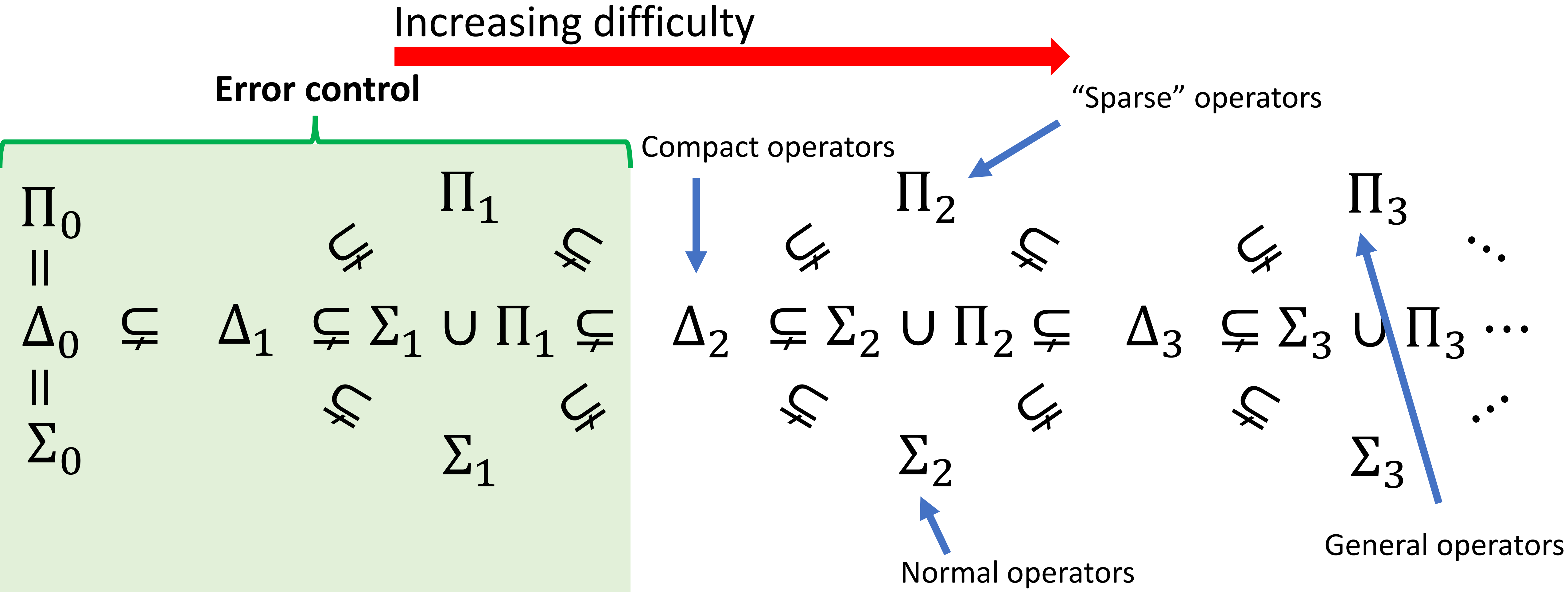
Increasing difficulty →

Error control

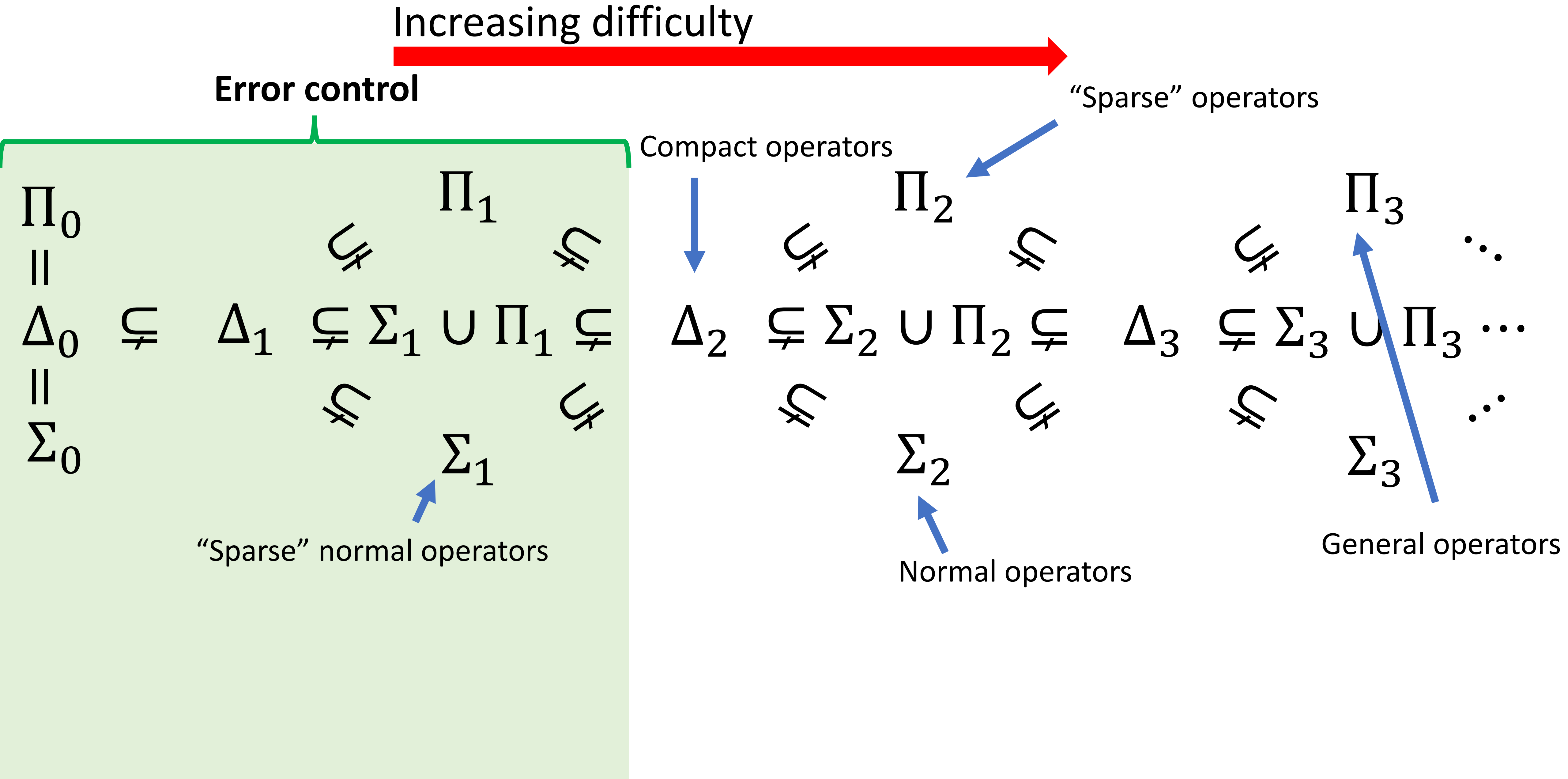
$$\begin{array}{ccccccccccccc}
\Pi_0 & & & & \Pi_1 & & & & \Pi_2 & & & & \Pi_3 \quad \cdots \\
\| & & & \subsetneq & & \subsetneq & & \subsetneq & & \subsetneq & & \subsetneq & \\
\Delta_0 & \subsetneq & \Delta_1 & \subsetneq & \Sigma_1 \cup \Pi_1 & \subsetneq & \Delta_2 & \subsetneq & \Sigma_2 \cup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

Three limits: SCI ≤ 3 ...

Sample: some results for bounded op. on $l^2(\mathbb{N})$



Sample: some results for bounded op. on $l^2(\mathbb{N})$



Sample: some results for bounded op. on $l^2(\mathbb{N})$

Self-adjoint quasiperiodic operators

Increasing difficulty

Error control

"Sparse" operators

Compact operators

$$\Pi_0 = \Delta_0 = \Sigma_0 \subsetneq \Delta_1 \subsetneq \Sigma_1 \cup \Pi_1 \subsetneq \Delta_2 \subsetneq \Sigma_2 \cup \Pi_2 \subsetneq \Delta_3 \subsetneq \Sigma_3 \cup \Pi_3 \cdots$$

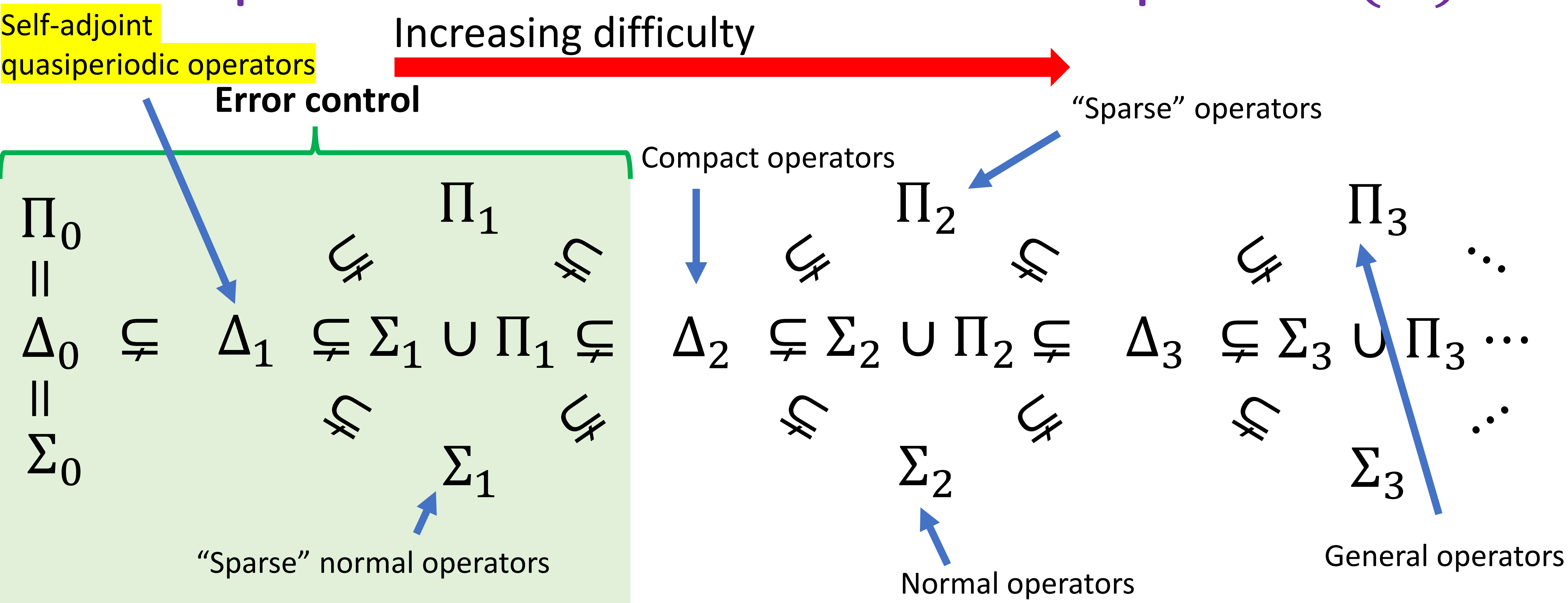
$$\Delta_1 \subsetneq \Pi_1 \subsetneq \Delta_2, \quad \Delta_1 \subsetneq \Sigma_1 \subsetneq \Delta_2, \quad \Delta_2 \subsetneq \Pi_2 \subsetneq \Delta_3, \quad \Delta_2 \subsetneq \Sigma_2 \subsetneq \Delta_3, \quad \Delta_3 \subsetneq \Pi_3 \cdots, \quad \Delta_3 \subsetneq \Sigma_3 \cdots$$

"Sparse" normal operators

Normal operators

General operators

Sample: some results for bounded op. on $l^2(\mathbb{N})$



Zoo of problems: spectral type (pure point, absolutely continuous, singularly continuous), Lebesgue measure and fractal dimensions of spectra, discrete spectra, essential spectra, eigenspaces + multiplicity, spectral radii, essential numerical ranges, geometric features of spectrum (e.g., capacity), spectral gap problem, resonances ...

- C., "*The foundations of infinite-dimensional spectral computations,*" PhD diss., University of Cambridge, 2020.

Why study these foundations?

- $\mathrm{SCI} > 1$ classifications $\Longrightarrow$ tells us assumptions needed to lower SCI.
- Σ_1 and Π_1 classifications $\Longrightarrow$ look-up table for computer-assisted proofs.
- Negative results prevent us from trying to prove too much.
- Much of computational literature does not prove sharp results.

Remarks:

- Can use with any model of computation.
- Existing hierarchies included as particular cases.

Example 1: Σ_1 algorithm for spectra

The three-limit algorithm

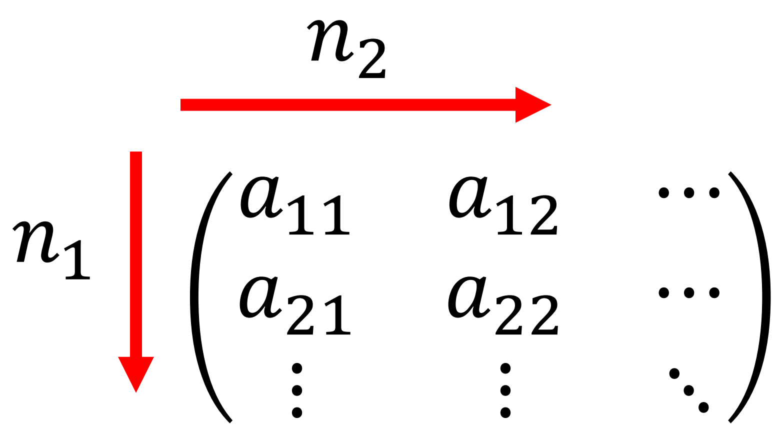
$$\sigma_{\inf}(T) = \inf\{\|Tv\| : v \in \mathfrak{D}(T), \|v\| = 1\}$$

• Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.

The three-limit algorithm

$$\sigma_{\inf}(T) = \inf\{\|Tv\| : v \in \mathfrak{D}(T), \|v\| = 1\}$$

$$\gamma_{n_1,n_2}(A,z) = \min\left\{\sigma_{\inf}\left(P_{n_1}[A-z]P_{n_2}\right), \sigma_{\inf}\left(P_{n_1}[A^* - \bar{z}]P_{n_2}\right)\right\}$$



- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.

The three-limit algorithm

$$\sigma_{\inf}(T) = \inf\{\|Tv\| : v \in \mathfrak{D}(T), \|v\| = 1\}$$

$$\gamma_{n_1,n_2}(A,z) = \min\{\sigma_{\inf}(P_{n_1}[A - z]P_{n_2}), \sigma_{\inf}(P_{n_1}[A^* - \bar{z}]P_{n_2})\}$$

$$\gamma_{n_1,n_2}(A,z) \uparrow \gamma_{n_2}(A,z) \coloneqq \min\{\sigma_{\inf}([A - z]P_{n_2}), \sigma_{\inf}([A^* - \bar{z}]P_{n_2})\}, \text{ as } n_1 \to \infty$$

$$\xrightarrow{\quad n_2 \quad}$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.

The three-limit algorithm

$$\sigma_{\inf}(T) = \inf\{\|Tv\| : v \in \mathfrak{D}(T), \|v\| = 1\}$$

$$\gamma_{n_1,n_2}(A,z) = \min\{\sigma_{\inf}(P_{n_1}[A-z]P_{n_2}), \sigma_{\inf}(P_{n_1}[A^*-\bar{z}]P_{n_2})\}$$

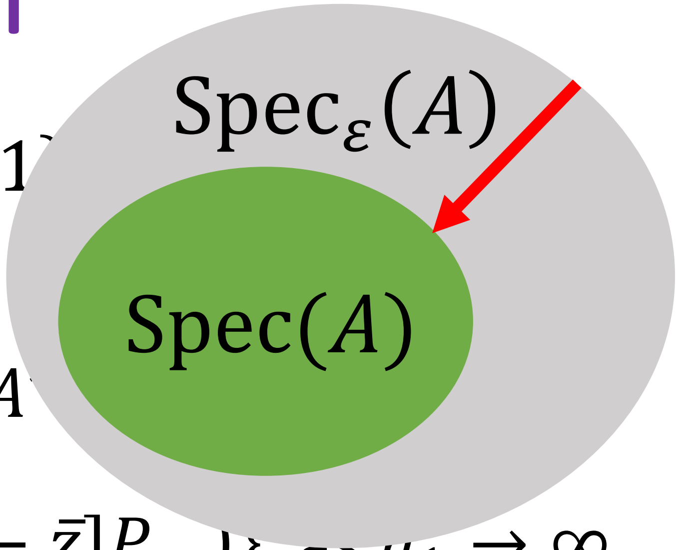
$$\gamma_{n_1,n_2}(A,z) \uparrow \gamma_{n_2}(A,z) := \min\{\sigma_{\inf}([A-z]P_{n_2}), \sigma_{\inf}([A^*-\bar{z}]P_{n_2})\}, \text{ as } n_1 \to \infty$$

$$\gamma_{n_2}(A,z) \downarrow \gamma(A,z) := \min\{\sigma_{\inf}(A-z), \sigma_{\inf}(A^*-\bar{z})\} = \|(A-z)^{-1}\|^{-1}, \text{ as } n_2 \to \infty$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.

The three-limit algorithm



$$\sigma_{\inf}(T) = \inf\{\|Tv\|: v \in \mathfrak{D}(T), \|v\| = 1$$

$$\gamma_{n_1,n_2}(A,z) = \min\{\sigma_{\inf}(P_{n_1}[A-z]P_{n_2}), \sigma_{\inf}(P_{n_1}[A$$

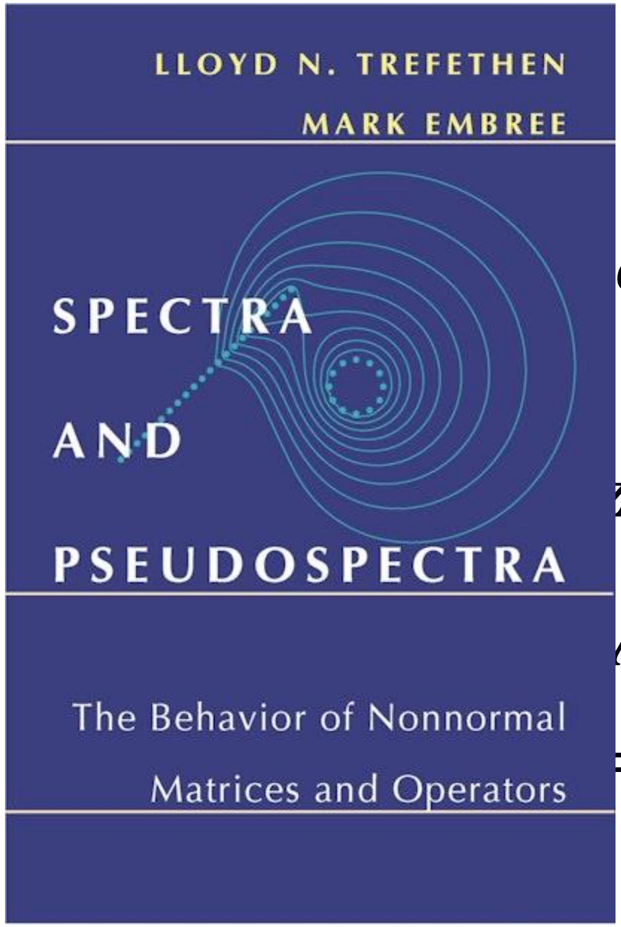
$$\gamma_{n_1,n_2}(A,z) \uparrow \gamma_{n_2}(A,z) := \min\{\sigma_{\inf}([A-z]P_{n_2}), \sigma_{\inf}([A^*-\bar{z}]P_{n_2})\}, \text{ as } n_1 \to \infty$$

$$\gamma_{n_2}(A,z) \downarrow \gamma(A,z) := \min\{\sigma_{\inf}(A-z), \sigma_{\inf}(A^*-\bar{z})\} = \|(A-z)^{-1}\|^{-1}, \text{ as } n_2 \to \infty$$

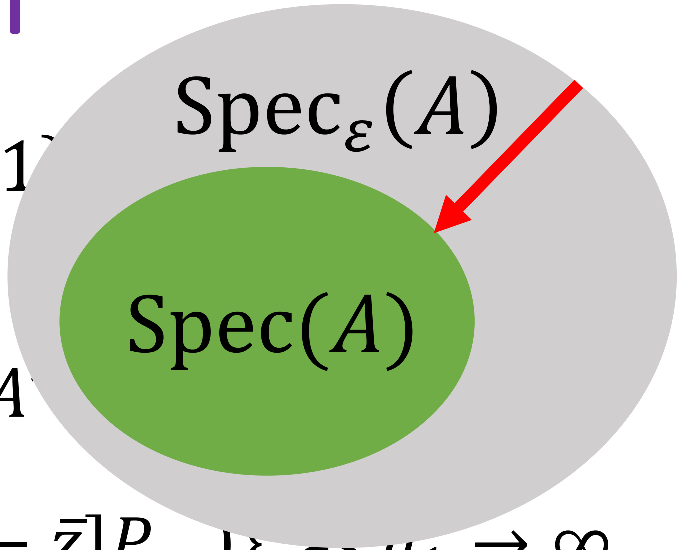
Approx. pseudospectrum: $\lim_{n_2\to\infty} \lim_{n_1\to\infty} \hat{\Gamma}_{n_1,n_2}(A,\varepsilon) = \mathrm{Spec}_\varepsilon(A) = \{z: \gamma(A,z) \le \varepsilon\}$

$$\Gamma_{n_1,n_2,n_3}(A) = \hat{\Gamma}_{n_1,n_2}(A, 1/n_3)$$

- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.



The three-limit algorithm



$\sigma_{\inf}(T) = \inf\{\|Tv\|: v \in \mathfrak{D}(T), \|v\| = 1$

$z) = \min\{\sigma_{\inf}\big(P_{n_1}[A - z]P_{n_2}\big), \sigma_{\inf}\big(P_{n_1}[A$

$A, z) \coloneqq \min\{\sigma_{\inf}\big([A - z]P_{n_2}\big), \sigma_{\inf}\big([A^* - \bar{z}]P_{n_2}\big)\}$, as $n_1 \to \infty$

$= \min\{\sigma_{\inf}(A - z), \sigma_{\inf}(A^* - \bar{z})\} = \|(A - z)^{-1}\|^{-1}$, as $n_2 \to \infty$

Approx. pseudospectrum: $\lim_{n_2 \to \infty} \lim_{n_1 \to \infty} \hat{\Gamma}_{n_1,n_2}(A, \varepsilon) = \mathrm{Spec}_\varepsilon(A) = \{z: \gamma(A, z) \le \varepsilon\}$

$$\Gamma_{n_1,n_2,n_3}(A) = \hat{\Gamma}_{n_1,n_2}(A, 1/n_3)$$

- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.

The three-limit algorithm

$$\sigma_{\inf}(T) = \inf\{\|Tv\|: v \in \mathfrak{D}(T), \|v\| = 1\}$$

$$\gamma_{n_1,n_2}(A,z) = \min\left\{\sigma_{\inf}\left(P_{n_1}[A-z]P_{n_2}\right), \sigma_{\inf}\left(P_{n_1}[A^*-\bar{z}]P_{n_2}\right)\right\}$$

$$\gamma_{n_1,n_2}(A,z) \uparrow \gamma_{n_2}(A,z) := \min\left\{\sigma_{\inf}\left([A-z]P_{n_2}\right), \sigma_{\inf}\left([A^*-\bar{z}]P_{n_2}\right)\right\}, \text{ as } n_1 \to \infty$$

$$\gamma_{n_2}(A,z) \downarrow \gamma(A,z) := \min\{\sigma_{\inf}(A-z), \sigma_{\inf}(A^*-\bar{z})\} = \|(A-z)^{-1}\|^{-1}, \text{ as } n_2 \to \infty$$

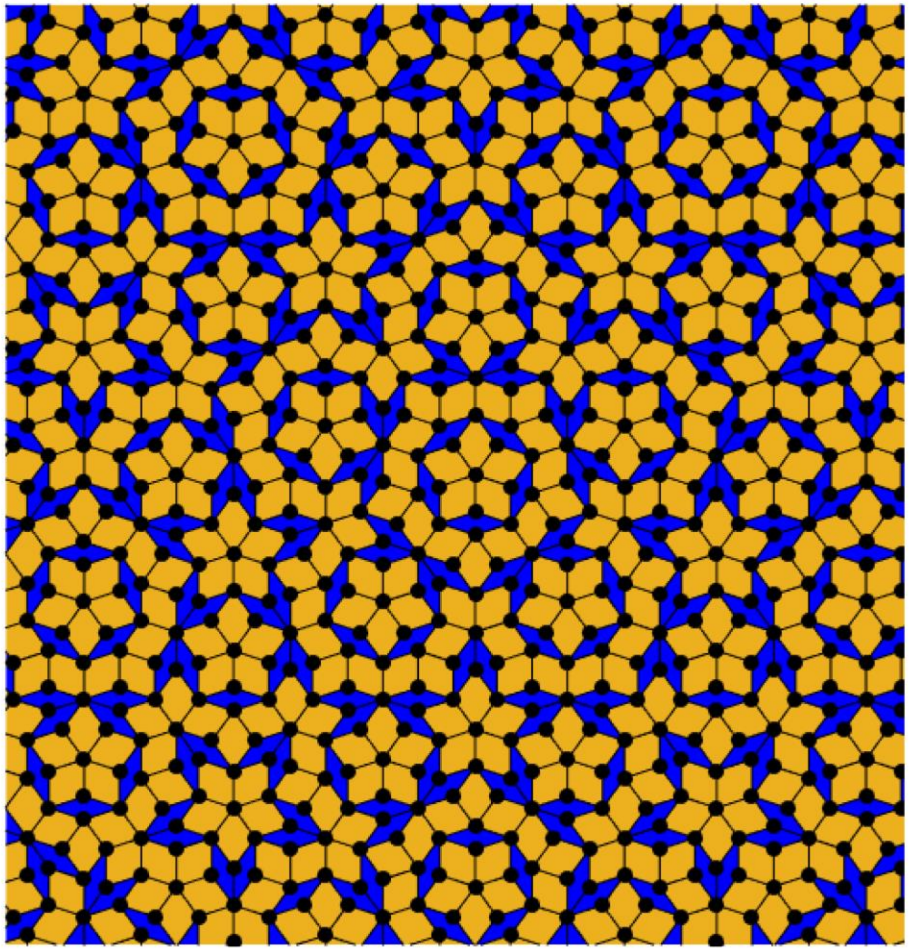
Approx. pseudospectrum: $\lim_{n_2\to\infty} \lim_{n_1\to\infty} \hat{\Gamma}_{n_1,n_2}(A,\varepsilon) = \mathrm{Spec}_{\varepsilon}(A) = \{z: \gamma(A,z) \le \varepsilon\}$

$$\Gamma_{n_1,n_2,n_3}(A) = \hat{\Gamma}_{n_1,n_2}(A, 1/n_3)$$

What assumptions are needed to reduce the number of limits?

- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, *"On the solvability complexity index hierarchy and towers of algorithms,"* preprint.

Example: quasicrystals



Aperiodicity $\Longrightarrow$ interesting physics but very hard to compute spectra!

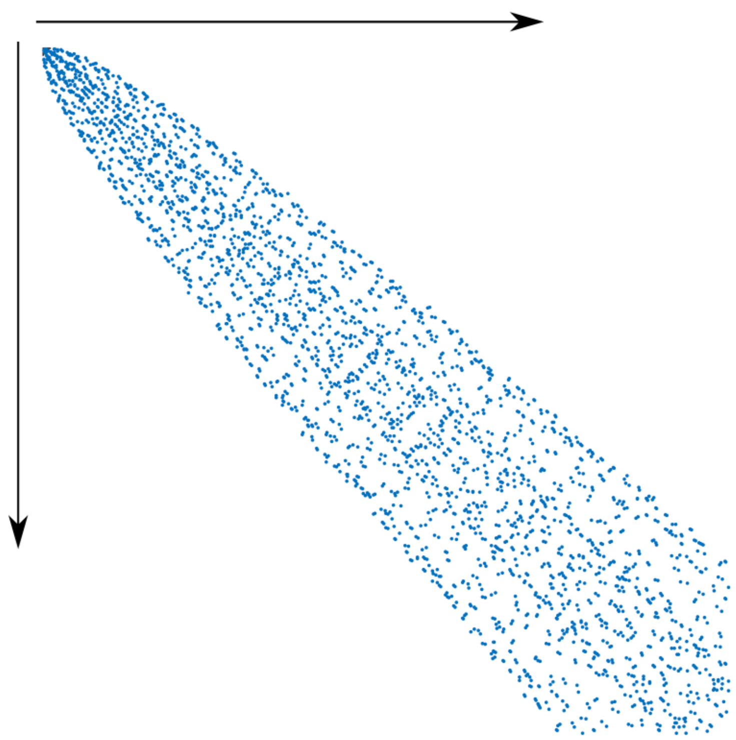
Example: quasicrystals

Model: Perpendicular magnetic field (of strength B).

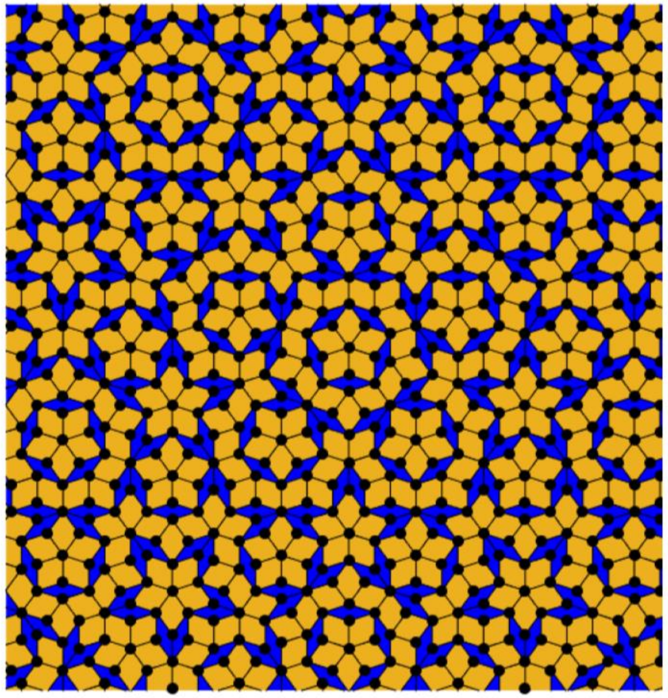
Matrix equation

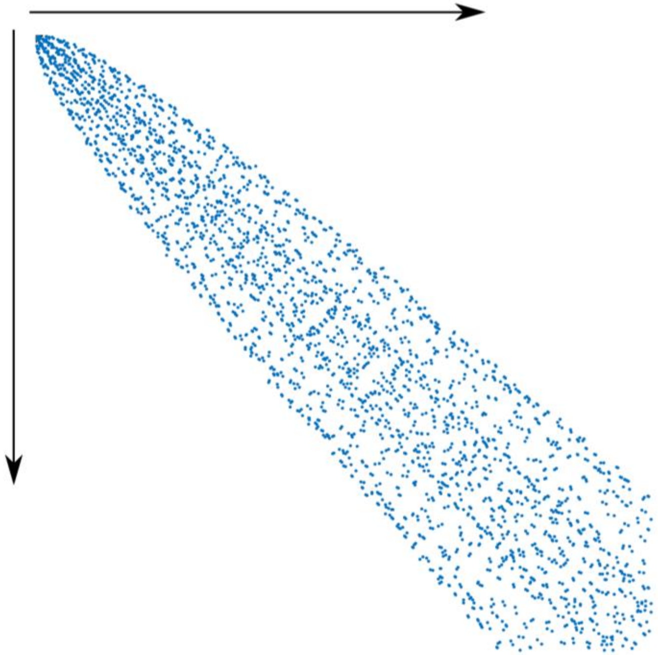
$$\left[A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{pmatrix} \right]_j = - \sum_{k \text{ connected to } j} e^{i\theta_{jk}(B)} x_k \,,$$

Matrix sparsity

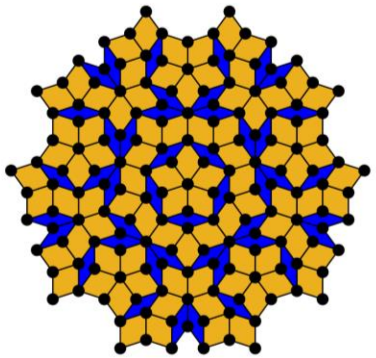


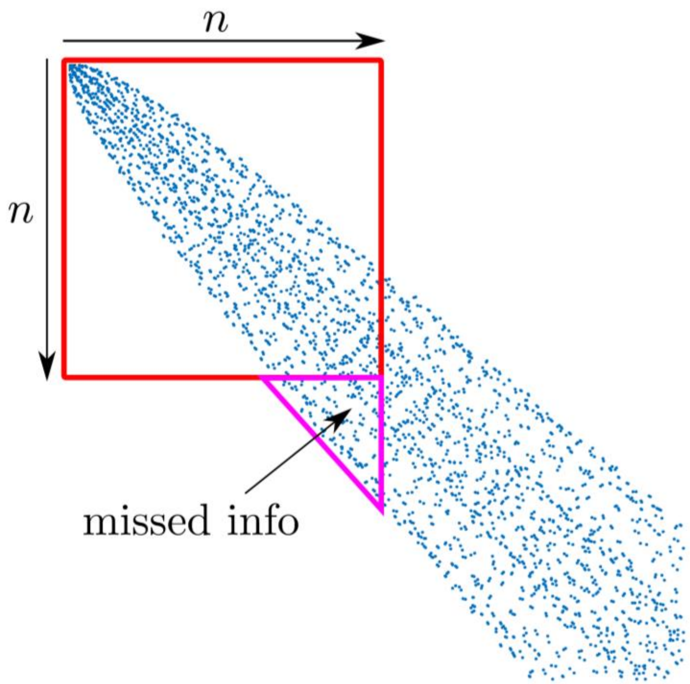
Example: quasicrystals





Example: quasicrystals

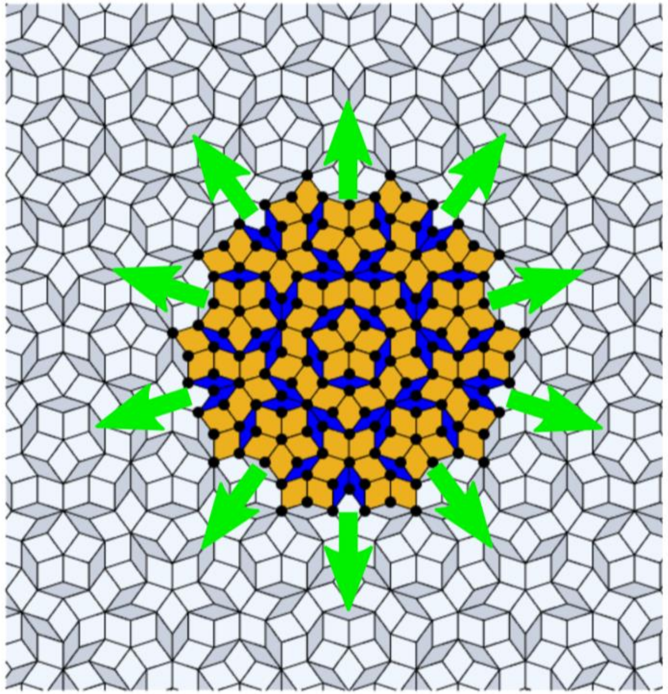


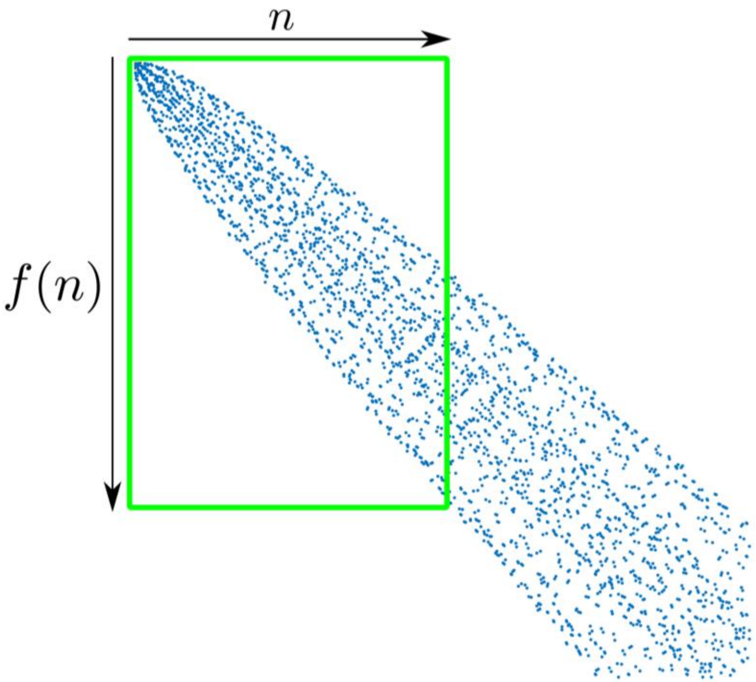


Typical approach: $n \times n$ truncation (possibly with BCs)

Problems: spectral pollution, which eigenvalues are reliable etc.

Example: quasicrystals





New approach: $f(n) \times n$ truncation.

Naturally captures interactions!

Sketch of algorithm

$$\sigma_{\inf}(T) = \inf\{\|Tv\| : v \in \mathfrak{D}(T), \|v\| = 1\}$$

$$\|(A - z)^{-1}\|^{-1} = \min\{\sigma_{\inf}(A - z), \sigma_{\inf}(A^* - \bar{z})\}$$

$$\sigma_{\inf}\left(P_{f(n)}[A - z]P_n\right) = \sigma_{\inf}([A - z]P_n) \downarrow \sigma_{\inf}(A - z)$$

Suppose we can relate $\|(A - z)^{-1}\|^{-1}$ to $\mathrm{dist}(z, \mathrm{Spec}(A))$, e.g., normal operators:

$$\sigma_{\inf}\left(P_{f(n)}[A - z]P_n\right) \downarrow \|(A - z)^{-1}\|^{-1} = \mathrm{dist}(z, \mathrm{Spec}(A))$$

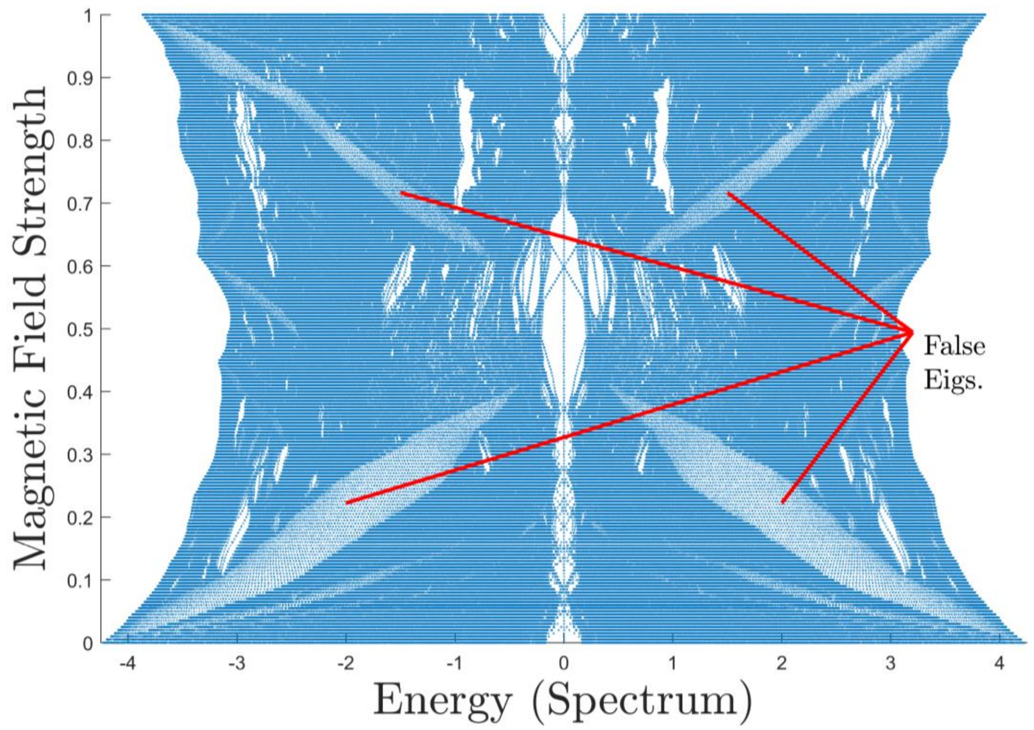
Final ingredient: local and adaptive search for local minimisers.

Error control!

Example: quasicrystals

Square truncations

Spectral pollution.

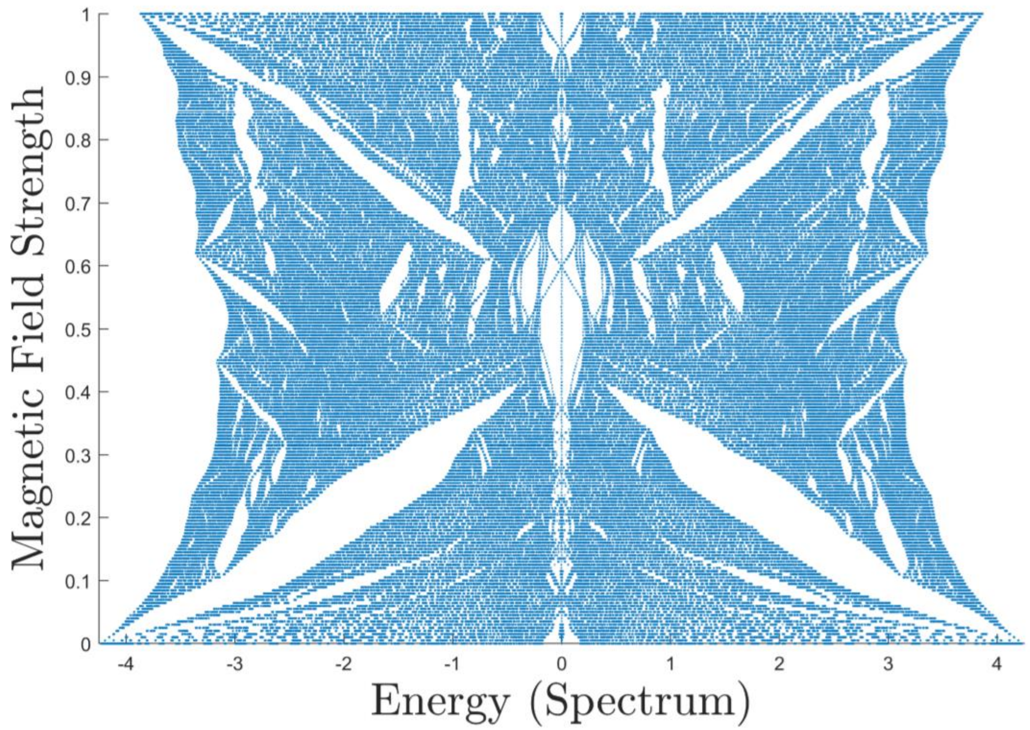


Does not converge

No error control

New method

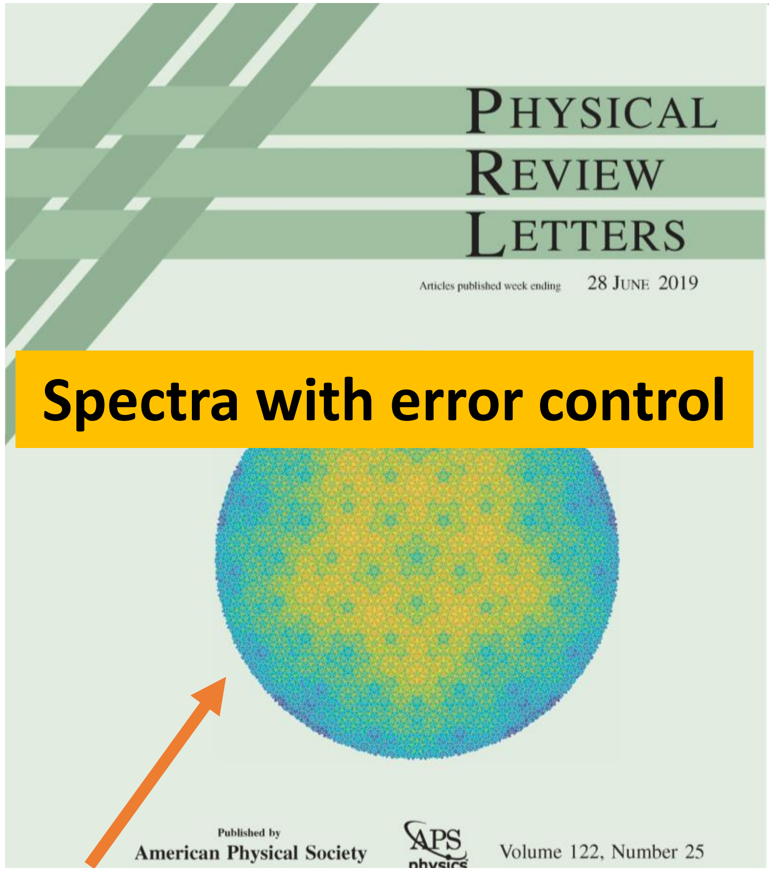
Convergent computation.



Converges

Error control

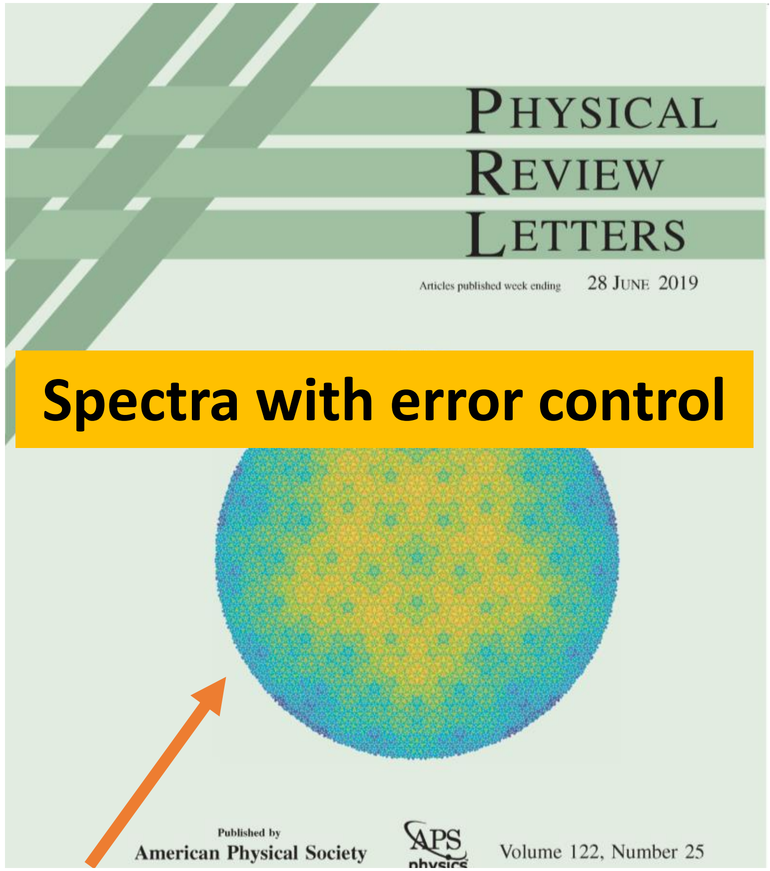
Is it right?
The importance of verification



E.g., ground state of quasicrystal

- C., Roman, Hansen, *"How to compute spectra with error control,"* **Phys. Rev. Lett.**, 2019.

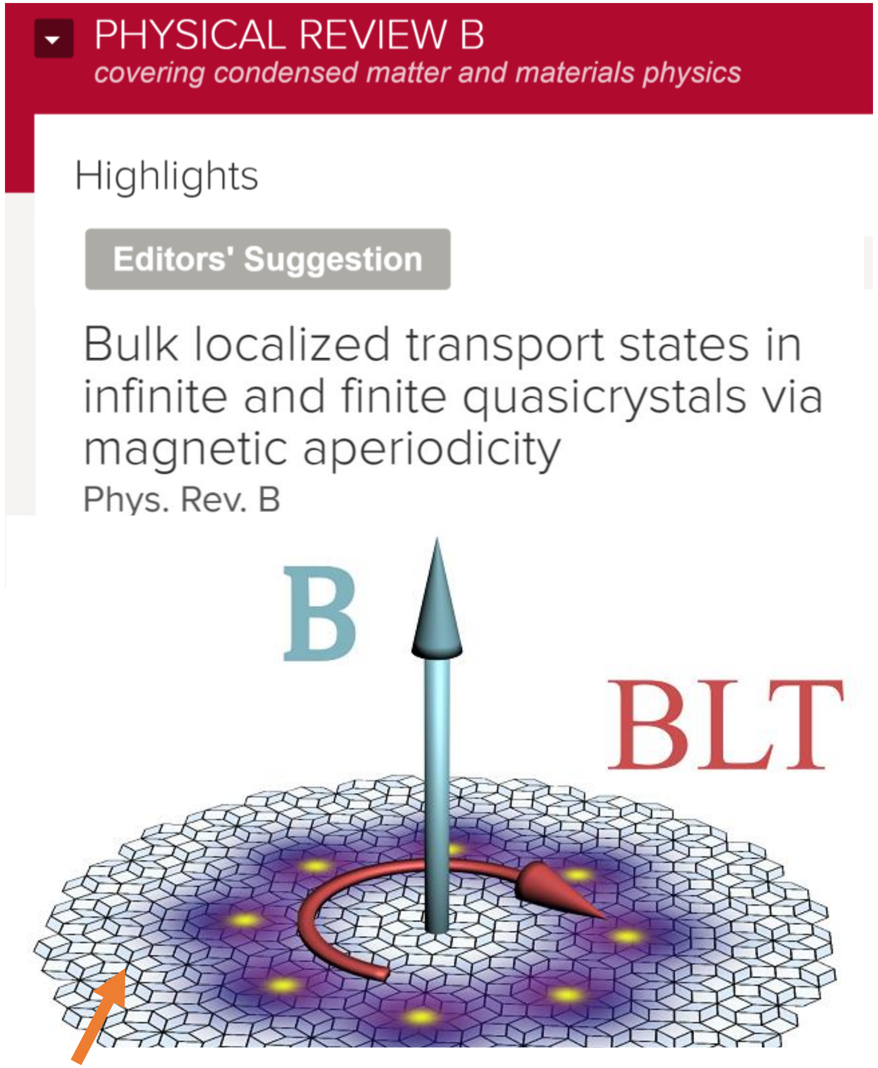
Is it right?
The importance of verification



Spectra with error control

E.g., ground state of quasicrystal

Certainty in computed spectral properties



E.g., new physical phenomena:
bulk localised transport states

- C., Roman, Hansen, *"How to compute spectra with error control,"* **Phys. Rev. Lett.**, 2019.
- Johnstone, C., Nielsen, Öhberg, Duncan, *"Bulk Localised Transport States in Infinite and Finite Quasicrystals via Magnetic Aperiodicity,"* **Phys. Rev. B**, 2022.

Example (local uniform convergence)

Theorem: Let Ω be class of self-adjoint diff. operators on $L^2(\mathbb{R}^d)$ of the form

$$T = \sum_{k \in \mathbb{Z}^d_{\geq 0}, |k| \leq N} c_k(x)\, \partial^k \qquad \text{s.t.}$$

- Smooth compactly supported functions form a core of T.
- $\{c_k\}$ are polynomially bounded and of locally bounded total variation.

Assume algorithm can:

- Point sample $\{c_k(q)\}$ for $q \in \mathbb{Q}^d$ to arbitrary prec.
- Evaluate a polynomial that bounds $\{c_k\}$ on $\mathbb{R}^d$.

Then…

- C., Hansen, *"The foundations of spectral computations via the solvability complexity index hierarchy,"* **J. Eur. Math. Soc.**, 2022

Example (local uniform convergence)

Theorem: Let Ω be class of self-adjoint diff. operators on $L^2(\mathbb{R}^d)$ of the form

$$T = \sum_{k \in \mathbb{Z}^d_{\geq 0}, |k| \leq N} c_k(x)\, \partial^k \qquad \text{s.t.}$$

- Smooth compactly supported functions form a core of T.
- $\{c_k\}$ are polynomially bounded and of locally bounded total variation.

Assume algorithm can:

- Point sample $\{c_k(q)\}$ for $q \in \mathbb{Q}^d$ to arbitrary prec.
- Evaluate a polynomial that bounds $\{c_k\}$ on $\mathbb{R}^d$.

Then

Verifiable

Not verifiable

(a) Know bound $\mathrm{TV}_{[-n,n]^d}(c_k) \leq b_n \Longrightarrow \{\mathrm{Sp}, \Omega\} \in \Sigma_1$.

(b) Only know asymp. bound $\mathrm{TV}_{[-n,n]^d}(c_k) = O(b_n) \Longrightarrow \{\mathrm{Sp}, \Omega\} \in \Delta_2 \backslash (\Sigma_1 \cup \Pi_1)$.

- C., Hansen, *“The foundations of spectral computations via the solvability complexity index hierarchy,”* **J. Eur. Math. Soc.**, 2022

Back to Schwinger: $-\Delta + V$ on $L^2(\mathbb{R}^d)$

Increasing difficulty →

Error control

Unbounded, sectorial V

Bounded V with locally bounded TV

Bounded V

$$\Pi_0 \qquad \Pi_1 \qquad \Pi_2 \qquad \Pi_3$$

$$\Pi_0 = \Delta_0 = \Sigma_0$$

$$\Delta_0 \subsetneq \Delta_1 \subsetneq \Sigma_1 \cup \Pi_1 \subsetneq \Delta_2 \subsetneq \Sigma_2 \cup \Pi_2 \subsetneq \Delta_3 \subsetneq \Sigma_3 \cup \Pi_3 \cdots$$

$$\Delta_1 \subsetneq \Pi_1 \subsetneq \Delta_2, \quad \Delta_1 \subsetneq \Sigma_1 \subsetneq \Delta_2, \quad \Delta_2 \subsetneq \Pi_2 \subsetneq \Delta_3, \quad \Delta_2 \subsetneq \Sigma_2 \subsetneq \Delta_3, \quad \Delta_3 \subsetneq \Pi_3 \cdots, \quad \Delta_3 \subsetneq \Sigma_3 \cdots$$

$$\Sigma_0 \qquad \Sigma_1 \qquad \Sigma_2 \qquad \Sigma_3$$

Self-adjoint, bounded V with locally bounded TV

- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, "*On the solvability complexity index hierarchy and towers of algorithms,*" preprint.

Back to Schwinger: $-\Delta + V$ on $L^2(\mathbb{R}^d)$

Increasing difficulty →

Error control

Unbounded, sectorial V

Bounded V with locally bounded TV

Bounded V

$$\Pi_0 = \Delta_0 = \Sigma_0 \subsetneq \Delta_1 \subsetneq \Sigma_1 \cup \Pi_1 \subsetneq \Delta_2 \subsetneq \Sigma_2 \cup \Pi_2 \subsetneq \Delta_3 \subsetneq \Sigma_3 \cup \Pi_3 \cdots$$

$$\Delta_1 \subsetneq \Pi_1 \subsetneq \Delta_2, \quad \Delta_1 \subsetneq \Sigma_1 \subsetneq \Delta_2, \quad \Delta_2 \subsetneq \Pi_2 \subsetneq \Delta_3, \quad \Delta_2 \subsetneq \Sigma_2 \subsetneq \Delta_3, \quad \Delta_3 \subsetneq \Pi_3 \cdots, \quad \Delta_3 \subsetneq \Sigma_3 \cdots$$

Self-adjoint, bounded V with locally bounded TV

NB: Most existing convergence results for spectra, even on bounded domains, prove Δ_2 results and miss the optimal Σ_1 convergence!

CHALLENGE: Can you get Σ_1 for your problem/method?

- Ben-Artzi, C., Hansen, Nevanlinna, Seidel, "*On the solvability complexity index hierarchy and towers of algorithms,*" preprint.

Example 2: Δ_2 alg. for spectral meas.

Spectral measures → diagonalisation

- **Fin.-dim.:** $B \in \mathbb{C}^{n\times n}$, $B^*B = BB^*$, o.n. basis of e-vectors $\{v_j\}_{j=1}^n$

$$v = \left[\sum_{j=1}^{n} v_j {v_j}^*\right] v, \qquad Bv = \left[\sum_{j=1}^{n} \lambda_j v_j {v_j}^*\right] v, \qquad \forall v \in \mathbb{C}^n$$

- **Inf.-dim.:** Operator $A: \mathcal{D}(A) \to \mathcal{H}$. Typically, no basis of e-vectors!
  Spectral theorem: (projection-valued) spectral measure E

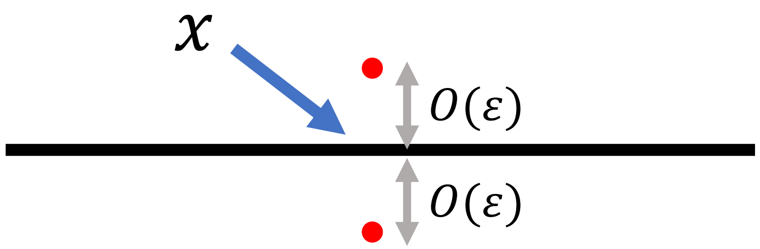
$$f = \left[\int_{\mathrm{Spec}(A)} 1 \,\mathrm{d}E(\lambda)\right] f, \qquad Af = \left[\int_{\mathrm{Spec}(A)} \lambda \,\mathrm{d}E(\lambda)\right] f, \qquad \forall f \in \mathcal{H}$$

- **Spectral measures:** $\mu_f(U) = \langle E(U)f, f\rangle$ ($\|f\| = 1$) prob. Measure on $\mathbb{R}$.

A two-limit algorithm (Stone's formula)

Smoothed spectral measure:

$$\mu_f^{\varepsilon}(x) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{\varepsilon \, \mathrm{d}\mu_f(\lambda)}{(x-\lambda)^2+\varepsilon^2} = \frac{\langle[(A-[x+i\varepsilon])^{-1}-(A-[x-i\varepsilon])^{-1}]f,f\rangle}{2\pi i}$$



$\varepsilon =$ "smoothing parameter"

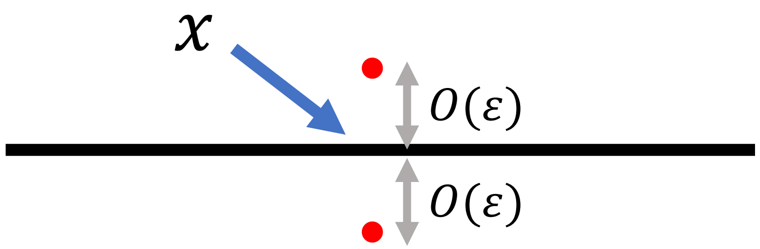
A two-limit algorithm (Stone's formula)

Smoothed spectral measure:

$$\mu_f^{\varepsilon}(x) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{\varepsilon \, \mathrm{d}\mu_f(\lambda)}{(x-\lambda)^2+\varepsilon^2} = \frac{\langle [(A-[x+i\varepsilon])^{-1} - (A-[x-i\varepsilon])^{-1}]f, f\rangle}{2\pi i}$$

Discretize RHS with size n_1, to get $\mu_{f,n_1}^{\varepsilon}$. Set

$$\Gamma_{n_1,n_2}(A) = \mu_{f,n_1}^{1/n_2}$$



$\varepsilon =$ "smoothing parameter"

Converges in weak sense.

Without extra assumptions, this is sharp!!

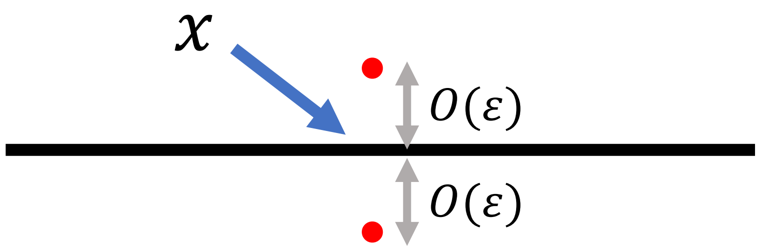
A two-limit algorithm (Stone's formula)

Smoothed spectral measure:

$$\mu_f^{\varepsilon}(x) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{\varepsilon \, \mathrm{d}\mu_f(\lambda)}{(x-\lambda)^2+\varepsilon^2} = \frac{\langle [(A-[x+i\varepsilon])^{-1} - (A-[x-i\varepsilon])^{-1}]f, f\rangle}{2\pi i}$$

Discretize RHS with size n_1, to get $\mu_{f,n_1}^{\varepsilon}$. Set

$$\Gamma_{n_1,n_2}(A) = \mu_{f,n_1}^{1/n_2}$$



$\varepsilon =$ "smoothing parameter"

Converges in weak sense.

Without extra assumptions, this is sharp!!

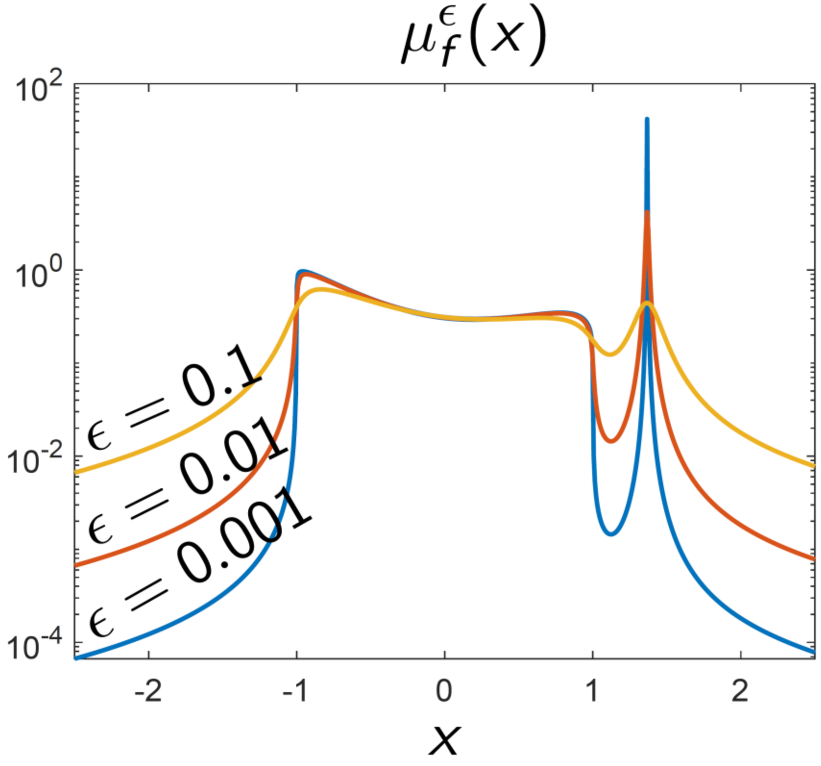
If we can compute RHS with error control (e.g., residuals), choose $n_1(\varepsilon)$.

Example: integral operator

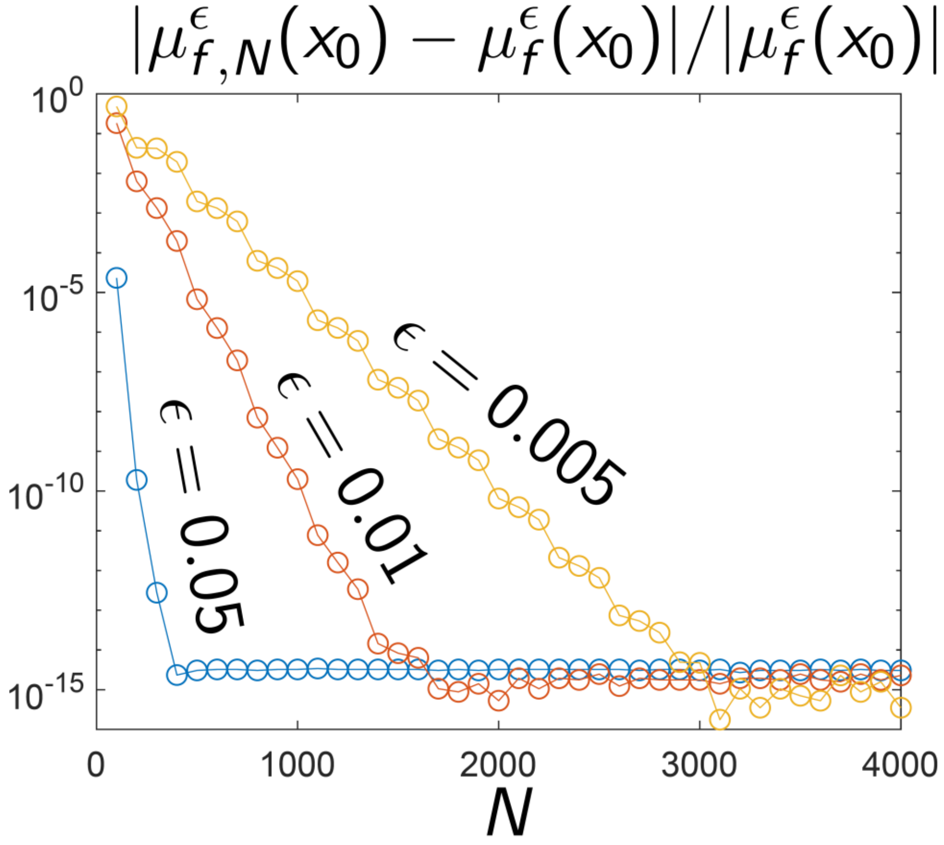
$$[Au](x) = xu(x) + \int_{-1}^{1} e^{-(x^2+y^2)} u(y)\mathrm{d}y$$

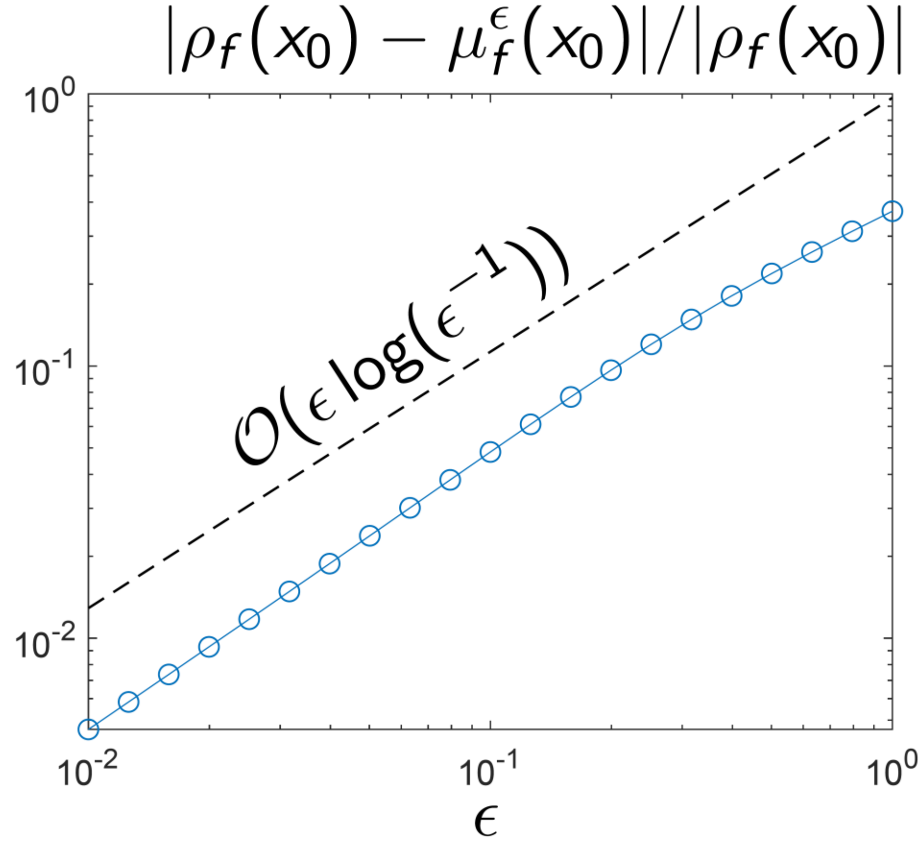
Discretize using adaptive Chebyshev collocation method.

Look at μ_f for $f(x) = \sqrt{3/2}x$



Example: integral operator





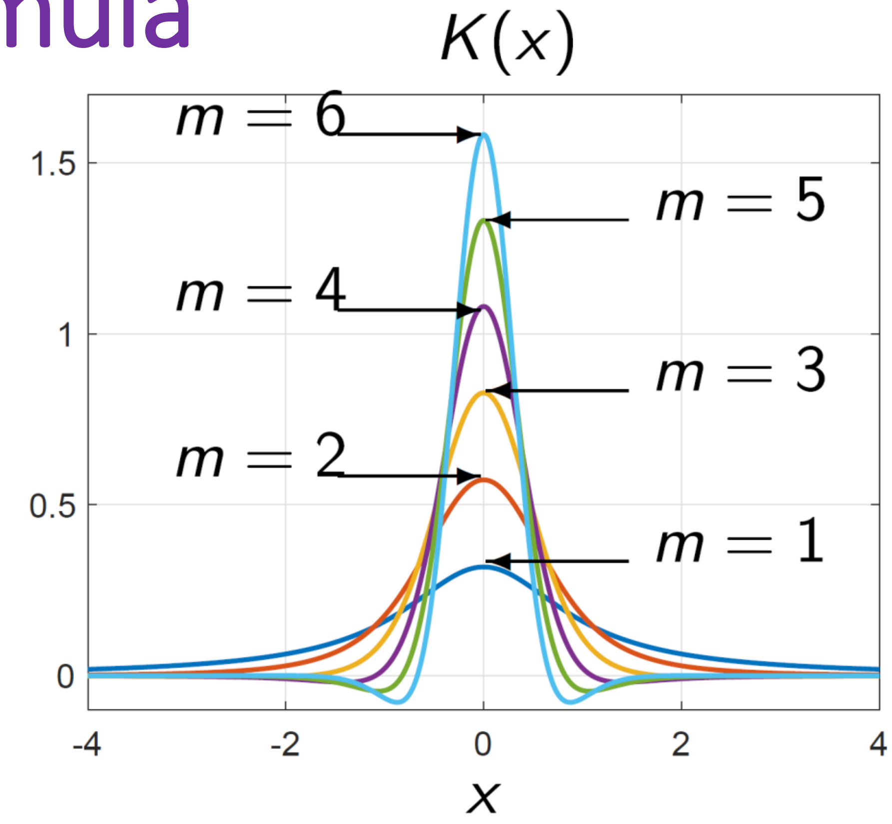
Slow convergence (more than five digits infeasible). **Can we do better?**

High-order versions of Stone's formula

mth order rational "smoothing" kernels:

$$K(x) = \frac{1}{2\pi i}\sum_{j=1}^{m} \frac{\alpha_j}{x-a_j} - \frac{\overline{\alpha_j}}{x-\overline{a_j}}, K_\varepsilon(x) = K(x/\varepsilon)/\varepsilon$$

$$[K_\varepsilon * \mu_f](x)$$
$$= \frac{-1}{2\pi i}\sum_{j=1}^{m}\langle[\alpha_j(A-[x-\varepsilon a_j])^{-1} - \overline{\alpha_j}(A-[x-\varepsilon \overline{a_j}])^{-1}]f, f\rangle$$



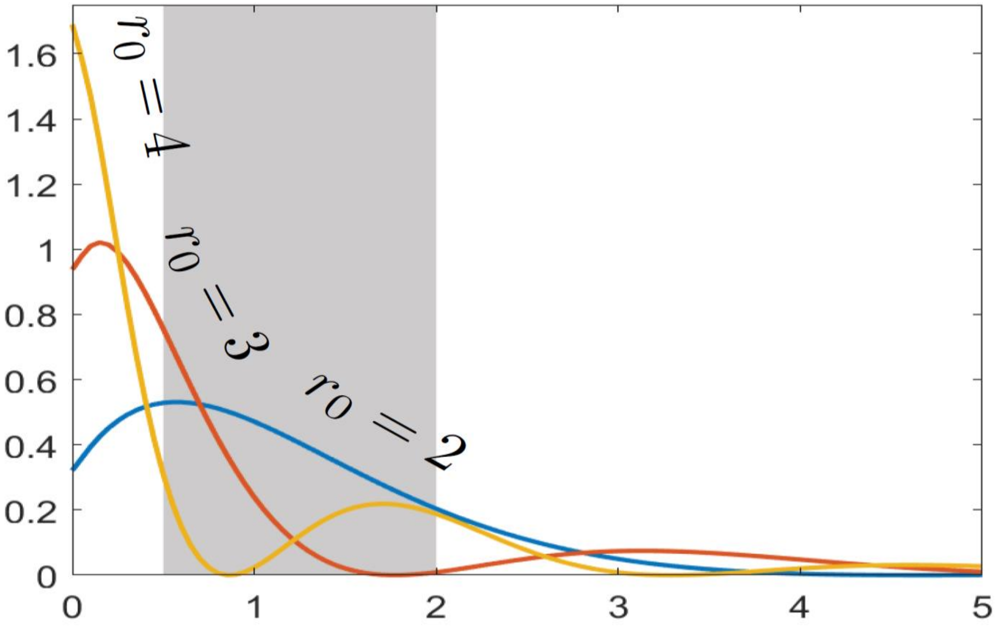
$\Longrightarrow$ **larger ε for a given accuracy $\Longrightarrow$ smaller $n_1(\varepsilon)$ for a given accuracy**

Demo: radial Schrödinger

$$[\mathcal{L}u](r) = -\frac{d^2u}{dr^2}(r) + \left(\frac{\ell(\ell+1)}{r^2} + \frac{1}{r}(e^{-r} - 1)\right) u(r), \qquad r > 0.$$

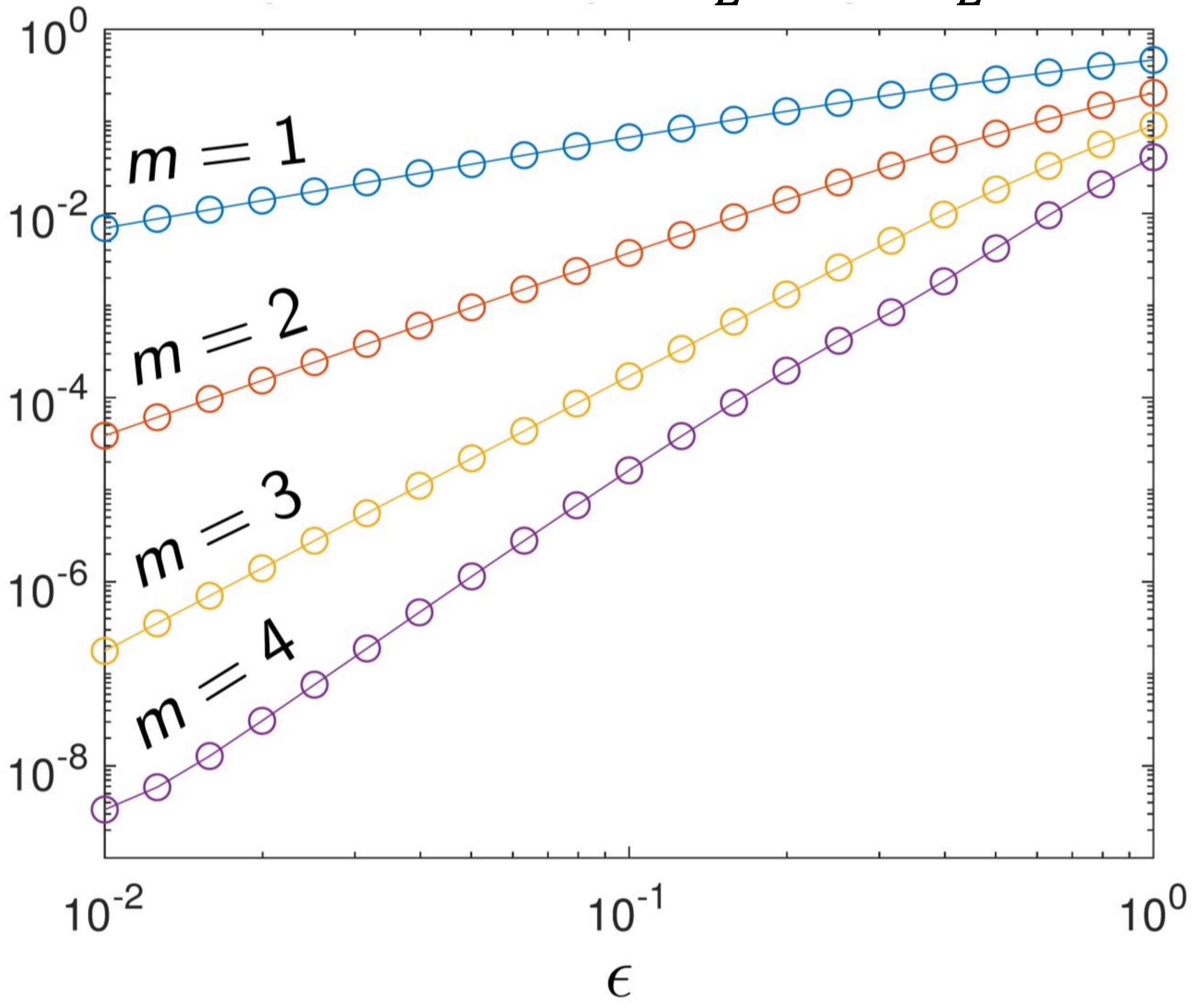
```
normf = sqrt(pi/8)*(2-igamma(1/2,8)/gamma(1/2)); % Normalization
f = @(r) exp(-(r-2).^2)/sqrt(normf);             % Measure wrt f(r)
V={@(r) 0, @(r) exp(-r)-1, 1};                   % Potential, l=1
[xi, wi] = chebpts(20,  [1/2 2]);                % Quadrature rule
mu = rseMeas(V, f, xi, 0.1, 'Order', 4)          % epsilon=0.1, m=4
ion_prob = wi * mu;                              % Ionization prob
```

Demo: radial Schrödinger



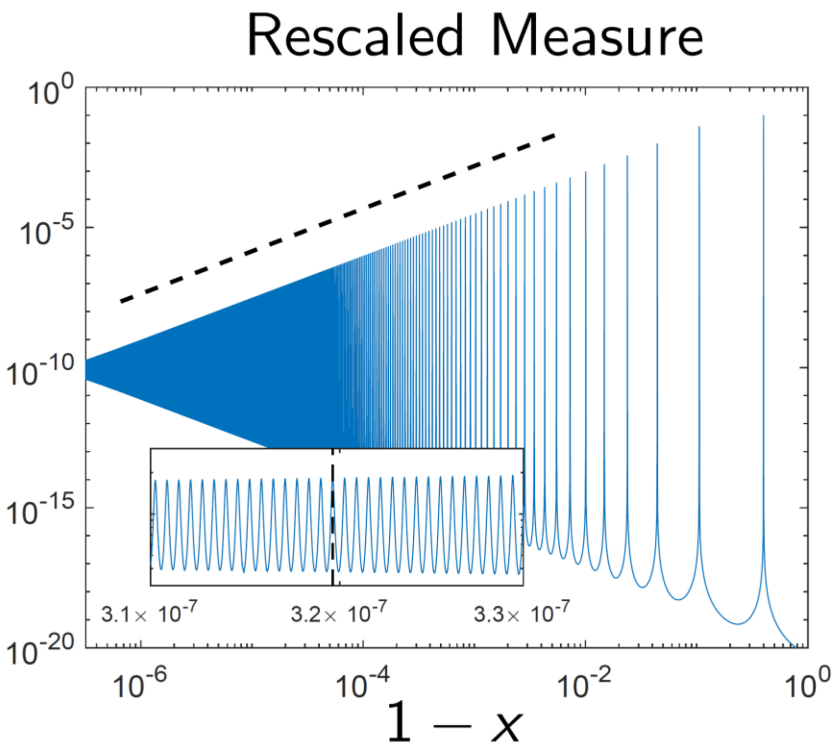
Wavefunction $\propto e^{-(r-r_0)^2}$

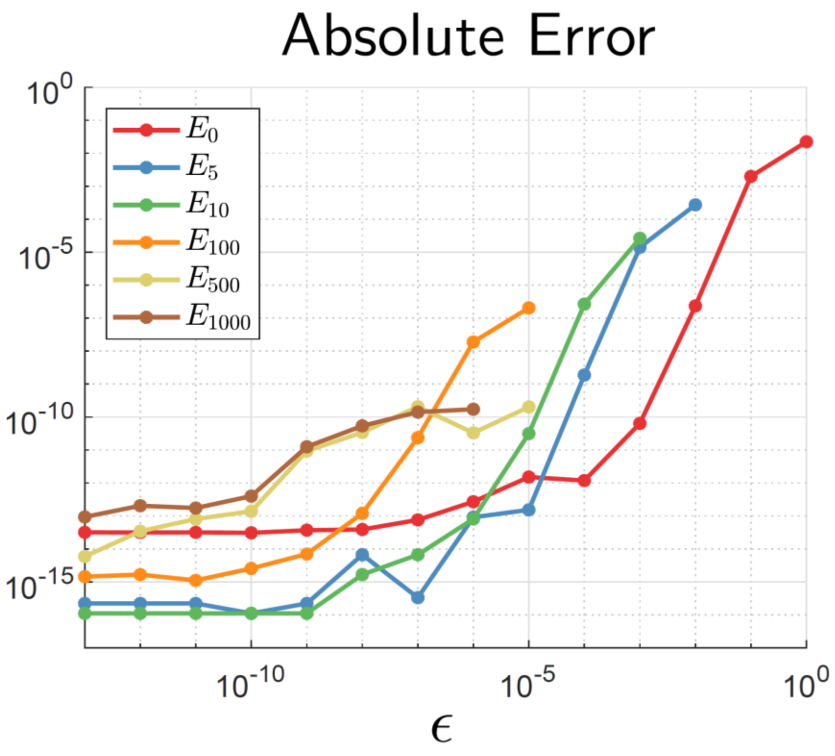
$\|\rho_f - [K_\varepsilon * \mu_f]\|_{L^1} / \|\rho_f\|_{L^1}$



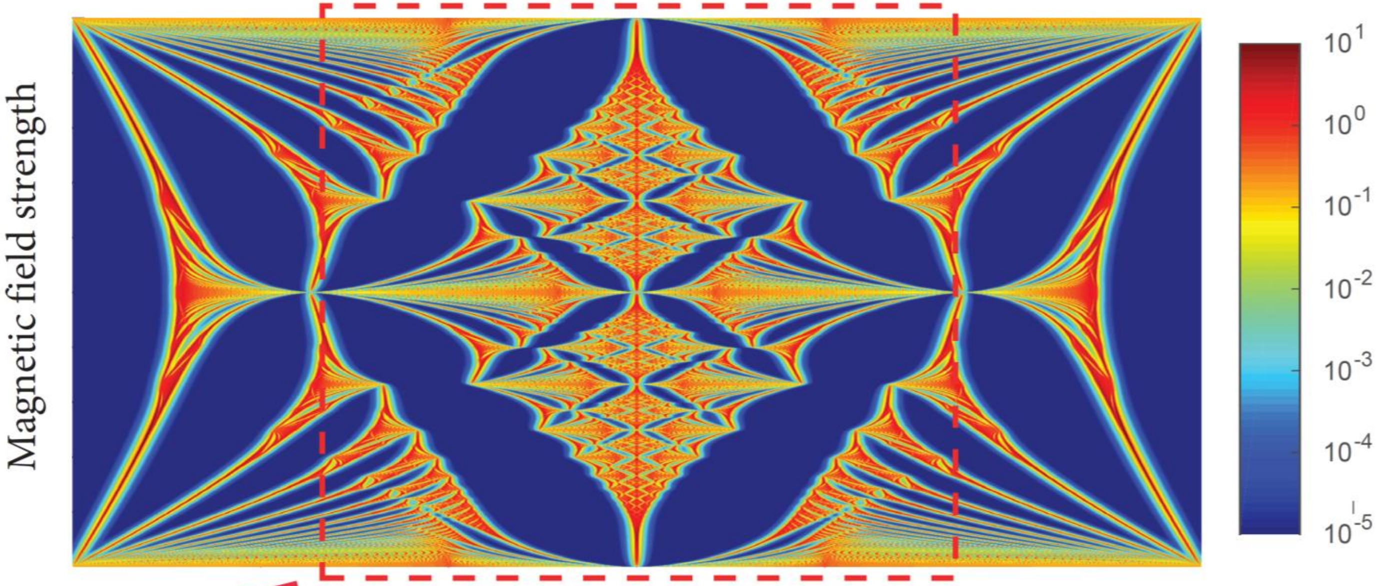
Eigenvalues of Dirac operator

$$\mathcal{D}_V = \begin{pmatrix} 1 + V(r) & -\frac{d}{dr} + \frac{\kappa}{r} \\ \frac{d}{dr} + \frac{\kappa}{r} & -1 + V(r) \end{pmatrix}$$





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.

Executive summary of theorems

Input an open (or closed) set

- **Generic assumptions:** Computing $(A, f, U) \hookrightarrow \mu_f(U)$ has $\mathrm{SCI} = 1$ but error control or rate *impossible* (even for discrete Schrödinger).
- If spectral measure μ_f is a.c. on interval I, with $\mathcal{C}^{n,\alpha}$ density ρ_f, then
$$\left\|\rho_f - \left[K_\varepsilon * \mu_f\right]\right\|_{L^\infty(I)} = \mathcal{O}(\varepsilon^{n+\alpha} + \varepsilon^m \log(1/\varepsilon))$$
- Weak convergence always $\mathcal{O}(\varepsilon^m \log(1/\varepsilon))$ for $\mathcal{C}^m$ test functions.
- Splitting into spectral type: $\mathrm{SCI} = 2$ or 3.

NB: Constants can be made **explicit**.

- C., *"Computing spectral measures and spectral types,"* **Comm. Math. Phys.**, 2021

Further areas

Other areas with SCI results

- PDEs e.g.:
  - Can you solve Schrödinger eq. on $L^2(\mathbb{R}^d)$ with error control?
  - Can you predict blow-up of non-linear PDEs?
- Optimization
- Inverse problems (e.g., imaging)
- Polynomial root-finding: Smale (settled by McMullen), *"Is there a purely iterative convergent algorithm for polynomial zero finding?"*
- Topology
- As well as … (computer-assisted proofs, AI, dynamical systems etc.)

Computer-assisted proof: Dirac-Schwinger conjecture

$E(Z)$ = ground state energy of

N: # of electrons, Z: charge of nucleus

$$H = \sum_{k=1}^{N}\left(-\Delta_{x_k} - Z|x_k|^{-1}\right) + \sum_{j<k}|x_j - x_k|^{-1}.$$

Theorem: $E(Z) = -c_0 Z^{7/3} + \frac{1}{8}Z^2 - c_1 Z^{5/3} + O(Z^{5/3-1/2835})$, as $Z \to \infty$

Proof involves spectral analysis, analytic number theory, …,
computer-assisted bound involving solutions of an ODE.

Fefferman and Seco implicitly prove Σ_1 classifications!

- Fefferman, Phong, "On the lowest eigenvalue of a pseudo-differential operator," **Proc. Natl. Acad. Sci. USA**, 1979.
- Fefferman, *"The N-body problem in quantum mechanics,"* **Comm. Pure Appl. Math.**, 1986.
- Fefferman, Seco, *"Interval arithmetic in quantum mechanics,"* **Applications of interval computations**, 1996.

Computer-assisted proof: Kepler conjecture (Hilbert's 18th problem)

Proof shows potential counterexamples would satisfy infeasible inequalities

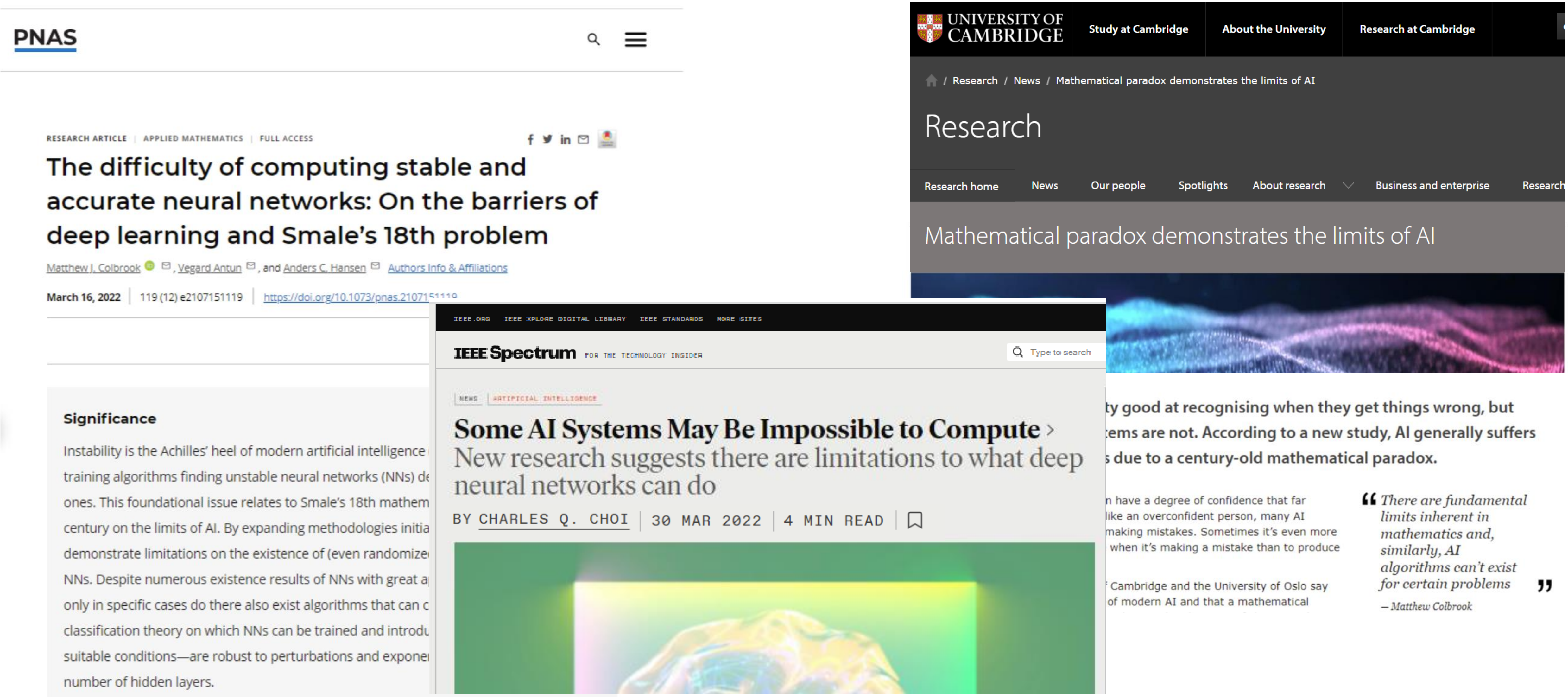
relaxed to $\approx$ 10,000s linear programs

These can't always be decided!

- Hales, *"A proof of the Kepler conjecture,"* **Ann. of Math.**, 2005.
- Hales et al., *"A formal proof of the Kepler conjecture,"* **Forum Math. Pi**, 2017.
- Bastounis, Hansen, Vlačić, *"The extended Smale's 9th problem,"* preprint.

Account of Flyspeck project (formal proof)

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 | https://doi.org/10.1073/pnas.2107151119

Significance

Instability is the Achilles' heel of modern artificial intelligence
training algorithms finding unstable neural networks (NNs) de
ones. This foundational issue relates to Smale's 18th mathem
century on the limits of AI. By expanding methodologies initia
demonstrate limitations on the existence of (even randomize
NNs. Despite numerous existence results of NNs with great a
only in specific cases do there also exist algorithms that can c
classification theory on which NNs can be trained and introdu
suitable conditions—are robust to perturbations and expone
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 enterprise | Research

Mathematical paradox demonstrates the limits of AI

ty good at recognising when they get things wrong, but
tems are not. According to a new study, AI generally suffers
s due to a century-old mathematical paradox.

n have a degree of confidence that far
ike an overconfident person, many AI
making mistakes. Sometimes it's even more
when it's making a mistake than to produce

Cambridge and the University of Oslo say
of modern AI and that a mathematical

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

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

IEEE Spectrum FOR THE TECHNOLOGY INSIDER

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

- 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**.

Example: Rigorous Koopmania!

- 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 $\{x^{(m)}, y^{(m)} = F(x^{(m)})\}_{m=1}^{M}$
- **Koopman operator** $\mathcal{K}$ acts on functions $g: \Omega \to \mathbb{C}$

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

- $\mathcal{K}$ is ***linear*** but acts on an ***infinite-dimensional*** space.
- Often spectral info encodes the features of the system we want.
- 35,000 papers over last decade, hardly anything on NA of this problem!

- 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.,** under minor rev.
- Code: https://github.com/MColbrook/Residual-Dynamic-Mode-Decomposition

Summary

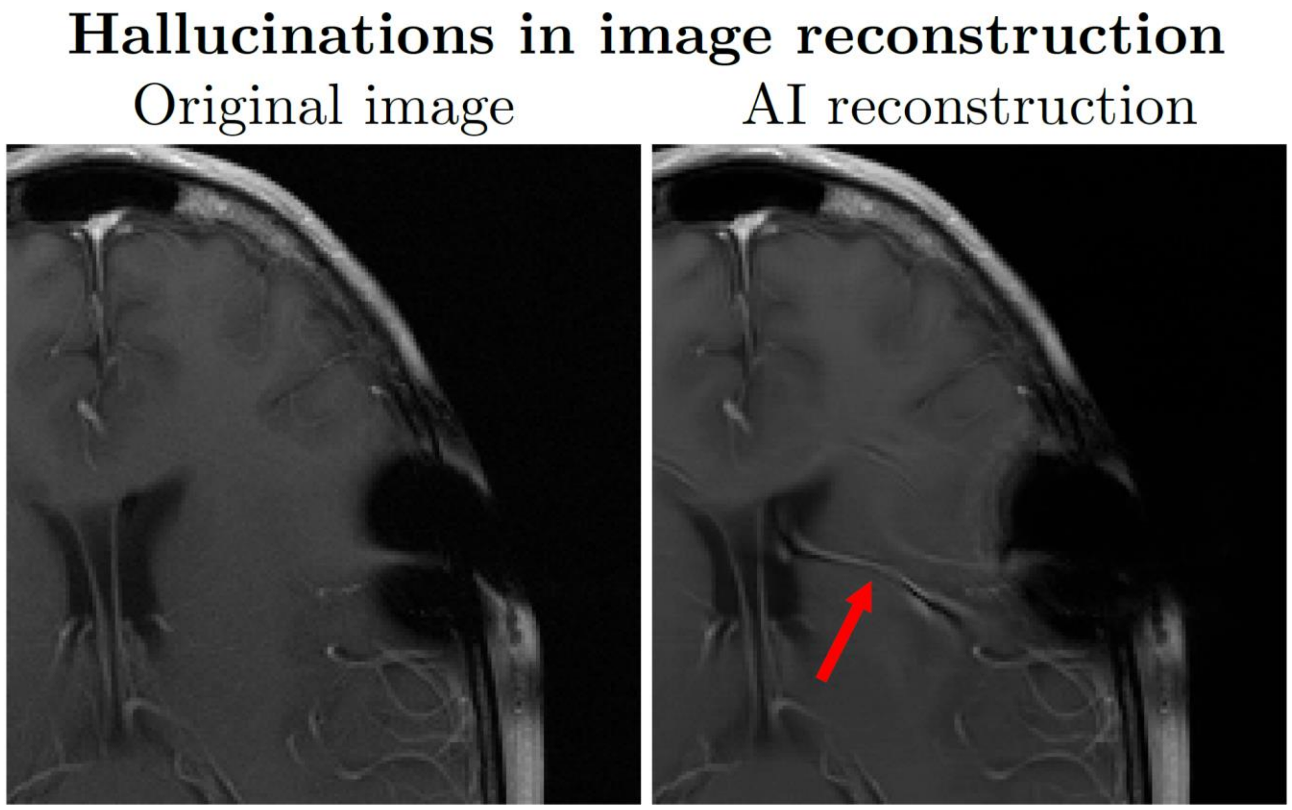
SCI hierarchy: a tool that allows us to

- Classify difficulty of continuous and discrete computational problems.
- Prove that algorithms are optimal (in any given computational model).
- Framework ⇒ find assumptions and methods for computational goals.

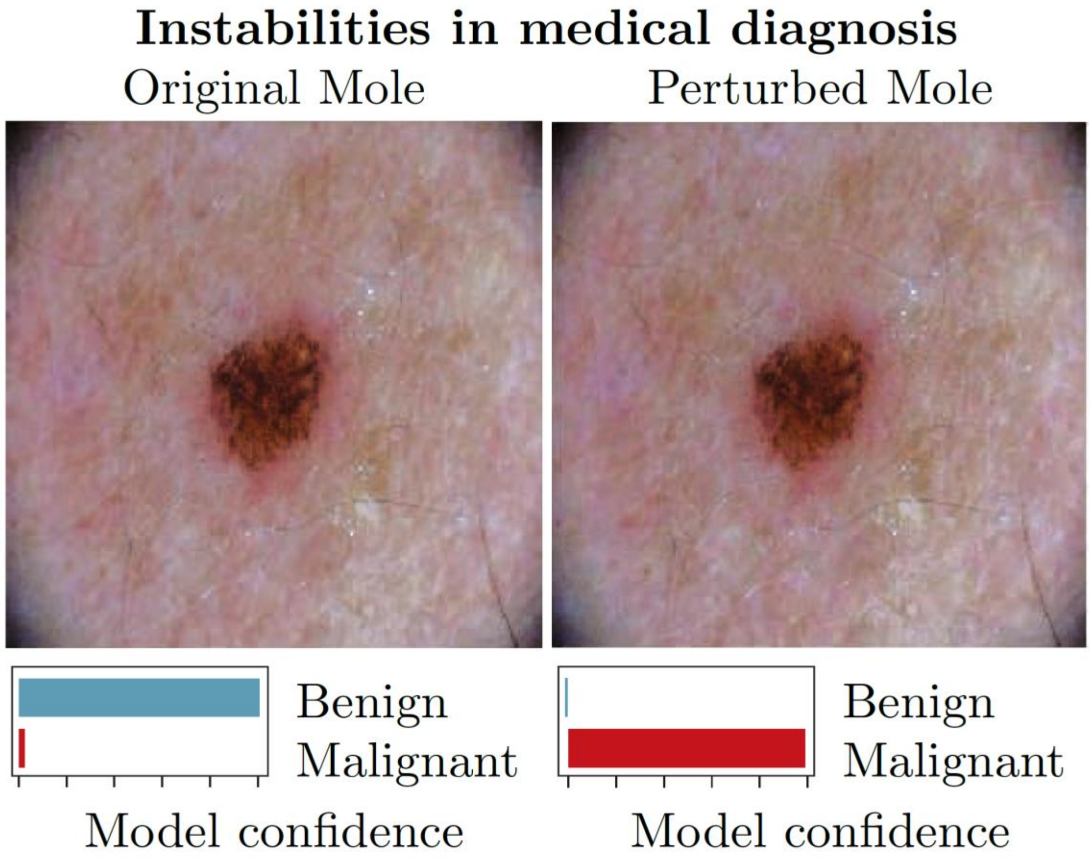
http://www.damtp.cam.ac.uk/user/mjc249/home.html: slides, papers, and code

Additional slides

Problem: hallucinations and instability



"**AI generated hallucination**", from Facebook and NYU's ***FastMRI challenge*** 2020



From Finlayson et al., "*Adversarial attacks on medical machine learning*," **Science**, 2019.

When can we make AI robust and trustworthy?

Example of the limits of deep learning

Paradox: "Nice" linear inverse problems where a *stable* and *accurate* neural network for image reconstruction exists, but it can never be trained!

E.g., suppose we want to solve (holds for much more general problems)

$$(P_1) \quad \mathrm{argmin}_{x \in \mathbb{C}^N} F_1^A(x) := \|x\|_{l^1_w}, \text{ such that } \|Ax - y\|_{l^2} \leq \epsilon,$$

$$(P_2) \quad \mathrm{argmin}_{x \in \mathbb{C}^N} F_2^A(x, y, \lambda) := \lambda \|x\|_{l^1_w} + \|Ax - y\|_{l^2}^2,$$

$$(P_3) \quad \mathrm{argmin}_{x \in \mathbb{C}^N} F_3^A(x, y, \lambda) := \lambda \|x\|_{l^1_w} + \|Ax - y\|_{l^2}.$$

$A \in \mathbb{C}^{m \times N}$ (modality, $m < N$), $\qquad S = \{y_j\}_{j=1}^{R}$ (samples)

Arises when given $y \approx Ax + e$.

Arbitrary precision of training data

In practice, A not known exactly or cannot be stored to infinite precision.

Assume access to: $\{y_{k,n}\}_{k=1}^{R}$ and A_n (rational approximations, e.g., floats) such that

$$\|y_{k,n} - y_k\| \leq 2^{-n}, \quad \|A_n - A\| \leq 2^{-n}, \quad \forall n \in \mathbb{N}.$$

Training set for $(A, \mathcal{S}) \in \Omega$:

$$\iota_{A,\mathcal{S}} := \{(y_{k,n}, A_n) \mid k = 1, \dots, R \text{ and } n \in \mathbb{N}\}.$$

In a nutshell: allow access to arbitrary precision training data.

Question: Given a collection Ω of $(A, \mathcal{S})$, does there exist a neural network approximating Ξ (solution map of (P_j)), and can it be trained by an algorithm?

Condition numbers

Given $\Omega \subseteq \mathbb{C}^n$, define

$$\mathrm{Act}(\Omega) = \{j : \exists x, y \in \Omega, x_j \neq y_j\}, \qquad \Omega^{\mathrm{Act}} = \{x : \exists y \in \Omega, x_{\mathrm{Act}(\Omega)^c} = y_{\mathrm{Act}(\Omega)^c}\}$$

- Condition of a mapping $\Xi : \widehat{\Omega} \rightrightarrows \mathbb{C}^m$ with $\Omega \subseteq \widehat{\Omega}$:

$$\mathrm{Cond}(\Xi, \Omega) = \sup_{x \in \Omega} \lim_{\varepsilon \to 0_+} \sup_{\substack{x+z \in \Omega^{\mathrm{Act}} \cap \widehat{\Omega} \\ 0 < \|z\|_\infty < \varepsilon}} \frac{\mathrm{dist}(\Xi(x+z), \Xi(x))}{\|z\|_\infty}$$

- For problems with constraints (e.g., basis pursuit P_1 or LPs)

$$\nu(A, y) = \inf\{\varepsilon \geq 0 : \|\hat{y} - y\|_2, \|\hat{A} - A\| \leq \varepsilon, (\hat{A}, \hat{y}) \in \Omega^{\mathrm{Act}} \text{ and infeasible}\}$$

$$C_{\mathrm{FP}}(A, y) = \frac{\max\{\|y\|_2, \|A\|\}}{\nu(A, y)}$$

- Renegar condition number

$$\mu(A, y) = \inf\{\varepsilon \geq 0 : \|\hat{y} - y\|_2, \|\hat{A} - A\| \leq \varepsilon, (\hat{A}, \hat{y}) \in \Omega^{\mathrm{Act}}, \Xi \text{ multivalued}\}$$

$$C_{\mathrm{RCC}}(A, y) = \frac{\max\{\|y\|_2, \|A\|\}}{\mu(A, y)}$$

Theorem: For any of prev. problems, integer $K \geq 3$ and $L \in \mathbb{N}$, $\exists$ a well-conditioned class $\Omega(K)$ of inputs s.t. simultaneously

1. No deterministic alg. can, given a training set $\iota_{A,S} \in \Omega_{\mathcal{T}}$, produce a neural network (NN) ϕ with

$$(1) \qquad \min_{y \in S} \inf_{x^* \in \Xi(A,y)} \|\phi(y) - x^*\|_2 \leq 10^{-K} \qquad \forall (A,S) \in \Omega(K).$$

For any $p > 1/2$, no random alg. (any model of comp.) can produce a NN ϕ s.t. (1) holds with prob. $\geq p$.

2. (a) $\exists$ deterministic alg. that , given a training set $\iota_{A,S} \in \Omega_{\mathcal{T}}$, produces a neural network (NN) ϕ with

$$(2) \qquad \max_{y \in S} \inf_{x^* \in \Xi(A,y)} \|\phi(y) - x^*\|_2 \leq 10^{-(K-1)} \qquad \forall (A,S) \in \Omega(K).$$

(b) However, for any probabilistic Turing Machine that produces such a NN, any $M \in \mathbb{N}$ and $p \in \left[0, \frac{N-m}{N+1-m}\right)$, there exists a training set $\iota_{A,S} \in \Omega_{\mathcal{T}}$ s.t. $\forall y \in S$

$$\mathbb{P}\left(\inf_{x^* \in \Xi(A,y)} \|\phi(y) - x^*\|_2 > 10^{-(K-1)} \text{ or size of training data to construct } \phi \text{ exceeds } M\right) > p.$$

3. $\exists$ deterministic alg. that, given a training set $\iota_{A,S} \in \Omega_{\mathcal{T}}$, produces a NN ϕ accessing at most L training samples of $\iota_{A,S}$ s.t.

$$(3) \qquad \max_{y \in S} \inf_{x^* \in \Xi(A,y)} \|\phi(y) - x^*\|_2 \leq 10^{-(K-2)} \qquad \forall (A,S) \in \Omega(K).$$

- 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**.

Theorem: For any of prev. problems, integer $K \geq 3$ and $L \in \mathbb{N}$, $\exists$ a well-conditioned class $\Omega(K)$ of inputs s.t. simultaneously

1. No deterministic alg. can, given a training set $\iota_{A,S} \in \Omega_{\mathcal{T}}$, produce a neural network (NN) ϕ with

$$(1) \qquad \min_{y \in S} \inf_{x^* \in \Xi(A,y)} \|\phi(y) - x^*\|_2 \leq 10^{-K} \qquad \forall (A,S) \in \Omega(K).$$

For any $p > 1/2$, no random alg. (any model of comp.) can produce a NN ϕ s.t. (1) holds with prob. $\geq p$.

2. (a) $\exists$ deterministic alg. that , given a training set $\iota_{A,S} \in \Omega_{\mathcal{T}}$, produces a neural network (NN) ϕ with

$$(2) \qquad \max \inf \|\phi(y) - x^*\|_2 \leq 10^{-(K-1)} \qquad \forall (A,S) \in \Omega(K).$$

Holds for any architecture, any precision of training data.

$\Longrightarrow$ Classification theory telling us what can and cannot be done

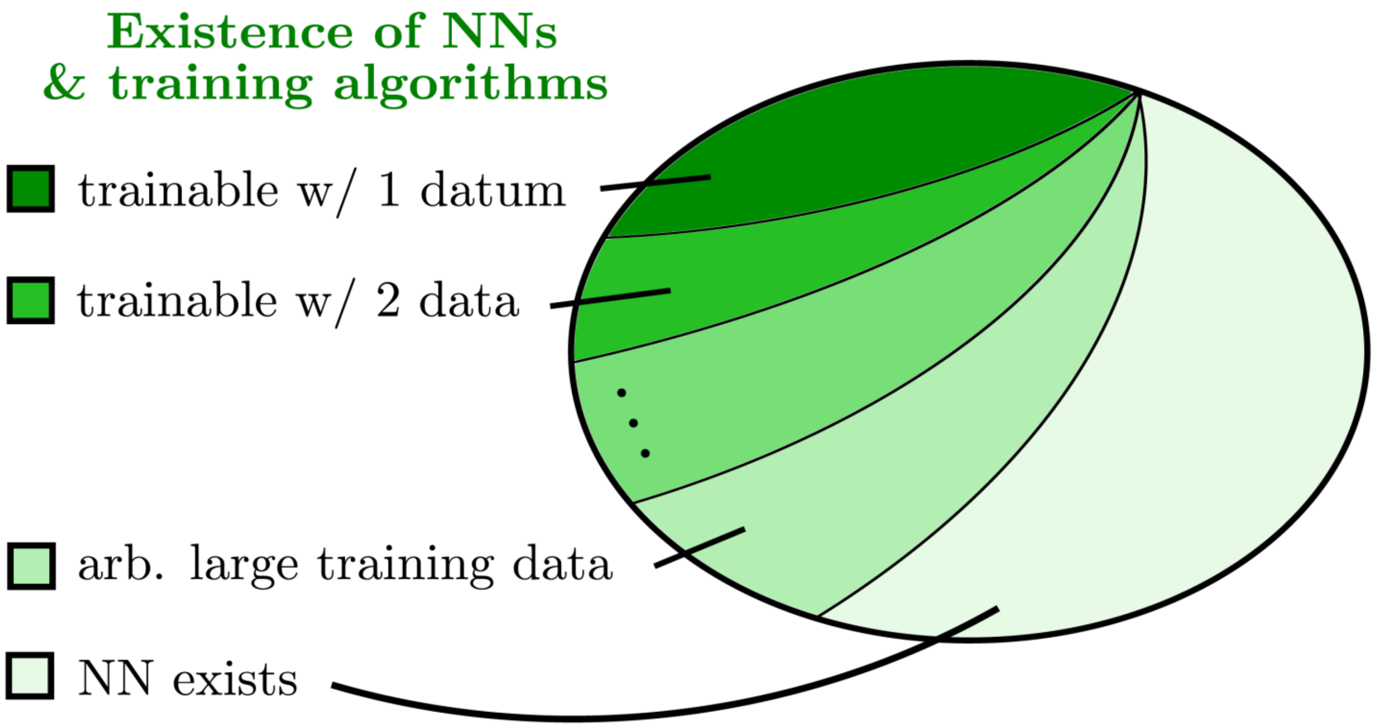
$$\mathbb{P}\left(\inf_{x^* \in \Xi(A,y)} \|\phi(y) - x^*\|_2 > 10^{-(K-1)} \text{ or size of training data to construct } \phi \text{ exceeds } M\right) > p.$$

3. $\exists$ deterministic alg. that, given a training set $\iota_{A,S} \in \Omega_{\mathcal{T}}$, produces a NN ϕ accessing at most L training samples of $\iota_{A,S}$ s.t.

$$(3) \qquad \max_{y \in S} \inf_{x^* \in \Xi(A,y)} \|\phi(y) - x^*\|_2 \leq 10^{-(K-2)} \qquad \forall (A,S) \in \Omega(K).$$

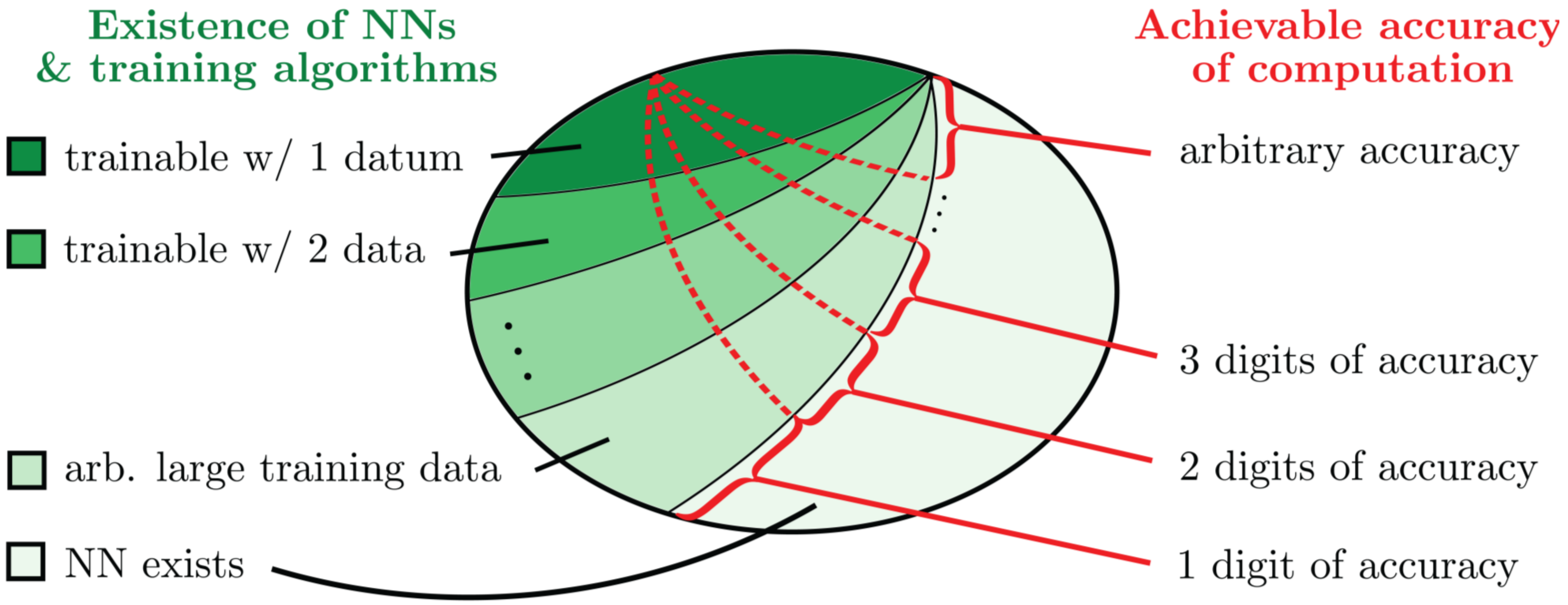
- C., Antun, Hansen, *"The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,"* **Proc. Natl. Acad. Sci. USA**.

The world of neural networks



Given a problem and conditions, where does it sit in this diagram?

The world of neural networks



Given a problem and conditions, where does it sit in this diagram?

Example counterpart theorem

Certain conditions: stable neural networks trained with exponential accuracy.
E.g., *approximate Łojasiewicz-type inequality*:

$$(1) \quad \min_{x \in \mathbb{C}^N} f(x) \quad \text{s.t.} \quad \|Ax - y\| \leq \varepsilon$$

$$\text{dist}(x, \text{solution}) \leq \alpha([f(x) - f^*] + [\|Ax - y\| - \varepsilon] + \delta)$$

Fast **I**terative **RE**started **NET**works (FIRENETs)
(unrolled primal-dual with novel restart scheme)

Theorem: Training algorithm that, under above assumption, produces *stable* neural networks φ_n of width $O(N)$, depth $O(n)$, guaranteed worst bound

$$\text{dist}(\varphi_n(y), \text{solution}) \lesssim e^{-n} + \delta$$

- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.
- C., "*WARPd: A linearly convergent first-order method for inverse problems with approximate sharpness conditions,*" **SIAM J. Imaging Sci.**, 2022.

Numerical example of GHA

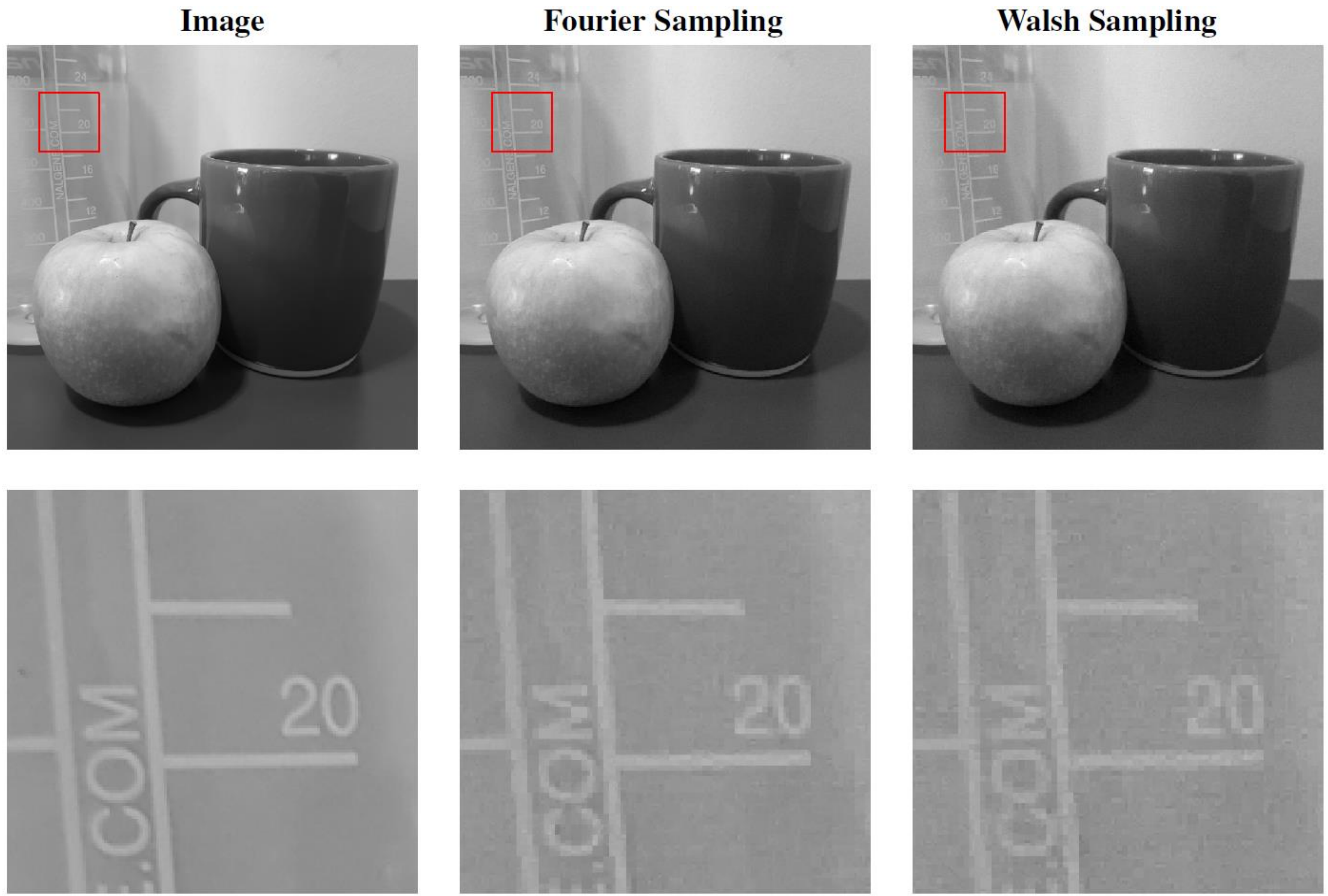
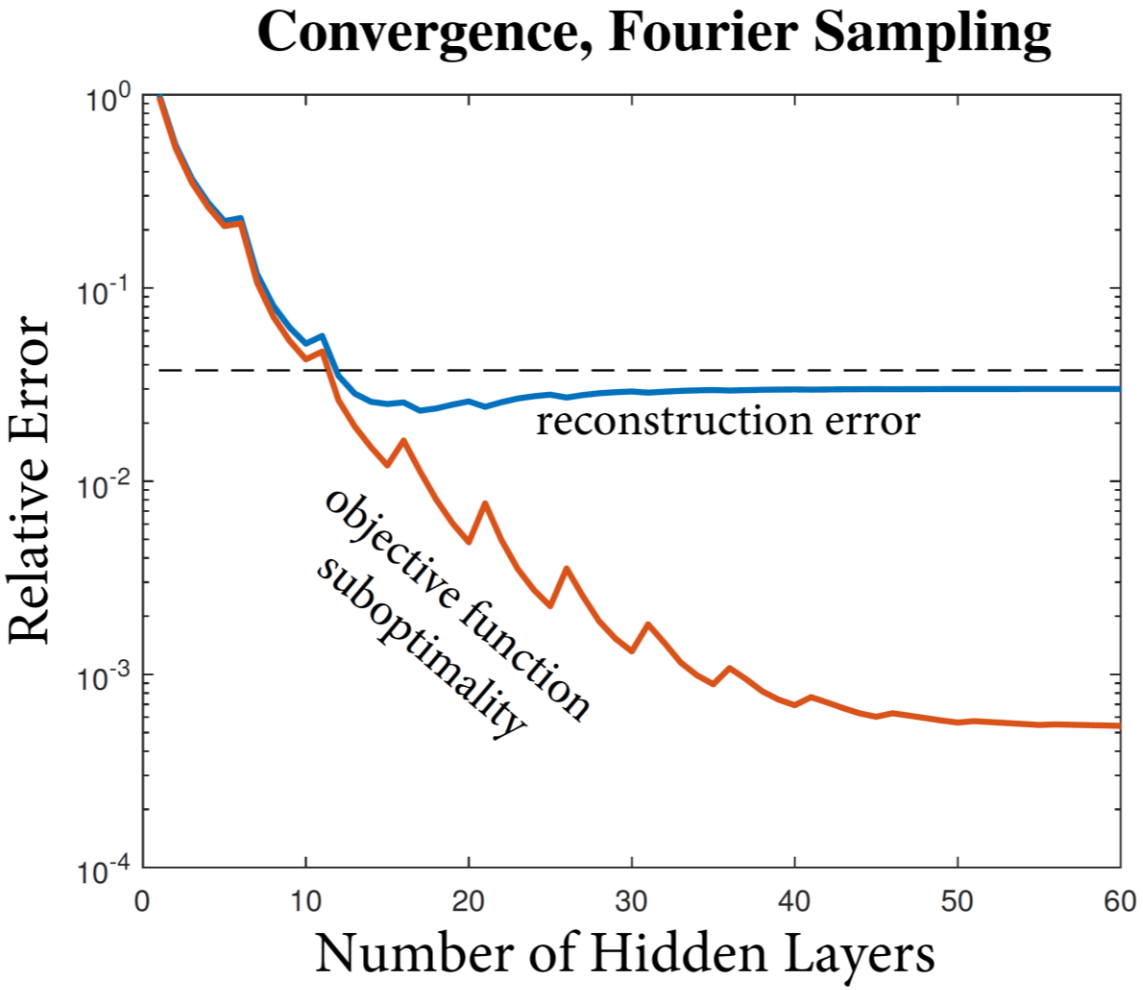
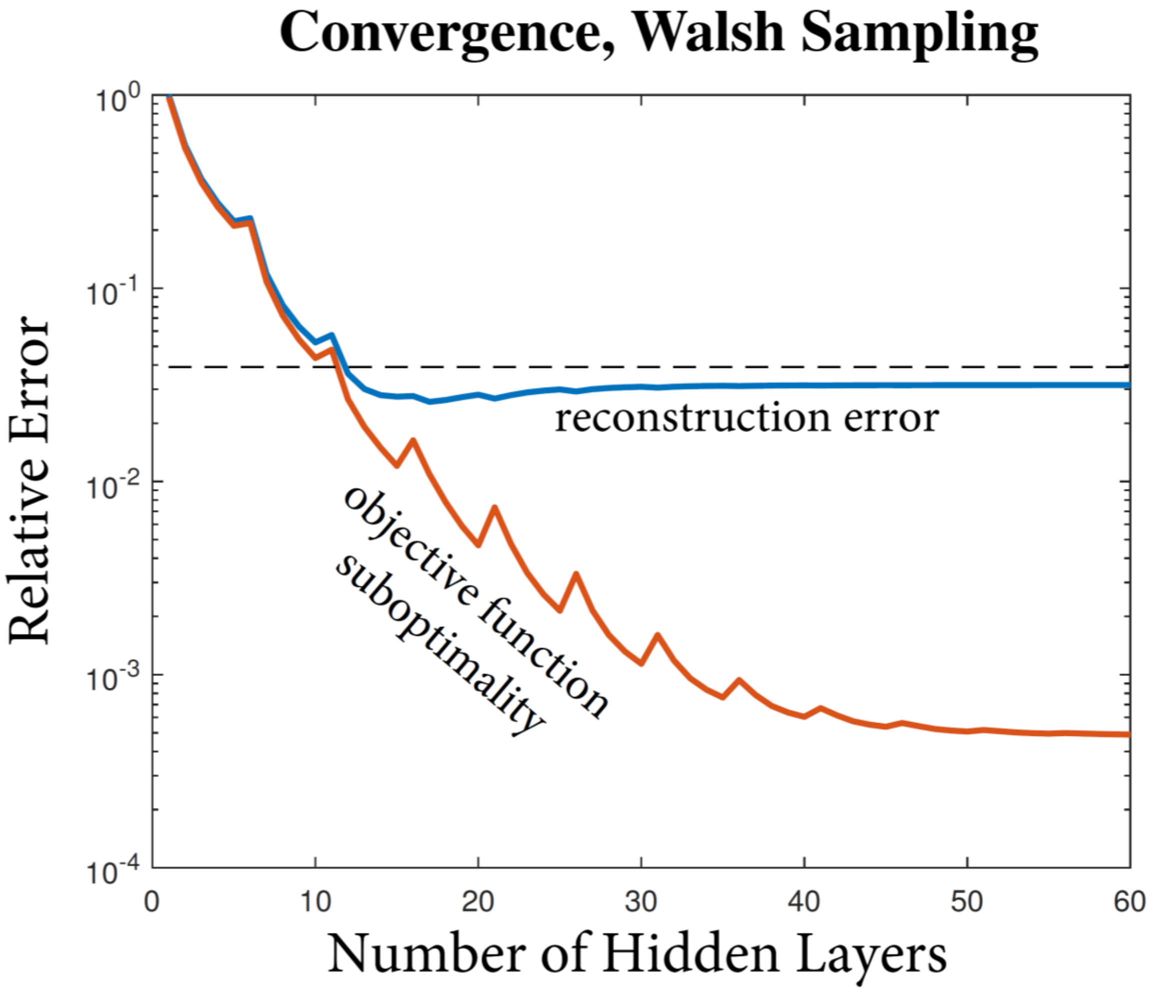


Figure: Images corrupted with 2% Gaussian noise and reconstructed using 15% sampling.

- 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**.

Numerical example of GHA

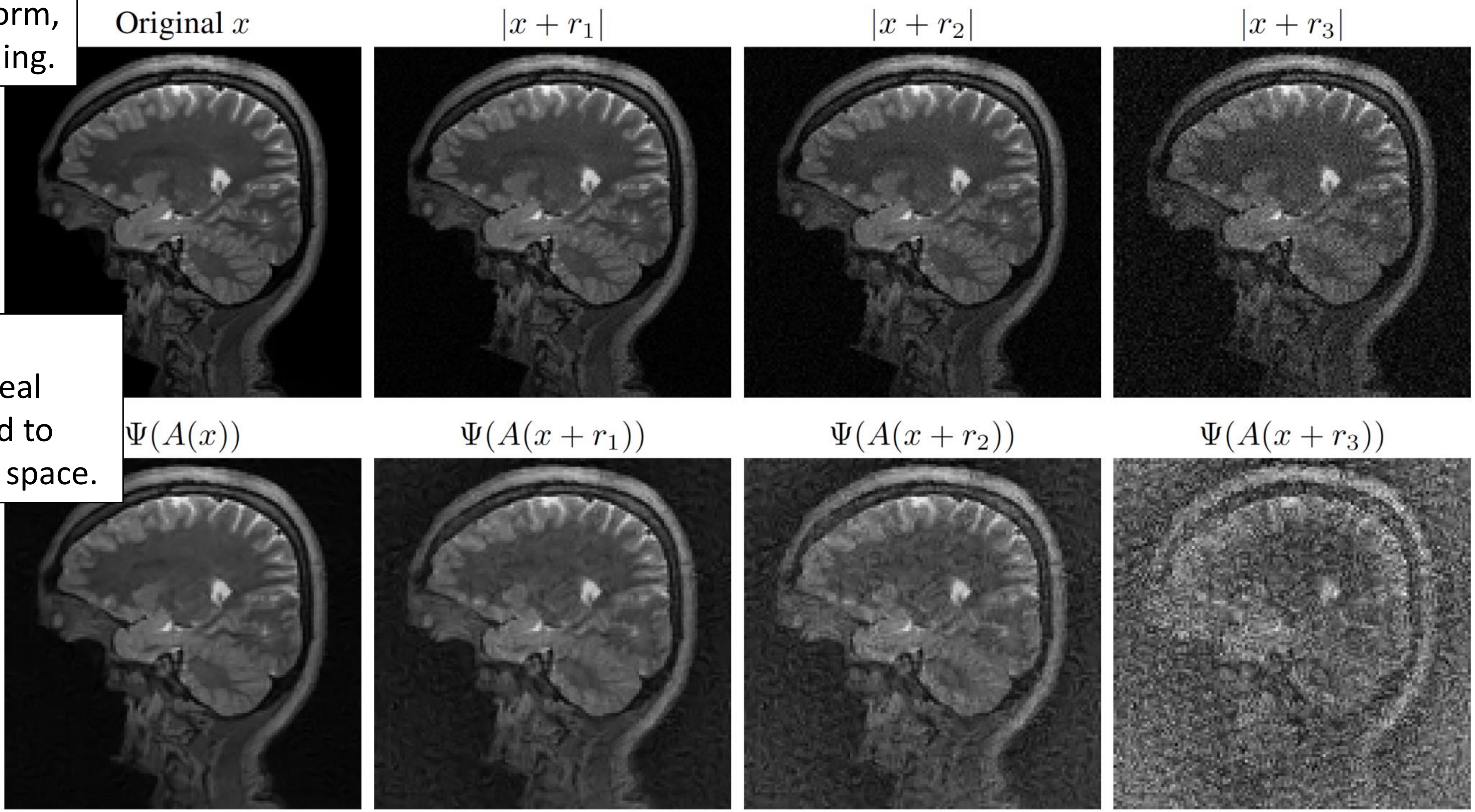




- C., Antun, Hansen, *"The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,"* **Proc. Natl. Acad. Sci. USA**.

Example of severe instability

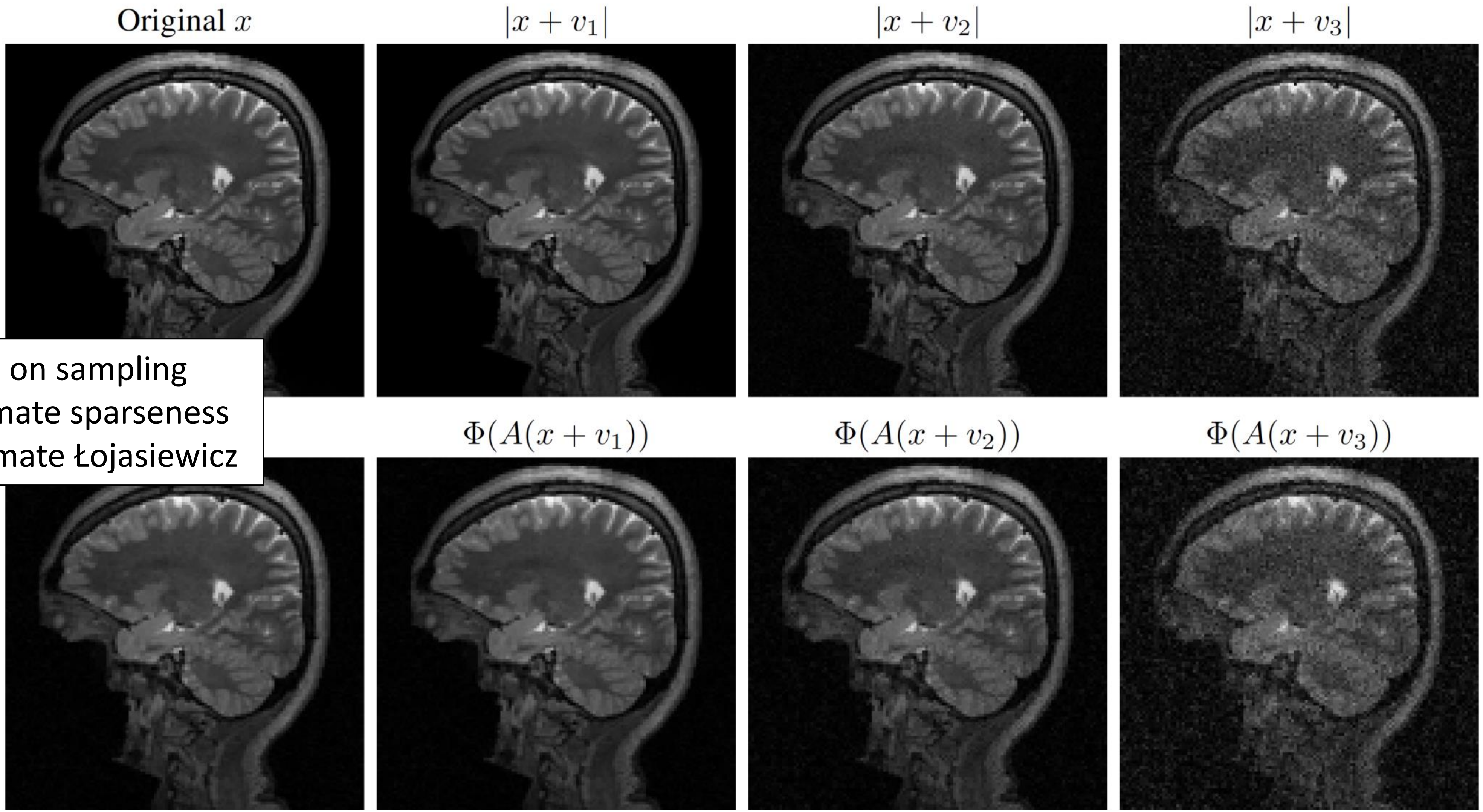
MRI: discrete 2D Fourier transform, 60% subsampling.



Perturbations computed in real space, mapped to measurement space.

- Zhu et al., *"Image reconstruction by domain-transform manifold learning,"* **Nature,** 2018.
- Antun et al., *"On instabilities of deep learning in image reconstruction and the potential costs of AI,"* **PNAS**, 2020.

FIRENET: provably stable (even to adversarial examples) and accurate



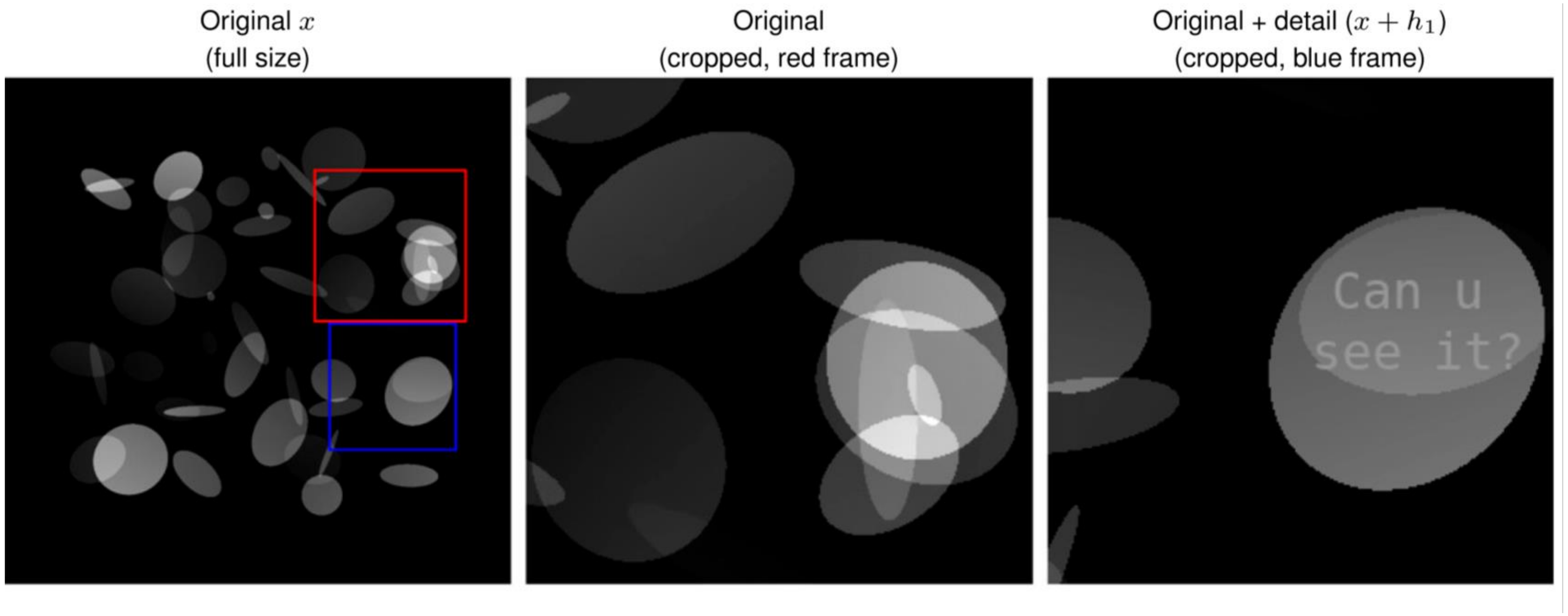
Assumptions on sampling and approximate sparseness give approximate Łojasiewicz

- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

Key pillars: stability and accuracy

MRI: discrete 2D Fourier transform, 15% subsampling.

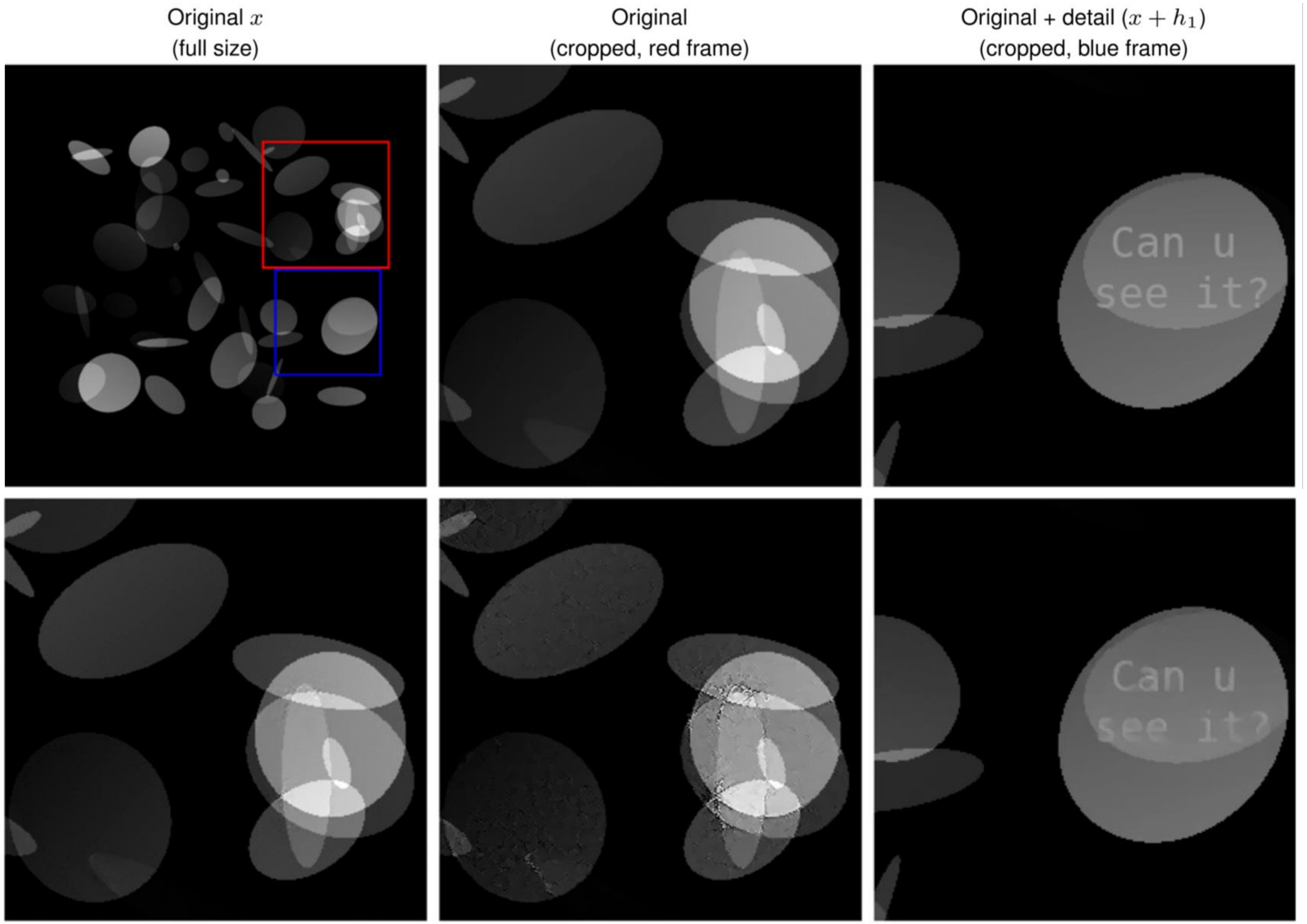
All networks trained on 5000 images of ellipses



- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

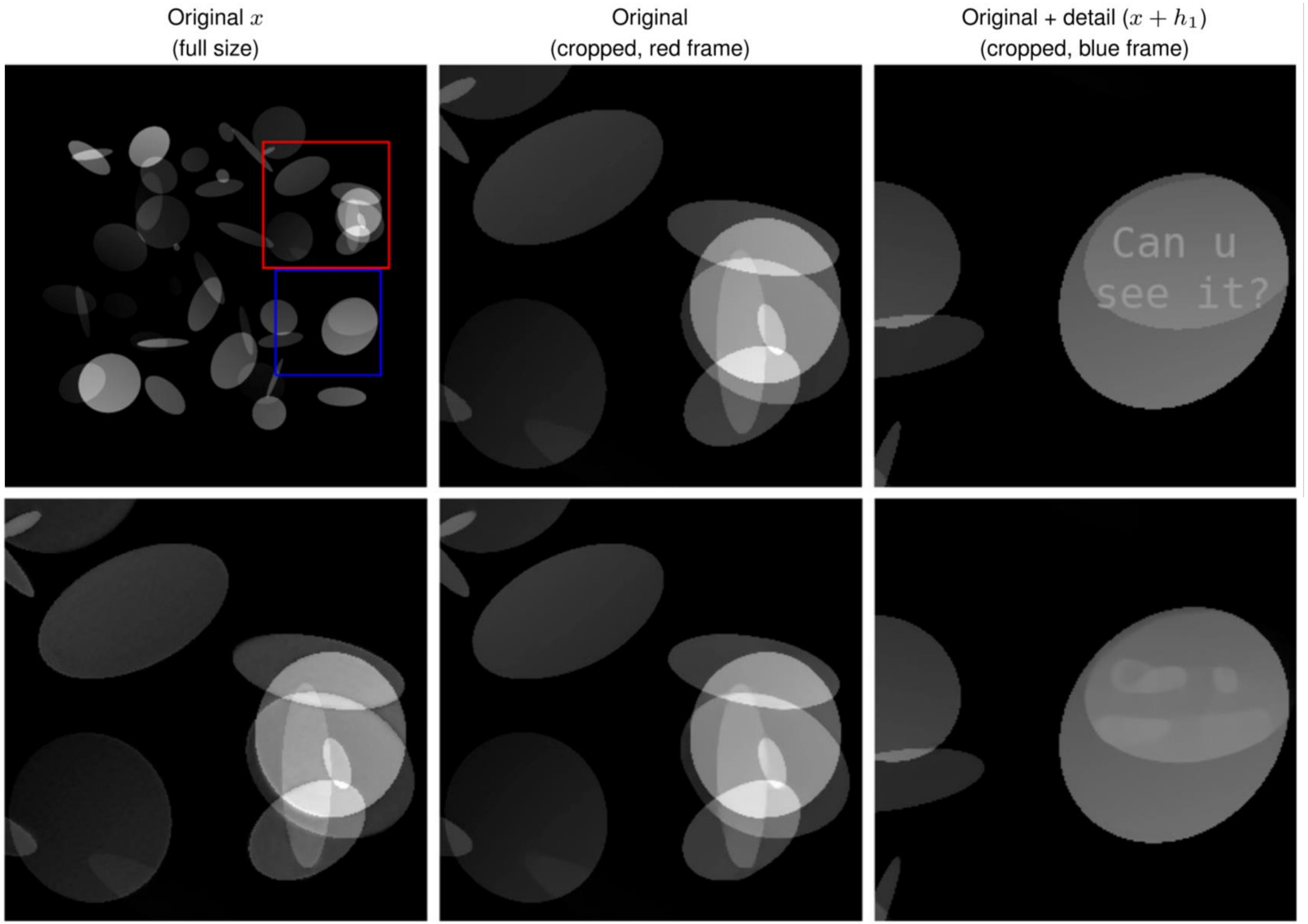
U-Net with no noise: accurate but unstable

U-Net: standard neural network architecture for imaging. Approx 4 million parameters.



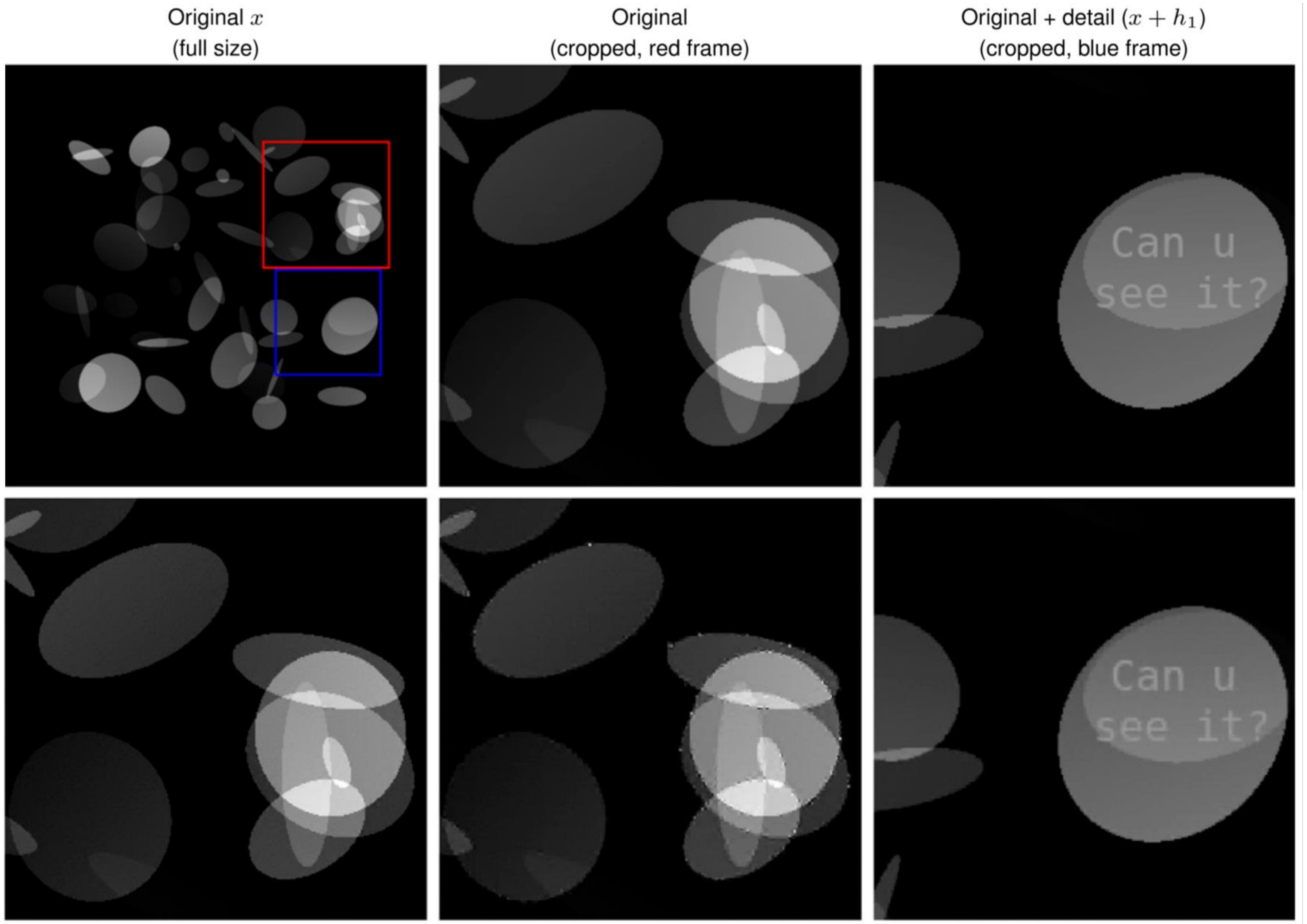
- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

U-Net with noise: stable but inaccurate



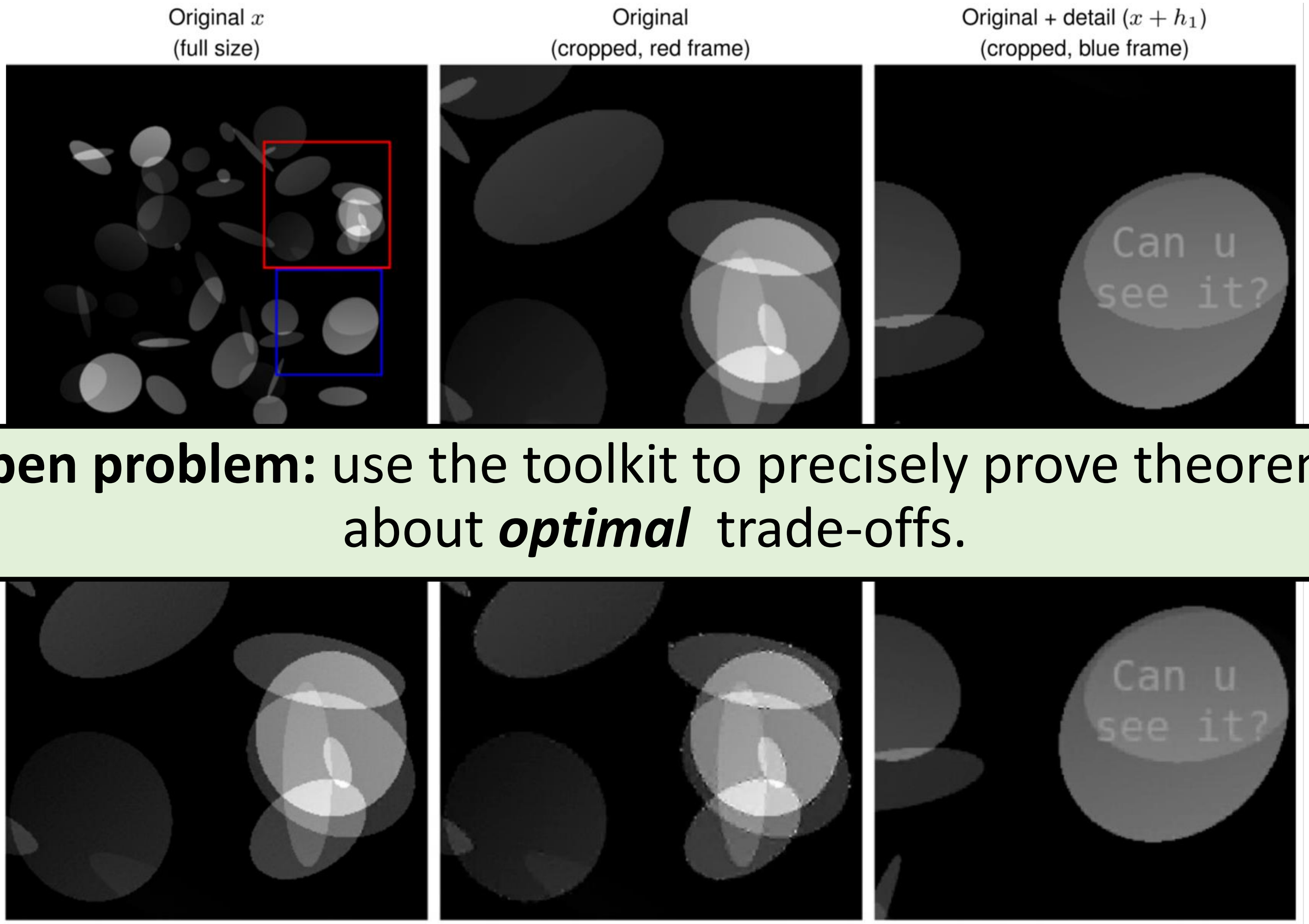
- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

FIRENET: balances stability and accuracy?



- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

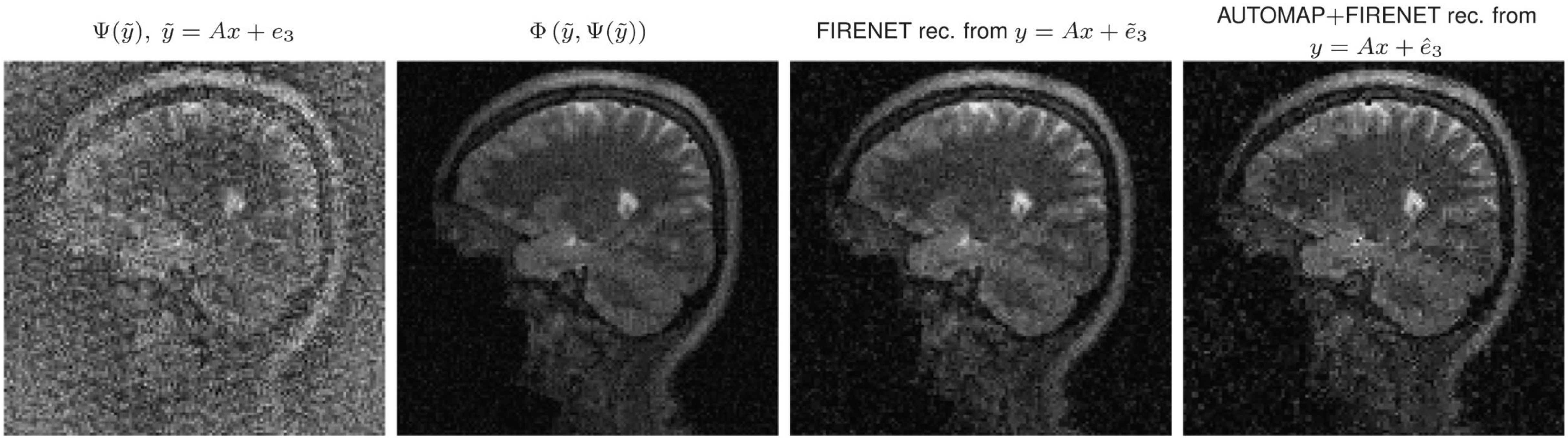
FIRENET: balances stability and accuracy?



Open problem: use the toolkit to precisely prove theorems about ***optimal*** trade-offs.

- C., Antun, Hansen, "*The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem,*" **PNAS**, 2022.

Stabilising unstable neural networks

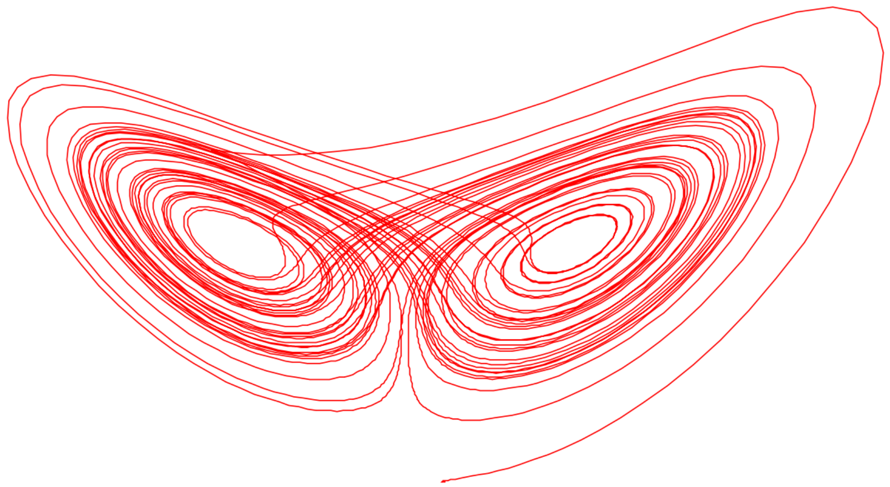


Data-driven dynamical systems

- 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 $\{x^{(m)}, y^{(m)} = F(x^{(m)})\}_{m=1}^{M}$
  - E.g., **data from** trajectories, experimental measurements, simulations, …
  - E.g., **used for** forecasting, control, design, understanding, …
- **Applications:** chemistry, climatology, electronics, epidemiology, finance, fluids, molecular dynamics, neuroscience, plasmas, robotics, video processing, …



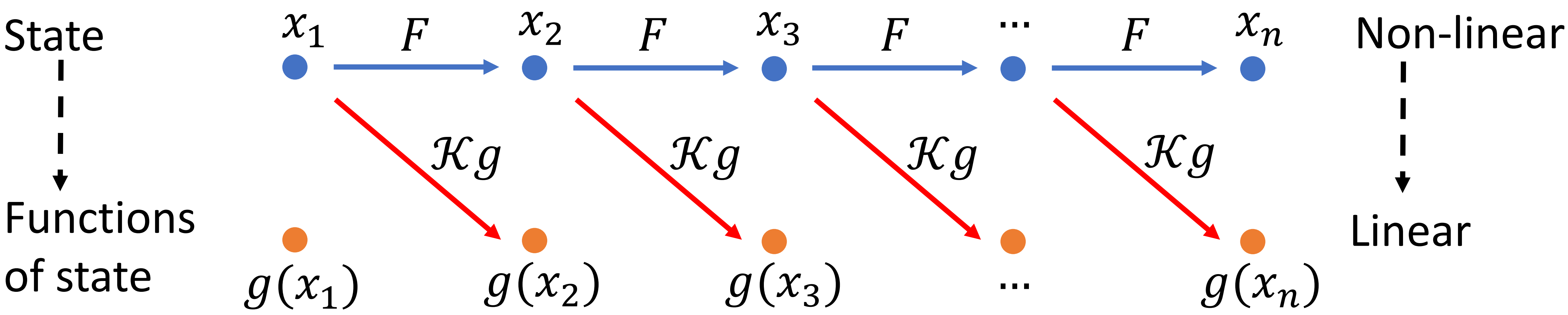
Can we develop verified methods?

Operator viewpoint

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

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

- $\mathcal{K}$ 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,"* **Proceedings of the National Academy of Sciences**, 1931.
- Koopman, v. Neumann, *"Dynamical systems of continuous spectra,"* **Proceedings of the National Academy of Sciences**, 1932.

Build the matrix

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

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

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

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

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

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

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

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

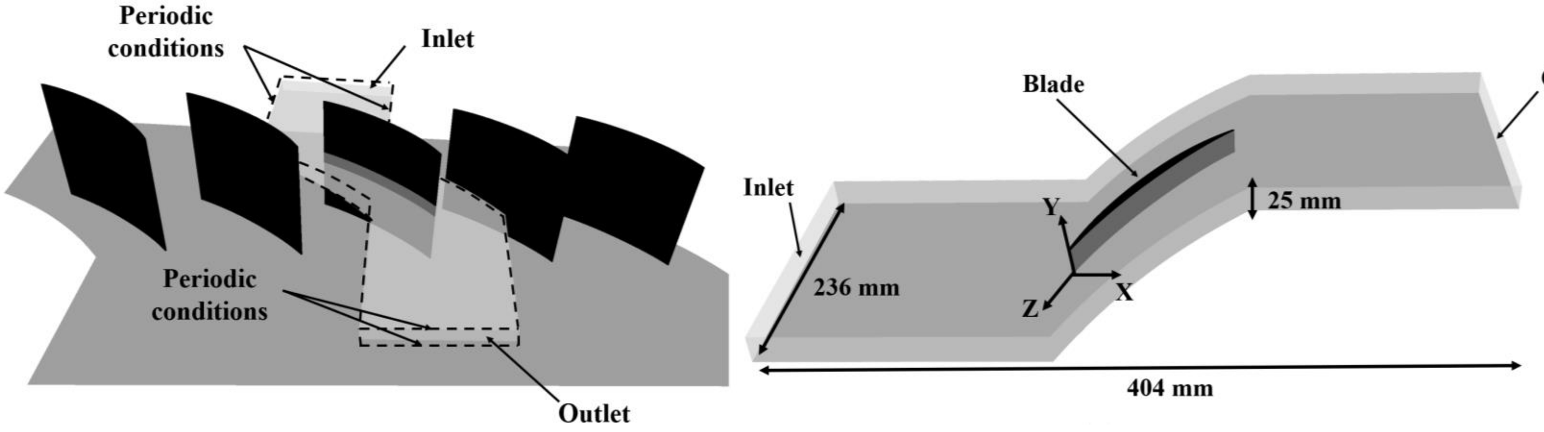
$$\langle \mathcal{K}\psi_k, \mathcal{K}\psi_j \rangle \approx \sum_{m=1}^{M} w_m \overline{\psi_j(y^{(m)})}\, \psi_k(y^{(m)}) = [\underbrace{\Psi_Y{}^* W \Psi_Y}_{K_2}]_{jk}$$

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,"* **Communications on Pure and Applied Mathematics**, under review.

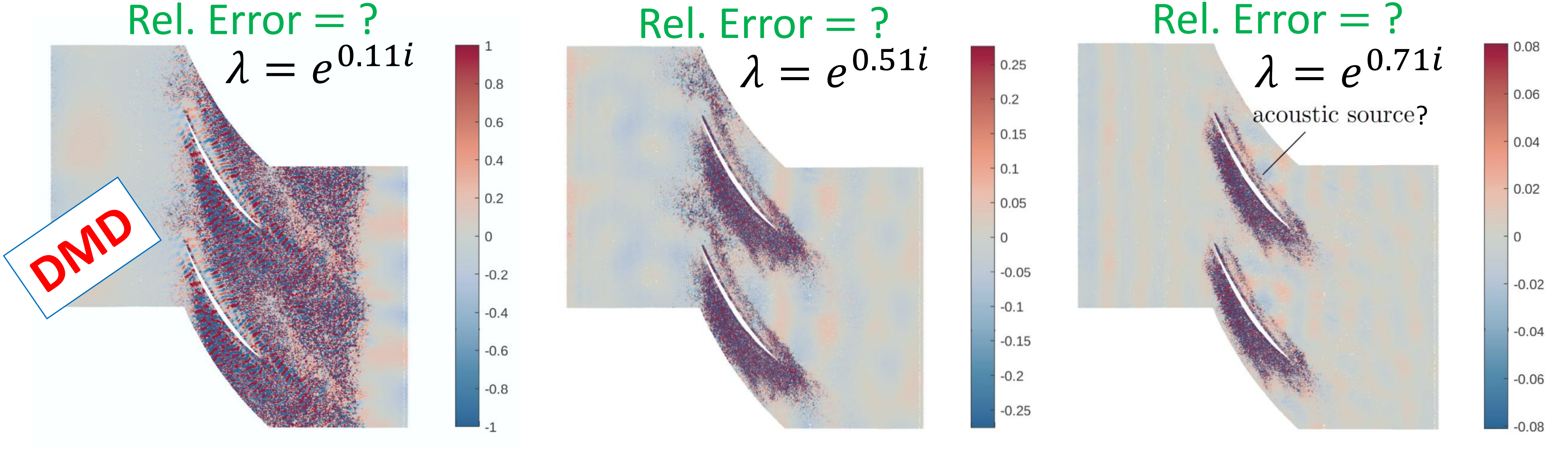
- Code: https://github.com/MColbrook/Residual-Dynamic-Mode-Decomposition

Example: Trustworthy computation for large d



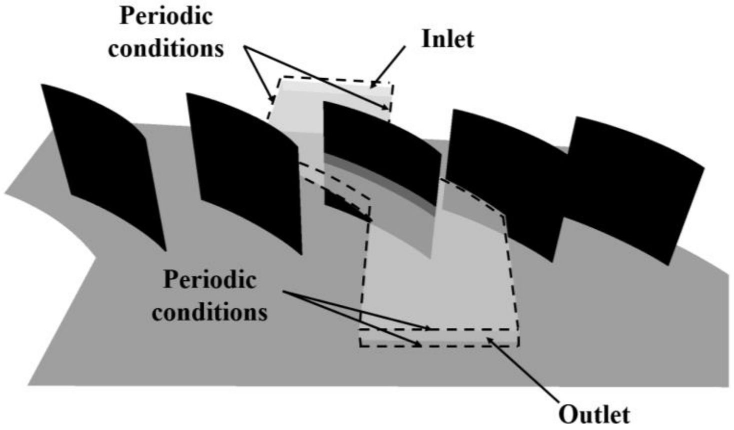
- Reynolds number $\approx 3.9 \times 10^5$
- Ambient dimension (d) $\approx 300{,}000$ (number of measurement points)

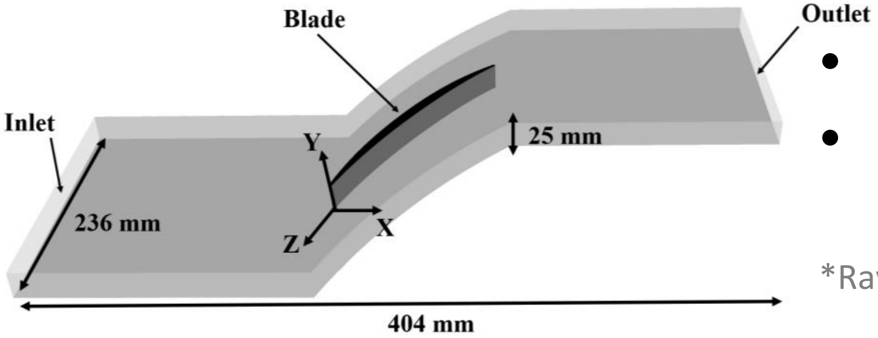
*Raw measurements provided by Stephane Moreau (Sherbrooke)



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

Example: Trustworthy computation for large d

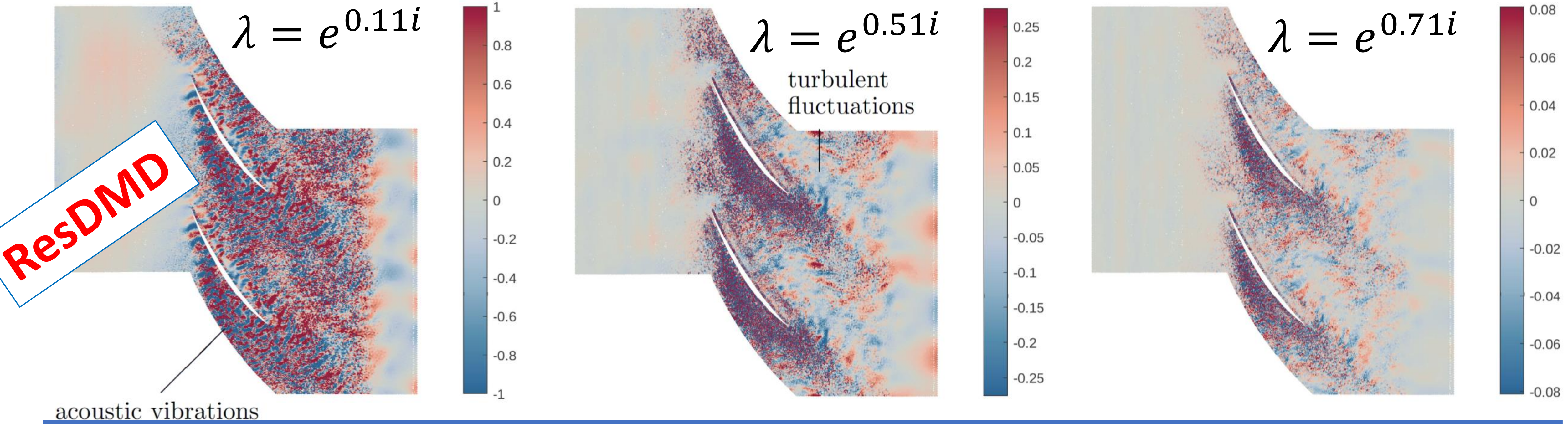




- Reynolds number $\approx 3.9 \times 10^5$
- Ambient dimension (d) $\approx 300{,}000$ (number of measurement points)

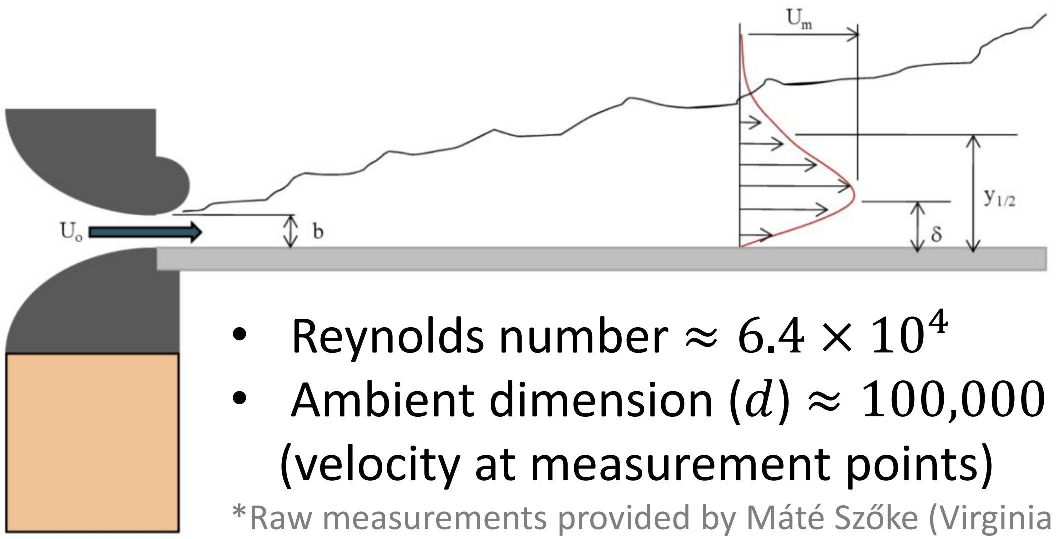
*Raw measurements provided by Stephane Moreau (Sherbrooke)

Rel. Error ≤ 0.0054 Rel. Error ≤ 0.0128 Rel. Error ≤ 0.0196



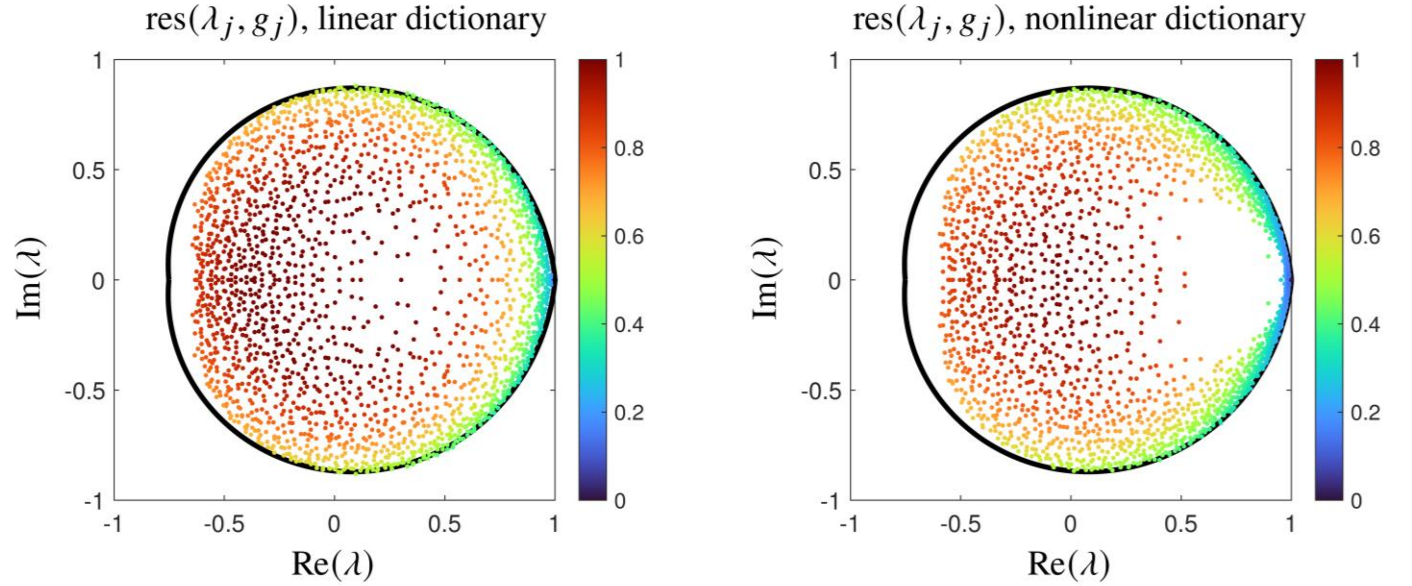
- C., Townsend, *"Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems,"* preprint.

Example: Verify the dictionary

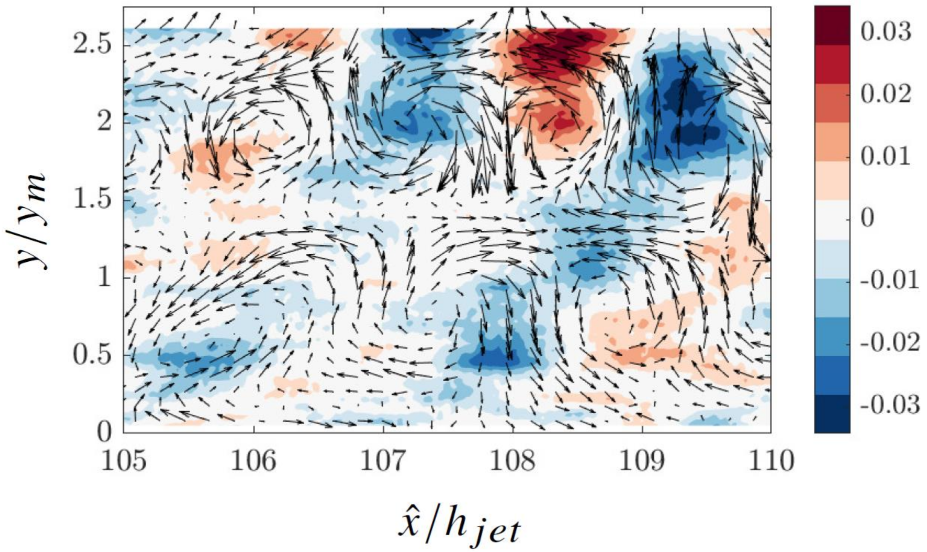


- Reynolds number $\approx 6.4 \times 10^4$
- Ambient dimension (d) $\approx 100{,}000$ (velocity at measurement points)

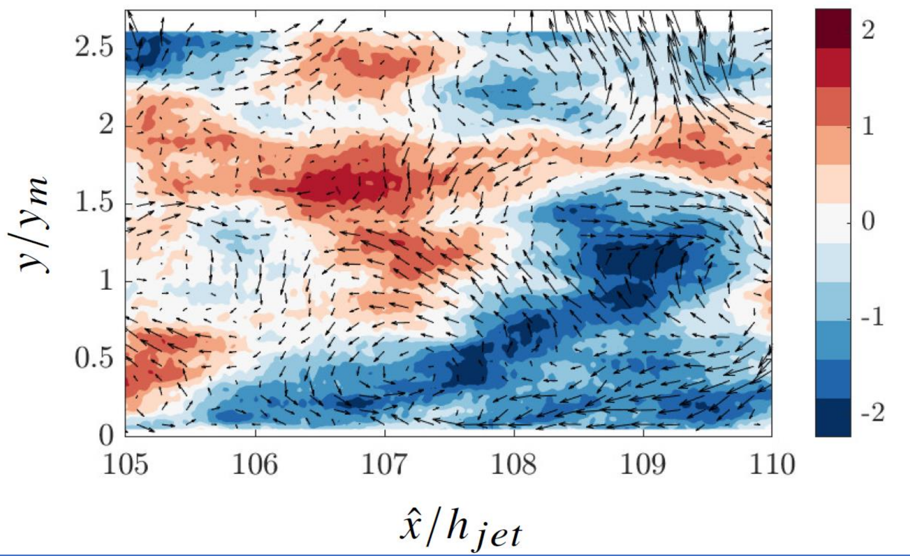
*Raw measurements provided by Máté Szőke (Virginia Te



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

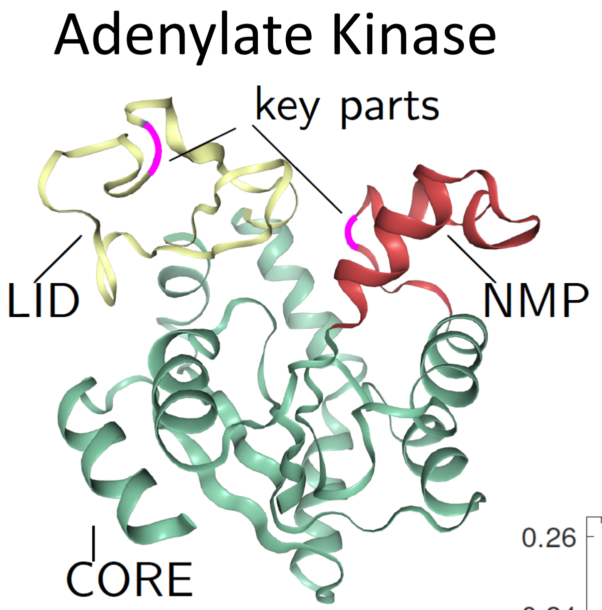


$\lambda = 0.8948 + 0.1065i$, error ≤ 0.1105



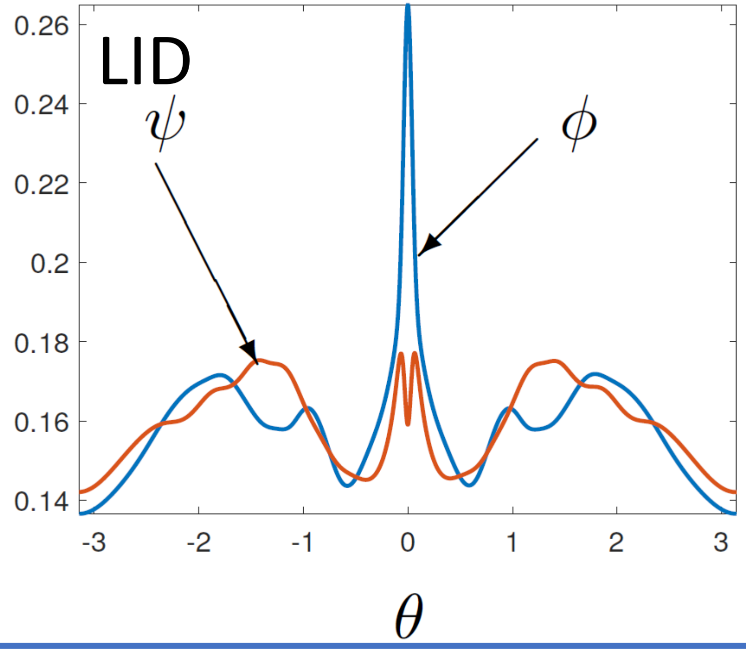
- C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition,*" **J. Fluid Mech.**, under minor rev.

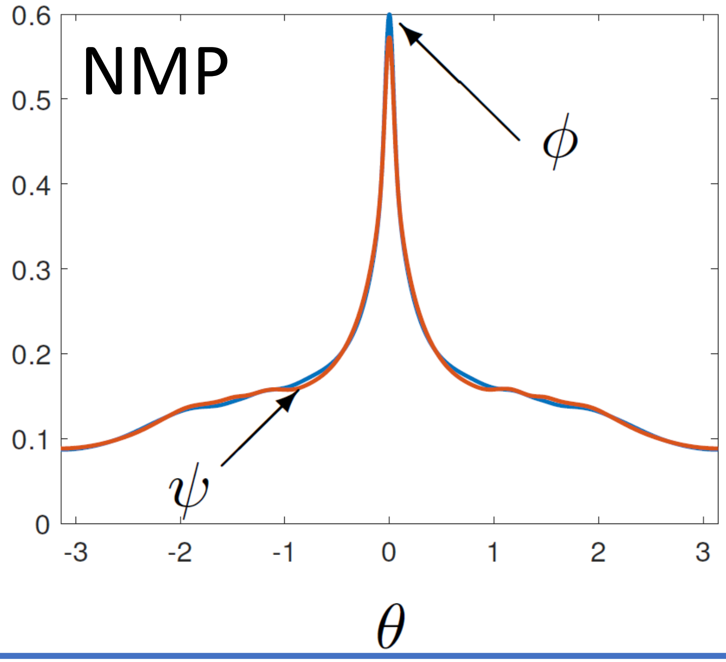
Example: molecular dynamics (Adenylate Kinase)



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

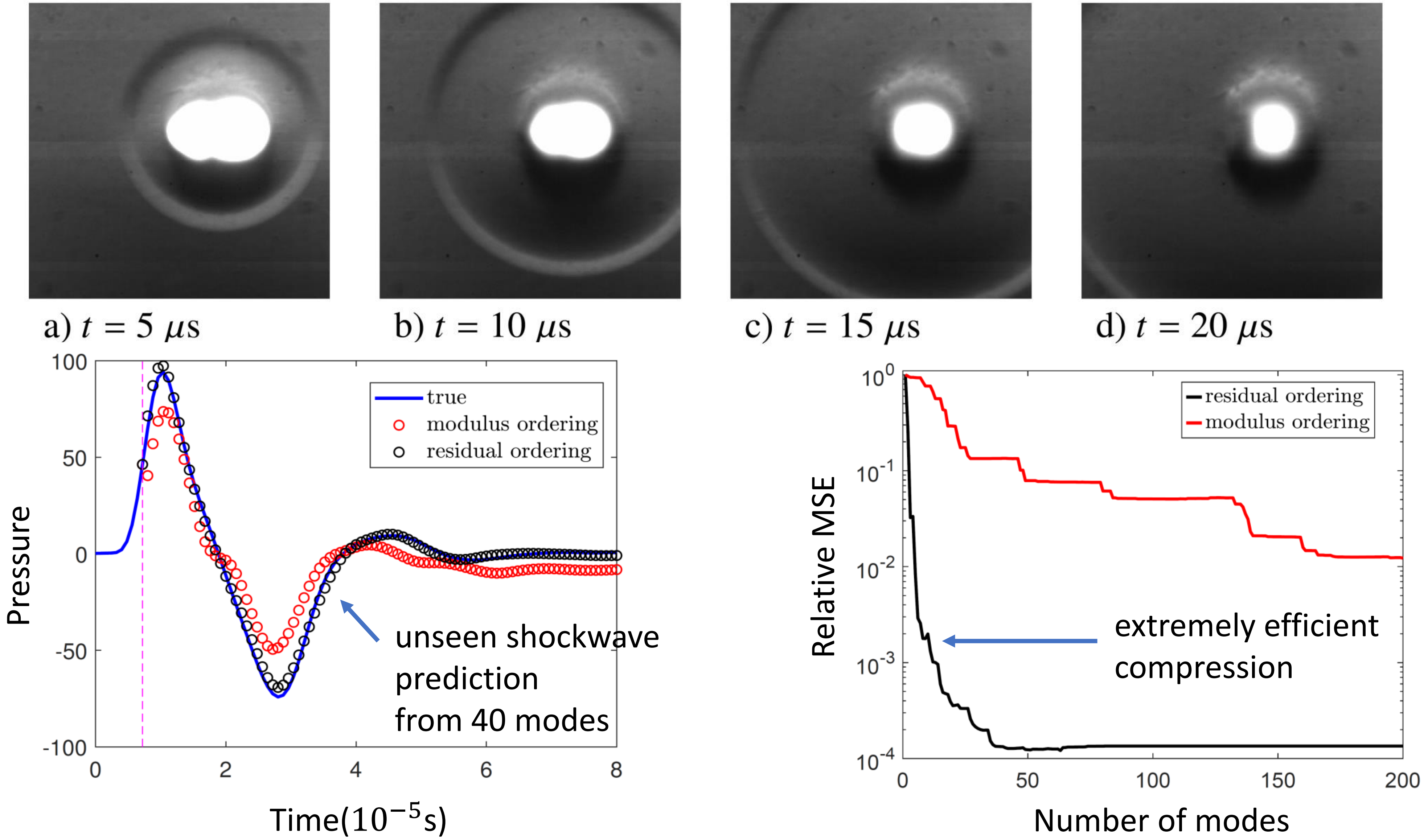
*Dataset: www.mdanalysis.org/MDAnalysisData/adk_equilibrium.html





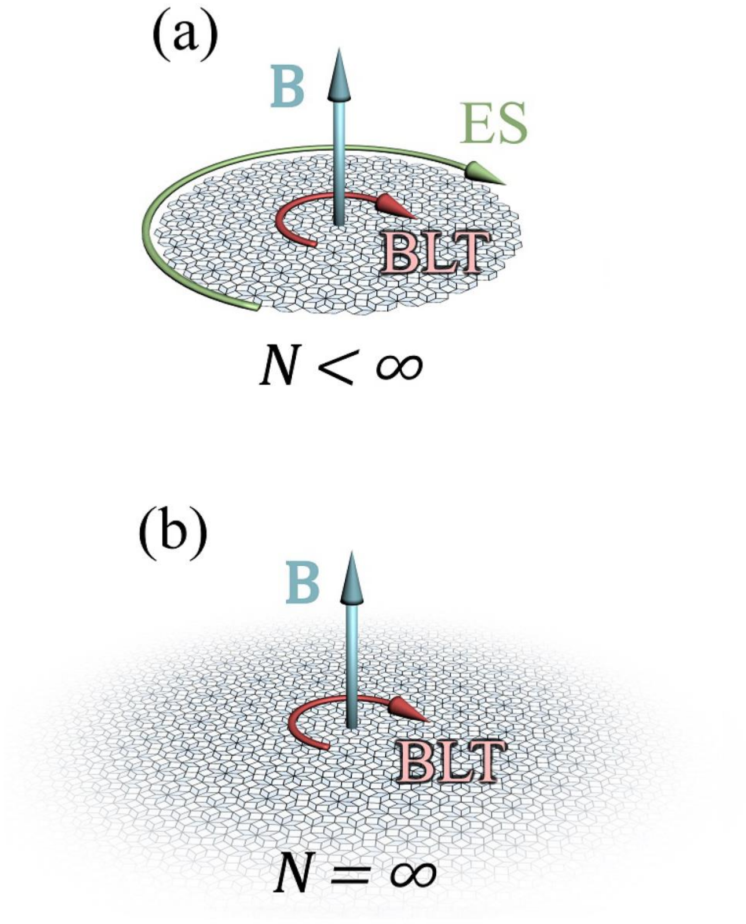
- C., Townsend, *"Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems,"* preprint.

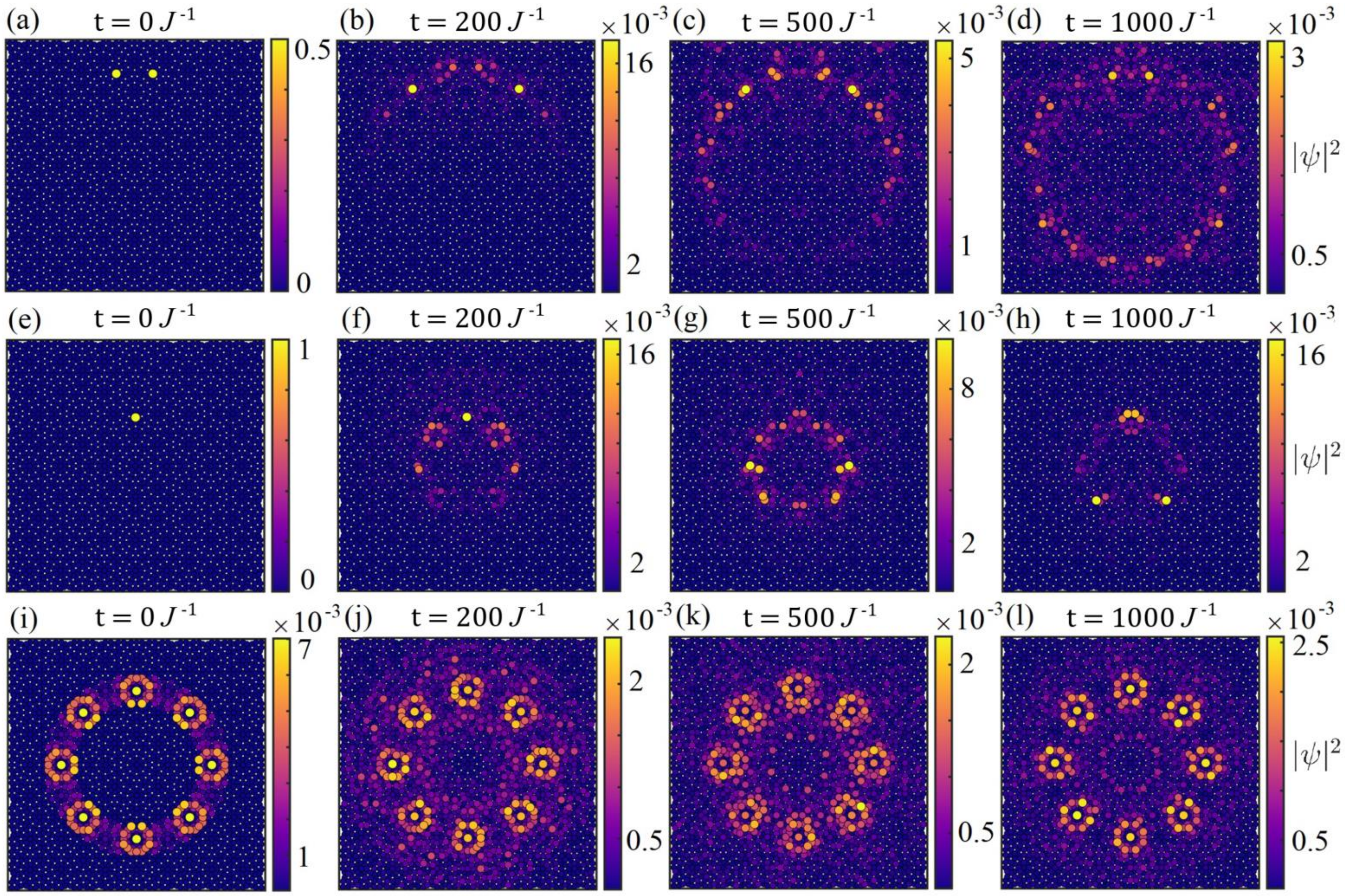
Example: Trustworthy Koopman mode decomposition



- C., Ayton, Szőke, "*Residual Dynamic Mode Decomposition,*" **J. Fluid Mech.,** under minor rev.

Bulk localised transport





- Johnstone, C., Nielsen, Öhberg, Duncan, *"Bulk Localised Transport States in Infinite and Finite Quasicrystals via Magnetic Aperiodicity,"* **Phys. Rev. B**, 2022.

PNAS
RESEARCH ARTICLE | APPLIED
The difficul
accurate ne
deep learni
Matthew J. Colbrook
March 16, 2022 | 119 (12)
Significance
Instability is the Achi
training algorithms f
ones. This foundatic
century on the limits
demonstrate limitati
NNs. Despite numer
only in specific cases
classification theory
suitable conditions–
number of hidden la
IEEE.ORG IEEE XPLORE DIGITAL LIBRARY IEEE STANDARDS MORE SITES
IEEE Spectrum FOR THE TECHNOLOGY INSIDER
NEWS | ARTIFICIAL INTELLIGENCE
Some AI Systems May Be Impossible to Compute
New research suggests there are limitations to wh ep
neural networks can do
BY CHARLES Q. CHOI | 30 MAR 2022
GETTY IMAGES/IEEE SPECTRUM
For engineers
For scientists
For mathematicians
Eurek
HOME NEWS RELEA
NEWS RELEASE 17-MAR-2022
Mathematical parad
the limits of AI
Peer-Reviewed Publication
UNIVERSITY OF CAMBRIDGE
Newsjournal of the Society for Industrial and Applied Mathematics
sinews.siam.org
siam news
Volume 55/ Issue 4
May 2022
Privacy with Synthetic Data
Differential Privacy Made Simple(r)
Original Data
Differential Privacy Preprocessing
Vine Copula Model
Synthetic Data
Extract Statistics
Apply Multiple Imputation Model
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 Limits of AI: Smale's 18th Problem
Hallucinations in image reconstruction
Original image
AI reconstruction
Instabilities in medical diagnosis
Original Mole
Perturbed Mole
Benign
Malignant
Model confidence