Sharp Spectral Rates for Koopman Operator Learning

Vladimir Kostic

Karim Lounici

Pietro Novelli

Massimiliano Pontil

iit CSML
ISTITUTO ITALIANO DI TECNOLOGIA

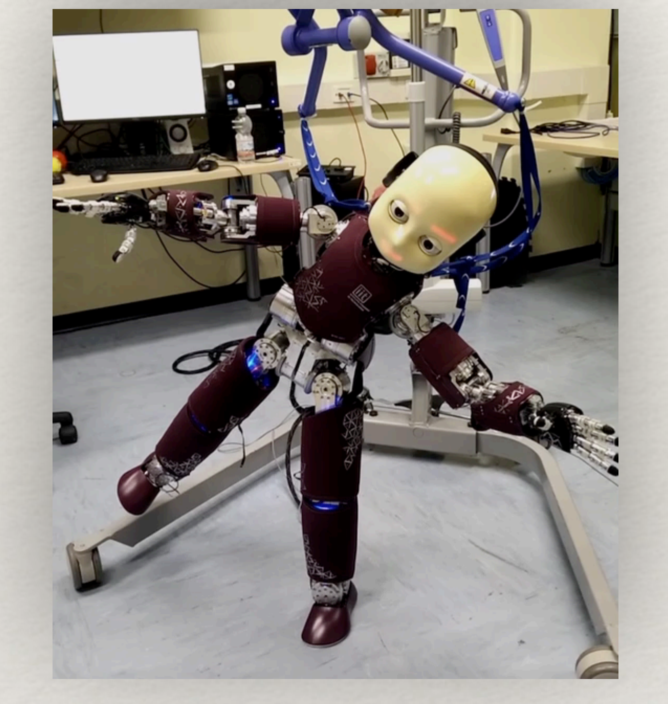
l'X ÉCOLE POLYTECHNIQUE

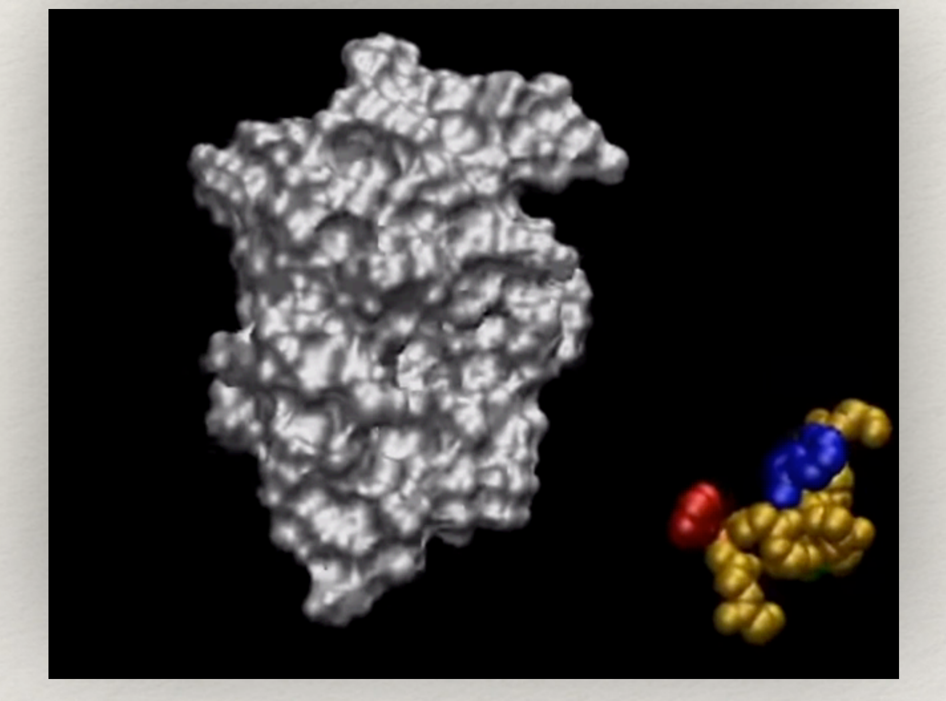
UCL

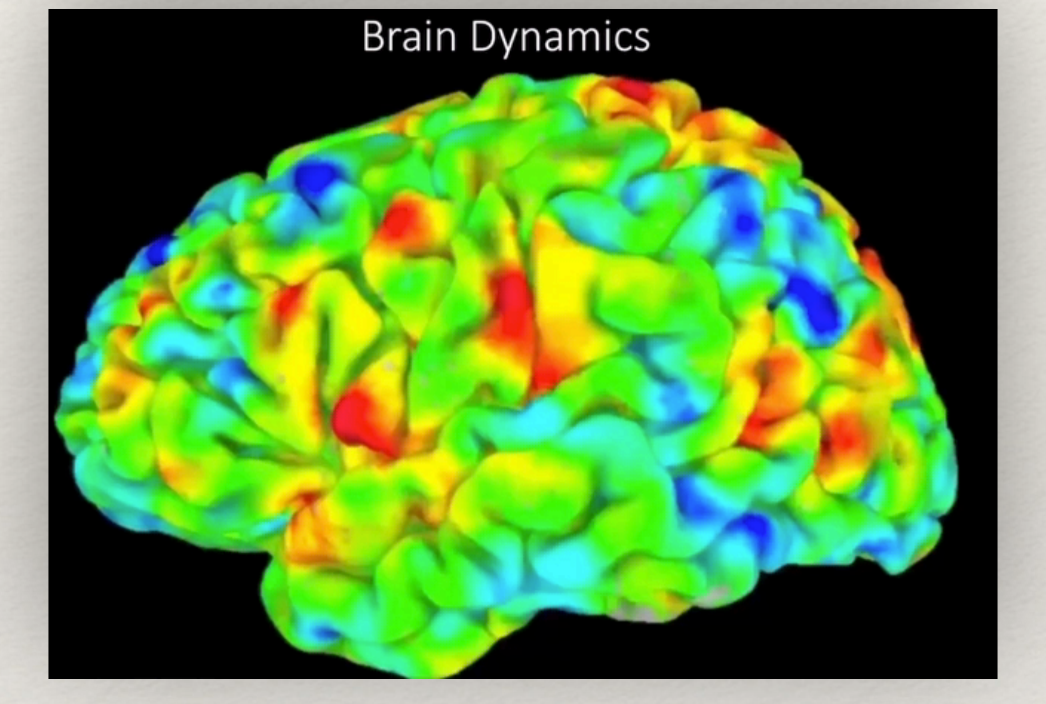
DMI DEPARTMAN ZA MATEMATIKU I INFORMATIKU

Dynamical Systems & ML

DS are backbone mathematical models of temporally evolving phenomena







Paradigm shift in Sci & Eng:

- Classical approach: ODE/PDE/SDE models + parameter fitting
- ML approach: Can we build dynamical models purely from the observed data?

This is remarkably elegant via transfer operators theory!

The Koopman/Transfer Operator

- Stochastic process $\boxed{(X_t)_{t\geq 0} \subseteq \mathcal{X}}$
- We focus on discrete time DS, i.e. time homogenous Markov process:

$$\boxed{\mathbb{P}\left[X_{t+1} \,|\, X_1, \ldots, X_t\right] = \mathbb{P}\left[X_{t+1} \,|\, X_t\right] \quad \text{independent of } t}$$

- Existence of the invariant measure π $\boxed{X_0 \sim \pi \quad \Longrightarrow \quad X_t \sim \pi, t \in \mathbb{N}}$
- The forward **transfer operator** evolves observables:

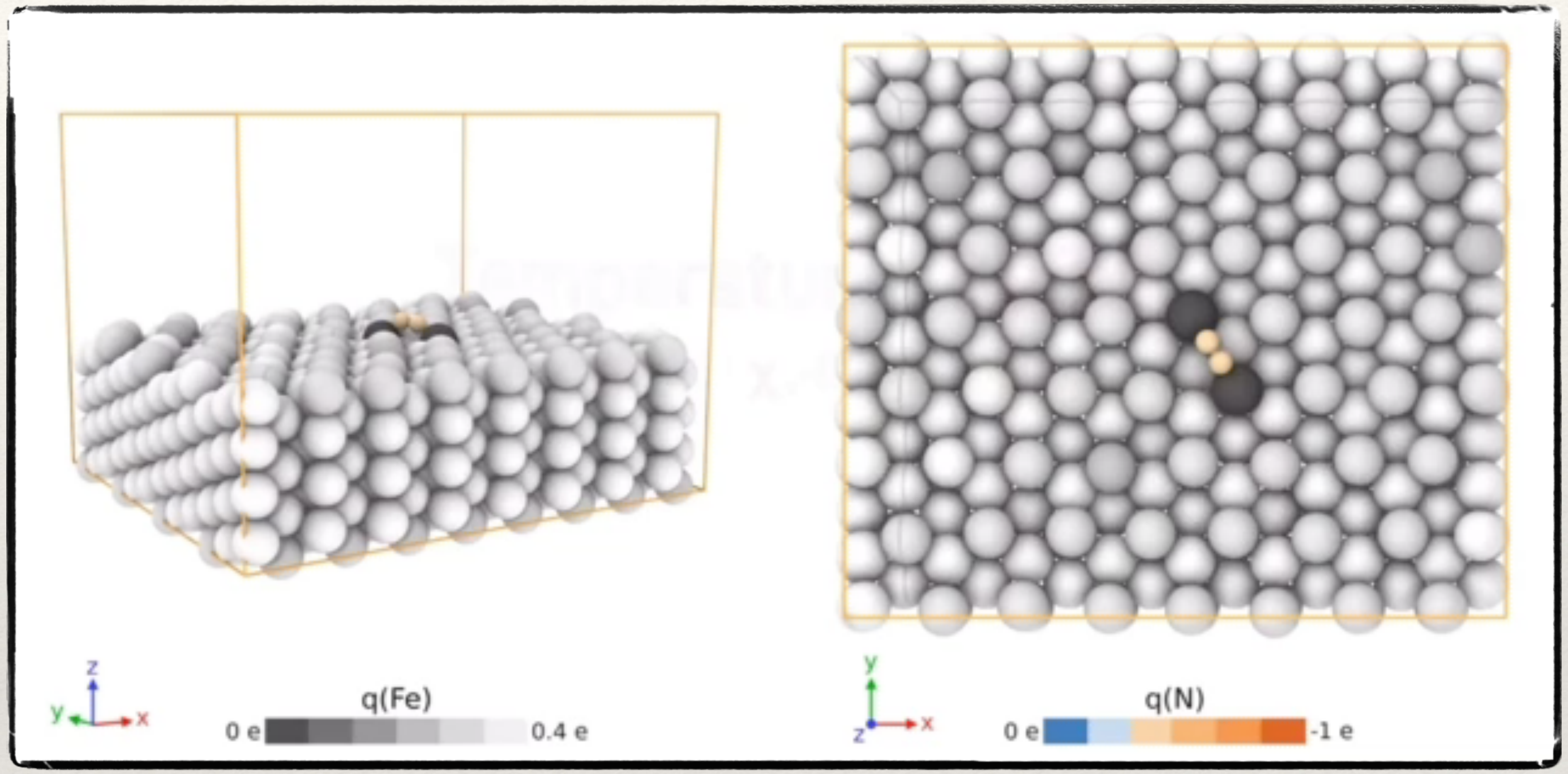
$$\boxed{A_\pi : L^2_\pi(\mathcal{X}) \to L^2_\pi(\mathcal{X}) \quad (A_\pi f)(x) = \mathbb{E}[\, f(X_{t+1}) \mid X_t = x\,]}$$

Spectral decomposition

The (overdamped) Langevin equation when discretised $X_{t+1} = F(X_t) + \text{noise}_t$

$$\underbrace{A_\pi = A_\pi^*}_{\text{compact}} \implies A_\pi = \sum_{i=1}^{\infty} \mu_i f_i \otimes f_i$$

where $A_\pi f_i = \mu_i f_i$ i.e. scalars μ_i and functions f_i are the eigenvalues and eigenfunctions



Source: *youtube.com/@luigi.bonati*

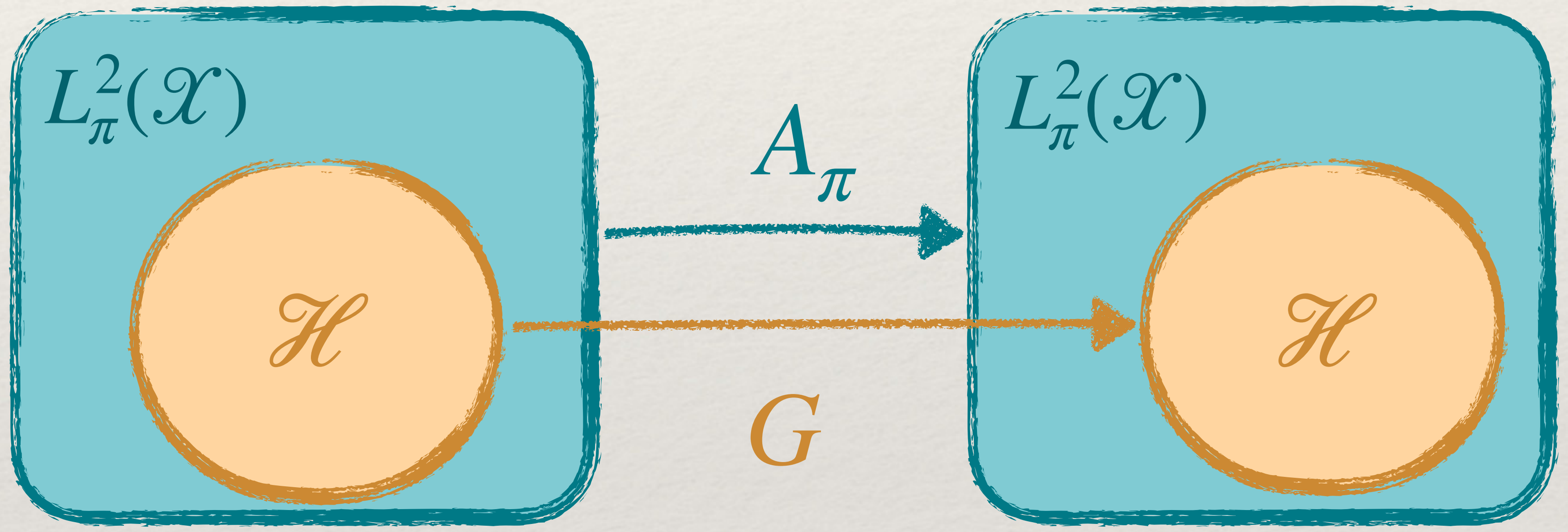
$$\mathbb{E}[\, f(X_t) \,|\, X_0 = x \,] = (A_\pi^t f)(x) = \sum_i \boxed{\mu_i^t} \boxed{f_i(x) \langle f_i, f \rangle}$$

the expectation of an observable is disentangled into **temporal** and **static** components

Learning the operator and its spectra

- Since we don't know $L^2_\pi(\mathcal{X})$ we restrict A_π to a chosen RKHS $\mathcal{H}$ and look for an operator $G : \mathcal{H} \to \mathcal{H}$ such that $A_\pi \langle w, \phi(\,\cdot\,)\rangle \approx \langle Gw, \phi(\,\cdot\,)\rangle$, that is

$$\boxed{\mathcal{R}(G) = \mathbb{E}_{X_t \sim \pi} \|\phi(X_{t+1}) - G^* \phi(X_t)\|^2}$$



$$G\psi_i = \lambda_i \psi_i \;\;\Rightarrow\;\; \|A_\pi \psi_i - \lambda_i \psi_i\|_{L^2_\pi(\mathcal{X})} \leq \underbrace{\|A_\pi|_{\mathcal{H}} - G\|_{\mathcal{H} \to L^2_\pi(\mathcal{X})}}_{\text{operator norm error } \mathcal{E}(G)} \;\; \underbrace{\frac{\|\psi_i\|_{\mathcal{H}}}{\|\psi_i\|_{L^2_\pi(\mathcal{X})}}}_{\eta(\psi_i) \text{ metric distorsion}}$$

Learning the operator and its spectra

- Given an iid sample $(x_i, y_i)_{i=1}^n$ learn $\hat{G} : \mathscr{H} \to \mathscr{H}$ via the empirical risk:

$$\boxed{\hat{\mathscr{R}}(\hat{G}) = \sum_{i=1}^n \|\phi(y_i) - \hat{G}^* \phi(x_i)\|^2 + \gamma \|\hat{G}\|_{\mathrm{HS}}^2}$$

- We considered three estimators:
  - Kernel ridge regression minimizes the regularized empirical risk
  - PCR minimizes the empirical risk on a feature subspace spanned by the principal components of the covariance operator
  - RRR minimizes the empirical risk with a rank constraint

Our Contributions

- For the choice of universal kernels, analysing metric distortion we conclude that low rank estimators are preferable, and we analyse two: PCR and RRR
- We derive minimax optimal operator norm learning rates for KRR, PCR and RRR
- We derive spectral learning rates for normal compact operators
- We show that spurious spectra can occur from the spectral bias even for normal operators
- From spectral bias of RRR estimator, we deduce model selection method

Our Contributions

- For the choice of universal kernels, analysing metric distorsion one sees that low rank estimators are preferable, and we analyse two PCR and RRR
- We derive minimax optimal operator norm learning rates for KRR, PCR and RRR
- We derive spectral learning rates for normal compact operators
- We show that spurious spectra can occur from the spectral bias even for normal operators
- Empirically estimating the spectral bias of RRR estimator, we deduce model selection method

Relationships $\mathscr{H} \sim A_\pi$ and $\mathscr{H} \sim L^2_\pi(\mathscr{X})$ are captured by $\alpha \in [1,2]$ and $\beta \in [0,1]$ we have

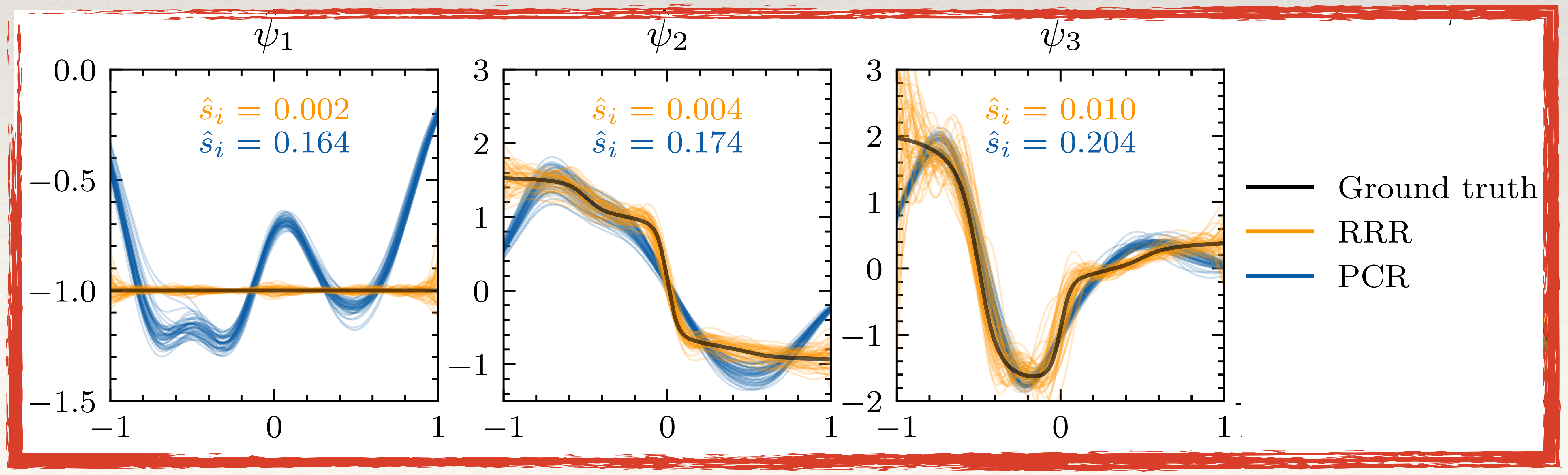
$$\varepsilon_n = n^{-\frac{\alpha}{2(\alpha+\beta)}}$$

With probability at least $1-\delta$ in the observed training data the estimation error is bounded by

$$\mathscr{E}(\hat{G}) \lesssim \varepsilon_n \ln(\delta^{-1})$$

Our Contributions

- For the choice of universal kernels, analysing metric distortion one sees that low rank estimators are preferable, and we analyse two PCR and RRR
- We derive minimax optimal operator norm learning rates for KRR, PCR and RRR
- We derive spectral learning rates for normal compact operators that reveal preference to RRR



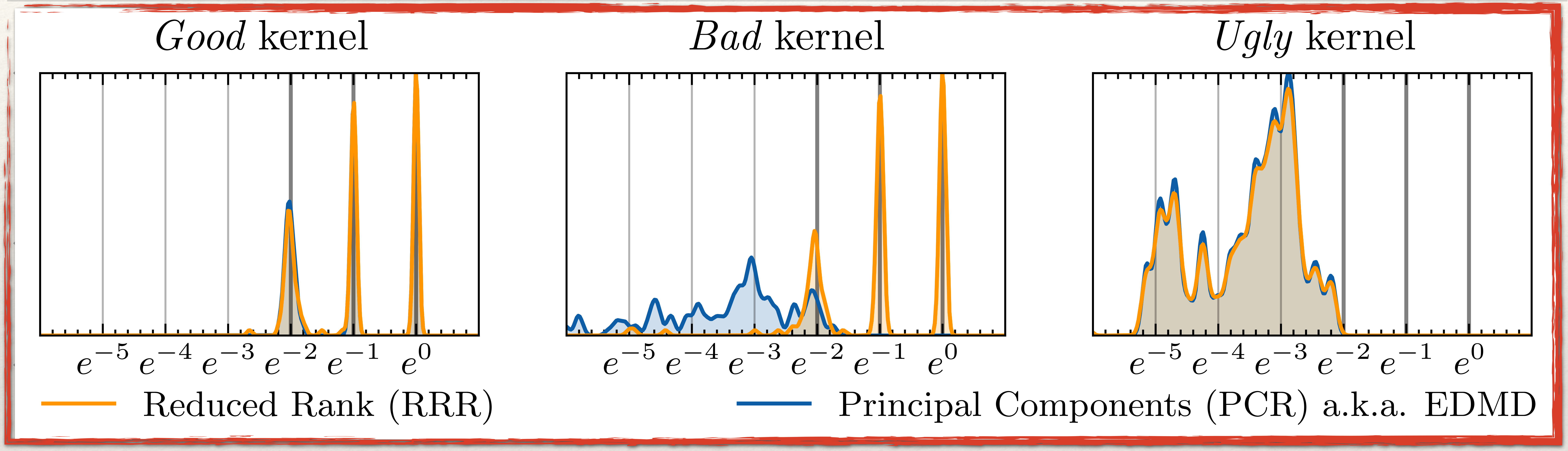
Relationships $\mathscr{H} \sim A_\pi$ and $\mathscr{H} \sim L^2_\pi(\mathscr{X})$ are captured by $\alpha \in [1,2]$ and $\beta \in [0,1]$ we have

$$\varepsilon_n = n^{-\frac{\alpha}{2(\alpha+\beta)}}$$

With probability at least $1 - \delta$ in the observed training data the eigenvalue error is bounded by

$$|\mu_i - \hat{\lambda}_i| \lesssim \frac{\sigma_{r+1}(A_{\pi|_{\mathscr{H}}})}{\sigma_r(A_{\pi|_{\mathscr{H}}})} + \varepsilon_n \ln(\delta^{-1})$$

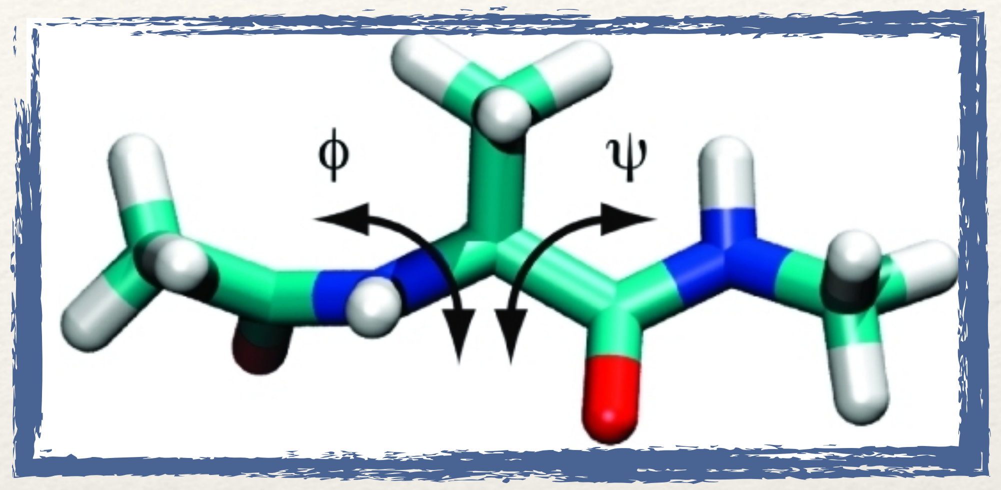
Our Contributions

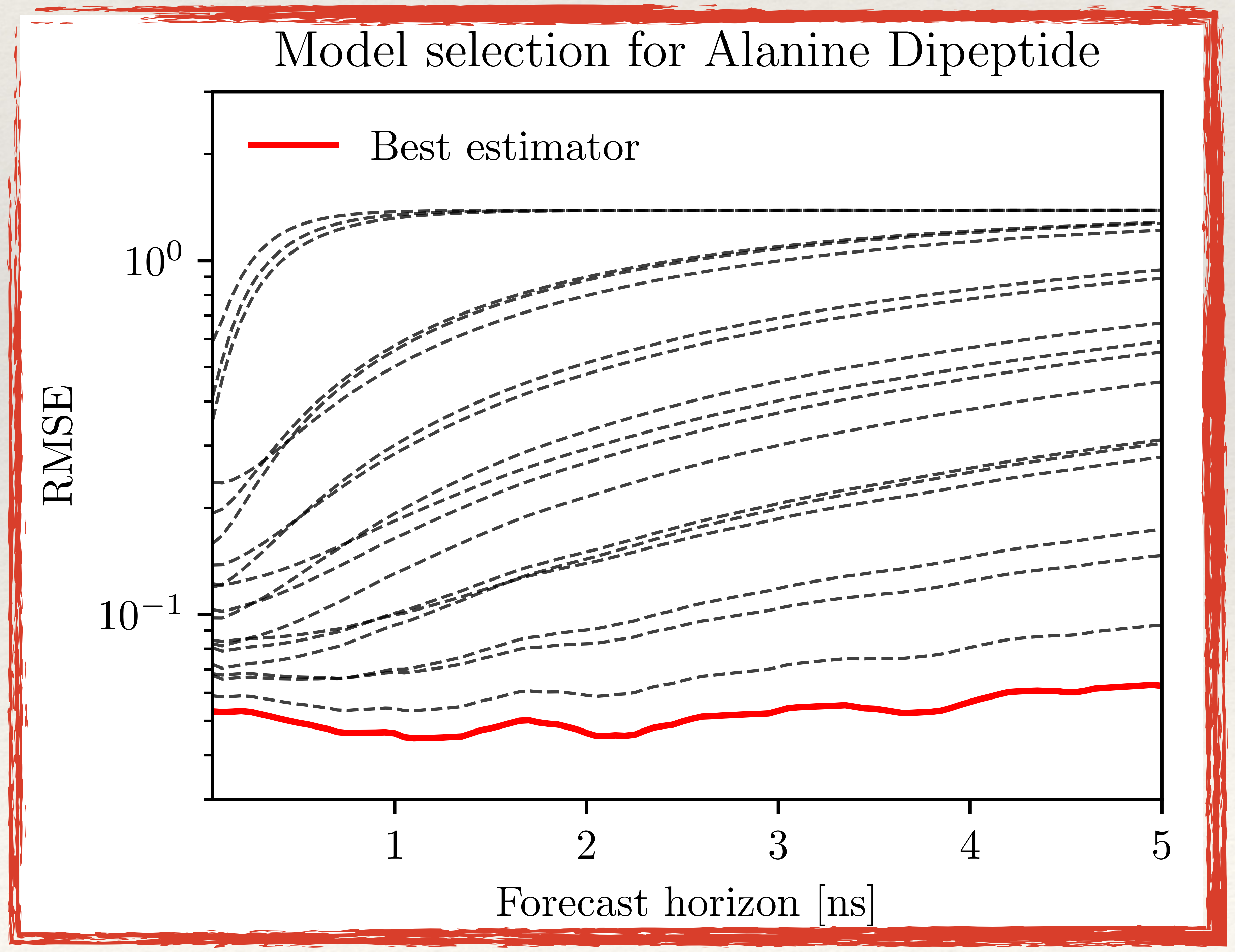


- We show that spurious spectra can occur from the spectral bias even for normal operators
- Empirically estimating the spectral bias of RRR estimator, we deduce model selection method

Our Contributions

- For the choice of universal kernels, analysing metric distortion one sees that low rank estimators are preferable, and we analyse two PCR and RRR
- We derive minimax optimal operator norm learning rates for KRR, PCR and RRR
- We derive spectral learning rates for normal compact operators
- We show that spurious spectra can occur from the spectral bias even for normal operators
- Empirically estimating the spectral bias of RRR estimator, we deduce model selection method





THANK YOU!

- While this was a high-level presentation, our paper is mathematically rigorous. Check it out or come see us at the poster session for many more details
- We have an available Python code:

https://github.com/CSML-IIT-UCL/kooplearn