Kernel Methods for Koopman Mode Analysis and Prediction: Ergodic and Skew-Product Systems

Dimitris Giannakis
Center for Atmosphere Ocean Science
Courant Institute of Mathematical Sciences
New York University

University of South Florida
March 10, 2017
Setting & objectives

Ergodic dynamical system $(A, B_A, \Phi_t, \alpha)$ observed through a vector-valued function $F : A \mapsto \mathbb{R}^n$

Given time-ordered observations $\{x_0, \ldots, x_{N-1}\}$ with $x_i = F(a_i)$, we seek to perform:

1. **Dimension reduction** with timescale separation and invariance under changes of observation modality
2. **Nonparametric prediction** of observables and probability measures on $A$
State- and observable-centric viewpoints

- **State space viewpoint**
  In data space, we observe the manifold $F(A) \subset \mathbb{R}^n$ and the vector field
  
  $$\bar{u}|_{x(t)} = \dot{x}(t) \quad \text{with} \quad x(t) = F(\Phi_t a)$$


  Associated with the dynamical system is a group of unitary operators $U_t : L^2(A, \alpha) \mapsto L^2(A, \alpha)$ s.t.

  $$U_t f(a) = f(\Phi_t a)$$

  The generator $u$ of $\{U_t\}$ is an unbounded, skew-adjoint operator $u : D(u) \mapsto L^2(A, \alpha)$, giving the directional derivative of functions along the dynamical flow

  $$uf(a) = \lim_{t \to 0} \frac{f(\Phi_t a) - f(a)}{t}, \quad \bar{u} = F_* u$$
Dimension reduction using Koopman eigenfunctions

- Eigenvalue problem for the generator $u$:
  \[ u(z_k) = \lambda_k z_k, \quad \lambda_k \in \mathbb{C}, \quad z_k \in D(u) \subset L^2(A, \alpha) \]

- The eigenvalues and eigenfunctions have the properties
  \[ \lambda_k = i\omega_k, \quad \omega_k \in \mathbb{R}, \quad \langle z_j, z_k \rangle = \delta_{jk} \]

- The eigenfunctions $z_k$ evolve as periodic observables,
  \[ U_t z_k(a) = e^{i\omega_k t} z_k(a) \]

- **Dimension reduction**: Represent the high-dimensional input data in $\mathbb{R}^n$ as vectors in $\mathbb{C}^r$ using the Koopman eigenfunctions as “coordinates”:
  \[ x = F(a) \rightarrow \pi(a) = \begin{pmatrix} z_1(a) \\ \vdots \\ z_r(a) \end{pmatrix}, \quad r \ll n \]

- The $z_k$ are independent of the observation map $F$
Nonparametric prediction

- Let \( \rho \in L^2(A, \alpha) \) be a **probability density** (wrt. \( \alpha \)) characterizing our knowledge about the state of the system at time \( t = 0 \).
- The **time-dependent expectation value** of an observable \( f \in L^2(A, \alpha) \) at time \( t > 0 \) is given by
  \[
  E_t f = \langle \rho, U_t f \rangle = \int_A U_t f \rho \, d\alpha
  \]
- Alternatively, we can consider the **Perron-Frobenius operator** \( U_t^* \) governing the evolution of probability densities in \( L^2(A, \alpha) \)
  \[
  E_t f = \langle U_t^* \rho, f \rangle
  \]
- If \( |f|^2 \in L^2(A, \alpha) \), we can also compute the **forecast variance** (useful for uncertainty quantification):
  \[
  \text{var}_t f = \langle \rho, |U_t f - E_t f|^2 \rangle
  \]
Data-driven techniques (Berry et al. 2015, G. et al. 2015, G. 2016)

- The **pointwise ergodic theorem** allows us to approximate inner products in $L^2(A, \alpha)$ by time averages,

$$\frac{1}{N} \sum_{j=0}^{N-1} f^*(a_j)g(a_j) \xrightarrow{\text{a.s.}} \int_A f^*(a)g(a) \, d\alpha(a) = \langle f, g \rangle$$

- **Kernel algorithms** from machine learning provide an orthonormal basis $\{\phi_0, \phi_1, \ldots\}$ of $L^2(A, \alpha)$ using the observed data $\{x_0, x_1, \ldots, x_{N-1}\}$ as inputs.

- We represent an observable $f \in L^2(A, \alpha)$ as an $l$-dimensional column vector

$$\vec{f} = (f_0, \ldots, f_{l-1})^\top, \quad f_i = \langle \phi_i, f \rangle \approx \frac{1}{N} \sum_{k=0}^{N-1} \phi_i(a_k)f(a_k)$$

and the Koopman operator as an $l \times l$ matrix

$$A_{ij}(t) = \langle \phi_i, U_t \phi_j \rangle \approx \frac{1}{N} \sum_{k=0}^{N-1} \phi_i(a_k)\phi_j(a_{k+t}), \quad i, j \in \{0, \ldots, l-1\}$$

- We can compute the action of $U_t f$ on arbitrary observables using matrix algebra

$$g = U_t f \longleftrightarrow \vec{g} = A(t)\vec{f}$$
Outline

1. Representation of Koopman operators in a data-driven orthonormal basis, and associated spectral decomposition and prediction techniques

2. “Space-time” Koopman mode analysis for skew-product systems and applications to Lagrangian tracers

Collaborators. Tyrus Berry, Shuddho Das, Matina Gkioulidou, John Harlim, Joanna Slawinska

Financial support. DARPA grant HR0011-16-C-0116, NSF grant DMS-1521775, ONR grant N00014-14-1-0150, ONR MURI grant N00014-12-1-0912, ONR YIP grant N00014-16-1-2649
Construct a **data-driven orthonormal basis** of \( L^2(A, \alpha) \) via the **diffusion maps algorithm** (Coifman & Lafon 2005) with a **variable-bandwidth kernel** (Berry & Harlim 2015) \( K_\epsilon : A \times A \mapsto \mathbb{R}_+ : \)

\[
K_\epsilon(a_i, a_j) = \exp \left( - \frac{\|x_i - x_j\|^2}{\epsilon \hat{\sigma}_\epsilon(x_i)^{-1/m} \hat{\sigma}_\epsilon(x_j)^{-1/m}} \right),
\]

\[
m = \dim A, \quad x_i = F(a_i), \quad \hat{\sigma}_\epsilon(x_i) = \frac{1}{N(\pi \epsilon)^{m/2}} \sum_{j=0}^{N-1} e^{-\|x_i - x_j\|^2 / \epsilon}
\]
Data-driven basis

Apply the diffusion maps normalization (Coifman & Lafon 2006; Berry & Sauer 2015):

\[\hat{q}_\epsilon(a_i) = \frac{1}{N} \sum_{j=0}^{N-1} K_\epsilon(a_i, a_j), \quad \hat{K}'_\epsilon(a_i, a_j) = \frac{K_\epsilon(a_i, a_j)}{\hat{q}_\epsilon(a_j)}\]

\[\hat{d}_\epsilon(a_i) = \frac{1}{N} \hat{K}'_\epsilon(a_i, a_j), \quad \hat{p}_\epsilon(a_i, a_j) = \frac{K'_\epsilon(a_i, a_j)}{\hat{d}_\epsilon(a_i)}\]

\(\hat{p}_\epsilon\) induces an averaging operator on \(L^2(A, \hat{\alpha})\) for the sampling measure \(\hat{\alpha} = N^{-1} \sum_{i=0}^{N-1} \delta_{a_i}:\)

\[\hat{P}_\epsilon f(b) = \int_A \hat{p}_\epsilon(b, a)f(a)\ d\hat{\alpha}(a) = \frac{1}{N} \sum_{j=0}^{N-1} \hat{p}_\epsilon(b, a_j)f(a_j)\]

By ergodicity, as \(N \to \infty\), \(\hat{P}_\epsilon f(b)\) converges \(\alpha\)-a.s. to \(P_\epsilon f(b)\), where

\[P_\epsilon f(b) = \int_A p_\epsilon(b, a)f(a)d\alpha(a)\]

is an averaging operator on \(L^2(A, \alpha)\).
Data-driven basis

Uniformly on $A$ (Coifman & Lafon 2006),

$$P_\epsilon f(a) = f(a) + \epsilon \Delta f(a) + O(\epsilon^2),$$

where $\Delta$ is the Laplace-Beltrami operator associated with the Riemannian metric $h = \sigma^2/m g$, $\sigma = d\alpha/d\text{vol}_g$, and $g$ is the Riemannian metric inherited from the observation map $F$

- $\frac{d\alpha}{d\text{vol}_h} = 1$ and the $\phi_k$ are orthogonal on $L^2(A, \alpha)$
- The corresponding eigenvalues $\eta_k$ are the extrema of a Dirichlet energy adapted to the sampling density of the data

$$E(f) = \int_A \| \text{grad}_h f \|^2_h d\alpha = \int_A \| \text{grad}_g f \|^2_g \sigma^{-2/m} d\alpha$$
Data-driven basis

• Basis of $L^2(A, \hat{\alpha})$:

$$\hat{P}_\epsilon \hat{\phi}_k = \hat{\lambda}_k \hat{\phi}_k, \quad \hat{\eta}_k = -\log \hat{\lambda}_k$$

• We also make use of the Sobolev space $H^1(A, h)$ and its data-driven counterpart $\hat{H}^1(A, \hat{\alpha})$:

$$f = \sum_k c_k \phi_k, \quad \|f\|_{H^1}^2 = \sum_{j=0}^1 \eta^j_k |c_k|^2$$

$$\hat{f} = \sum_k \hat{c}_k \hat{\phi}_k, \quad \|\hat{f}\|_{\hat{H}^1}^2 = \sum_{j=0}^1 \hat{\eta}^j_k |\hat{c}_k|^2$$

• The rescaled eigenfunctions $\varphi_k$ and $\hat{\phi}_k$ provide orthogonal bases for $H^1(A, \hat{\alpha})$ and $\hat{H}^1(A, \hat{\alpha})$, respectively:

$$\varphi_k = \frac{\phi_k}{\eta^{1/2}_k}, \quad \hat{\varphi}_k = \frac{\hat{\phi}_k}{\hat{\eta}^{1/2}_k}$$
Properties of Koopman eigenfunctions on manifolds

\[ u(z_k) = i\omega_k z_k, \quad k = (k_1, \ldots, k_r) \in \mathbb{Z}^r, \quad \omega_k \in \mathbb{R} \]

- **Group structure:**

\[
\begin{align*}
 u(z_j) &= i\omega_j z_j \\
 u(z_k) &= i\omega_k z_k \\
\end{align*}
\]

\[ \Rightarrow u(z_j z_k) = i(\omega_j \omega_k) z_j z_k \]

- We can generate the eigenvalues and eigenfunctions recursively through

\[
\begin{align*}
 \omega_k &= k_1 \Omega_1 + \ldots + k_r \Omega_r, \\
 z_k &= \zeta_1^{k_1} \cdots \zeta_r^{k_r},
\end{align*}
\]

where \( \{\zeta_1, \ldots, \zeta_r\} \), \( r \leq m \), is a maximal set of eigenfunctions corresponding to **rationally independent frequencies** \( \{\Omega_1, \ldots, \Omega_r\} \)

- For a suitable normalization, \( |z_k(a)| = 1 \) for all \( a \in A \)

- The map \( \pi : A \mapsto \mathbb{C}^r, \pi(a) = (\zeta_1(a), \ldots, \zeta_r(a)) \) maps the dynamics on \( A \) into a rotation \( R_t \) on \( \pi(A) = \mathbb{T}^r \),

\[
\pi \circ \Phi_t = R_t \circ \pi,
\]

\[
R_t(\theta_1, \ldots, \theta_r) = (\theta_1 + \Omega_1 t, \ldots, \theta_r + \Omega_r t) \mod 2\pi
\]

- The space \( \mathcal{D} = \text{span} z_k \) is a **closed invariant subspace** of \( L^2(A, \alpha) \) associated with the point spectrum of \( u \)}
Example: rotation on a torus

\[ A = \mathbb{T}^2, \quad \Phi_t(\theta_1, \theta_2) = (\theta_1 + t, \theta_2 + \Omega t) \mod 2\pi, \quad \Omega \in \mathbb{R} \setminus \mathbb{Q} \]

\[ u = \frac{\partial}{\partial \theta_1} + \Omega \frac{\partial}{\partial \theta_2}, \quad \alpha = \text{Lebesgue} \]

- In this case, the Koopman eigenfunctions are **Fourier functions**

\[ u(z_k) = i\omega_k z_k, \quad k = (k_1, k_2), \]

\[ \omega_k = k_1 + \Omega k_2, \]

\[ z_k = \zeta_1^{k_1} \zeta_2^{k_2}, \quad \zeta_1(\theta_1, \theta_2) = e^{i\theta_1}, \quad \zeta_2(\theta_1, \theta_2) = e^{i\theta_2} \]
Example: rotation on a torus

\[ A = \mathbb{T}^2, \quad \Phi_t(\theta_1, \theta_2) = (\theta_1 + t, \theta_2 + \Omega t) \mod 2\pi, \quad \Omega \in \mathbb{R} \setminus \mathbb{Q} \]

\[ u = \frac{\partial}{\partial \theta_1} + \Omega \frac{\partial}{\partial \theta_2}, \quad \alpha = \text{Lebesgue} \]

- In this case, the Koopman eigenfunctions are **Fourier functions**
  \[ u(z_k) = i\omega_k z_k, \quad k = (k_1, k_2), \]
  \[ \omega_k = k_1 + \Omega k_2, \]
  \[ z_k = \zeta_1^{k_1} \zeta_2^{k_2}, \quad \zeta_1(\theta_1, \theta_2) = e^{i\theta_1}, \quad \zeta_2(\theta_1, \theta_2) = e^{i\theta_2} \]

- The set \( \{\omega_k\} \) of frequencies is **dense** on the real line
- In particular, for suitable \( k_1, k_2 \in \mathbb{Z}, |k_j| \gg 1, k_1 k_2 < 0, \omega_k \) can be made arbitrarily small
- However, the **Dirichlet energy** of such eigenfunctions is arbitrarily large:
  \[ E(z_k) = \int_A \|\text{grad } z_k\|^2 d\alpha \propto (k_1^2, k_2^2) \]

- Our approach is to select Koopman eigenfunctions of high smoothness by ordering them wrt. \( E(z_k) \)
Eigenvalue problem for the Koopman generator

We compute approximate Koopman eigenfunctions through a Galerkin method for the regularized generator

\[ L = u - \epsilon \Delta, \quad \epsilon > 0 \]

**Continuous problem.** Find \( z \in H^1(A, h) \) and \( \lambda \in \mathbb{C} \) s.t.

\[ \langle \psi, u(z) \rangle + \epsilon \langle \text{grad}_h \psi, \text{grad}_h z \rangle = \lambda \langle \psi, z \rangle, \quad \forall \psi \in H^1(A, h) \]

**Discrete approximation.** Set \( \hat{H}_l = \text{span}\{ \hat{\phi}_0, \ldots, \hat{\phi}_{l-1} \} \subseteq \hat{H}^1(A) \). Find \( \hat{z} \in \hat{H}_l \) and \( \hat{\lambda} \in \mathbb{C} \) s.t.

\[ \langle \psi, \hat{u}(\hat{z}) \rangle_{\hat{\alpha}} + \langle \text{grad}_h \psi, \text{grad}_h \hat{z} \rangle_{\hat{\alpha}} = \hat{\lambda} \langle \psi, \hat{z} \rangle_{\hat{\alpha}}, \quad \forall \psi \in \hat{H}_l. \]
Eigenvalue problem for the Koopman generator

\[ \langle \psi, \hat{u}(\hat{z}) \rangle_\hat{\alpha} + \langle \hat{\text{grad}}_h \psi, \hat{\text{grad}}_h \hat{z} \rangle_\hat{\alpha} = \hat{\lambda} \langle \psi, \hat{z} \rangle_\hat{\alpha}, \]

\[ \hat{z} = \sum_{i=0}^{l-1} c_i \hat{\varphi}_i, \quad \psi = \sum_{i=0}^{l-1} w_i \hat{\varphi}_i \]

- \( \hat{u} \) is a \textbf{finite-difference approximation} of \( u \), e.g.,

\[
\langle \psi, \hat{u}(\hat{z}) \rangle_\hat{\alpha} = \sum_{i,j=0}^{l-1} w_i c_j \int_M \hat{\varphi}_i \hat{\nu}(\hat{\varphi}_j) \, d\hat{\alpha}
\]

\[
= \sum_{i,j=0}^{l-1} w_i c_j \left[ \frac{1}{N} \sum_{k=1}^{N-2} \hat{\varphi}_i(a_k) \frac{\hat{\varphi}_j(a_{k+1}) - \hat{\varphi}_j(a_{k-1})}{2 \delta t} \right]
\]

- By construction of the \{\( \hat{\varphi}_i \)\} basis,

\[
\langle \hat{\text{grad}}_h \psi, \hat{\text{grad}}_h \hat{z} \rangle_\hat{\alpha} = \sum_{i,j=0}^{l-1} w_i c_j \delta_{ij}
\]

- Scheme remains well-conditioned at large spectral order \( l \)
Variable-speed rotation on $\mathbb{T}^2$

\[ v = \sum_{\mu=1}^{2} v^\mu \frac{\partial}{\partial \theta^\mu} \]
\[ v^1 = 1 + \beta \cos \theta^1 \]
\[ v^2 = \tilde{\omega}(1 - \beta \sin \theta^2) \]

\[ \tilde{\omega} = \sqrt{30}, \quad \beta = \sqrt{1/2} \]
Numerical Koopman eigenfunctions for variable-speed rotation on $\mathbb{T}^2$

$\zeta_1, \Omega_1 = 0.71$

$\zeta_2, \Omega_2 = 3.87$
Vector field decomposition

In systems with pure point spectra \((r = m)\), define the vector fields \(u_i : C^\infty(A) \rightarrow C^\infty(A)\) through their action on the eigenfunctions:

\[
u_i(\zeta_1^{k_1} \cdots \zeta_i^{k_i} \cdots \zeta_m^{k_m}) = ik_i \Omega_i \zeta_1^{k_1} \cdots \zeta_i^{k_i} \cdots \zeta_r^{k_r}\]

The \(u_i\) are linearly independent, nowhere vanishing, mutually commuting vector fields

\[
u = \sum_{i=1}^{m} u_i, \quad [u_i, u_j] = 0\]

- Due to their vanishing commutator, the \(u_i\) can be thought of as dynamically independent components
- These vector fields can be realized in data space through the pushforward map \(F_* : TA \rightarrow T\mathbb{R}^n \simeq \mathbb{R}^n\)

\[
u_i = F_*(u_i) = u_i(F) = \sum_k \hat{F}_k u(z_k), \quad \hat{F}_k = \langle z_k, F \rangle\]
Recovering convective organization in the atmosphere

**Top:** Extraction of *convectively coupled equatorial waves* from satellite observations of brightness temperature (Slawinska & G. 2016)
Forecasting densities and expectation values

The **Perron-Frobenius operator** $U^*_t : L^2(A, \alpha) \mapsto L^2(A, \alpha)$ governs the evolution of probability densities relative to $\mu$

$$\rho_0 = \sum_k c_k(0) \phi_k, \quad c_k(0) = \langle \phi_k, \rho_0 \rangle,$$

$$\rho_t = U^*_t \rho_0 = \sum_k c_k(t) \phi_k, \quad c_k(t) = \sum_j A_{kj}(t) c_k(0), \quad A_{kj}(t) = \langle \phi_k, U_t \phi_j \rangle$$

Approximate $A_{kj}(t)$ using

$$\hat{A}_{kj}(t) = \langle \hat{\phi}_k, \hat{\phi}_l \rangle \hat{\mu} = \frac{1}{N} \sum_{i=0}^{N-1} \phi_k(a_i) \phi_j(a_{i+t})$$

The time-dependent expectation value of an observable $f \in L^2(A, \alpha)$ is

$$\mathbb{E}_t f = \sum_k \hat{f}_k c^*_k(t), \quad \hat{f}_k = \langle z_k, f \rangle$$
Threshold-based event prediction

- Of interest is to predict the observable $\chi_S$, where
  \[ S = \{ a \in A : |f(a)| > f_0 \}, \quad f \in L^2(A, \alpha) \]

- The quantity
  \[ \mathbb{E}_t \chi_S = \langle \rho_t, \chi_S \rangle = \int_A (U_t^* \rho_0) \chi_S \, d\alpha \]

is equal to the probability that $|f|$ exceeds the threshold $f_0$ at forecast time $t$, given the current knowledge of the state of the system at time $t = 0$ as represented by $\rho_0$

**Top:** Probabilistic prediction of the AL index characterizing auroral activity and geomagnetic substorms (G. et al. 2017)
Connections with delay-coordinate maps

Another approach for dimension reduction with timescale separation is to use diffusion eigenfunctions from delay-coordinate mapped data (G. & Majda 2012; Berry et al. 2013)

\[ F_s : A \mapsto \mathbb{R}^{sd}, \quad F_s(a) = X = (F(a), F(\Phi_{-\delta t} a), \ldots, F(\Phi_{-(s-1)\delta t} a)), \]

\[ K_\epsilon(a_i, a_j) = \exp \left( -\frac{\|X_i - X_j\|^2}{\epsilon \hat{\sigma}_{s,\epsilon}^{-1/m}(X_i) \hat{\sigma}_{s,\epsilon}^{-1/m}(X_j)} \right), \]

Interannual and decadal patterns of Indo-Pacific SST recovered via kernel methods with delay embeddings (Slawinska & G. 2016; G. & Slawinska 2016)
Connections with delay-coordinate maps

The induced Riemannian metric for $s$ delays is given by

$$g_s(u, w) = \frac{1}{s} \sum_{i=0}^{s-1} g(D\Phi_{-i}\delta t u, D\Phi_{-i}\delta t w)$$

**Theorem.** In systems of dimension $m$ having a generator $u$ with pure point spectrum, $\lim_{s \to \infty} g_s = \bar{g} = \sum_{i,j=1}^{m} B_{ij} \beta_i \otimes \beta_j$, where $B_{ij} = \langle u_i, u_j \rangle = \int_{A} g(u_i, u_j) \, d\alpha$ are the Hodge inner products of vector fields $v_i$ s.t. $u = \sum_{i=1}^{m} u_i$, $[u_i, u_j] = 0$, and $\beta_i$ are dual vector fields s.t. $\beta_i(u_j) = \delta_{ij}$

Moreover, $\bar{g}$ is flat and has uniform volume form relative to the invariant measure $\alpha$ of the dynamics

**Corollary.** The Laplace-Beltrami operator $\Delta_{\bar{g}}$ associated with $\bar{g}$ is given by $\Delta_{\bar{g}} = \sum_{i,j=1}^{n} B_{ij}^{-1} u_i \circ u_j$. Moreover, because $[u_i, u] = 0$, $\Delta_{\bar{g}}$ and the Koopman generator $u$ commute
Efficiency of Galerkin schemes and timescale separation

Because they commute, $u$ and $\Delta_{\bar{g}}$ have common eigenfunctions

- For sufficiently large $s$, eigenfunctions of $\Delta_s$ should provide an efficient approximation space for the eigenvalue problem for $u$

- The fact that the eigenfunctions of $u$ are periodic observables explains (at least in pure point spectrum systems) why the eigenfunctions of $\Delta_s$ have timescale separation
Removing i.i.d. measurement noise

For data $\tilde{x}_i = x_i + \xi_i$, $x_i = F(a_i)$, corrupted by i.i.d. noise $\xi_i$,

$$\mathbb{E} \|\tilde{x}_i - \tilde{x}_j\|^2 = \|x_i - x_j\|^2 + 2R^2, \quad R^2 = \text{var} \xi_i$$

Performing delay-coordinate maps, and taking the limit $s \to \infty$,

- $\mathbb{E} \|\tilde{X}_i - \tilde{X}_j\|^2 \xrightarrow{a.s.} \|X_i - X_j\|^2 + 2R^2$
- Because $\bar{g}$ is flat, the bias term in the pairwise distance produces a
  multiplicative bias in the kernel which cancels to $O(\epsilon^2)$ in the diffusion
  maps normalization

$$\tilde{K}_\epsilon(a_i, a_j) = \exp \left( - \frac{\|\tilde{X}_i - \tilde{X}_j\|^2}{\epsilon \sigma_{s,\epsilon}^{-1/m}(\tilde{X}_i) \sigma_{s,\epsilon}^{-1/m}(\tilde{X}_j)} \right) \to e^{-\frac{2R^2}{\epsilon(c^2 + O(\epsilon^2))}} K_\epsilon(a_i, a_j),$$

$$\sum_j \tilde{P}_\epsilon(a_i, a_j) f(a_j) = \sum_j \frac{\tilde{K}_\epsilon(a_i, a_j)}{\sum_k \tilde{K}_\epsilon(a_j, a_k)} f(a_j) \to \int_A P_\epsilon(a_i, b) f(b) \, d\alpha + O(\epsilon^2)$$

- Since $\int_A P_\epsilon(a_i, b) f(b) \, d\alpha(b) = f(a_i) - \epsilon \Delta_{\bar{g}} f(a_i) + O(\epsilon^2)$, this bias does
  not affect the convergence of the eigenfunctions of $P_\epsilon$ to the
  eigenfunctions $\phi_k$ of $\Delta_{\bar{g}}$ as $\epsilon \to 0$, and the denoised $\phi_k$ can be employed
  in the Galerkin scheme for Koopman eigenfunctions
Koopman eigenfunctions from noisy data

Koopman eigenfunctions for the variable-speed flow on $\mathbb{T}^2$ recovered from data from data corrupted with i.i.d. Gaussian noise in $\mathbb{R}^3$ with SNR $\approx 1$
Model for time-dependent fluid flows. Ergodic dynamical system \((A, \mathcal{A}, \Phi_t, \alpha)\) observed through a function \(F : A \mapsto \mathcal{X}\) taking values in the space of divergence-free vector fields \(\mathcal{X}\) on a domain \((\mathcal{X}, \mathcal{X}, \xi)\).

Given time-ordered velocity field snapshots \(\{v_0, \ldots, v_{N-1}\}\) with \(v_i = F(a_i)\), we seek to

1. Recover **coherent patterns** associated with the motion of Lagrangian tracers in \(X\)
2. Perform **equation-free forecasting** of observables and densities defined on the tracers
Example: “switching” Gaussian vortex

- Base dynamics is rotation on the circle $A$ s.t. $\Phi_t(\theta) = \theta + \beta t$
- $X = \mathbb{T}^2$ is a doubly-periodic domain; $\nu = F(\theta)$ is the velocity field with the Gaussian streamfunction

$$\zeta(\theta, x) = \Gamma \cos \theta e^{\kappa [\cos x^1 + \cos x^2]} + \Gamma \sin \theta e^{\kappa [\cos(x^1 - \pi) + \cos x^2]}$$

$$\nu = -\frac{\partial \zeta}{\partial x^2} \frac{\partial}{\partial x^1} + \frac{\partial \zeta}{\partial x^1} \frac{\partial}{\partial x^2}$$
Example: “switching” Gaussian vortex

**Movie.** Evolution of an ensemble of 2000 Lagrangian tracers with initial velocities and positions drawn from the Gaussian distributions

\[
\rho_A(\theta) = \frac{1}{l_0(\sigma)} e^{\sigma \cos \theta}, \quad \rho_X(x) = \frac{1}{l_0(\sigma)^2} e^{\sigma [\cos(x^1 - \pi) + \cos x^2]},
\]

respectively, with \( \sigma = 3 \) and the vortex parameters \( \beta = 1, \Gamma = 2, \kappa = 0.5 \).
Coherent structures

Identification of coherent structures in fluid flows and general dynamical systems includes:

- **Geometric approaches:** Invariant manifolds (Meiss 1992; Wiggins 1992), hyperbolic material surfaces (Haller 2001), finite-time Lyapunov exponents (Haller 2000; Shadden et al. 2005), coherent Lagrangian vortices (Haller & Beron-Vera 2013)

- **Operator-theoretic/probabilistic approaches:** Almost-invariant sets (Dellnitz & Junge 1992; Froyland & Dellnitz 2005; Schütte et al. 2010); coherent sets (Froyland et al. 2010); ergodic quotients (Budišić & Mezić 2012); dynamic Laplacians (Froyland 2015)

- In this work, we identify coherent structures using eigenfunctions of Koopman operators on the “space-time manifold” $M = A \times X$ with low oscillatory frequency and small Dirichlet energy

- As in the ergodic case, we represent these operators in a smooth orthonormal basis learned from velocity field data using the diffusion maps algorithm
Dynamics on the product space $A \times X$

**Ergodic dynamics on $A$ (evolution of vel. field)**, $\Phi_t : A \rightarrow A$, with invariant measure $\alpha$

**Nonautonomous dynamics on $X$ (tracer motion)**, $\Psi_t : A \times X \rightarrow X$, preserving the Lebesgue measure $\xi$

- Ccycle property, $\Psi_s(\Phi_t(a), \Psi_t(a, x)) = \Psi_{s+t}(a, x)$
- Given $f : X \rightarrow \mathbb{C}$, $\lim_{t \rightarrow 0} [f(\Psi_t(a, \cdot)) - f(\cdot)]/t = v|_af$

**Autonomous dynamics on $M = A \times X$, $\Omega_t : M \rightarrow M$, $\Omega_t(a, x) = (\Phi_t(a), \Psi_t(a, x))$** with invariant measure $\mu = \alpha \times \xi$
Koopman operators

**Evolution of observables on** \( A \): \( U_t : L^2(A, \alpha) \mapsto L^2(A, \alpha) \), \( U_t f = f \circ \Phi_t \)

- \( U_t \) is unitary on \( L^2(A, \alpha) \)
- The generator \( u = \left. \frac{dU_t}{dt} \right|_{t=0} \) is a vector field on \( A \) with \( \text{div}_\alpha u = 0 \)

**Evolution of observables on** \( M \): \( W_t : L^2(M, \mu) \mapsto L^2(M, \mu) \), \( W_t f = f \circ \Omega_t \)

- \( W_t \) is unitary on \( L^2(M, \mu) \)
- The generator \( w = \left. \frac{dW_t}{dt} \right|_{t=0} \) is a vector field on \( M \) with \( \text{div}_\mu w = 0 \) and \( w = u + v \)
- Analogy with material derivative in fluid dynamics: \( \frac{D}{Dt} = \frac{\partial}{\partial t} + v \cdot \nabla \) and \( \frac{D}{Dt} \leftrightarrow w, \frac{\partial}{\partial t} \leftrightarrow u, v \cdot \nabla \leftrightarrow v \)

We identify **coherent spatiotemporal patterns** \( z \in L^2(M, \mu) \) through (approximate) eigenfunctions of \( w \) at small corresponding eigenvalue:

\[ w(z) = ivz, \quad v \in \mathbb{R} \]

- Eigenfunctions corresponding to \( v = 0 \) are conserved observables on the tracers with \( z \circ \Omega_t = z \)
Koopman operators and coherent structures

1. Equip $A$ and $X$ with Riemannian metrics $h_A$ and $h_X$, respectively, s.t.
   \[ \frac{d\alpha}{d\text{vol } h_A} = 1, \quad \frac{d\xi}{d\text{vol } h_X} = 1; \]
   construct the metric $h = h_A + h_X$ on $M$ and the associated Laplace-Beltrami operator
   \[ \Delta = \Delta_A + \Delta_X, \quad \Delta \geq 0 \]

2. Solve the eigenvalue problem for the regularized generator
   \[ L_\epsilon = w - \epsilon \Delta; \]
   i.e., find $z \in H^1(M, h)$ and $\lambda \in \mathbb{C}$ s.t.
   \[ \langle f, w(z) \rangle + \epsilon \langle \text{grad}_h \psi, \text{grad}_h z \rangle = \lambda \langle f, z \rangle, \quad \forall f \in H^1(M, h) \]

3. Order the solutions $(\lambda_k, z_k)$ in order of increasing Dirichlet energy
   \[ E(z_k) = \int_M \|\text{grad}_h z_k\|^2 \, d\mu \]
   - $\text{Re}(\lambda_k)$ and $\nu_k := \text{Im}(\lambda_k)$ measure the growth rate (guaranteed $\leq 0$) and oscillatory frequency of pattern $z_k$, respectively
   - For smooth coherent structures $|\lambda_k|$ and $E(z_k)$ are both small
Galerkin method in a tensor product basis

Let \( \{ \omega_{ij} \} \) be an orthonormal basis of \( L^2(M, \mu) \) consisting of eigenfunctions of \( \Delta \):

\[
\Delta \omega_{ij} = \eta_{ij} \omega_{ij}, \quad \eta_{ij} = \eta_{A,i} + \eta_{X,j}, \quad \omega_{ij} = \phi_i \psi_j \\
\Delta_A \phi_i = \eta_{A,i} \phi_i, \quad \Delta_X \psi_j = \eta_{X,j} \psi_j
\]

The normalized eigenfunctions \( \varpi_{ij} = \omega_{ij}/\eta_{ij}^{1/2} \) are orthogonal on \( H^1(M, h) \)

**Finite-dimensional approximation of Koopman eigenvalue problem:**
Find \( z = \sum_{i=1}^{n_A} \sum_{j=1}^{n_X} c_{ij} \varpi_{ij} \) and \( \lambda \in \mathbb{C} \) s.t. for all \( f = \sum_{i=1}^{n_A} \sum_{j=1}^{n_X} f_{ij} \varpi_{ij} \),

\[
\langle f, w(z) \rangle + \epsilon \langle \text{grad}_h \psi, \text{grad}_h z \rangle = \lambda \langle f, z \rangle
\]

- Solutions are obtained from the matrix generalized eigenvalue problem
  \( \sum_{k,l} A_{ijkl} c_{kl} = \lambda B_{ij} c_{ij} \) with \( A_{ijkl} = w_{ijkl} - \epsilon \delta_{ik} \delta_{jl}, \ w_{ijkl} = \langle \varpi_{ij}, w(\varpi_{kl}) \rangle, \ B_{ij} = \langle \varpi_{ij}, \varpi_{ij} \rangle = 1/\eta_{ij} \)
- In certain cases, the matrix \( w_{ijkl} \) need not be formed explicitly
- By construction of the \( \{ \varpi_{ij} \} \) basis, \( \langle \text{grad}_h \varpi_{ij}, \text{grad}_h \varpi_{kl} \rangle = \delta_{ik} \delta_{jl} \), and the scheme remains well-conditioned at large \( n_A, n_X \)
- The Dirichlet energy of the solutions is \( E(z) = \sum_{ij} |c_{ij}|^2 \)
Switching vortex example

\[ \Phi_t(\theta) = \theta + \beta t, \]
\[ \zeta(\theta, x) = \Gamma \cos \theta e^{\kappa[\cos x^1 + \cos x^2]} + \Gamma \sin \theta e^{\kappa[\cos(x^1 - \pi) + \cos x^2]}, \]
\[ \nu = -\frac{\partial \zeta}{\partial x^2} \frac{\partial}{\partial x^1} + \frac{\partial \zeta}{\partial x^1} \frac{\partial}{\partial x^2} \]

- Build an orthonormal basis \( \{\omega_{ijk} = \phi_i \psi_{jk}\} \) of \( L^2(\mathcal{M}, \mu) \) from Fourier functions \( \phi_i(\theta) = e^{ii\theta}, \psi_{jk}(x) = e^{i(jx^1 + kx^2)} \); the Laplacian eigenvalues on \( \mathcal{M} \) are \( \eta_{ijk} = i^2 + j^2 + k^2 \)
- In this basis, the matrix elements of the generator can be evaluated analytically using integral identities for modified Bessel functions:

\[
\langle \omega_{ijk}, w(\omega_{lmn}) \rangle = \langle \omega_{ijk}, u(\omega_{lmn}) \rangle + \langle \omega_{ijk}, \nu(\omega_{lmn}) \rangle,
\]
\[
\langle \omega_{ijk}, u(\omega_{lmn}) \rangle = il\beta \delta_{il} \delta_{jm} \delta_{kn},
\]
\[
\langle \omega_{ijk}, \nu(\omega_{lmn}) \rangle = C_{jkmn} \delta_{i,l+1} + C_{jkmn}^* \delta_{i,l-1},
\]
\[
C_{jkmn} = (1 - i(-1)^{j-m}) \left[ nl_{|n-k|}(\kappa) (l_{|m-j+1|}(\kappa) - l_{|m-j-1|}(\kappa)) - ml_{|m-j|}(\kappa) (l_{|n-k+1|}(\kappa) - l_{|n-k-1|}(\kappa)) \right]
\]
Koopman eigenfunctions for switching vortex

**Movie.** Real parts of the leading Koopman eigenfunctions (ordered in order of increasing Dirichlet energy) for vortex parameters $\beta = 1$, $\Gamma = 1$, and $\kappa = 0.5$ (“sloshing” regime), computed using $n_A = n_X = 2 \times 32 + 1$ Fourier modes and $\epsilon = 0.0001$
Koopman eigenfunctions for switching vortex

**Movie.** Real parts of the leading Koopman eigenfunctions (ordered in order of increasing Dirichlet energy) for vortex parameters $\beta = 1$, $\Gamma = 2$, and $\kappa = 0.5$ ("fingering" regime), computed using $n_A = n_X = 2 \times 32 + 1$ Fourier modes and $\epsilon = 0.001$
Predicting the evolution of observables

Given an observable $f \in L^2(M, \mu)$, the quantity

$$W_t f(a, x) = e^{tw} f(a, x) = f(\Omega_t(a, x))$$

is equal to the value of $f$ measured at time $t$ on the tracer released at time $t = 0$ at the point $x \in X$ for the flow state $a \in A$

In the switching vortex example, observables of interest could be

$$f_1(a, x) = e^{ix^1}, \quad f_2(a, x) = e^{ix^2}$$

- Knowledge of $W_t f_1$ and $W_t f_2$ is sufficient to uniquely determine the position of Lagrangian tracers in $X$ for arbitrary initial conditions in $A$
Spectral approximation of the action of the Koopman operator on functions

Fix spectral order parameters $n_A, n_X$, and form the orthonormal set \( \{\omega_{ij} = \phi_i \psi_j\} \) with $1 \leq i \leq n_A$, $1 \leq j \leq n_X$

Approximate $W_t f$ by $\hat{W}_t f := e^{t \hat{L}_\epsilon} f$, $\hat{L}_\epsilon = PL_\epsilon P$, where $L_\epsilon = w - \epsilon \Delta$ is the regularized generator and $P$ the projection operator from $L^2(M, \mu)$ to the $n_A n_X$-dimensional approximation space span$\{\omega_{ij}\}$

Expanding $Pf = \sum_{i=1}^{n_A} \sum_{j=1}^{n_X} c_{ij} \omega_{ij}$, we have $\hat{W}_t f = \sum_{i=1}^{n_A} \sum_{j=1}^{n_X} c_{ij}(t) \omega_{ij}$, and

\[
\bar{c}(t) = e^{tL} \bar{c}, \quad \bar{c}(t) = [c_{ij}(t)], \quad \bar{c} = [c_{ij}], \quad L = [\langle \omega_{ij}, L_\epsilon \omega_{kl} \rangle]
\]

- We compute $e^{tL_\epsilon} \bar{c}$ without explicit computation of matrix exponentials using **Leja interpolation** (Caliari et al. 2004, 2016; Kandolf et al. 2014)
Leja interpolation

Let \( f : \mathbb{C} \mapsto \mathbb{C} \) be analytic in an open set \( S \subseteq \mathbb{C} \), and \( \{\zeta_i\}_{i=0}^n \) a set of points in \( S \).

The **Newton interpolation polynomial** of \( f \) is the degree-\( n \) polynomial

\[
p_n(z) = f(\zeta_0) + \sum_{i=1}^{n} f[\zeta_0, \ldots, \zeta_i] \prod_{j=0}^{i-1}(z - \zeta_j),
\]

where \( f[\zeta_0, \ldots, \zeta_i] \) are the **divided differences** defined recursively as

\[
f[\zeta_i] = f(\zeta_i), \quad f[\zeta_i, \ldots, \zeta_j] = \frac{f[\zeta_{i+1}, \ldots, \zeta_k] - f[\zeta_i, \ldots, \zeta_{j-1}]}{\zeta_j - \zeta_i}
\]

- Procedure is generally ill-conditioned for arbitrary points
- The **Leja interpolation points** are computed recursively by choosing \( \zeta_0 \) arbitrarily in \( K \) and \( \zeta_j \in \text{argmax}_{z \in K} \prod_{i=0}^{j-1}|z - \zeta_i| \) for \( j > 0 \)
- For Leja points, the condition number of the map \( \{\zeta_i\}_{i=0}^n \mapsto p_n \) grows sub-exponentially with \( n \)

To compute $e^{tL}\vec{c}$ for $L + L^* \leq 0$:

1. Construct a rectangle $[\alpha, \beta] \times [-\gamma, \gamma] \subset \mathbb{C}$ (e.g., via Gershgorin disks) bounding $\sigma(L)$
2. Compute Leja points $\{\zeta_i\}$ on the interval $\frac{\alpha + \beta}{2} + i[-\gamma, \gamma]
3. Approximate $e^{tL}\vec{c}$ via $p_n(tL)\vec{c}$, where $p_n$ is the Newton interpolation polynomial associated with $\{\zeta_i\}$

- Superlinear convergence holds, $\limsup_{n \to \infty} \|e^{tL}\vec{c} - p_n(tL)\vec{c}\|_2^{1/n} \to 0$
- In practice, a substepping procedure is used to compute $e^{tL}\vec{c} = (e^{tL/s})^s\vec{c}$; in each substep, the algorithm is terminated at the smallest $n$ s.t. $\|p_{n+1}(tL)\vec{c}_s - p_n(tL)\vec{c}_s\| \leq \delta$
Movie. Imaginary parts of the observables $W_t f_1$ and $W_t f_2$ for $f_1(a, x) = e^{ix_1}$ and $f_2(a, x) = e^{ix_2}$ and the vortex parameters parameters $\beta = 1$, $\Gamma = 2$, and $\kappa = 0.5$, computed using $n_A = n_X = 2 \times 32 + 1$ Fourier modes, $\epsilon = 0.0001$, and matrix exponential relative tolerance $\delta = 10^{-8}$
Predicting the evolution of probability measures

Let $\nu$ be a probability measure on $(M, \mathcal{A} \otimes \mathcal{X})$ s.t. $\rho = \frac{d\nu}{d\mu} \in L^2(M, \mu)$

The evolution of the density $\rho$ under the flow $\Omega_t$ is given by $\rho_t = W_t^* \rho_0$, where $W_t^* : L^2(M, \mu) \mapsto L^2(M, \mu)$ is the Perron-Frobenius operator; that is, $\rho_t = \frac{d\nu_t}{dt}$ where $\nu_t = \Omega_{*t} \nu$

Another object of interest is the marginal density $\sigma_t = \int_A \rho_t(a, \cdot) \, d\alpha(a)$

- $\sigma_t$ represents the density of tracers in the physical domain $X$ given an initial configuration $\sigma = \int_A \rho(a, \cdot) \, d\alpha$ and uncertainty in the state of the velocity field

Similarly to the Koopman operator case, we expand $P \rho = \sum_{i=1}^{n_A} \sum_{j=1}^{n_X} c_{ij} \omega_{ij}$, $c_{ij} = \langle \omega_{ij}, \rho \rangle$, and approximate $W_t \rho$ via $\hat{W}_t^* \rho := e^{tL_\epsilon^*} \rho$, where $L_\epsilon^* = PL_\epsilon P = P(-w - \epsilon \Delta)P$

- We have $\hat{W}_t^* \rho = \sum_{i=1}^{n_A} \sum_{j=1}^{n_X} c_{ij}(t) \omega_{ij}$, where

$$c(t) = e^{tL^*} c, \quad \bar{c}(t) = [c_{ij}(t)], \quad \bar{c} = [c_{ij}], \quad L^* = [[L_\epsilon \omega_{ij}, \omega_{kl}]$$
Density evolution for switching vortex

**Movie.** Evolution of the marginal density $\sigma_t$ associated with an initial density $\rho(a,x) = \rho_A(a)\rho_X(x)$ given by a product of circular Gaussians,

$$\rho_A(a) = \frac{1}{I_0(\tilde{\kappa})} e^{\tilde{\kappa} \cos a}, \quad \rho_X(x) = \frac{1}{I_0^2(\tilde{\kappa})} e^{-\kappa(\cos x^1 + \cos x^2)},$$

with $\tilde{\kappa} = 20$ and the vortex parameters parameters $\beta = 1$, $\Gamma = 2$, and $\kappa = 0.5$, computed using $n_A = n_X = 2 \times 32 + 1$ Fourier modes, $\epsilon = 0.0001$, and relative tolerance $\delta = 10^{-8}$.
Lorenz 96 system (Lorenz 1996)

The **L96 system** is a low-order deterministic model originally introduced for modeling **atmospheric circulation** on a constant_latitude ring:

\[
\dot{s}_i = s_{i-1}(s_{i+1} - s_{i-2}) - s_i + F, \quad i \in \{0, \ldots, d\}, \quad s_{d+1} = s_0
\]

- Energy-preserving nonlinear advection, \( s_{i-1}(s_{i+1} - s_{i-2}) \)
- Damping, \(-s_i\)
- Forcing ("solar heating"), \(F\)

For \(d = 40\), the forcing values \(F = 4, 5, \) and \(8\) correspond to **quasiperiodic**, **weakly chaotic**, and **strongly chaotic** dynamical regimes.
Tracer flows driven by Lorenz 96 systems (Qi & Majda 2016)

Use the L96 system to drive the incompressible flow on $X = \mathbb{T}^2$ with velocity field

$$v|_s(x) = v^1(s)\frac{\partial}{\partial x^1} + v^2(s, x^1)\frac{\partial}{\partial x^2},$$

$$v^1(s) = \hat{s}_0, \quad v^2(s, x^1) = \sum_{k \neq 0} \hat{s}_k e^{ikx^1},$$

$$\hat{s}_k = \sum_j e^{-2\pi ikj/(d+1)} s_j$$

- $\text{div } v = \frac{\partial v^1}{\partial x^1} + \frac{\partial v^2}{\partial x^2} = 0$ since $v^1$ is independent of $x^1$ and $v^2$ is independent of $x^2$

- Model describes turbulent advection of tracers in a cross-sweep flow along the $x^1$ direction; develops non-Gaussian statistics

Here, we consider that the state space $A$ of the base dynamics is the attractor of the L96 system with the corresponding flow map $\Phi_t$

- We build a tensor product basis of $L^2(M, \mu)$ using diffusion eigenfunctions for $L^2(A, \alpha)$ and Fourier functions for $L^2(X, \xi)$
Tracer evolution – quasiperiodic regime

**Movie.** Imaginary parts of the observables $W_t f_1$ and $W_t f_2$ for 
$f_1(a, x) = e^{i x_1}$ and $f_2(a, x) = e^{i x_2}$ and L96 parameters $d = 40$, $F = 4$, 
computed using $n_A = 65$ diffusion eigenfunctions, $n_X = 2 \times 32 + 1$ Fourier 
modes, $\epsilon = 0.0001$, and matrix exponential relative tolerance $\delta = 10^{-8}$
Density evolution – quasiperiodic regime

**Movie.** Evolution of the marginal density $\sigma_t$ associated with an initial density $\rho(a, x) = \rho_A(a)\rho_X(x)$,

$$
\rho_A(a) = p_\tilde{\epsilon}(b, a), \quad \rho_X(x) = \frac{1}{l_0^2(\kappa)} e^{-\kappa (\cos x^1 + \cos x^2)},
$$

with $\kappa = 3$ and the L96 parameters $d = 40$ and $F = 4$, computed using $n_A = n_X = 2 \times 32 + 1$ Fourier modes, $\epsilon = 0.0001$, and relative tolerance $\delta = 10^{-8}$ $\tilde{\kappa} = 3$.
Conclusions

Koopman and Perron-Frobenius operator techniques combined with kernel algorithms from machine learning provide a promising approach for pattern extraction and model-free statistical prediction in ergodic and skew-product systems.

- The approach relies on representation of the Koopman group and its generator in a smooth data-driven basis, requiring no a priori knowledge of the state space manifold and/or the equations of motion.
- Approximate Koopman eigenfunctions are selected on the basis of a Dirichlet energy criterion.
- Method is able to evolve full probability densities (important for UQ).

Future work includes:

- Efficiency improvements by replacing the tensor product basis by adapted and/or multiresolution bases.
- Implementation of the Lagrangian tracer scheme in domains with complex geometries.
References