# MATH 146
# Current Problems in Applied Mathematics: Dynamical Systems and Data Science

Dimitris Giannakis
dimitrios.giannakis@dartmouth.edu

Dartmouth College
Winter Term 2022

Section 1

# Introduction

# Ergodic theory

Ergodic theory studies the statistical behavior of measurable actions of groups or semigroups on spaces.

**Definition 1.1.**
A left action, or flow, of a (semi)group $G$ on a set $\Omega$ is a map $G \times \Omega \to \Omega$ with the following properties:

1. $\Phi(e, \omega) = \omega$, for the the identity element $e \in G$ and all $\omega \in \Omega$.
2. $\Phi(gh, \omega) = \Phi(g, \Phi(h, \omega))$, for all $g, h \in G$ and $\omega \in \Omega$.

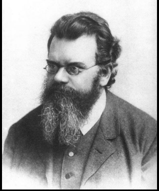
The set $\Omega$ is called the state space.

In this course, $G$ will be an abelian group or semigroup that represents the time domain. Common choices include:

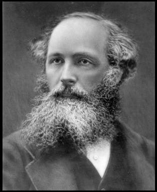
$$\mathbb{N}, \qquad \mathbb{Z}, \qquad \mathbb{R}_+, \qquad \mathbb{R}.$$

We write $\Phi^g \equiv \Phi(g, \cdot)$, $n \in \mathbb{N}, \mathbb{Z}$, and $t \in \mathbb{R}_+, \mathbb{R}$.

# Ergodic theory



Ludwig Boltzmann



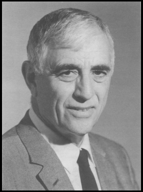
James Clerk Maxwell

Ergodic theory has its origin in the mid 19th century with the work of Boltzmann and Maxwell on statistical mechanics.

The term ergodic is an amalgamation of the Greek words ergo (έργο), which means *work*, and odos (οδός), which means *street*.

# Ergodic theory


George David Birkhoff


Bernard Osgood Koopman


John von Neumann

The mathematical foundations of the subject were established by Koopman, von Neumann, Birkhoff, and many others, in work dating to the 1930s.

Modern ergodic theory is a highly diverse subject with connections to functional analysis, harmonic analysis, probability theory, topology, geometry, number theory, and other mathematical disciplines.

# Observables and ergodic hypothesis

Rather than studying the flow $\Phi$ directly, ergodic theory focuses on its induced action on linear spaces of observables, e.g.,

$$\mathcal{F} = \{f : \Omega \to \mathcal{Y}\},$$

for a vector space $\mathcal{Y}$ (oftentimes, $\mathcal{Y} = \mathbb{R}$ or $\mathbb{C}$).

Drawing on intuition from mechanical systems, Boltzmann postulated that time averages of observables should well-approximate expectation values with respect to a reference distribution, $\mu$.

This is encapsulated in the ergodic hypothesis,

$$\underbrace{\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\Phi^n(\omega))}_{\text{time average}} = \underbrace{\int_\Omega f \, d\mu}_{\text{space average}} ,$$

which is stipulated to hold for typical initial conditions $\omega \in \Omega$ and observables $f : \Omega \to \mathcal{Y}$ in a suitable class.

# Operator-theoretic perspective

**Definition 1.2.**

1. For every $g \in G$, the composition operator, or Koopman operator is the linear map $U^g : \mathcal{F} \to \mathcal{F}$ defined as
$$U^g = f \circ \Phi^g.$$
2. The transfer operator $P^g : \mathcal{F}' \to \mathcal{F}'$ is the transpose of $U^g$, defined as
$$P^g \mu = \mu \circ U^g.$$

Koopman and transfer operators allow the study of nonlinear dynamics using techniques from linear operator theory.

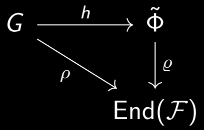
A central theme of this course is that operator-theoretic techniques also provide a bridge between dynamical systems theory and data science.

# Connections with representation theory

Observe that the set $\tilde{\Phi} = \{\Phi^g \mid g \in G\}$ equipped with composition of maps forms a group.

1. $h : G \to \tilde{\Phi}$ with $h(g) = \Phi^g$ is a group homomorphism.
2. $\varrho : \tilde{\Phi} \to \mathrm{End}(\mathcal{F})$ with $\varrho(\Phi^{)}) = U^g$ is a representation.

Using operator-theoretic techniques, we study the dynamics through the induced representation $\rho : G \to \mathrm{End}(\mathcal{F})$, where $\rho = \varrho \circ h$:



# Examples

## Circle rotation in continuous time

- $G = \mathbb{R}$, $\Omega = S^1$.
- Frequency parameter $\alpha \in \mathbb{R}$.
- $\Phi^t(\theta) = \theta + \alpha t \mod 2\pi$.

# Examples

Rational circle rotation in discrete time

- $G = \mathbb{Z}$, $\Omega = S^1$.
- Rotation angle $A \in [0, 2\pi)$, $A/(2\pi) \in \mathbb{Q}$.
- $\Phi^1(\theta) \equiv \Phi(\theta) = \theta + A \mod 2\pi$.

# Examples

Irrational circle rotation in discrete time

- $G = \mathbb{Z}$, $\Omega = S^1$.
- Rotation angle $A \in [0, 2\pi)$, $A/(2\pi) \notin \mathbb{Q}$.
- $\Phi^1(\theta) \equiv \Phi(\theta) = \theta + A \mod 2\pi$.

# Examples

## Doubling map

- $G = \mathbb{N}$, $\Omega = S^1$.
- $\Phi(\theta) = 2\theta \mod 2\pi$.

# Examples

Rational torus flow

- $G = \mathbb{R}$, $\Omega = \mathbb{T}^2$.
- Frequency vector $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$, $\alpha_1/\alpha_2 \in \mathbb{Q}$.
- $\Phi^t(\theta) = \theta + \alpha t \mod 2\pi$.

# Examples

## Irrational torus flow

- $G = \mathbb{R}$, $\Omega = \mathbb{T}^2$.
- Frequency vector $\alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$, $\alpha_1/\alpha_2 \notin \mathbb{Q}$.
- $\Phi^t(\theta) = \theta + \alpha t \mod 2\pi$.

# Examples

Arnold cat map

- $G = \mathbb{Z}$, $\Omega = \mathbb{T}^2$.
- $\Phi(\theta) = A\theta \mod 2\pi$, $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$.

# Examples

Lorenz 63 system

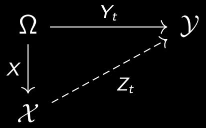
- $G = \mathbb{R}$, $\Omega = \mathbb{R}^3$.
- $\Phi^t(x)$ is the solution map of the initial-value problem

$$\dot{x}(t) = v(x(t)), \quad x(0) = x$$

with

$$v(y) = (v_1, v_2, v_3)$$
$$v_1 = \sigma(x_2 - x_1), \quad v_2 = x_1(\rho - x_3), \quad v_3 = x_1 x_2 - \beta x_3,$$
$$\rho = 28, \sigma = 10, \quad \beta = 8/3.$$

# Dynamical systems and data science




**Given.** Time-ordered samples $(x_0, y_0), (x_1, y_1), \dots, (x_{N-1}, y_{N-1})$ of observables $X : \Omega \to \mathcal{X}$ (covariate) and $Y : \Omega \to \mathcal{Y}$ (response), where $\mathcal{Y}$ is a vector space and

$$x_n = X(\omega_n), \quad y_n = Y(\omega_n), \quad \omega_n = \Phi^{t_n}(\omega_0), \quad t_n = (n-1)\,\Delta t.$$

**Problem 1 [forecasting].** Using the data $(x_n, y_n)$, construct ("learn") a function $Z_t : \mathcal{X} \to \mathcal{Y}$ that predicts $Y$ at a lead time $t \geq 0$. That is, $Z_t$ should have the property that $Z_t \circ X$ is closest to $Y_t := Y \circ \Phi^t$ among all functions in a suitable class.

**Problem 2 [coherent pattern extraction].** Using the data $x_n$, identify a collection of observables $\zeta_j : \Omega \to \mathcal{Y}$ which have the property of evolving coherently under the dynamics in a suitable sense.

# Dynamical systems and data science

In this course, we explore various approaches for pointwise approximation (for Problem 1) and spectral analysis (for Problem 2) of Koopman/transfer operators.

A major requirement is that the approximations are refinable, i.e., the learned functions $Z_t$ and $\varphi_j$ should have well-controlled limits as $N \to \infty$.

**Challenges.** Linear operators on infinite-dimensional function spaces can exhibit qualitatively new features which are not present in finite-dimensional linear algebra, including:

1. Discontinuous (unbounded) linear maps.
2. Elements of the spectrum which are not eigenvalues (e.g., continuous spectrum).

# Further reading


[1] T. Berry, D. Giannakis, and J. Harlim, "Bridging data science and dynamical systems theory," *Notices Amer. Math. Soc.*, vol. 67, no. 9, pp. 1336–1349, 2020. DOI: 10.1090/noti2151.

[2] F. Cucker and S. Smale, "On the mathematical foundations of learning," *Bull. Amer. Math. Soc.*, vol. 39, no. 1, pp. 1–49, 2001. DOI: 10.1090/S0273-0979-01-00923-5.

[3] M. Dellnitz and O. Junge, "On the approximation of complicated dynamical behavior," *SIAM J. Numer. Anal.*, vol. 36, p. 491, 1999. DOI: 10.1137/S0036142996313002.

[4] T. Eisner, B. Farkas, M. Haase, and R. Nagel, *Operator Theoretic Aspects of Ergodic Theory*, ser. Graduate Texts in Mathematics. Springer, 2015, vol. 272.

[5] B. O. Koopman, "Hamiltonian systems and transformation in Hilbert space," *Proc. Natl. Acad. Sci.*, vol. 17, no. 5, pp. 315–318, 1931. DOI: 10.1073/pnas.17.5.315.

# Further reading

[6] I. Mezić, "Spectral properties of dynamical systems, model reduction and decompositions," *Nonlinear Dyn.*, vol. 41, pp. 309–325, 2005. DOI: 10.1007/s11071-005-2824-x.

# Section 2

## Measure-preserving transformations; Ergodic theorems

# Measure-preserving dynamical systems

**Definition 2.1.**
Let $(\Omega, \Sigma, \mu)$ be a measure space.

1. A measurable map $T : \Omega \to \Omega$ is said to be measure-preserving if $T_*\mu = \mu$, i.e.,
$$\mu(T^{-1}(S)) = \mu(S), \quad \forall S \in \Sigma.$$
Conversely, we say that $\mu$ is an invariant measure for $T$.
2. A measure-preserving map $T : \Omega \to \Omega$ is said to be invertible measure-preserving if $T$ is bijective and $T^{-1}$ is also measure-preserving.
3. A measurable action $\Phi : G \times \Omega \to \Omega$ is $\mu$-preserving if $\Phi^g : \Omega \to \Omega$ is $\mu$-preserving for every $g \in G$.

# Recurrence

**Theorem 2.2 (Poincaré).**

*Let $T : \Omega \to \Omega$ be a measure-preserving transformation of the probability space $(\Omega, \Sigma, \mu)$. Let $S \in \Sigma$ be a measurable set with $\mu(S) > 0$. Then, under iteration by $T$, almost every point of $S$ returns to $S$ infinitely often. That is, for $\mu$-a.e. $\omega \in S$, there exists a sequence $n_1 < n_2 < n_3 < \cdots$ of natural numbers, increasing to infinity, such that $T^{n_j}(\omega) \in S$ for all $j$.*

# Ergodicity

**Definition 2.3.**
Let $(\Omega, \Sigma, \mu)$ be a probability space.

1. A measurable map $T : \Omega \to \Omega$ is said to be ergodic if for every $T$-invariant set, i.e., every $S \in \Sigma$ such that $T^{-1}(S) = S$ we have either $\mu(S) = 0$ or $\mu(S) = 1$.
2. A measurable action $\Phi : G \times \Omega \to \Omega$ is ergodic if for every $S \in \Sigma$ such that $\Phi^{-g}(S) = S$ for all $g \in G$ we have either $\mu(S) = 0$ or $\mu(S) = 1$.

# Measure-theoretic characterization of ergodicity

**Theorem 2.4.**
*Let $T : \Omega \to \Omega$ be a measure-preserving transformation of the probability space $(\Omega, \Sigma, \mu)$. Then, the following are equivalent.*

1. *$T$ is ergodic.*
2. *The only measurable sets $S \in \Sigma$ such that $\mu(T^{-1}(S) \triangle S) = 0$ have either $\mu(S) = 0$ or $\mu(S) = 1$.*
3. *For every $S \in \Sigma$ with $\mu(S) > 0$, we have $\mu(\bigcup_{n=1}^{\infty} T^{-1}(S)) = 1$.*
4. *For every $R, S \in \Sigma$ with $\mu(R) > 0$ and $\mu(S) > 0$, there exists $n > 0$ with $\mu(T^{-n}(R) \cap S) > 0$.*

# Measure-theoretic characterization of ergodicity

**Theorem 2.5.**
*Let* $(\Omega, \Sigma, \mu)$ *be a probability space.*

1. *A measure-preserving action* $\Phi : \mathbb{N} \times \Omega \to \Omega$ *is ergodic iff*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(\Phi^{-n}(R) \cap S) = \mu(R)\mu(S), \quad \forall R, S \in \Sigma.$$

2. *A measure-preserving action* $\Phi : \mathbb{R}_+ \times \Omega \to \Omega$ *is ergodic iff*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \mu(\Phi^{-t}(R) \cap S)\, dt = \mu(R)\mu(S), \quad \forall R, S \in \Sigma.$$

# Koopman operators on $L^p$ spaces

**Definition 2.6.**
A measurable map $T : \Omega \to \Omega$ on a measure space $(\Omega, \Sigma, \mu)$ is said to be nonsingular if it preserves null sets, i.e., if whenever $\mu(S) = 0$ we have $T_*\mu(S) = \mu(T^{-1}(S)) = 0$.

**Notation.**

- $\mathbb{L}(\Sigma) = \{f : \Omega \to \mathbb{R} : f \text{ is } \Sigma\text{-measurable}\}$.
- $L(\mu) = \{[f]_\mu : f \in \mathbb{L}(\Sigma)\}$.
- $L^p(\mu) = \{[f]_\mu \in L(\mu) : \int_\Omega |f|^p \, d\mu < \infty\}$ .
- $L^\infty(\mu) = \{[f]_\mu \in L(\mu) : \operatorname{esssup}_\mu |f| < \infty\}$.

# Koopman operators on $L^p$ spaces

**Proposition 2.7.**

*With notation as above, the following hold.*

1. *If $T$ is measurable, then the composition map $U : f \mapsto f \circ T$ maps $\mathbb{L}(\Sigma)$ into itself.*
2. *If $T$ is nonsingular, then $\mathcal{U} : L(\mu) \to L(\mu)$ with $\mathcal{U}[f]_\mu = [Uf]_\mu$ is a well-defined linear map.*
3. *If $T$ is nonsingular, then $L^\infty(\mu)$ is invariant under $\mathcal{U}$, i.e.,*

$$\mathcal{U} L^\infty(\mu) \subseteq L^\infty(\mu).$$

4. *If $T$ is measure-preserving, then $\mathcal{U}$ is an isometry of $L^p(\mu)$, $1 \leq p \leq \infty$, i.e.,*

$$\|\mathcal{U}[f]_\mu\|_{L^p(\mu)} = \|[f]_\mu\|_{L^p(\mu)}.$$

5. *If $T$ is invertible measure-preserving, then $\mathcal{U}$ is an isomorphism of $L^p(\mu)$, $1 \leq p \leq \infty$, i.e., $\mathcal{U}$ and $\mathcal{U}^{-1}$ are both isometries.*

Henceforth, we abbreviate $[f]_\mu \equiv f$, $U \equiv \mathcal{U}$.

# Koopman operators on $L^2$

**Notation.**

- $\langle f_1, f_2 \rangle_{L^2(\mu)} = \int_\Omega f_1 f_2 \, d\mu$.

The Koopman operator induced by a $\mu$-preserving map $T : \Omega \to \Omega$ preservers Hilbert space inner products,

$$\langle Uf_1, Uf_2 \rangle_{L^2(\mu)} = \langle f_1, f_2 \rangle_{L^2(\mu)}.$$

If, in addition, $T$ is invertible measure-preserving, then $U$ is a unitary operator,

$$U^* = U^{-1}.$$

# Duality of $L^p$ spaces

**Notation.**
For a probability space $(\Omega, \Sigma, \mu)$, we let:

- $M_q(\Omega, \mu) = \left\{ \text{measures } \nu \ll \mu \text{ with density } \frac{d\nu}{d\mu} \in L^q(\mu) \right\}$.
- Duality pairing: $\langle \cdot, \cdot \rangle_\mu : L^p(\mu)^* \times L^p(\mu) \to \mathbb{R}$, $\langle \alpha, f \rangle_\mu = \alpha f$.

For $1 \leq p < \infty$, we can identify functionals in $L^p(\mu)^*$ with measures in $M_q(\Omega, \mu)$, $\frac{1}{p} + \frac{1}{q} = 1$, through the map $\iota_q : M_q(\Omega, \mu) \to L^p(\mu)^*$,

$$(\iota_q \nu) f = \int_\Omega f \rho \, d\mu, \quad \rho = \frac{d\nu}{d\mu}.$$

Equipping $M_q(\Omega, \mu)$ with the norm

$$\|\nu\|_{M_q(\Omega, \nu)} = \left\| \frac{d\nu}{d\mu} \right\|_{L^q(\mu)},$$

$\iota_q$ becomes an isomorphism of Banach spaces. Thus, we have

$$L^p(\mu)^* \simeq M_q(\Omega, \mu) \simeq L^q(\mu), \quad 1 \leq p < \infty, \quad \frac{1}{p} + \frac{1}{q} = 1.$$

# Transfer operators on $L^p$

**Definition 2.8.**
With the notation of Proposition 2.7, the transfer operator $P : L^1(\mu) \to L^1(\mu)$ is is the unique operator satisfying

$$\int_S Pf \, d\mu = \int_{T^{-1}(S)} f \, d\mu, \quad \forall f \in L^1(\mu).$$

We define $P : L^p(\mu) \to L^p(\mu)$, $1 < p \leq \infty$ by restriction of $P : L^1(\mu) \to L^1(\mu)$.

**Proposition 2.9.**
*Under the identification $L^1(\mu)^* \simeq L^\infty(\mu)$, the transpose $P' : L^1(\mu)^* \to L^1(\mu)^*$ of the transfer operator $P : L^1(\mu) \to L^1(\mu)$ is identified with the Koopman operator $U : L^\infty(\mu) \to L^\infty(\mu)$; that is,*

$$\int_\Omega f(Pg) \, d\mu = \int_\Omega (Uf) g \, d\mu, \quad \forall f \in L^\infty(\mu), \quad \forall g \in L^1(\mu).$$

# Duality between Koopman and transfer operators

**Proposition 2.10.**

*Let $1 \leq p < \infty$. Then, under the identification $L^p(\mu)^* \simeq L^q(\mu)$, $\frac{1}{p} + \frac{1}{q} = 1$, the following hold:*

1. *The transpose $U' : L^p(\mu)^* \to L^p(\mu)^*$ of the Koopman operator $U : L^p(\mu) \to L^p(\mu)$ is identified with the transfer operator $P : L^q(\mu) \to L^q(\mu)$; that is,*

$$\langle f, Ug \rangle_\mu = \langle Pf, g \rangle_\mu, \quad \forall f \in L^q(\mu), \quad \forall g \in L^p(\mu).$$

2. *The transpose $P' : L^p(\mu)^* \to L^p(\mu)^*$ of the transfer operator $P : L^p(\mu) \to L^p(\mu)$ is identified with the Koopman operator $U : L^q(\mu) \to L^q(\mu)$; that is,*

$$\langle f, Pg \rangle_\mu = \langle Uf, g \rangle_\mu, \quad \forall f \in L^q(\mu), \quad \forall g \in L^p(\mu).$$

# Duality between Koopman and transfer operators

**Corollary 2.11.**

1. *For* $1 < p < \infty$, $U : L^p(\mu) \to L^p(\mu)$ *and* $P : L^p(\mu) \to L^p(\mu)$ *satisfy*
$$U = U'', \quad P = P''.$$
2. *In the Hilbert space case,* $p = 2$, *we have* $P = U^*$.
3. *For* $1 \leq p \leq \infty$, $P$ *has unit operator norm,* $\|P\|_{L^p(\mu)} = 1$.

**Lemma 2.12.**
*With the notation of Proposition 2.8, if* $T : \Omega \to \Omega$ *is invertible measure-preserving then* $P : L^p(\mu) \to L^p(\mu)$ *is the inverse of* $U : L^p(\mu) \to L^p(\mu)$, $P = U^{-1}$.

# Spectral characterization of ergodicity

Observe that the Koopman operator $U : \mathcal{F} \to \mathcal{F}$ on any function space $\mathcal{F}$ has an eigenvalue equal to 1 with a constant corresponding eigenfunction, $\mathbb{1} : \Omega \to \mathbb{R}$,

$$U\mathbb{1} = \mathbb{1}, \quad \mathbb{1}(\omega) = 1.$$

**Theorem 2.13.**
*Let $T : \Omega \to \Omega$ be a measure-preserving transformation of a probability space $(\Omega, \Sigma, \mu)$. Then, $\mu$ is ergodic iff the eigenvalue equal to 1 of the associated Koopman operator $U$ on $L(\mu)$ (and thus on any of the $L^p(\mu)$ spaces with $1 \leq p \leq \infty$) is simple, i.e.,*

$$Uf = f \implies f = \text{const. } \mu\text{-a.e.}$$

# Spectral characterization of ergodicity

**Theorem 2.14.**

1. *Let* $\Phi : \mathbb{N} \times \Omega \to \Omega$ *be a measure-preserving action and* $U^n$, $n \in \mathbb{N}$, *the associated Koopman operators on any of* $L(\mu)$ *or* $L^p(\mu)$, $1 \leq p \leq \infty$. *Then* $\Phi$ *is ergodic iff* $U^n f = f$ *for all* $n \in \mathbb{N}$ *implies that* $f$ *is constant* $\mu$*-a.e.*
2. *Let* $\Phi : \mathbb{R}_+ \times \Omega \to \Omega$ *be a measure-preserving action and* $U^t$, $t \in \mathbb{R}_+$, *the associated Koopman operators on any of* $L(\mu)$ *or* $L^p(\mu)$, $1 \leq p \leq \infty$. *Then,* $\Phi$ *is ergodic iff* $U^t f = f$ *for all* $t \in \mathbb{R}_+$ *implies that* $f$ *is constant* $\mu$*-a.e.*

# Pointwise ergodic theorem

**Theorem 2.15 (Birkhoff).**

*Let $T : \Omega \to \Omega$ be a measure-preserving transformation of a probability space $(\Omega, \Sigma, \mu)$ with associated Koopman operator $U : L^1(\mu) \to L^1(\mu)$. Then, for every $f \in L^1(\mu)$ and $\mu$-a.e. $\omega \in \Omega$,*

$$f_N(\omega) := \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(\omega))$$

*converges to a function $\bar{f} \in L^1(\mu)$ that satisfies*

$$U\bar{f} = \bar{f}, \quad \int_\Omega f \, d\mu = \int_\Omega \bar{f} \, d\mu.$$

*In particular, if $T$ is ergodic, then for $\mu$-a.e. $\omega \in \Omega$,*

$$\bar{f}(\omega) = \int_\Omega f \, d\mu.$$

# Mean ergodic theorem

**Theorem 2.16 (von Neumann).**

*Let $T : \Omega \to \Omega$ be a measure-preserving transformation of a probability space $(\Omega, \Sigma, \mu)$ with associated Koopman operator $U : L^2(\mu) \to L^2(\mu)$. Let $\Pi : L^2(\mu) \to L^2(\mu)$ be the orthogonal projection onto the eigenspace of $U$ corresponding to eigenvalue 1. Then, the sequence of operators $U_N = N^{-1} \sum_{n=0}^{N-1} U^n$ converges strongly to $\Pi$, i.e.,*

$$\lim_{N\to\infty} U_N f = \Pi f, \quad \forall f \in L^2(\mu).$$

*In particular, if $T$ is ergodic, $\Pi$ is the projection onto the 1-dimensional subspace of $L^2(\mu)$ containing $\mu$-a.e. constant functions, i.e.,*

$$\Pi f = \langle \mathbb{1}, f \rangle_{L^2(\mu)} \mathbb{1} = \left( \int_\Omega f \, d\mu \right) \mathbb{1}.$$

# Topological dynamics

Of particular interest is the case where $(G, \tau_G)$ and $(\Omega, \tau_\Omega)$ are topological spaces and $\Phi : G \times \Omega \to \Omega$ is a continuous, and thus Borel-measurable, action. We let $\mathfrak{B}(\Omega)$ denote the Borel $\sigma$-algebra of $\Omega$.

**Definition 2.17.**
The support of a measure $\mu : \mathfrak{B}(\Omega) \to [0, \infty]$ is the set

$$\operatorname{supp} \mu := \{\omega \in \Omega : \mu(N_\omega) > 0, \forall N_\omega \in \tau_\Omega\}.$$

**Lemma 2.18.**
*With notation as above, the following hold.*

1. $\operatorname{supp} \mu$ *is a closed (and thus Borel-measurable) subset of* $\Omega$.
2. *If* $\Omega$ *is Hausdorff, and* $\mu$ *is a Radon measure, every Borel-measurable set* $S \subset \Omega \setminus \operatorname{supp} \mu$ *has* $\mu(S) = 0$.
3. *If* $\mu$ *is invariant under a continuous map* $T : \Omega \to \Omega$, *then* $\operatorname{supp} \mu$ *is also invariant,*

$$T^{-1}(\operatorname{supp} \mu) \subseteq \operatorname{supp} \mu.$$

# Existence of invariant measures

**Theorem 2.19 (Krylov-Bogoliubov).**

*Let $(\Omega, \tau_\Omega)$ be a compact metrizable space and $T : \Omega \to \Omega$ a continuous map. Then, there exists an invariant Borel probability measure under $T$.*

# Existence of dense orbits

**Theorem 2.20.**
*Let* $(\Omega, \tau_\Omega)$ *be a compact metrizable space,* $T : \Omega \to \Omega$ *a continuous map, and* $\mu$ *an ergodic, invariant Borel probability measure with* $\operatorname{supp} \mu = \Omega$*. Then,* $\mu$*-a.e.* $\omega \in \Omega$ *has a dense orbit* $\{T^n(\omega)\}_{n=0}^{\infty}$*.*

# Geometry of invariant measures

**Theorem 2.21.**
*Let $T : \Omega \to \Omega$ be a continuous map on a compact metrizable space. Let $\mathcal{M}(\Omega; T)$ denote the set of $T$-invariant Borel probability measures on $\Omega$. Then, the following hold:*

1. *$\mathcal{M}(\Omega; T)$ is a weak-$^*$ compact, convex space.*
2. *$\mu$ is an extreme point of $\mathcal{M}(\Omega; T)$ iff it is ergodic.*
3. *If $\mu$ and $\nu$ are distinct, ergodic measures in $\mathcal{M}(\Omega; T)$, then they are mutually singular.*

# Equidistributed sequences

**Definition 2.22.**
Let $T : \Omega \to \Omega$ be a continuous map on a compact metrizable space $(\Omega, \tau_\Omega)$ and $\mu$ a Borel probability measure. A sequence $\omega_0, \omega_1, \dots$ with $\omega_n = T^n(\omega_0)$ is said to be $\mu$-equidistributed if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\omega_n) = \int_\Omega f \, d\mu, \quad \forall f \in C(\Omega).$$

**Remark.**
$\mu$-equidistribution of $\omega_0, \omega_1, \dots$ is equivalent to weak-$^*$ convergence of the sampling measures $\mu_N = N^{-1} \sum_{n=0}^{N-1} \delta_{\omega_n}$ to the measure $\mu$.

# Basin of a measure

**Definition 2.23.**
With the notation of Definition 2.22 the basin of $\mu$ is the set

$$\mathcal{B}(\mu) = \{\omega_0 \in \Omega : \omega_0, \omega_1, \ldots \text{ is } \mu\text{-equidistributed}\}.$$

By the pointwise ergodic theorem (Theorem 2.15), if $\Omega$ is a metrizable space and $\mu$ is an ergodic invariant measure with compact support, then $\mu$-a.e. $\omega \in \Omega$ lies in $\mathcal{B}(\mu)$.

# Observable measures

**Definition 2.24.**
With the notation of Definition 2.23, let $\nu$ be a reference Borel probability measure on $\Omega$. The measure $\mu$ is said to be $\nu$-observable if there exists a Borel set $S \in \mathfrak{B}(\Omega)$ with $\nu(S) > 0$ such that $\nu$-a.e. $\omega \in S$ lies in $\mathcal{B}(\mu)$.

Intuitively, we think of $\nu$ as the measure from which we draw initial conditions. $\nu$-observability of $\mu$ then means that the statistics of observables with respect to $\mu$ can be approximated from experimentally accessible initial conditions.

# Koopman operators on spaces of continuous functions

**Proposition 2.25.**

*Let $T : \Omega \to \Omega$ be a continuous map on a locally compact Hausdorff space. Then, the Koopman operator $U : f \mapsto f \circ T$ is well-defined as a linear map from $C(\Omega)$ into itself. Moreover:*

1. *$U$ is a contraction, i.e.,*

$$\|Uf\|_{C(\Omega)} \leq \|f\|_{C(\Omega)}, \quad \forall f \in C(\Omega),$$

   *with equality if $T$ is invertible.*

2. *$U$ has operator norm $\|U\| = 1$.*

3. *$U$ has the properties*

$$U(fg) = (Uf)(Ug), \quad U(f^*) = (Uf)^*, \quad \forall f, g \in C(\Omega),$$

   *i.e., it is a ${}^*$-homomorphism of the $C^*$-algebra $C(\Omega)$.*

# Transfer operators on Borel measures

**Notation.**

- $M(\Omega)$: Space of signed Borel measures on topological space $(\Omega, \tau_\Omega)$.

**Definition 2.26.**
Let $T : \Omega \to \Omega$ be a continuous map on a compact metrizable space. The transfer operator $P : C(\Omega)^* \to C(\Omega)^*$ is the transpose (dual) operator to the Koopman operator $U : C(\Omega) \to C(\Omega)$,

$$P\alpha = \alpha \circ U.$$

# Unique ergodicity

**Definition 2.27.**
Let $T : \Omega \to \Omega$ be a continuous map on a compact metrizable space $(\Omega, \tau_\Omega)$. $T$ is said to be uniquely ergodic if there is only one $T$-invariant Borel probability measure.

**Theorem 2.28.**
*With notation as above, the following are equivalent.*

1. *$T$ is uniquely ergodic.*
2. *For every $f \in C(\Omega)$, $N^{-1}\sum_{n=0}^{N-1} f(T^n(\omega))$ converges to a constant, uniformly with respect to $\omega \in \Omega$.*
3. *For every $f \in C(\Omega)$, $N^{-1}\sum_{n=0}^{N-1} f(T^n(\omega))$ converges pointwise to a constant.*
4. *There exists an invariant Borel probability measure $\mu$ such that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(\omega)) = \int_\Omega f\, d\mu, \quad \forall \omega \in \Omega.$$

# Strong and weak continuity of continuous-time (semi)flows

**Theorem 2.29.**
*Let $\Phi^t : \Omega \to \Omega$, $t \geq 0$, be a continuous-time, continuous, semiflow on a compact metrizable space $\Omega$ with associated Koopman operators $U^t : C(\Omega) \to C(\Omega)$. Then, as $t \to 0$, $U^t$ converges strongly to the identity,*

$$\lim_{t \to 0} \| U^t f - f \|_{C(\Omega)} = 0, \quad \forall f \in C(\Omega).$$

**Theorem 2.30.**
*Let $\Phi^t : \Omega \to \Omega$, $t \geq 0$, be a continuous-time, measurable semiflow with invariant probability measure $\mu$ and associated Koopman operators $U^t : L^p(\mu) \to L^p(\mu)$. Then, the following hold as $t \to 0$:*

1. *For $1 \leq p < \infty$, $U^t$ converges strongly to the identity,*

$$\lim_{t \to 0} \| U^t f - f \|_{L^p(\mu)} = 0, \quad \forall f \in L^p(\mu).$$

2. *For $p = \infty$, $U^t$ converges in weak-$^*$ sense to the identity,*

$$\lim_{t \to 0} \int_\Omega g (U^t f) \, d\mu = \int_\Omega g f \, d\mu, \quad \forall f \in L^\infty(\mu), \quad \forall g \in L^1(\mu).$$

# Mixing

Recall from Theorem 2.4 that a measure-preserving transformation is ergodic iff

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}(R) \cap S) = \mu(R)\mu(S), \quad \forall R, S \in \Sigma.$$

**Definition 2.31.**
Let $T : \Omega \to \Omega$ be a measure-preserving transformation of the probability space $(\Omega, \Sigma, \mu)$.

1. $T$ is said to be weak-mixing if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |\mu(T^{-n}(R) \cap S) - \mu(R)\mu(S)| = 0, \quad \forall R, S \in \Sigma.$$

2. $T$ is said to be strong-mixing, or mixing, if

$$\lim_{n\to\infty} \mu(T^{-n}(R) \cap S) = \mu(R)\mu(S), \quad \forall R, S \in \Sigma.$$

# Mixing

**Theorem 2.32.**
*Let $T : \Omega \to \Omega$ be a measure-preserving transformation of the probability space $(\Omega, \Sigma, \mu)$. Then, the following are equivalent.*

1. *$T$ is weak-mixing.*
2. *There is a subset $\mathcal{N} \subset \mathbb{N}$ of zero density such that*

$$\lim_{\substack{n \to \infty \\ n \notin \mathcal{N}}} \mu(T^{-n}(R) \cap S) = \mu(R)\mu(S), \quad \forall R, S \in \Sigma.$$

# Observable-centric characterization of ergodicity and mixing

Let $T : \Omega \to \Omega$ be a measure-preserving transformation of the probability space $(\Omega, \Sigma, \mu)$. Let $U : L^2(\mu) \to L^2(\mu)$ be the associated Koopman operator on $L^2$.

For $f, g \in L^2(\mu)$, define the cross-correlation function $C_{fg} : \mathbb{N} \to \mathbb{R}$, where

$$C_{fg}(n) = \langle f, U^n g \rangle_{L^2(\mu)},$$

and the autocorrelation function $C_f = C_{ff}$.

Consider also the expectation values $\bar{f} = \int_\Omega f \, d\mu$ and $\bar{g} = \int_\Omega g \, d\mu$.

**Theorem 2.33.**
*With notation as above, the following are equivalent.*

1. *$T$ is ergodic.*
2. *For all $f, g \in L^2(\mu)$, $\lim_{n\to\infty} N^{-1} \sum_{n=0}^{N-1} C_{fg}(n) = \bar{f}\bar{g}$.*
3. *For all $f \in L^2(\mu)$, $\lim_{n\to\infty} N^{-1} \sum_{n=0}^{N-1} C_f(n) = \bar{f}^2$.*

# Observable-centric characterization of ergodicity and mixing

**Theorem 2.34.**
*With notation as above, the following are equivalent.*

1. $T$ *is weak-mixing.*
2. *For all* $f, g \in L^2(\mu)$, $\lim_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} |C_{fg}(n) - \bar{f}\bar{g}| = 0.$
3. *For all* $f \in L^2(\mu)$, $\lim_{N\to\infty} N^{-1} \sum_{n=0}^{N-1} |C_f(n) - \bar{f}^2| = 0.$

**Theorem 2.35.**
*With notation as above, the following are equivalent.*

1. $T$ *is mixing.*
2. *For all* $f, g \in L^2(\mu)$, $\lim_{N\to\infty} C_{fg}(n) = \bar{f}\bar{g}.$
3. *For all* $f \in L^2(\mu)$, $\lim_{N\to\infty} C_f(n) = \bar{f}^2.$

# Spectral characterization of mixing

**Theorem 2.36.**
*Let* $T : \Omega \to \Omega$ *be a measure-preserving transformation of the probability space* $(\Omega, \Sigma, \mu)$, *and* $U : L^2(\mu) \to L^2(\mu)$ *the corresponding Koopman operator. Then,* $T$ *is weak-mixing iff the only eigenvalue of* $U$ *is the eigenvalue equal to 1.*

# Mixing and product flows

**Theorem 2.37.**
*Let $T : \Omega \to \Omega$ be a measure-preserving transformation of the probability space $(\Omega, \Sigma, \mu)$. Then, the following are equivalent.*

1. *$T$ is weak-mixing.*
2. *$T \times T$ is ergodic with respect to the product measure $\mu \times \mu$.*
3. *$T \times T$ is weak-mixing with respect to the product measure $\mu \times \mu$.*

# Further reading


[1] V. Baladi, *Positive Transfer Operators and Decay of Correlations*, ser. Advanced Series in Nonlinear Dynamics. Singapore: World Scientific, 2000, vol. 16.

[2] N. Edeko, M. Gerlach, and V. Kühner, "Measure-preserving semiflows and one-parameter Koopman semigrpoups," *Semigr. Forum*, vol. 98, pp. 48–63, 2019. DOI: 10.1007/s00233-018-9960-3.

[3] T. Eisner, B. Farkas, M. Haase, and R. Nagel, *Operator Theoretic Aspects of Ergodic Theory*, ser. Graduate Texts in Mathematics. Springer, 2015, vol. 272.

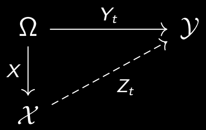
[4] P. Walters, *An Introduction to Ergodic Theory*, ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 1981, vol. 79.

Section 3

# Forecasting

# Setting

Recall the forecasting problem from Section 1:




**Given.** Time-ordered samples $(x_0, y_0), (x_1, y_1), \dots, (x_{N-1}, y_{N-1})$ of observables $X : \Omega \to \mathcal{X}$ (covariate) and $Y : \Omega \to \mathcal{Y}$ (response), where $\mathcal{Y}$ is a vector space and

$$x_n = X(\omega_n), \quad y_n = Y(\omega_n), \quad \omega_n = \Phi^{t_n}(\omega_0), \quad t_n = (n-1)\,\Delta t.$$

**Goal.** Using the data $(x_n, y_n)$, construct ("learn") a function $Z_t : \mathcal{X} \to \mathcal{Y}$ that predicts $Y$ at a lead time $t \geq 0$. That is, $Z_t$ should have the property that $Z_t \circ X$ is closest to $Y_t := Y \circ \Phi^t$ among all functions in a suitable class.

# General assumptions and notation

Throughout this section we assume:

1. $\Phi^t : \Omega \to \Omega$, $t \geq 0$, is a continuous, measure-preserving, ergodic semiflow on a compact metrizable space $\Omega$, with a Borel invariant probability measure $\mu$.
2. $X : \Omega \to \mathcal{X}$ is a continuous map into a metrizable space $\mathcal{X}$.
3. $Y : \Omega \to \mathcal{Y}$ is a continuous map into a Banach space $\mathcal{Y}$ (typically, $\mathcal{Y} = \mathbb{R}$).
4. The discrete-time $\Phi^{\Delta t} : \Omega \to \Omega$ is ergodic.

**Notation.**

- $M_p(\Omega; \mu) = \left\{ \text{measures } \nu \ll \mu \text{ with density } \frac{d\nu}{d\mu} \in L^p(\mu) \right\}$.
- $M_C(\Omega; \mu) = \left\{ \text{measures } \nu \ll \mu \text{ with density } \frac{d\nu}{d\mu} \in C(\Omega) \right\}$.
- $\mathcal{X}_\Omega = X(\Omega)$: Image of state space in covariate space.
- $\mu_{\mathcal{X}} = X_* \mu$: Pushforward of invariant measure into covariate space.

# Probabilistic initial conditions

We first consider the case where we assign to each initial condition $x \in \mathcal{X}$ with $x = X(\omega)$ a probability measure $p_x \in M_2(\Omega; \mu)$ with continuous density.

We let $\rho_x = \frac{dp_x}{d\mu} \in C(\Omega)$ be the density of $p_x$ relative to $\mu$.

**Algorithm 3.1 (construction of the density $\rho_x$).**

1. Pick a continuous, strictly positive kernel function $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$, e.g.,
$$\kappa(x, x') = \exp\left(-\frac{d^2_{\mathcal{X}}(x, x')}{\epsilon^2}\right), \quad \epsilon > 0.$$
2. Normalize $\kappa$ to a continuous Markov kernel $\rho : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$,
$$\rho(x, x') = \frac{\kappa(x, x')}{v(x)}, \quad v(x) = \int_{\mathcal{X}} \kappa(x, \cdot)\, d\mu_{\mathcal{X}}.$$
3. Set $\rho_x(\omega) = \rho(x, X(\omega))$.

# Target function

Assuming $Y \in L^2(\mu)$, define the target function $Z_t \in C(\mathcal{X}_\Omega)$ where

$$Z_t(x) = \mathbb{E}_{p_x}(U^t Y) = \langle \rho_x, U^t Y \rangle_{L^2(\mu)} \equiv \langle P^t \rho_x, Y \rangle_{L^2(\mu)}.$$

**Notation.**

- $\phi_0, \phi_1, \ldots$: Orthonormal basis functions of $L^2(\mu)$.
- $\Pi_L : L^2(\mu) \to L^2(\mu)$: Orthogonal projection onto $\mathrm{span}\{\phi_0, \ldots, \phi_{L-1}\}$.
- $U_L^{(t)} = \Pi_L U^t \Pi_L$: Finite-rank approximation of the Koopman operator.

**Proposition 3.2.**

*With notation as above, as $L \to \infty$ $U_L^{(t)}$ converges to $U^t$ weakly. As a result, $Z_{t,L} = \mathbb{E}_{p_x}(U_L^{(t)} Y)$ satisfies*

$$\lim_{L \to \infty} Z_{t,L}(x) = Z_t(x),$$

*where the convergence is uniform with respect to $x \in \mathcal{X}_\Omega$ and $t$ in compact sets.*

# Target function

**Algorithm 3.3 (evaluation of the target function).**

1. Represent $U_L^{(t)}$ by the $L \times L$ matrix $\boldsymbol{U}^{(t)} = [U_{ij}^{(t)}]$ with

$$U_{ij}^{(t)} = \langle \phi_i, U^t \phi_j \rangle_{L^2(\mu)}, \quad i, j = 0, \ldots, L-1.$$

2. Represent $\rho_x$ by the column vector $\hat{\boldsymbol{\rho}}_x = (\hat{\rho}_{x,0}, \ldots, \hat{\rho}_{x,L-1})^\top \in \mathbb{R}^L$ with

$$\hat{\rho}_{x,i} = \langle \phi_i, \rho_x \rangle_{L^2(\mu)}.$$

3. Represent $Y$ by the column vector $\hat{\boldsymbol{y}} = (\hat{y}_0, \ldots, \hat{y}_{L-1})^\top \in \mathbb{R}^L$ with

$$\hat{y}_i = \langle \phi_i, Y \rangle_{L^2(\mu)}.$$

4. Compute $Z_{t,L}(x)$ as the matrix–vector product

$$Z_{t,L}(x) = \hat{\boldsymbol{\rho}}_x^\top \boldsymbol{U}^{(t)} \hat{\boldsymbol{y}}.$$

# Shift operator

**Notation.**

- $\mathcal{B}(\mu)$ : Basin of the invariant measure.
- $\mu_N = N^{-1} \sum_{n=0}^{N-1} \delta_{\omega_n}$: Sampling measure.
- $\mu_{\mathcal{X},N} := X_* \mu_N = N^{-1} \sum_{n=0}^{N-1} \delta_{x_n}$: Sampling measure in data space.
- $\{e_{0,N}, \ldots, e_{N-1,N}\}$, $e_{j,N}(\omega_n) = N^{1/2} \delta_{jn}$: Orthonormal basis of $L^2(\mu_N)$.
- $\iota : C(\Omega) \to L^2(\mu)$, $\iota f = [f]_\mu$ : Inclusion map.
- $\iota_N : C(\Omega) \to L^2(\mu_N)$, $\iota f = [f]_{\mu_N}$: Restriction map.

# Shift operator

**Definition 3.4.**
For $q \in \mathbb{N}$ we define the shift operator $\hat{U}_N^q : L^2(\mu_N) \to L^2(\mu_N)$ as

$$(\hat{U}_N^q f)(\omega_n) = \begin{cases} f(\omega_{n+1}), & 0 \leq n \leq N-1-q, \\ 0, & N-q \leq n \leq N-1. \end{cases}$$

**Remark 3.5.**
Intuitively, $\hat{U}_N^q$ should be related to $U^{q\,\Delta t}$. However, it is *not* a composition operator. In fact, it is a nilpotent operator, $\hat{U}_N^{N-q+1} = 0$.

**Lemma 3.6.**
*The following hold:*

1. $U^t \circ \iota = \iota \circ \mathcal{U}^t$.
2. *For every* $q \in \mathbb{N}$ *and* $f \in C(\Omega)$,

$$(\hat{U}_N^q \circ \iota_N) f = (\iota_N \circ \mathcal{U}^{q\,\Delta t}) f + r_N,$$

*where* $r_N$ *is a remainder satisfying* $\lim_{N\to\infty} \|r_N\|_{L^2(\mu_N)} = 0$.

# Koopman operator approximation in a continuous basis

**Theorem 3.7.**
*Let* $\{\phi_{0,N}, \dots, \phi_{N-1,N}\}$ *be an orthonormal basis of* $L^2(\mu_N)$ *such that*

$$\phi_{j,N} = \iota_N \varphi_{j,N}, \quad \varphi_{j,N} \xrightarrow[N\to\infty]{C(\Omega)} \varphi_j,$$

*where* $\phi_j = \iota \varphi_j$ *are orthonormal basis vectors of* $L^2(\mu)$*. Let* $\Pi_{L,N} : L^2(\mu_N) \to L^2(\mu_N)$ *be the orthogonal projection onto* $\operatorname{span}\{\phi_{0,N}, \dots, \phi_{L-1,N}\}$*. Assume that the initial state* $\omega_0$ *lies in the basin* $\mathcal{B}(\mu)$*, and set* $q, L \in \mathbb{N}$*. Then, the* $L \times L$ *matrix representations* $\hat{\boldsymbol{U}}_N^{(q)} = [\hat{U}_{ij,N}^{(q)}]$ *and* $\boldsymbol{U}^{(q\,\Delta t)} = [U_{ij}^{(q\,\Delta t)}]$ *of* $\hat{U}_{L,N}^{(q)}$ *and* $U_L^{q\,\Delta t}$*, respectively, with*

$$\hat{U}_{ij,N}^{(q)} = \langle \phi_{i,N}, \hat{U}_N^q \phi_{j,N} \rangle_{L^2(\mu_N)}, \quad U_{ij}^{(q\,\Delta t)} = \langle \phi_i, U^{\Delta t} \phi_j \rangle_{L^2(\mu)}$$

*satisfy* $\lim_{N\to\infty} \hat{\boldsymbol{U}}_N^{(q)} = \boldsymbol{U}^{(q\,\Delta t)}$*, in any matrix norm.*

# Discrete density

We assign to each initial condition $x \in \mathcal{X}$ with $x = X(\omega)$ a probability measure $p_{x,N} \in M_2(\Omega; \mu_N)$ with continuous density.

We let $\rho_{x,N} = \frac{dp_{x,N}}{d\mu_N} \in C(\Omega)$ be the density of $p_{x,N}$ relative to $\mu_N$.

**Algorithm 3.8 (construction of the discrete density $\rho_{x,N}$).**

1. Pick a continuous, strictly positive kernel function $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ as in Algorithm 3.1, e.g.,

$$\kappa(x, x') = \exp\left(-\frac{d^2_{\mathcal{X}}(x, x')}{\epsilon^2}\right), \quad \epsilon > 0.$$

2. Normalize $\kappa$ to a continuous Markov kernel $\rho_N : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ with respect to $\mu_{\mathcal{X},N}$,

$$\rho_N(x, x') = \frac{\kappa(x, x')}{v_N(x)}, \quad v_N(x) = \int_{\mathcal{X}} \kappa(x, \cdot)\, d\mu_{\mathcal{X},N}.$$

3. Set $\rho_{x,N}(\omega) = \rho_N(x, X(\omega))$.

# Discrete density

**Proposition 3.9.**

*With the notation of Algorithm 3.8, the following hold as $N \to \infty$ for every initial state $\omega_0 \in \mathcal{B}(\mu)$:*

1. $\rho_{x,N}$ *converges to* $\rho_x$ *in the* $C(\Omega)$ *norm, uniformly with respect to* $x \in \mathcal{X}_\Omega$.
2. $p_{x,N}$ *converges to* $p_x$ *in the weak-$^*$ topology of* $M(\Omega)$.

# Target function based on samples

We consider a forecast lead time $t = q\,\Delta t$, $q \in \mathbb{N}$.

**Algorithm 3.10 (evaluation of the sample-based target function).**

1. Represent $\hat{U}^{(q)}_{L,N}$ by the $L \times L$ matrix $\hat{\boldsymbol{U}}^{(q)}_N = [\hat{U}^{(q)}_{ij,N}]$ with

$$\hat{U}^{(q)}_{ij,N} = \langle \phi_{i,N}, \hat{U}^q_N \phi_{j,N} \rangle_{L^2(\mu_N)}, \quad i, j = 0, \ldots, L-1.$$

2. Represent $\rho_{x,N}$ by the column vector $\hat{\boldsymbol{\rho}}_{x,N} = (\hat{\rho}_{x,N,0}, \ldots, \hat{\rho}_{x,N,L-1})^\top \in \mathbb{R}^L$ with

$$\hat{\rho}_{x,N,i} = \langle \phi_{i,N}, \rho_{x,N} \rangle_{L^2(\mu_N)}.$$

3. Represent $Y$ by the column vector $\hat{\boldsymbol{y}}_N = (\hat{y}_{N,0}, \ldots, \hat{y}_{N,L-1})^\top \in \mathbb{R}^L$ with

$$\hat{y}_{N,i} = \langle \phi_{i,N}, Y \rangle_{L^2(\mu_N)}.$$

4. Compute $Z_{t,L,N}(x)$ as the matrix–vector product

$$Z_{t,L,N}(x) = \hat{\boldsymbol{\rho}}^\top_{x,N} \hat{\boldsymbol{U}}_N \hat{\boldsymbol{y}}_N.$$

# Approximation in a continuous basis

**Corollary 3.11.**

*With the notation of Algorithm 3.10, the target function* $Z_{t,L,N} : \mathcal{X} \to \mathbb{R}$*,*

$$Z_{t,L,N}(x) = \hat{\boldsymbol{\rho}}_{x,N}^{\top} \hat{\boldsymbol{U}}_N^{(q)} \hat{\boldsymbol{y}}_{\boldsymbol{N}},$$

*satisfies*

$$\lim_{N \to \infty} Z_{t,L,N}(x) = Z_{t,L}(x),$$

*uniformly with respect to* $x \in \mathcal{X}_\Omega$ *and* $t$ *in compact sets.*

# Measures of forecast skill

**Definition 3.12.**
For forecast lead time $t \geq 0$ and a target function $Z_t : \mathcal{X} \to \mathbb{R}$ such that $\tilde{Y}_t := Z_t \circ X$ lies in $L^2(\mu)$, we define:

1. Mean: $\bar{Y} = \mathbb{E}_\mu Y$, $\bar{\tilde{Y}}_t = \mathbb{E}_\mu \tilde{Y}_t$.
2. Anomaly relative to mean: $Y' = Y - \bar{Y}$, $\tilde{Y}'_t = Y_t - \bar{\tilde{Y}}_t$.
3. Standard deviation: $\operatorname{std} Y = \|Y'\|_{L^2(\mu)}$, $\operatorname{std} \tilde{Y}_t = \|\tilde{Y}'_t\|_{L^2(\mu)}$.
4. Root mean square error (RMSE): $\mathrm{RMSE}_t = \|\tilde{Y}_t - Y_t\|_{L^2(\mu)}$.
5. Normalized RMSE: $\mathrm{NRMSE}_t = \mathrm{RMSE}_t \,/\, \operatorname{std} Y$.
6. Anomaly correlation: $\mathrm{AC}_t = \langle \tilde{Y}'_t, Y_t \rangle_{L^2(\mu)} / (\operatorname{std} \tilde{Y}_t \operatorname{std} Y)$.

**Remark 3.13.**
In practice, we estimate the skill scores in Definition 3.12 by approximating integrals with respect to $\mu$ by integrals (time averages) with respect to a sampling measure $\hat{\mu}_{\hat{N}}$ associated with a trajectory that is *independent* of the training trajectory $\omega_0, \omega_1, \ldots$.

# Mixing and loss of predictability

**Proposition 3.14.**

*With notation as above, suppose that the system is mixing. Then, for any $x \in \mathcal{X}$ and $L \in \mathbb{N}$, the long-time limit of the target function $Z_{t,L}(x)$ is a constant $\check{Y}$ independent of $x$,*

$$\lim_{t\to\infty} Z_{t,L}(x) = \tilde{Y}.$$

*In addition, if* $\operatorname{span}\{\phi_0, \dots, \phi_{L-1}\}$ *includes the constant functions, we have $\check{Y} = \bar{Y}$. In that case,*

$$\lim_{t\to\infty} \mathrm{NRMSE}_t = 1, \quad \lim_{t\to\infty} \mathrm{AC}_t = 0.$$

# Estimating the forecast uncertainty

**Definition 3.15.**
Let $\tilde{Y}_t = Z_t \circ \mathcal{X}$ be the pullback of target function onto $\Omega$. We define the forecast variance associated with the initial condition $x \in \mathcal{X}$ as

$$\beta_t(x) = \mathbb{E}_{p_x}(\tilde{Y}_t')^2.$$

We can approximate $\beta_t : \mathcal{X} \to \mathbb{R}$ using an analogous approach as in the construction of $Z_{t,L}$ and $Z_{t,L,N}$, treating $Y_t'$ as the response variable.

# Kernels and kernel integral operators

For our purposes, a kernel is a bivariate function $k : \Omega \times \Omega \to \mathbb{R}$ that captures a notion of similarity or correlation between points in $\Omega$.

Given a continuous kernel $k \in C(\Omega \times \Omega)$ and a Borel probability measure $\nu \in M(\Omega)$, there is an associated kernel integral operator $K : L^2(\nu) \to C(\Omega)$, where

$$Kf(\omega) = \int_\Omega k(\omega, \cdot) f \, d\nu.$$

**Notation.**

- When we wish to make the dependence of $K$ on $\nu$ explicit, we will use the notation $K_\nu \equiv K$.

**Lemma 3.16.**
*Under our general assumptions, $K$ is a compact operator.*

**Corollary 3.17.**
*The operators $G : L^2(\nu) \to L^2(\nu)$ and $\tilde{G} : C(\Omega) \to C(\Omega)$ with*

$$G = \iota \circ K, \quad \tilde{G} = K \circ \iota$$

*are compact.*

# Types of kernels

**Definition 3.18.**

1. A kernel $k : \Omega \times \Omega \to \mathbb{R}$ on a set $\Omega$ is said to be positive-definite if for any finite sequence $\omega_1, \ldots, \omega_n \in \Omega$ and numbers $c_1, \ldots, c_n \in \mathbb{R}$, we have
$$\sum_{i,j=1}^{n} c_i c_j k(\omega_i, \omega_j) \geq 0.$$
2. A kernel $k : \Omega \times \Omega \to \mathbb{R}$ on a set $\Omega$ is said to be strictly positive-definite if for any finite sequence $\omega_1, \ldots, \omega_n$ of distinct points in $\Omega$ and numbers $c_1, \ldots, c_n \in \mathbb{R}$, at least one of which is nonzero, we have
$$\sum_{i,j=1}^{n} c_i c_j k(\omega_i, \omega_j) > 0.$$

# Types of kernels

**Definition 3.19.**
Let $\Omega$ be a topological space and $k : \Omega \times \Omega \to \mathbb{R}$ be a Borel-measurable, bounded kernel.

1. $k$ is said to be integrally positive-definite if for every finite, signed Borel measure $\nu$ on $\Omega$, we have

$$\int_\Omega \int_\Omega k(\omega, \omega')\, d\nu(\omega)\, d\nu(\omega') \geq 0.$$

2. $k$ is said to be strictly integrally positive-definite if for every nonzero, finite, signed Borel measure $\nu$ on $\Omega$, we have

$$\int_\Omega \int_\Omega k(\omega, \omega')\, d\nu(\omega)\, d\nu(\omega') > 0.$$

**Remark 3.20.**
If $k$ is (strictly) integrally positive-definite, $G_\nu : L^2(\nu) \to L^2(\nu)$ is a (strictly) positive operator. We will then say that $k$ is $L^2(\nu)$-(strictly)-positive.

# Types of kernels

**Theorem 3.21.**
*Suppose that $\Omega$ is a compact metrizable space. Then, with the notation of Definition 3.19, the following hold:*

1. *$k$ is integrally positive-definite iff it is positive-definite.*
2. *If $k$ is strictly integrally positive-definite then it is strictly positive-definite.*

# Kernel eigenfunctions

**Proposition 3.22.**

*Let $K : L^2(\nu) \to C(\Omega)$, $G : L^2(\nu) \to L^2(\mu)$, $\tilde{G} : C(\Omega) \to C(\Omega)$ be the operators from Corollary 3.17. Then:*

1. *There exists an orthonormal basis $\{\phi_0, \phi_1, \ldots\}$ of $L^2(\nu)$ which are eigenfunctions of $G$ corresponding to real eigenvalues $\lambda_0, \lambda_1, \ldots$, i.e.,*

$$G\phi_j = \lambda_j \phi_j.$$

2. *Every nonzero eigenvalue $\lambda_j$ has finite multiplicity, and the eigenvalues can be ordered in a sequence $|\lambda_0| \geq |\lambda_1| \geq \cdots \searrow 0$ with no accumulation point other than 0.*
3. *For every $\lambda_j \neq 0$, the continuous function $\varphi_j := \lambda_j^{-1} K \phi_j$ is an eigenfunction of $\tilde{G}$ corresponding to the same eigenvalue $\lambda_j$,*

$$\tilde{G}\varphi_j = \lambda_j \varphi_j.$$

# Data-driven basis

**Algorithm 3.23 (data-driven basis).**

Set $\nu = \mu_N$, $K_N \equiv K_{\mu_N}$, $G_N \equiv G_{\mu_N}$, $\tilde{G}_N \equiv \tilde{G}_{\mu_N}$. Assume $k$ is symmetric.

1. Represent $G_N$ by the $N \times N$ kernel matrix $\boldsymbol{K} = [K_{ij}]$ with

$$K_{ij} = \langle e_{i,N}, G_N e_{j,N} \rangle_{L^2(\mu_N)} = k(\omega_i, \omega_j).$$

2. Solve the matrix eigenvalue problem

$$\boldsymbol{K}\boldsymbol{\phi}_j = \lambda_{j,N}\boldsymbol{\phi}_j, \quad \boldsymbol{\phi}_j = (\phi_{0j}, \dots, \phi_{N-1,j})^\top, \quad \|\boldsymbol{\phi}_j\|_2 = \sqrt{N}.$$

3. Reconstruct the eigenvectors $\phi_{j,N} \in L^2(\mu_N)$,

$$\phi_{j,N} = \sum_{i=0}^{N-1} \phi_{ij,N} e_{i,N}, \quad G_N \phi_{j,N} = \lambda_{j,N}\phi_{j,N}.$$

4. For $\lambda_{j,N} \neq 0$, compute the continuous extensions $\varphi_{j,N} \in C(\Omega)$,

$$\varphi_{j,N} = \frac{1}{\lambda_{j,N}} K_N \phi_{j,N} = \frac{1}{\lambda_{j,N} N} \sum_{i=0}^{N-1} k(\cdot, \omega_i)\, \phi_{ij,N}, \quad \tilde{G}_N \varphi_{j,N} = \lambda_{j,N}\varphi_{j,N}.$$

# Data-driven basis

Set $\nu = \mu$, $K \equiv K_\mu$, $G \equiv G_\mu$, $\tilde{G} \equiv \tilde{G}_\mu$,

$$G\phi_j = \lambda_j \phi_j, \quad \varphi_j = \frac{1}{\lambda_j} K\phi_j, \quad \tilde{G}\varphi_j = \lambda_j \varphi_j.$$

**Strategy.**

- Use the weak-$^*$ convergence of $\mu_N$ to $\mu$ to deduce spectral convergence of $\tilde{G}_N$ to $\tilde{G}$. This implies convergence of the *nonzero* $\lambda_{j,N}$ to $\lambda_j$ and convergence of $\varphi_{j,N}$ to $\varphi_j$ in a suitable sense.
- Use an integrally strictly positive-definite kernel $k$. Then, $\lambda_j > 0$, and the $\varphi_{j,N}$ converge to an orthonormal basis of $L^2(\mu)$. The assumptions of Theorem 3.7 are satisfied.

# Compact convergence

**Definition 3.24.**
Let $(F, \|\cdot\|_F)$ be a Banach space, and $A_1, A_2, \ldots$ a sequence of bounded operators on $F$. We say that $A_n$ converges compactly if it converges to a (bounded) operator $A$ pointwise, and for every bounded sequence $f_n \in F$, the sequence $(A - A_n)f_n$ has a convergent subsequence in $F$.

**Notation.**

- $B(F)$: Banach space of bounded operators on $F$.
- $\sigma(A)$: Spectrum of an operator $A \in B(F)$.

# Compact convergence

**Theorem 3.25.**
*With the notation of Definition 3.24, let $A_n$ converge to $A$ compactly. Let $\lambda \in \sigma(A)$ be an isolated eigenvalue of $A$ with finite multiplicity $m$, and $\Pi$ the spectral projection of $A$ corresponding to $\lambda$. Let $S \subseteq \mathbb{C}$ be an open neighborhood of $\lambda$ such that $\sigma(A) \cap S = \{\lambda\}$. Then, the following hold.*

1. *There exists $n_* \in \mathbb{N}$ such that for all $n > n_*$, $\sigma(A_n) \cap S$ is an isolated subset of $\sigma(A_n)$, consisting of at most $m$ distinct eigenvalues whose multiplicities sum up to $m$. Moreover, every sequence $\lambda_n \in \sigma(A_n) \cap S$ satisfies $\lim_{n\to\infty} \lambda_n = \lambda$.*
2. *The spectral projections $\Pi_n$ of $A_n$ corresponding to $\sigma(A_n) \cap S$ (which are well-defined for $n > n_*$) converge strongly to $\Pi$. In particular, for every eigenfunction $\phi$ of $A$ at eigenvalue $\lambda$, there exists a sequence of eigenfunctions $\phi_n$ of $A_n$ at eigenvalue $\lambda_n$ such that $\lim_{n\to\infty} \phi_n = \phi$ in the norm of $F$.*

# Compact convergence

**Theorem 3.26.**
*Suppose that* $k : \Omega \times \Omega \to \mathbb{R}$ *is a continuous kernel. Then, under our general assumptions,* $\tilde{G}_N$ *converges compactly to* $\tilde{G}$.

**Corollary 3.27.**
*If* $k$ *is integrally strictly positive-definite, the conditions of Theorem 3.7 are satisfied.*

# Kernels on covariate space

In practice, we construct $k : \Omega \times \Omega \to \mathbb{R}$ as a pullback of a continuous kernel $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ in covariate space, i.e.,

$$k(\omega, \omega') = \kappa(X(\omega), X(\omega')).$$

All computations involving $G_N$ and $K_N$ can be expressed using covariate data.

- $G_N$ is represented by the $N \times N$ matrix

$$\boldsymbol{K} = [K_{ij}], \quad K_{ij} = k(\omega_i, \omega_j) = \kappa(x_i, x_j).$$

- The continuous extension $\varphi_{j,N}$ of $\phi_{j,N}$ is given by

$$\varphi_{j,N}(\omega) = \frac{1}{\lambda_{j,N}} K_N \phi_j(\omega) = \frac{1}{\lambda_{j,N} N} \sum_{n=0}^{N-1} \kappa(F(\omega), x_n).$$

# Kernels on covariate space

Every eigenfunction $\varphi_{j,N}$ or $\varphi_j$ corresponding to nonzero eigenvalue is of the form

$$\varphi_{j,N} = \varphi_{j,N}^{(\mathcal{X})} \circ X, \quad \varphi_j = \varphi_j^{(\mathcal{X})} \circ X,$$

for continuous functions $\varphi_{j,N}^{(\mathcal{X})}, \varphi_j^{(\mathcal{X})} \in C(\mathcal{X}_\Omega)$.

- The corresponding $\phi_j = \iota\varphi_j$ form an orthonormal *set* in $L^2(\mu)$, but if $X$ is not injective they might not form an orthonormal basis (even if $\kappa$ is integrally positive definite).

**Remark.**
The kernel $\kappa$ used to compute the basis vectors $\phi_{j,N}$ and $\phi_j$ need not (and in general, will not) be the same as the kernel used to assign the initial density $\rho_x$ via Algorithm 3.8.

# Data-driven forecast

The construction and evaluation of the data-driven target function $Z_{t,L,N}(x)$ for $x \in \mathcal{X}$ and $t = q\,\Delta t$, can be summarized as follows.

**Algorithm 3.28 (data-driven target function).**

1. Apply Algorithm 3.23 using the covariate training data $x_n$ and a kernel $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ to compute the basis vectors $\phi_{j,N}$.
2. Apply Algorithm 3.8 using the covariate training data $x_n$ and a strictly positive kernel $\tilde{\kappa} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ to compute the initial density $\rho_x$.
3. Set a spectral resolution parameter $L$ (number of eigenfunctions). Apply Algorithm 3.10 using $\phi_{j,N}$ from Step 1, $\rho_x$ from Step 2, and the response training data $y_n$ to compute $Z_{t,L,N}(x)$.

# Conditional expectation

**Notation.**
For a measure space $(\Omega, \Sigma, \mu)$ and a sub-$\sigma$-algebra $\Sigma' \subseteq \Sigma$, we let:

- $\mathbb{L}(\Sigma') = \{f : \Omega \to \mathbb{R} : f \text{ is } \Sigma'\text{-measurable}\} \subseteq \mathbb{L}(\Sigma)$.
- $L(\mu, \Sigma') = \{[f]_\mu : f \in \mathbb{L}(\Sigma')\} \subseteq L(\mu)$.
- $L^p(\mu, \Sigma') = L^p(\mu) \cap L(\mu, \Sigma)$.

**Definition 3.29.**
Let $(\Omega, \Sigma, \mu)$ be a probability space. Given $f \in L^1(\mu)$ and a sub-$\sigma$-algebra $\Sigma' \subseteq \Sigma$, the conditional expectation of $f$ on $\Sigma'$ is the unique element $g \in L^1(\mu, \Sigma')$ such that

$$\int_E f \, d\mu = \int_E g \, d\mu, \quad E \in \Sigma'.$$

We write $g \equiv \mathbb{E}(f \mid \Sigma')$.

# Conditional expectation

**Lemma 3.30.**
*With the notation of Definition 3.29, the following hold.*

1. $L^p(\mu, \Sigma')$ *is a closed subspace of* $L^p(\mu)$.
2. *If* $f \in L^p(\mu)$, *then* $\mathbb{E}(f \mid \Sigma') \in L^p(\mu, \Sigma')$.
3. *The map* $\Pi_{\Sigma'} : L^p(\mu) \to L^p(\mu)$ *with*

$$\Pi_{\Sigma'} f = \mathbb{E}(f \mid \Sigma')$$

   *is a linear projection onto* $L^p(\mu, \Sigma')$ *with norm 1.*
4. $\Pi_{\Sigma'} : L^2(\mu) \to L^2(\mu)$ *is the orthogonal projection onto* $L^2(\mu, \Sigma')$.

# Conditional expectation

**Corollary 3.31.**

*For any* $f \in L^2(\mu)$,

$$\mathbb{E}(f \mid \Sigma') = \Pi_{\Sigma'} f$$

*is the unique element of* $L^2(\mu, \Sigma')$ *that minimizes the distance from* $f$ *to* $L^2(\mu, \Sigma')$, *i.e.,*

$$\|f - \mathbb{E}(f \mid \Sigma')\|_{L^2(\mu)} < \|f - g\|_{L^2(\mu)}, \quad \forall g \in L^2(\mu, \Sigma') \setminus \{f\}.$$

# Conditional expectation on measurable maps

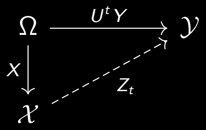
**Definition 3.32.**
Let $(\Omega, \Sigma, \mu)$ be a probability space and $X : (\Omega, \Sigma) \to (\mathcal{X}, \Sigma_{\mathcal{X}})$ a measurable map. We define the conditional expectation of $f \in L^1(\mu)$ on $X$ as

$$\mathbb{E}(f \mid X) = \mathbb{E}(f \mid \Sigma_X), \quad \Sigma_X = X^{-1}(\Sigma_{\mathcal{X}}).$$

**Remark 3.33.**

1. $\Sigma_X$ is the $\sigma$-algebra generated by $X$, i.e., the smallest sub-$\sigma$-algebra of $\Sigma$ such that $X : (\Omega, \Sigma_X) \to (\mathcal{X}, \Sigma_{\mathcal{X}})$ is measurable.
2. Every $f \in \mathbb{L}(\Sigma_X)$ is of the form $f = g \circ X$ for $g \in \mathbb{L}(\Sigma_{\mathcal{X}})$.
3. Every $f \in L^p(\mu, \Sigma_X)$ is of the form $f = g \circ X$ for $g \in L^p(\mu_{\mathcal{X}})$, where $\mu_{\mathcal{X}} = X_* \mu$.

# Ideal target function




In light of Corollary 3.31 and Remark 3.33, the ideal target function in the sense of $L^2(\mu)$ error (RMS error; see Definition 3.12) is $Z_t \in L^2(\mu_{\mathcal{X}})$ such that

$$\mathbb{E}(U^t Y \mid X) = Z_t \circ X.$$

That is, $Z_t$ satisfies

$$\|U^t Y - Z_t \circ X\|_{L^2(\mu)} \leq \|U^t Y - \tilde{Y}_t\|_{L^2(\mu)}, \quad \forall \tilde{Y}_t \in L^2(\mu, \Sigma_X).$$

# Conditional probability

**Notation.**

- $\chi_S : \Omega \to \{0,1\}$: Characteristic function of a set $S \subseteq \Omega$.

**Definition 3.34.**
Let $(\Omega, \Sigma, \mu)$ be a probability space. For every sub-$\sigma$-algebra $\Sigma' \subseteq \Sigma$ and measurable set $S \in \Sigma$, we define the conditional probability $\mathbb{P}(S \mid \Sigma') \in L^1(\mu, \Sigma')$ as

$$\mathbb{P}(S \mid \Sigma') = \mathbb{E}(\chi_S \mid \Sigma').$$

**Remark 3.35.**
The map $S \in \Sigma \mapsto \mathbb{P}(S \mid \Sigma')$ defines a vector measure on $\Sigma'$, i.e., an $L^1(\mu)$-valued map such that for any sequence $S_n$ of disjoint measurable sets in $\Sigma$,

$$\mathbb{P}\left(\bigcup_n S_n \mid \Sigma'\right) = \sum_n \mathbb{P}(S_n \mid \Sigma').$$

# Regular conditional probability

**Definition 3.36.**
With the notation of Definition 3.34 we say that $\mathbb{P}(S \mid \Sigma')$ is a regular conditional probability if there is a map $p : \Omega \times \Sigma \to \mathbb{R}$ such that:

1. For every $\omega \in \Omega$, $p(\omega, \cdot)$ is a probability measure on $\Sigma$.
2. For every $S \in \Sigma$, the map $\omega \mapsto p(\omega, S)$ is a representative of the conditional probability $\mathbb{P}(S \mid \Sigma') \in L^1(\mu)$.

The map $p$ is called a Markov kernel.

If $\Sigma' = \Sigma_X$ is the sub-$\sigma$-algebra generated by a measurable map $X : (\Omega, \Sigma) \to (\mathcal{X}, \sigma_{\mathcal{X}})$ then we have

$$p(\omega, \cdot) = p_{\mathcal{X}}(X(\omega), \cdot)$$

for a Markov kernel $p_{\mathcal{X}} : \mathcal{X} \times \Sigma \to \mathbb{R}$.

# Regular conditional probability

**Theorem 3.37.**
*For a compact metrizable space $\Omega$ equipped with its Borel $\sigma$-algebra every conditional probability $\mathbb{P}(\cdot \mid \Sigma')$ is a regular conditional probability.*

**Proposition 3.38.**
*If $\mathbb{P}(\cdot \mid \Sigma')$ is a regular conditional probability with Markov kernel $p$, then for every $Y \in L^1(\mu)$ we have*

$$\mathbb{E}(Y \mid \Sigma') = \int_\Omega Y(\omega)\, p(\cdot, d\omega),$$

*where the equality holds $\mu$-a.e.*

# Conditional density

**Definition 3.39.**
With the notation of Definition 3.34, if $p(\omega, \cdot) \ll \mu$, we say that a function $\rho : \Omega \times \Omega \to \mathbb{R}$ is a conditional density of $p$ if

$$\rho(\omega, \cdot) = \frac{dp(\omega, \cdot)}{d\mu}, \quad \mu\text{-a.e.}$$

**Proposition 3.40.**
*If it exists, $\rho(\omega, \cdot)$ lies in $L^\infty(\mu)$.*

If $\Sigma' = \Sigma_X$ is the sub-$\sigma$-algebra generated by $X : \Omega \to \mathcal{X}$, then we have $\rho(\omega, \cdot) = \rho_{\mathcal{X}}(X(\omega), \cdot)$ for a function $\rho_{\mathcal{X}} : \mathcal{X} \times \Omega \to \mathbb{R}$.

# Conditional density

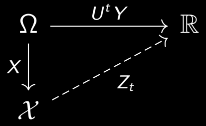
**Remark 3.41.**
If $p$ has a conditional density, then for every $Y \in L^1(\mu)$, we have

$$\mathbb{E}(Y \mid \Sigma') = \int_\Omega Y(\omega)\rho(\cdot,\omega)\, d\mu(\omega).$$

In particular, for $Y \in L^2(\mu)$ and $\mu$-a.e. $\omega' \in \Omega$,

$$\mathbb{E}(Y \mid \Sigma')(\omega') = \langle \rho(\omega',\cdot), Y\rangle_{L^2(\mu)}.$$

# Hypothesis space




**Goal.** Construct the ideal target function $Z_t : \mathcal{X} :\to \mathbb{R}$ such that $Z_t \circ X = \mathbb{E}(U^t Y \mid X)$, $\mu$-a.e.

**Strategy.** Approximate $Z_t$ in a hypothesis space $\mathcal{H}$ of continuous functions on $\mathcal{X}_\Omega$ such that:

1. $\mathcal{H}$ is a convex subset of a Hilbert space $\mathcal{K}$ of continuous functions on $\mathcal{X}_\Omega$.
2. The inclusion map $\iota : \mathcal{K} \to L^2(\mu_\mathcal{X})$ is compact.
3. $H \equiv \iota\mathcal{H}$ is closed in $L^2(\mu_\mathcal{X})$.

**Proposition 3.42.**

*Under the assumptions stated above, there is a unique minimizer $Z_{t,\mathcal{H}}$ of the square error functional $\mathcal{E}_t : \mathcal{H} \to \mathbb{R}_+$, where*

$$\mathcal{E}_t(f) = \|\iota f - Z_t\|^2_{L^2(\mu_\mathcal{X})}.$$

# Reproducing kernel Hilbert spaces

**Definition 3.43.**
A Hilbert space $(\mathcal{K}, \langle \cdot, \cdot \rangle_{\mathcal{K}})$ of complex-valued functions functions on a set $\mathcal{X}$ is called a reproducing kernel Hilbert space (RKHS) if for every $x \in \mathcal{X}$ the pointwise evaluation functional $\delta_x : \mathcal{K} \to \mathbb{C}$ is continuous.

By the Riesz representation theorem, for every $x \in \mathcal{X}$, there exists a unique function $k_x \in \mathcal{K}$ such that

$$f(x) = \langle k_x, f \rangle_{\mathcal{K}}, \quad \forall f \in \mathcal{K}.$$

The function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ with $k(x, x') = k_x(x')$ is called the reproducing kernel of $\mathcal{K}$.

**Proposition 3.44.**
*$k$ is a positive-definite kernel on $\mathcal{X}$.*

# Reproducing kernel Hilbert spaces

**Theorem 3.45 (Moore-Aronszajn).**

*Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ be a positive-definite kernel on a set $\mathcal{X}$. Then, there is a unique RKHS $\mathcal{K}$ on $\mathcal{X}$ with $k$ as its reproducing kernel. Explicitly, $\mathcal{K}$ is the completion of the inner product space $(\mathcal{K}_0, \langle \cdot, \cdot \rangle_{\mathcal{K}_0})$ with*

$$\mathcal{K}_0 = \mathrm{span}\{k_x : x \in \mathcal{X}\}, \quad \left\langle \sum_{i=1}^{m} a_i k_{x_i}, \sum_{j=1}^{n} b_j k_{x_j} \right\rangle_{\mathcal{K}_0} = \sum_{i=1}^{m} \sum_{j=1}^{n} \bar{a}_i k(x_i, x_j) b_j.$$

# Reproducing kernel Hilbert spaces

**Lemma 3.46.**
*Let* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ *be a positive-definite kernel on a set* $\mathcal{X}$ *with associated RKHS* $\mathcal{K}$*. Then, for any subset* $S \subseteq \mathcal{X}$*, the restriction*

$$\mathcal{K}|_S = \{f|_S : f \in \mathcal{K}\}$$

*is an RKHS with reproducing kernel* $k|_{S \times S}$*.*

**Notation.**

- If $\nu$ is a probability measure on $\mathcal{X}$ we write $\mathcal{K}_\nu \equiv \mathcal{K}|_{\operatorname{supp} \nu}$ and $k_\nu \equiv k|_{\operatorname{supp} \nu \times \operatorname{supp} \nu}$.

# Mercer kernels

**Theorem 3.47 (Mercer).**
*Let $\mathcal{X}$ be a compact Hausdorff space and $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ a continuous, positive-definite kernel with associated RKHS $\mathcal{K}$. Let $\nu$ be a Borel probability measure on $\mathcal{X}$. Consider the the corresponding self-adjoint integral operator $G_\nu : L^2(\nu) \to L^2(\nu)$, $G_\nu = \iota_\nu \circ K_\nu$ (see Corollary 3.17), and its eigendecomposition as in Proposition 3.22,*

$$G_\nu \phi_j = \lambda_j \phi_j, \quad \langle \phi_i, \phi_j \rangle_{L^2(\nu)} = \delta_{ij}, \quad \lambda_0 \geq \lambda_1 \geq \cdots \searrow 0.$$

*Then, the kernel $k_\nu$ admits the series expansion*

$$k_\nu(x, x') = \sum_{j:\lambda_j > 0} \lambda_j \overline{\varphi_j(x)} \varphi_j(x'),$$

*where $\varphi_j = \lambda_j^{-1} K_\nu \phi_j$ is the continuous representative of $\phi_j$, and the convergence is uniform with respect to $(x, x') \in \operatorname{supp} \nu \times \operatorname{supp} \nu$.*

# Mercer kernels

**Corollary 3.48.**

1. *$\mathcal{K}$ is a subspace of $C(\mathcal{X})$.*
2. *Upon restriction to* $\operatorname{supp}\nu$*, the range of $K_\nu : L^2(\nu) \to C(\mathcal{X})$ is a dense subspace of $\mathcal{K}_\nu$.*
3. *The functions $\psi_j = \lambda_j^{-1/2} K_\nu \phi_j$ form an orthonormal set in $\mathcal{K}$, and their restrictions to* $\operatorname{supp}\nu$ *form an orthonormal basis of $\mathcal{K}_\nu$.*
4. *The operator $G_\nu : L^2(\nu) \to L^2(\nu)$ is of trace class, and we have*

$$\|G_\nu\|_1 = \operatorname{tr} G_\nu = \int_{\mathcal{X}} k(x,x)\, d\nu(x).$$

# Inclusion operators

**Proposition 3.49.**

*Viewing $K_\nu$ as an operator from $L^2(\nu)$ to $\mathcal{K}$, the adjoint $K_\nu^* : \mathcal{K} \to L^2(\nu)$ coincides with the inclusion map $\iota_\nu : C(\mathcal{X}) \to L^2(\nu)$, i.e.,*

$$K_\nu^* f = \iota_\nu f, \quad \forall f \in \mathcal{K}.$$

*In particular, we have $G_\nu = K_\nu^* K_\nu$.*

**Corollary 3.50.**

1. *$\mathcal{K}_\nu$ embeds compactly into $L^2(\nu)$.*
2. *Every element of $\operatorname{ran} K_\nu^*$ has a representative in $\mathcal{K}$ (and thus in $C(\mathcal{X})$).*
3. *$\operatorname{ran} K^*$ is closed iff $\mathcal{K}_\nu$ is finite-dimensional.*

# Universal kernels

**Definition 3.51.**
A positive-definite kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ on a locally compact Hausdorff space is said to be:

1. $C_0$-universal if $k(x, \cdot)$ lies in $C_0(\mathcal{X})$ for all $x \in \mathcal{X}$, and the corresponding RKHS $\mathcal{K}$ is dense in $C_0(\mathcal{X})$.
2. $C$-universal if $\mathcal{X}$ is compact, $k$ is continuous, and the corresponding RKHS $\mathcal{K}$ is dense in $C(\mathcal{X})$.
3. $L^p$-universal if for every Borel probability measure $\nu$ on $\mathcal{X}$, $\mathcal{K}$ is a dense subspace of $L^p(\nu)$ for some $p \in [1, \infty)$.

**Theorem 3.52.**
*On a compact Hausdorff space, $C$-universality, $L^p$-universality, and strict integral-positiveness are equivalent notions. Moreover, every kernel having these properties is strictly positive-definite.*

# Radial kernels

**Definition 3.53.**
A bounded, continuous kernel $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ is said to be radial if there exists a positive, finite Borel measure $\alpha$ on $[0, \infty)$ such that

$$k(x, x') = \int_{[0,\infty)} e^{-s\|x-x'\|_2^2} \, d\alpha(s), \quad \forall x, x' \in \mathbb{R}^d.$$

**Theorem 3.54.**
*A radial, strictly-positive definite kernel on $\mathbb{R}^d$ is $C_0$-universal.*

**Theorem 3.55.**
*The radial basis function (RBF) kernel on $\mathbb{R}^d$,*

$$k(x, x') = \exp\left(-\frac{\|x - x'\|_2^2}{\epsilon^2}\right), \quad \epsilon > 0,$$

is strictly positive-definite (and radial).

# Feature maps

**Definition 3.56.**
A feature map on a set $\mathcal{X}$ is a map $F : \mathcal{X} \to \mathcal{F}$, where $\mathcal{F}$ is a Hilbert space, called feature space.

**Lemma 3.57.**
*If $F : \mathcal{X} \to \mathcal{F}$ is a feature map, then $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ with*

$$k(x, x') = \langle F(x), F(x') \rangle_{\mathcal{F}}$$

*is a positive-definite kernel.*

**Definition 3.58.**
Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ be a positive-definite kernel on a a set $\mathcal{X}$. We say that a feature map $F : \mathcal{X} \to \mathcal{F}$ is associated to $k$ if

$$k(x, x') = \langle F(x), F(x') \rangle_{\mathcal{F}}.$$

# Feature maps

**Proposition 3.59.**

*Let* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ *be a positive-definite kernel on a set* $\mathcal{X}$ *with associated RKHS* $\mathcal{K}$*.*

1. $F : \mathcal{X} \to \mathcal{K}$ *with* $F(x) = k(x, \cdot)$ *is a feature map associated to* $k$*.*
2. *If* $k$ *is strictly positive-definite, then* $F$ *is injective. Moreover,* $F(x)$ *and* $F(x')$ *are linearly independent whenever* $x$ *and* $x'$ *are distinct.*

# Feature maps

**Lemma 3.60.**
*Let* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ *be a continuous, positive-definite kernel on a compact Hausdorff space* $\mathcal{X}$ *with associated RKHS* $\mathcal{K}$*. Let* $\{\psi_0, \psi_1, \ldots\}$ *be the orthonormal basis of* $\mathcal{K}_\nu$ *from Corollary 3.48. Then,* $F : \operatorname{supp} \nu \to \ell^2$ *with*

$$F(x) = (\psi_0(x), \psi_1(x), \ldots)$$

*is a feature map associated to* $k_\nu$*.*

# Moore-Penrose pseudoinverse

**Theorem 3.61.**
*Let $A : H_1 \to H_2$ be a linear map between two finite-dimensional Hilbert spaces. Then, there exists a unique linear map $A^+ : H_2 \to H_1$, called the Moore-Penrose pseudoinverse of $A$, with the following properties:*

1. $\ker A^+ = \operatorname{ran} A^\perp$.
2. $\operatorname{ran} A^+ = \ker A^\perp$.
3. $AA^+ f = f$ *for all* $f \in \operatorname{ran} A$.

**Theorem 3.62.**
*With notation as above $A^+ : H_2 \to H_1$ is the pseudoinverse of $A$ iff the following conditions hold:*

1. $AA^+A = A$.
2. $A^+AA = A^+$.
3. $(AA^+)^* = AA^+$.
4. $(A^+A)^* = A^+A$.

# Moore-Penrose pseudoinverse

**Proposition 3.63.**

*With the notation of Theorem 3.61, the following hold.*

1. *If* $\operatorname{ran} A = H_2$, *then* $A^+ = A^*(AA^*)^{-1}$ *and* $AA^+ = I$.
2. *If* $\operatorname{ran} A^* = H_1$, *then* $A^+ = (A^*A)^{-1}A^*$ *and* $A^+A = I$.

# Pseudoinverse in infinite-dimensional Hilbert spaces

**Theorem 3.64.**
*Let $H_1$ and $H_2$ be Hilbert spaces, and $A : D(A) \to H_2$ a closed linear map with dense domain $D(A) \subseteq H_1$. Then, there exists a unique, densely defined, closed operator $A^+ : D(A^+) \to H_1$ with domain $D(A^+) \subseteq H_2$, called the pseudoinverse of $A$, such that*

1. $\ker A^+ = \operatorname{ran} A^\perp$.
2. $\overline{\operatorname{ran} A^+} = \ker A^\perp$.
3. $AA^+ f = f$ *for all* $f \in \operatorname{ran} A$.

**Theorem 3.65.**
*With notation as above, the following hold.*

1. $(A^+)^+ = A$.
2. $(A^+)^* = (A^*)^+$.
3. $A^+$ *is bounded iff* $\operatorname{ran} A$ *is closed.*

# Pseudoinverse in infinite-dimensional Hilbert spaces

**Theorem 3.66.**
*With the notation of Theorem 3.64, given $g \in D(A^+)$, $f = A^+ g$ has the properties*

1. $\|Af - g\|_{H_2} = \inf_{h \in D(A)} \|Ah - g\|_{H_2}$.
2. $\|f\|_{H_1} < \|h\|_{H_1}$ *for all* $h \neq f$ *attaining the infinum above.*

We refer to $g$ as the best approximate solution to the equation $Af = g$.

# Nyström operator

**Definition 3.67.**
Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ be a continuous, positive-definite kernel on a compact Hausdorff space $\mathcal{X}$ with corresponding RKHS $\mathcal{K}$. Let $\nu$ be a Borel probability measure on $\mathcal{X}$. We define the Nyström operator associated with $k$ and $\nu$ as

$$\mathcal{N}_\nu = (K_\nu^*)^+.$$

# Nyström operator

**Proposition 3.68.**

*With the notation of Definition 3.67, the following hold.*

1. $\operatorname{ran} \mathcal{N}_\nu = \mathcal{K}_\nu$. *In particular, $\mathcal{N}_\nu$ has closed range.*
2. *The domain $D(\mathcal{N}_\nu) \subseteq L^2(\nu)$ is given by*

$$D(\mathcal{N}_\nu) = \left\{ f = \sum_j c_j \phi_j : \sum_{j:\lambda_j > 0} \frac{|c_j|^2}{\lambda_j} < \infty \right\}.$$

3. *$\mathcal{N}_\nu$ is bounded iff $\mathcal{K}_\nu$ is finite-dimensional.*
4. *For every $f \in D(\mathcal{N}_\nu)$ we have*

$$\mathcal{N}_\nu f = \sum_{j:\lambda_j > 0} \frac{c_j}{\lambda_j^{1/2}} \psi_j, \quad c_j = \langle \phi_j, f \rangle_{L^2(\nu)}.$$

# Nyström operator

**Proposition 3.69.**

*With the notation of Definition 3.67, the following hold.*

1. *For every* $f \in \mathcal{K}_\nu$,
$$K_\nu^* \mathcal{N}_\nu^+ f = f.$$
2. *For every* $f \in D(\mathcal{N}_\nu)$,
$$K_\nu^* \mathcal{N}_\nu^+ f = \Pi_\nu f,$$
*where* $\Pi_\nu : L^2(\nu) \to L^2(\nu)$ *is the orthogonal projection onto* $\ker K_\nu^\perp$.

# Truncated Nyström operator

**Definition 3.70.**
With the notation of Definition 3.67, and for $L \in \mathbb{N}$ such that $\lambda_{L-1} > 0$, we define the truncated Nyström operator $\mathcal{N}_{\nu,L} : L^2(\nu) \to \mathcal{K}_\nu$ as

$$\mathcal{N}_{\nu,L} = \mathcal{N}_\nu \circ \Pi_{\nu,L},$$

where $\Pi_{\nu,L} : L^2(\nu) \to L^2(\nu)$ is the orthogonal projection onto $\operatorname{span}\{\phi_0, \dots, \phi_{L-1}\}$.

**Lemma 3.71.**
*The following hold as $L \to \infty$.*

1. *$\mathcal{N}_{\nu,L}$ converges to $\mathcal{N}_\nu$ strongly on $D(\mathcal{N}_\nu)$.*
2. *$K_\nu^* \mathcal{N}_{\nu,L}$ converges to $\Pi_\nu$ strongly on $L^2(\nu)$.*

**Corollary 3.72.**
*For every $g \in L^2(\nu)$, $f_L = \mathcal{N}_{\nu,L} g$ is a sequence of continuous functions that converges to $\Pi_\nu g$ in $L^2(\nu)$ norm.*

# Approximating the conditional expectation

**Theorem 3.73.**
*Let $\kappa : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ be a continuous, positive-definite kernel on the covariate space $\mathcal{X}$ with associated RKHS $\mathcal{K}^{(\mathcal{X})}$. Let $k : \Omega \times \Omega \to \mathbb{C}$ with $k(\omega, \omega') = \kappa(X(\omega), X(\omega'))$ be the pullback of $\kappa$ to $\Omega$. Define*

$$\tilde{Y}_{t,L} = \mathcal{N}_{\mu,L} U^t Y.$$

*Then, the following hold:*

1. *$\tilde{Y}_{t,L}$ is the pullback of a function $Z_{t,L} \in \mathcal{K}^{(\mathcal{X})}$, i.e.,*

$$\tilde{Y}_{t,L} = Z_{t,L} \circ X.$$

2. *$Z_{t,L}$ is the minimizer of the error functional $\mathcal{E}_{t,L} \equiv \mathcal{E}_t$ from Proposition 3.42 for the hypothesis space $\mathcal{H}_L \equiv \mathcal{H}$..*
3. *As $L \to \infty$, $\mathcal{E}_{t,L}(Z_{t,L})$ converges to 0 and $\tilde{Y}_{t,L}$ converges in $L^2(\mu)$ norm to the conditional expectation $\mathbb{E}(U^t Y \mid X)$.*

# Approximating the conditional expectation

Assume the notation of Theorem 3.73, set a lead time $t = q\,\Delta t$, $q \in \mathbb{N}$.

**Algorithm 3.74 (data-driven conditional expectation).**

1. Apply Algorithm 3.23 using the covariate training data $x_n$ and the kernel $\kappa^{(\mathcal{X})}$ to compute the basis vectors $\phi_{l,N}$ of $L^2(\mu_N)$.
2. Fix a spectral resolution parameter $L$, and compute the expansion coefficients of $\hat{U}_N^q Y$ in the $\phi_{l,N}$ basis,

$$\hat{y}_{t,N,l} = \langle \phi_{l,N}, \hat{U}_N^q Y \rangle_{L^2(\mu_N)}.$$

3. Compute the target function $Z_{t,L,N} : \mathcal{X} \to \mathbb{R}$, where

$$Z_{t,L,N} = \sum_{l=0}^{L-1} \hat{y}_{t,N,l} \varphi_l^{(\mathcal{X})}.$$

# Further reading


[1] R. Alexander and D. Giannakis, "Operator-theoretic framework for forecasting nonlinear time series with kernel analog techniques," *Phys. D*, vol. 409, p. 132 520, 2020. DOI: 10.1016/j.physd.2020.132520.

[2] F. J. Beutler and W. L. Root, "The operator pseudoinverse in control and systems identification," in *Generalized Inverses and Applications*, M. Zuhair Nashed, Ed., University of Wisconsin-Madison, Academic Press, 1973, pp. 397–494. DOI: 10.1016/B978-0-12-514250-2.50014-7.

[3] F. Cucker and S. Smale, "On the mathematical foundations of learning," *Bull. Amer. Math. Soc.*, vol. 39, no. 1, pp. 1–49, 2001. DOI: 10.1090/S0273-0979-01-00923-5.

[4] U. von Luxburg, M. Belkin, and O. Bousquet, "Consitency of spectral clustering," *Ann. Stat.*, vol. 26, no. 2, pp. 555–586, 2008. DOI: 10.1214/009053607000000640.

# Further reading


[5] V. I. Paulsen and M. Raghupathi, *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*, ser. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2016, vol. 152.

[6] B. K. Sriperumbudur, K. Fukumizu, and G. R. Lanckriet, "Universality, characteristic kernels and RKHS embedding of measures," *J. Mach. Learn. Res.*, vol. 12, pp. 2389–2410, 2011.

# Section 4

## Spectral theory

# Setting and objectives

**General assumptions**

- $\Phi : G \times \Omega \to \Omega$: Continuous-time, continuous flow on compact, metrizable space $\Omega$.
- $\mu$: Ergodic invariant Borel probability measure.
- $X : \Omega \to \mathbb{X}$ continuous observation map into metric space $\mathcal{X}$.
- $U^t : \mathcal{F} \to \mathcal{F}$: Koopman operator on Banach space $\mathcal{F}$ of complex-valued observables.

**Given.** Time-ordered samples

$$x_n = X(\omega_n), \quad \omega_n = \Phi^{t_n}(\omega_0), \quad t_n = (n-1)\,\Delta t.$$

**Goal.** Using the data $x_n$, identify a collection of observables $\zeta_j : \Omega \to \mathcal{Y}$ which have the property of evolving coherently under the dynamics in a suitable sense.

# Setting and objectives

We recall the following facts from Section 2 (see Proposition 2.7 and Theorems 2.29, 2.30).

**Theorem 4.1.**

1. $\{U^t : C(\Omega) \to C(\Omega)\}_{t\in\mathbb{R}}$ *is a strongly continuous group of isometries.*
2. $\{U^t : L^p(\mu) \to L^p(\mu)\}_{t\in\mathbb{R}}$, $p \in [0,\infty)$ *is a strongly continuous group of isometries. Moreover,* $U^t : L^2(\mu) \to L^2(\mu)$ *is unitary.*
3. $\{U^t : L^\infty(\mu) \to L^\infty(\mu)\}_{t\in\mathbb{R}}$ *is a weak-$^*$ continuous group of isometries.*

**Notation.**

- $\mathcal{F}$: Any of the $C(\Omega)$ or $L^p(\mu)$ spaces with $1 \leq p \leq \infty$.
- $\mathcal{F}_0$: Any of the $C(\Omega)$ or $L^p(\mu)$ spaces with $1 \leq p < \infty$.
- $C_0$ (semi)group $\equiv$ strongly continuous (semi)group.
- $C_0^*$ (semi)group $\equiv$ weak-$^*$ continuous (semi)group.

# Generator of $C_0$ semigroups

**Definition 4.2.**
Let $\{S^t\}_{t\geq 0}$ be a $C_0$ semigroup on a Banach space $E$. The generator $A : D(A) \to E$ of the semigroup $\{S^t\}_{t\geq 0}$ is defined as

$$Af = \lim_{t\to 0} \frac{S^t f - f}{t}, \quad f \in D(A),$$

where the limit is taken in the norm of $E$, and the domain $D(A) \subseteq E$ consists of all $f \in E$ for which the limit exists.

# Generator of $C_0$ semigroups

**Theorem 4.3.**
*With the notation of Definition 4.2, the following hold.*

1. *$A$ is closed and densely defined.*
2. *For all $f \in D(A)$ and $t \geq 0$, the function $t \mapsto S^t f$ is continuously differentiable, and satisfies*

$$\frac{d}{dt} S^t f = A S^t f = S^t A f.$$

3. *$A$ uniquely characterizes the semigroup $\{S^t\}$, i.e., if $\{\tilde{S}^t\}$ is another $C_0$ semigroup on $E$ with the same generator $A$, then $S^t = \tilde{S}^t$ for all $t \geq 0$.*

# Generator of $C_0^*$ semigroups

**Definition 4.4.**
Let $\{S^t\}_{t\geq 0}$ be a $C_0^*$ semigroup on a Banach space $E$ with predual $E_*$. The generator $A : D(A) \to E$ of the semigroup $\{S^t\}_{t\geq 0}$ is defined as the weak-* limit

$$\langle g, Af \rangle = \lim_{t\to 0} \frac{\langle g, S^t f - f \rangle}{t}, \quad f \in D(A), \quad \forall g \in E_*,$$

where the domain $D(A) \subseteq E$ consists of all $f \in E$ for which the limit exists.

**Theorem 4.5.**
*With the notation of Definition 4.4, the following hold.*

1. *$A$ is weak-$^*$ closed and densely defined.*
2. *For all $f \in D(A)$ and $t \geq 0$, the function $t \mapsto S^t f$ is weak-$^*$ continuously differentiable, and satisfies*

$$\left\langle g, \frac{d}{dt} S^t f \right\rangle = \langle g, AS^t f\rangle = \langle g, S^t Af\rangle.$$

3. *$A$ uniquely characterizes the semigroup $\{S^t\}$, i.e., if $\{\tilde{S}^t\}$ is another $C_0^*$ semigroup on $E$ with the same generator $A$, then $S^t = \tilde{S}^t$ for all $t \geq 0$.*

# Generator of unitary $C_0$ groups

**Theorem 4.6 (Stone).**

*Let $\{S^t\}_{t\geq 0}$ be a unitary $C_0$ group on a Hilbert space $H$. Then, the generator $A : D(A) \to H$ is skew-adjoint, i.e.,*

$$A^* = -A.$$

*Conversely, if $A : D(A) \to H$ is skew-adjoint, it is the generator of a unitary evolution group.*

# Generator of Koopman evolution groups

**Corollary 4.7.**

*Under our general assumptions the following hold:*

1. *The Koopman evolution groups* $U^t : \mathcal{F}_0 \to \mathcal{F}_0$ *are uniquely characterized by their generator* $V : D(V) \to \mathcal{F}_0$*, where*

$$Vf = \lim_{t \to 0} \frac{U^t f - f}{t}.$$

   *Moreover, for* $\mathcal{F}_0 = L^2(\mu)$*,* $V$ *is skew-adjoint.*

2. *The Koopman evolution group* $U^t : L^\infty(\mu) \to L^\infty(\mu)$ *is uniquely characterized by its generator* $V : D(V) \to \mathcal{F}_0$*, where*

$$Vf = \lim_{t \to 0} \frac{U^t f - f}{t}$$

   *in weak-*$^*$ *sense.*

# Generator of Koopman evolution groups

**Theorem 4.8 (ter Elst & Lemańczyk).**

*Let $(\Omega, \Sigma)$ be a compact metrizable space equipped with its Borel $\sigma$-algebra $\Sigma$. Let $\mu$ be a Borel probability measure on $\Omega$ and $U^t : L^2(\mu) \to L^2(\mu)$ a $C_0$ unitary evolution group with generator $V : D(V) \to L^2(\mu)$. Then, the following are equivalent.*

1. *For every $t \in \mathbb{R}$ there exists a $\mu$-a.e. invertible, measurable, and measure-preserving flow $\Phi^t : \Omega \to \Omega$ such that $U^t f = f \circ \Phi^t$.*
2. *The space $\mathfrak{A}(V) = D(V) \cap L^\infty(\mu)$ is an algebra with respect to function multiplication, and $V$ is a derivation on $\mathfrak{A}$:*

$$V(fg) = (Vf)g + f(Vg), \quad \forall f, g \in \mathfrak{A}(V).$$

# Point spectrum

**Definition 4.9.**
Let $A : D(A) \to E$ be an operator on a Banach space with domain $D(A) \subseteq E$. The point spectrum of $A$, denoted as $\sigma_p(A) \subseteq \mathbb{C}$ is defined as the set of its eigenvalues. That is, $\lambda \in \mathbb{C}$ is an element of $\sigma_p(A)$ iff there is a nonzero vector $u \in E$ (an eigenvector) such that

$$Au = \lambda u.$$

**Notation.**

- We use the notation $\sigma_p(A; E)$ when we wish to make explicit the Banach space on which $A$ acts.

# Eigenvalues and eigenfunctions

**Definition 4.10.**
Let $A : D(A) \to E$ be the generator of a $C_0$ semigroup $\{S^t\}_{t \geq 0}$ on a Banach space $E$. We say that $\lambda \in \mathbb{C}$ is an eigenvalue of the semigroup if $\lambda$ is an eigenvalue of $A$, i.e., there exists a nonzero $u \in D(A)$ such that

$$Au = \lambda u.$$

**Lemma 4.11.**
*With notation as above, $\lambda$ is an eigenvalue of $\{S^t\}$ if and only if $z$ is an eigenvector of $S^t$ for all $t \geq 0$, i.e., there exist $\Lambda^t \in \mathbb{C}$ such that*

$$S^t u = \Lambda^t u, \quad \forall t \geq 0.$$

*In particular, we have $\Lambda^t = e^{\lambda t}$.*

# Point spectra for measure-preserving flows

**Theorem 4.12.**
*Let $\Phi^t : \Omega \to \Omega$ a be a measure-preserving flow of a probability space $(\Omega, \Sigma, \mu)$. Let $U^t : L^p(\mu) \to L^p(\mu)$ be the associated Koopman operators on $L^p(\mu)$, $p \in [1, \infty]$, and $V : D(V) \to L^p(\mu)$ the corresponding generators. Then, the following hold.*

1. *For every $p, q \in [1, \infty]$ and $t \in \mathbb{R}$, $\sigma_p(U^t, L^p(\mu)) = \sigma_p(U^t, L^q(\mu))$.*
2. *$\sigma_p(V, L^p(\mu)) = \sigma_p(V, L^q(\mu))$.*
3. *$\sigma_p(U^t)$ is a subgroup of $S^1$.*
4. *$\sigma_p(V)$ is a subgroup of $i\mathbb{R}$.*

**Corollary 4.13.**
*Every eigenfunction of $V$ lies in $L^\infty(\mu)$, and thus in $L^p(\mu)$ for every $p \in [1, \infty]$.*

Given $\lambda = i\alpha \in \sigma_p(V)$, we say that $\alpha$ is an eigenfrequency of $V$.

# Generating frequencies

**Definition 4.14.**
Assume the notation of Theorem 4.12.

1. We say that $\{ia_0, ia_1, \ldots\} \subseteq \sigma_p(V)$ is a generating set if for every $i\alpha \in \sigma_p(V)$ there exist $j_1, j_2, \ldots, j_n \in \mathbb{Z}$ and $k_1, k_2, \ldots, k_n \in \mathbb{N}$ such that
$$\alpha = j_1\alpha_{k_1} + j_2\alpha_{k_2} + \ldots + j_n\alpha_{k_n}.$$
2. We say that $\sigma_p(V)$ is finitely generated if it has a finite generating set.
3. A generating set is said to be minimal if it does does not have any proper subsets which are generating sets.

**Lemma 4.15.**

1. *The elements of a minimal generating set are rationally independent.*
2. *If a minimal generating set has at least two elements, then $\sigma_p(V)$ is a dense subset of the imaginary line.*

# Generating frequencies

**Lemma 4.16.**
*Let $g_1, g_2, \ldots$ be eigenfunctions corresponding to the eigenvalues of the generating set in Definition 4.14, i.e., $Vg_j = i\alpha_j g_j$. Then, for every $i\alpha \in \sigma_p(V)$ with $\alpha = j_1\alpha_{k_1} + j_2\alpha_{k_2} + \ldots + j_n\alpha_{k_n}$,*

$$z = g_{k_1}^{j_1} g_{k_2}^{j_2} \cdots g_{k_n}^{j_n}$$

*is an eigenfunction of $V$ corresponding to the eigenfrequency $\alpha$.*

# Invariant subspaces

**Notation.**

- $H_p = \overline{\operatorname{span}\{u \in L^2(\mu) \colon u \text{ is an eigenfunction of } V\}}$.
- $H_c = H_p^{\perp}$.
- $\{z_0, z_1, \dots\}$: Orthonormal eigenbasis of $H_p$, $V z_j = i\alpha_j z_j$.

**Theorem 4.17.**
*Let $\Phi^t : \Omega \to \Omega$ be a measure-preserving flow on a completely metrizable space with an invariant probability measure $\mu$.*

1. *$H_p$ and $H_c$ are $U^t$-invariant subspaces.*
2. *Every $f \in H_p$ satisfies*

$$U^t f = \sum_{j=0}^{\infty} \hat{f}_j e^{i\alpha_j t} z_j, \quad \hat{f}_j = \langle z_j, f \rangle_{L^2(\mu)}.$$

3. *Every $f \in H_c$ satisfies*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T |\langle g, U^t f \rangle_{L^2(\mu)}| = 0, \quad \forall g \in L^2(\mu).$$

# Pure point spectrum

**Definition 4.18.**
With the notation of Theorem 4.17, we say that a measure-preserving flow $\Phi^t : \Omega \to \Omega$ has pure point spectrum if $H_p = L^2(\mu)$.

**Remark 4.19.**
For a system with pure point spectrum:

1. The spectrum of $V$ is not necessarily discrete.
2. The continuous spectrum is not necessarily empty.

# Point spectra for ergodic flows

**Proposition 4.20.**

*With the notation of Theorem 4.12, assume that* $\Phi^t : \Omega \to \Omega$ *is ergodic.*

1. *Every eigenvalue* $\lambda \in \sigma_p(V)$ *is simple.*
2. *Every corresponding eigenfunction* $z \in L^p(\mu)$ *normalized such that* $\|z\|_{L^p(\mu)} = 1$ *for any* $p \in [1, \infty]$ *satisfies* $|z| = 1$ $\mu$*-a.e.*

# Factor maps

**Definition 4.21.**
Let $T_1 : \Omega_1 \to \Omega_1$ and $T_2 : \Omega_2 \to \Omega_2$ be measure-preserving transformations of the probability spaces $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$. We say that $T_2$ is a factor of $T_1$ if there exists a $T_1$-invariant set $S_1 \in \Sigma_1$ with $\mu_2(S_+1) = 1$, a $T_2$-invariant set $S_2 \in \Sigma_2$ with $\mu_2(S_2) = 1$, and a measure-preserving, surjective map $\varphi : S_1 \to S_2$ such that

$$T_2 \circ \varphi = \varphi \circ T_1.$$

Such a map $\varphi$ is called a factor map and satisfies the following commutative diagram:

$$\begin{array}{ccc} M_1 & \xrightarrow{T_1} & M_1 \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\varphi} \\ M_2 & \xrightarrow{T_2} & M_2 \end{array} .$$

# Metric isomorphisms

**Definition 4.22.**
With the notation of Definition 4.21, we say that $T_1$ and $T_2$ are measure-theoretically isomorphic or metrically isomorphic if there is a factor $\varphi : S_1 \to S_2$ with a measurable inverse.

**Theorem 4.23 (von Neumann).**
*Let $\Phi^t : \Omega \to \Omega$ be a measure-preserving flow on a completely metrizable probability space $(\Omega, \Sigma, \mu)$ with pure point spectrum. Then, $\Phi^t$ is metrically isomorphic to a translation on a compact abelian group $\mathcal{G}$. Explicitly, $\mathcal{G}$ can be chosen as the character group of the point spectrum $\sigma_p(V)$.*

# Metric isomorphisms

**Corollary 4.24.**

*If $\sigma_p(V)$ is finitely generated, then $\Phi^t$ is metrically isomorphism to an ergodic rotation on the d-torus, where d is the number of generating frequencies of $\sigma_p(V)$. Explicitly, supposing that $\{i\alpha_1, \ldots, i\alpha_d\}$ is a minimal generating set of $\sigma_p(V)$ with corresponding unit-norm eigenfunctions $z_1, \ldots, z_d$ we have*

$$R^t \circ \varphi = \varphi \circ \Phi^t,$$

*where $R^t : \mathbb{T}^d \to \mathbb{T}^d$ is the torus rotation with frequencies $\alpha_1, \ldots, \alpha_d$, and*

$$\varphi(\omega) = (z_1(\omega), \ldots, z_d(\omega)), \quad \mu\text{-a.e.}$$

# Spectral isomorphisms

**Definition 4.25.**
With the notation of Definition 4.22, let $U_1 : L^2(\mu_1) \to L^2(\mu_1)$ and $U_2 : L^2(\mu_2) \to L^2(\mu_2)$ be the Koopman operators associated with $T_1$ and $T_2$, respectively. We say that $T_1$ and $T_2$ are spectrally isomorphic if there exists a unitary map $\mathcal{U} : L^2(\mu_1) \to L^2(\mu_2)$ such that

$$U_2 \circ \mathcal{U} = \mathcal{U} \circ U_1.$$

**Theorem 4.26 (von Neumann).**
*Two measure-preserving flows with pure point spectra are metrically isomorphic iff they are spectrally isomorphic.*

# Dynamics-invariant kernels

$$k : \Omega \times \Omega \to \mathbb{R}, \quad G : L^2(\mu) \to L^2(\mu), \quad Gf = \int_\Omega k(\cdot, \omega) f(\omega) \, d\mu(\omega)$$

- $k$: Bounded, symmetric kernel.
- $G$ is self-adjoint, compact.

**Proposition 4.27.**
*If $k$ is invariant under the product flow,*

$$k(\Phi^t(\omega), \Phi^t(\omega')) = k(\omega, \omega'),$$

*then $G$ commutes with the Koopman operator,*

$$[U^t, G] = U^t G - G U^t = 0.$$

# Dynamics-invariant kernels

$$k : M \times M \to \mathbb{R}, \quad G : L^2(\mu) \to L^2(\mu), \quad Gf = \int_\Omega k(\cdot, \omega) f(\omega)\, d\mu(\omega)$$

**Corollary 4.28.**

*Every eigenspace $W$ of $G$ with nonzero corresponding eigenvalue is a finite-dimensional, $U^t$-invariant subspace of $H_p$, and $V|_W$ is unitarily diagonalizable.*

# Kernels from delay-coordinate maps

$$S_Q(\omega,\omega') = \frac{1}{Q}\sum_{q=0}^{Q-1} \left\| X(\Phi^{q\,\Delta t}(\omega)) - X(\Phi^{q\,\Delta t}(\omega')) \right\|^2.$$

By the mean ergodic theorem,

$$S_Q \xrightarrow[Q\to\infty]{} \bar{S},$$

in $L^2(\mu\times\mu)$, where $\bar{S}$ is a $U^t \otimes U^t$ invariant function.

**Proposition 4.29.**

*Fix a continuous kernel shape function $h:\mathbb{R}_+\to\mathbb{R}_+$. Then:*

1. $\bar{k}(\omega,\omega') := h(\bar{S}(\omega,\omega'))$ *satisfies the assumptions of Proposition 4.27.*
2. $G_Q : L^2(\mu)\to L^2(\mu)$ *with*

$$G_Q f = \int_\Omega k_Q(\cdot,\omega) f \, d\mu(\omega)$$

*converges to $G$ in $L^2(\mu)$ operator norm.*

# Finite-difference approximation of the generator

$$V_{\Delta t,N} : L^2(\mu_N) \to L^2(\mu_N), \; V_{\Delta t,N} = \frac{\tilde{V}_{\Delta t,N} - \tilde{V}^*_{\Delta t,N}}{2}, \; \tilde{V}_{\Delta t,N} = \frac{\hat{U}_N - \mathrm{Id}}{\Delta t}$$

Explicitly, we have

$$\tilde{V}_{\Delta t,N} f(\omega_n) = \begin{cases} (f(\omega_{n+1}) - f(\omega_n))/\Delta t, & 0 \le n \le N-2, \\ -f(\omega_{N-1})/\Delta t, & n = N-1. \end{cases}$$

# Finite-difference approximation of the generator

$$V_{\Delta t,N} : L^2(\mu_N) \to L^2(\mu_N),\ V_{\Delta t,N} = \frac{\tilde{V}_{\Delta t,N} - \tilde{V}^*_{\Delta t,N}}{2},\ \tilde{V}_{\Delta t,N} = \frac{\hat{U}_N - \mathrm{Id}}{\Delta t}$$

**Lemma 4.30.**
*For $f \in C^1(\Omega)$ and $g \in C(\Omega)$,*

$$\lim_{\Delta t \to 0} \lim_{N \to \infty} \langle g, V_{\Delta t,N} f \rangle_{L^2(\mu_N)} = \langle g, Vf \rangle_{L^2(\mu)}.$$

**Corollary 4.31.**
*With the notation of Section 3, if $k$ is $C^1$, then for every $i, j \in \mathbb{N}$ such that $\lambda_i, \lambda_j \neq 0$,*

$$\lim_{\Delta t \to 0} \lim_{N \to \infty} \langle \phi_{i,N} V_{N,\Delta t} \phi_{j,N} \rangle_{L^2(\mu_N)} = \langle \phi_i, V \phi_j \rangle_{L^2(\mu)}.$$

# Markov normalization

$$p_\nu(\omega, \omega') = \frac{\tilde{k}(\omega, \omega')}{\rho_\nu(\omega)}, \quad \tilde{k}_\nu(\omega, \omega') = \frac{k(\omega, \omega')}{\sigma_\nu(\omega')},$$
$$\rho_\nu(\omega) = \int_\Omega \tilde{k}_\nu(\omega, \omega') \, d\nu(\omega'), \quad \sigma_\nu(\omega') = \int_\Omega k(\omega', \omega'') \, d\nu(\omega'')$$

- Assume: $k \geq 0$, $k, k^{-1} \in L^\infty(\nu \times \nu)$.
- $p$ is a Markov kernel with respect to $\nu$, i.e.,

$$p \geq 0, \quad \int_\Omega p(\omega, \cdot) \, d\nu = 1, \quad \nu\text{-a.e. } \omega \in M.$$

# Markov normalization

$$p_\nu(\omega, \omega') = \frac{\tilde{k}(\omega, \omega')}{\rho_\nu(\omega)}, \quad \tilde{k}_\nu(\omega, \omega') = \frac{k(\omega, \omega')}{\sigma_\nu(\omega')},$$

$$\rho_\nu(\omega) = \int_\Omega \tilde{k}_\nu(\omega, \omega')\, d\nu(\omega'), \quad \sigma_\nu(\omega') = \int_\Omega k(\omega', \omega'')\, d\nu(\omega'')$$

**Set:** $k = k_Q$, $\nu = \mu_N$ or $\nu = \mu$. We get Markov operators $G_{Q,N} : L^2(\mu_N) \to L^2(\mu_N)$, $G_Q : L^2(\mu) \to L^2(\mu)$ with continuous transition kernels:

$$G_{Q,N} f = \int_\Omega p_{Q,\mu_N}(\cdot, \omega) f(\omega)\, d\mu_N(\omega), \quad Gf = \int_\Omega p_{Q,\mu}(\cdot, \omega) f(\omega)\, d\mu_N(\omega),$$

Large-data limit: As $N \to \infty$, $G_{Q,N}$ converges spectrally to $G_Q$ in the sense of Theorem 3.25.

# Markov normalization

$$p_\nu(\omega,\omega') = \frac{\tilde{k}(\omega,\omega')}{\rho_\nu(\omega)}, \quad \tilde{k}_\nu(\omega,\omega') = \frac{k(\omega,\omega')}{\sigma_\nu(\omega')},$$
$$\rho_\nu(\omega) = \int_\Omega \tilde{k}_\nu(\omega,\omega')\,d\nu(\omega'), \quad \sigma_\nu(\omega') = \int_\Omega k(\omega',\omega'')\,d\nu(\omega'')$$

**Set:** $k = \bar{k}$, $\nu = \mu$. We get a self-adjoint Markov operator $G : L^2(\mu) \to L^2(\mu)$ that commutes with the Koopman operator:

$$Gf = \int_\Omega \bar{p}_\mu(\cdot,\omega) f(\omega)\,d\mu(\omega).$$

Infinite-delay limit: As $Q \to \infty$ $G_Q$ converges in operator norm, and thus spectrally, to $G$.

**Remark.**
By Corollary 4.28, every eigenfunction $\phi_j$ of $G$ corresponding to nonzero eigenvalue lies in the domain of the generator $V$.

# Diffusion regularization

$$\Delta : D(\Delta) \to \tilde{H}_p, \quad \Delta = (I - G)^{-1}$$

$$\Delta \phi_j = \eta_j \phi_j, \quad \eta_j = 1 - \frac{1}{\lambda_j}$$

- $\tilde{H}_p = \overline{\text{ran}\, G} \subseteq H_p$.
- $D(\Delta) \equiv \tilde{H}_p^2 = \{f \in \tilde{H}_p : \sum_j \eta_j |\langle \phi_j, f \rangle_{L^2(\mu)}|^2 < \infty\}$.

**Proposition 4.32.**

1. *For every $\epsilon > 0$,*

$$\mathcal{L}_\epsilon = V - \epsilon \Delta,$$

   *is a well-defined dissipative operator on $\tilde{H}_p^2$, i.e.,* $\text{Re}\langle f, \mathcal{L}_\epsilon f \rangle \leq 0$.

2. *Let $z$ be an eigenfunction of $V$ lying in $H_p^2$ with corresponding eigenvalue $i\omega$. Then, we have*

$$\Delta z = \eta z, \quad L_\epsilon z = \gamma z, \quad \gamma = -\epsilon \eta + i\omega.$$

# Petrov-Galerkin method

**Infinite-dimensional variational problem**

*Find* $z_j \in \tilde{H}_p^2$ *and* $\gamma_j \in \mathbb{C}$*, such that for all* $f \in \tilde{H}_p$*,*

$$\langle f, V z_j \rangle_{L^2(\mu)} - \epsilon \langle f, \Delta z \rangle_{L^2(\mu)} = \gamma_j \langle f, z \rangle_{L^2(\mu)}.$$

- The above is a well-defined variational eigenvalue problem, i.e., it satisfies the appropriate boundedness and coercivity conditions.
- We order the solutions $z_j$ in order of increasing Dirichlet energy,

$$E_j = \langle z_j, \Delta z_j \rangle_{L^2(\mu)} = \operatorname{Re} \gamma_j / \epsilon.$$

# Petrov-Galerkin method

**Data-driven approximation**

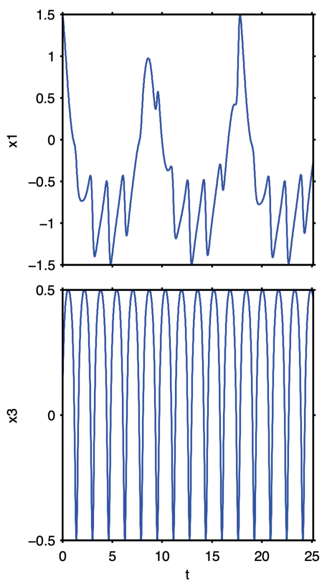
*Find* $z_j \in \tilde{H}^2_{p,L,Q,N}$ *and* $\gamma \in \mathbb{C}$*, such that for all* $f \in \tilde{H}_{p,L,Q,N}$*,*

$$\langle f, V z_j \rangle_{L^2(\mu_N)} - \epsilon \langle f, \Delta z_j \rangle_{L^2(\mu_N)} = \gamma_j \langle f, z_j \rangle_{L^2(\mu_N)}.$$

- $\tilde{H}_{p,L,Q,N} = \operatorname{span}\{\phi_{0,Q,N}, \dots, \phi_{L-1,Q,N}\} \subseteq L^2(\mu_N)$, where $\phi_{j,Q,N}$ are eigenfunctions of $G_{Q,N}$.
- $H^2_{p,L,Q,N}$ defined analogously to $\tilde{H}^2_p$.
- The data-driven scheme converges in the iterated limit

$$\lim_{L\to\infty} \lim_{Q\to\infty} \lim_{\Delta t\to 0} \lim_{N\to\infty}.$$

# Variable-speed rotation on $\mathbb{T}^2$

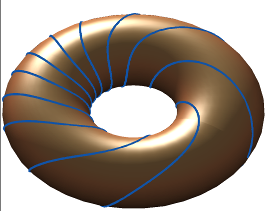



$$\dot{\omega}(t) = \vec{V}(\omega(t))$$

$$\vec{V}(\omega) = (V_1, V_2), \quad \omega = (\theta_1, \theta_2)$$

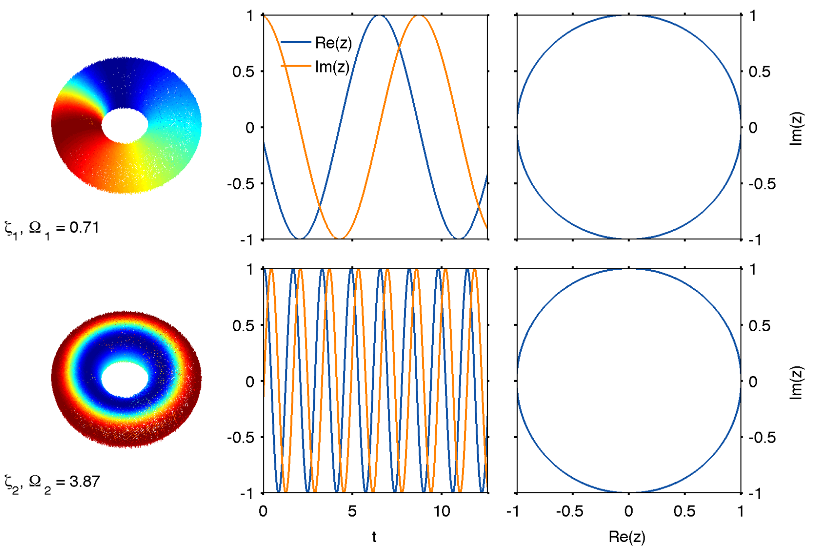
$$V_1 = 1 + \beta \cos \theta_1$$

$$V_2 = \alpha(1 - \beta \sin \theta_2)$$

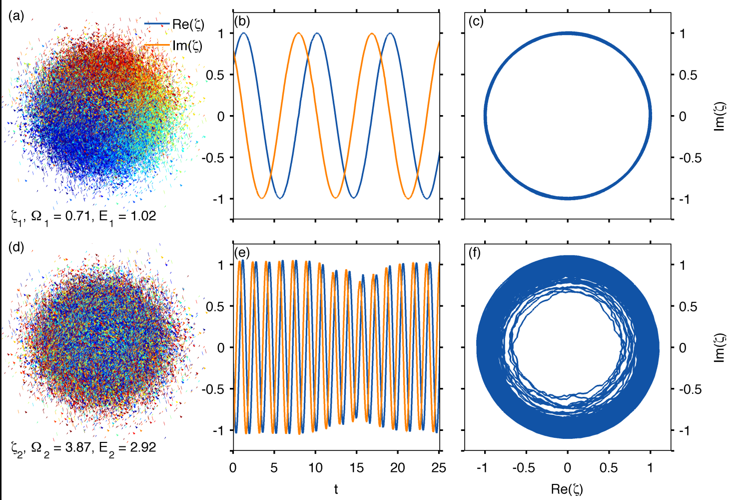


$\alpha = \sqrt{30}$, $\beta = \sqrt{1/2}$

# 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 $\simeq 1$.

# Approximate Koopman eigenfunctions

**Definition 4.33.**
An observable $z \in L^2(\mu)$ is said to be an $\epsilon$-approximate Koopman eigenfunction if there exists $\nu_t \in \mathbb{C}$ such that

$$\|U^t z - \nu_t z\|_{L^2(\mu)} < \epsilon \|z\|_{L^2(\mu)}.$$

- A Koopman eigenfunction is an $\epsilon$-approximate eigenfunction for every $\epsilon > 0$.
- We seek $z \in L^2(\mu)$ which is an $\epsilon$-approximate eigenfunction for “small” $\epsilon$, and $t$ lying in a “large” time interval.

# Approximate eigenfunctions from delay-coordinate maps

**Theorem 4.34.**
*Let $\phi$ and $\psi$ be mutually-orthogonal, unit-norm, real eigenfunctions of $G_Q$ corresponding to nonzero eigenvalues $\kappa$ and $\lambda$, respectively, with $\kappa \geq \lambda$. Assume that $\kappa, \lambda$ are simple if distinct and twofold-degenerate if equal. Define*

$$z = \frac{1}{\sqrt{2}}(\phi + i\psi), \quad \alpha_t = \langle z, U^t z \rangle, \quad \nu = \langle \psi, V\phi \rangle,$$

*where $\omega$ is real, and set $T = (Q-1)\,\Delta t$, $\delta_T = (\kappa - \lambda)/\sqrt{2}$, $\tilde{\delta}_T = \delta_T/\kappa$,*

$$\gamma_T = \min_{u \in \sigma(G_Q) \setminus \{\kappa, \lambda\}} \{\min\{|\kappa - u|, |\lambda - u|\}\}.$$

*Then, the following hold for every $t \geq 0$:*

# Approximate eigenfunctions from delay-coordinate maps

**Theorem 4.34.**

1. *$\alpha_t$ lies in the $\tilde\epsilon_t$-approximate point spectrum of $U^t$, and $z$ is a corresponding $\tilde\epsilon_t$-approximate eigenfunction for the bound*

$$\tilde\epsilon_t = s_t + \sqrt{S_t},$$

*where*

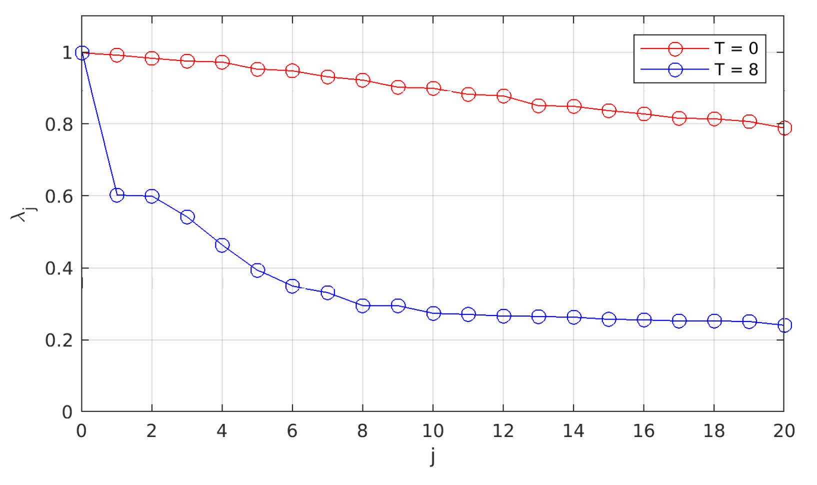
$$s_t = \frac{1}{\gamma_T}\left(\frac{C_1 t}{T} + 3\delta_T\right), \quad S_t = \frac{C_2(1+\tilde\delta_T)}{\lambda}\int_0^t s_u\,du.$$

*Here, $C_1$ and $C_2$ are constants that depend only on the observation map $F$ and generator $V$.*

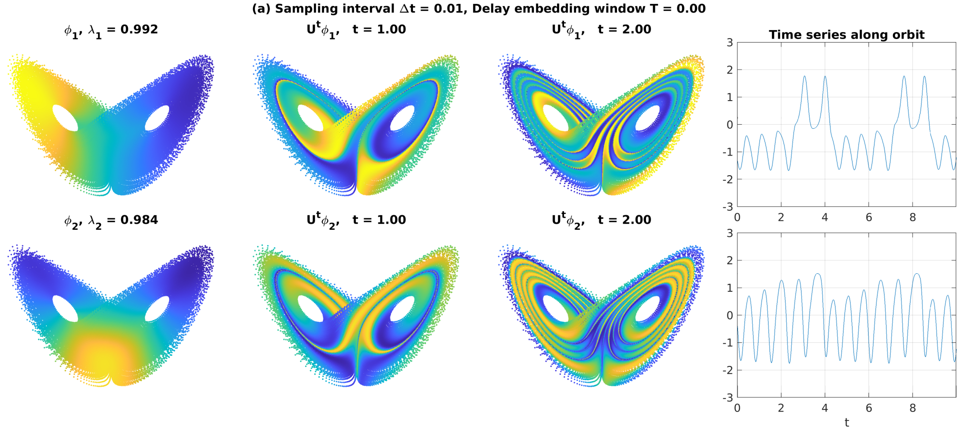
2. *The modulus $|\nu|$ is independent of the choice of the real orthonormal basis $\{\phi, \psi\}$ for the eigenspace(s) corresponding to $\kappa$ and $\lambda$. Moreover, the phase factor $e^{i\nu t}$ is related to the autocorrelation function $\alpha_t$ according to the bound*

$$|\alpha_t - e^{i\nu t}| \le 2\sqrt{S_t}.$$

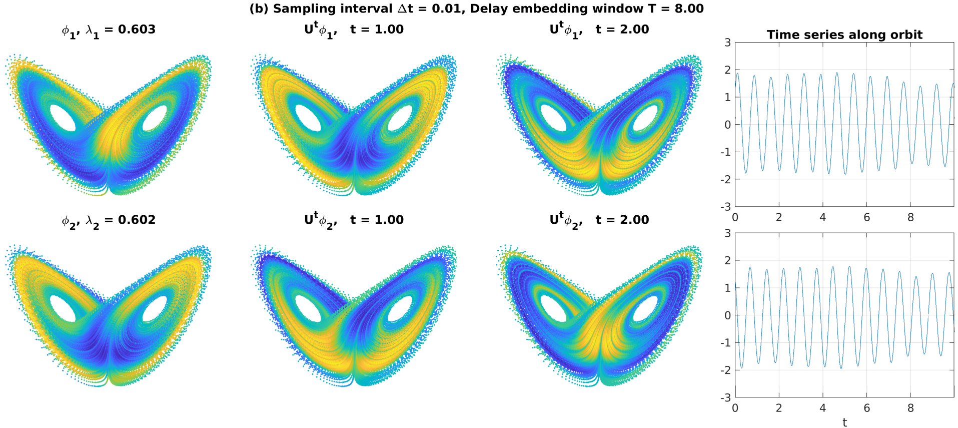
# Application to L63 system



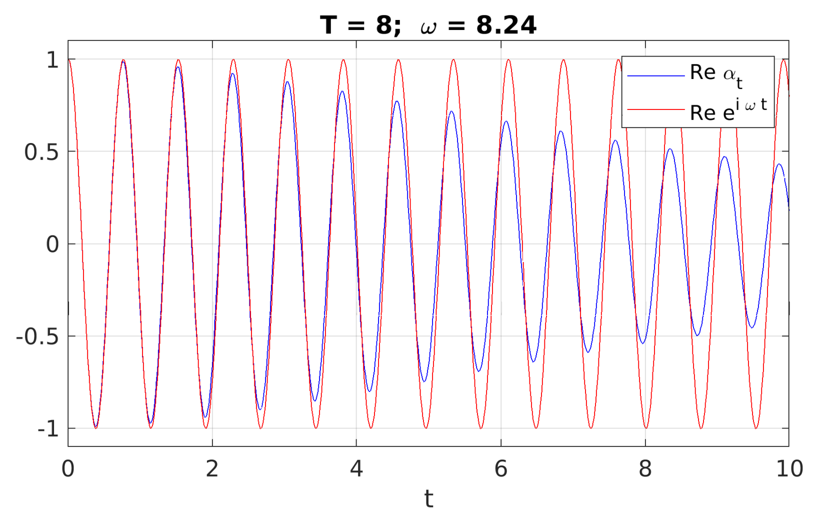
# Application to L63 system



# Application to L63 system



# Application to L63 system



# Spectrum

**Definition 4.35.**
Let $A : D(A) \to F$ be a densely-define operator on a Banach space $F$ over $\mathbb{C}$ with domain $D(A) \subseteq F$.

1. The spectrum of $A$, denoted as $\sigma(A)$ is the set of complex numbers $\lambda$ such that $A - \lambda I$ has no bounded inverse.
2. The resolvent set of $A$, denoted as $\rho(A)$, is the complement of $\sigma(A)$ in $\mathbb{C}$.
3. For every $\lambda \in \rho(A)$ the resolvent $R_A(\lambda)$ is the bounded operator given by $\rho(A) = (A - \lambda I)^{-1}$.
4. The spectral radius of $A$ is defined as $r_\sigma(A) = \sup_{\lambda \in \sigma(A)|\lambda|}$.

## Spectrum

**Theorem 4.36.**
*With the notation of Definition 4.35, the following hold.*

1. $\sigma(A)$ *is a closed subset of* $\mathbb{C}$.
2. *If* $A$ *is not closed, then* $\sigma(A) = \mathbb{C}$.
3. *If* $D(A) = F$ *and* $A$ *is bounded, then* $r_\sigma(A) \le \|A\|$.

# Projection-valued measures

**Definition 4.37.**
Let $(H, \langle \cdot, \cdot \rangle_H)$ be a Hilbert space over $\mathbb{C}$. A map $E : \mathfrak{B}(\mathbb{C}) \to B(H)$ is called a projection-valued measure (PVM) if:

1. For every $S \in \mathfrak{B}(\mathbb{C})$, $E(S)$ is an orthogonal projection.
2. $E(\mathbb{C}) = I$.
3. For every $f, g \in H$, the map $\varepsilon_{fg} : \mathfrak{B}(\mathbb{C}) \to \mathbb{C}$ with

$$\varepsilon_{fg}(S) = \langle f, E(S)g \rangle_H$$

   is a complex measure.

# Projection-valued measures

**Theorem 4.38.**
*With the notation of Definition 4.37, let $f : \mathbb{C} \to \mathbb{C}$ be a Borel-measurable function. Then, there exists a unique operator $E_f : D(E_f) \to H$ with domain*

$$D(E_f) = \left\{ h \in H : \int_{\mathbb{C}} |f|^2 \, d\varepsilon_{hh} < \infty \right\},$$

*such that*

$$\langle g, E_f h \rangle_H = \int_{\mathbb{C}} f \, d\varepsilon_{gh}, \quad \forall g \in H, \quad \forall h \in D(E_f).$$

**Notation.**

- $\int_{\mathbb{C}} f \, dE \equiv E_f$.
- If $A = \int_{\mathbb{C}} \mathrm{Id} \, dE$, then $f(A) \equiv E_f$.

# Spectral theorem for skew-adjoint operators

**Theorem 4.39.**
*Let $A : D(A) \to H$ be skew-adjoint.*

1. $\sigma(A)$ *is a subset of the imaginary line.*
2. *There exists a unique PVM* $E_A : \mathfrak{B}(\mathbb{C}) \to \mathbb{C}$ *such that*

$$A = \int_{\mathbb{R}} i\alpha \, dE(\alpha).$$

3. $i \operatorname{supp} E_A = \sigma(A)$.
4. *If* $\{U^t : H \to H\}_{t \in \mathbb{R}}$ *is the* $C_0$ *unitary group generated by* $A$*, then*

$$U^t = e^{tA} \equiv \int_{\mathbb{R}} e^{i\alpha t} \, dE(\alpha).$$

# Unitary Koopman evolution group

$$U^t : L^2(\mu) \to L^2(\mu), \quad U^t f = f \circ \Phi^t, \quad U^{t*} = U^{-t}$$

Generator: $V : D(V) \to L^2(\mu)$,

$$D(V) \subset L^2(\mu), \quad V^* = -V, \quad Vf = \lim_{t\to 0} \frac{U^t f - f}{t}.$$

Spectral measure: $E : \mathfrak{B}(\mathbb{R}) \to B(L^2(\mu))$,

$$V = \int_{\mathbb{R}} i\omega \, dE(\alpha), \quad U^t = \int_{\mathbb{R}} e^{i\alpha t} \, dE(\omega).$$

# Unitary Koopman evolution group

$$U^t : L^2(\mu) \to L^2(\mu), \quad U^t f = f \circ \Phi^t, \quad U^{t*} = U^{-t}$$

**Theorem 4.40.**
*There is a $U^t$-invariant orthogonal splitting $L^2(\mu) = H_p \oplus H_c$ such that:*

1. *$H_p$ has an orthonormal basis $\{z_j\}$ consisting of eigenfunctions of the generator,*
$$V z_j = i\alpha_j z_j, \quad \alpha_j \in \mathbb{R}.$$
2. *For every $f \in H_c$ and $g \in L^2(\mu)$,*
$$\lim_{T\to\infty} \frac{1}{T} \int_0^T |\langle g, U^t f\rangle_{L^2(\mu)}| \, dt = 0.$$
3. *$E = E_p + E_c$, where:*
   - *$E_p$ is a purely atomic measure taking values in $B(H_p)$.*
   - *$E_c$ is a continuous measure taking values in $B(H_c)$.*

# Compactification schemes for the Koopman generator

**Given:**

Positive-definite, $C^1$ kernel $k : \Omega \times \Omega \to \mathbb{R}$.

Integral operators $K : L^2(\mu) \to \mathcal{K}$, $G = K^* K$.

**Pre-smoothing:**

$$A : L^2(\mu) \to L^2(\mu), \quad A = VG.$$

- $\operatorname{ran} G \subseteq \operatorname{ran} K^* \subset D(V)$.
- $A = VG$ is a Hilbert-Schmidt integral operator on $L^2(\mu)$ with kernel $k' \in C(X \times X)$, $k'(\cdot, \omega) = Vk(\cdot, \omega)$, i.e.,

$$Af = \int_\Omega k'(\cdot, \omega) f(\omega) \, d\mu(\omega).$$

# Compactification schemes for the Koopman generator

**Given:**

Positive-definite, $C^1$ kernel $k : \Omega \times \Omega \to \mathbb{R}$.

Integral operators $K : L^2(\mu) \to \mathcal{K}$, $G = K^* K$.

**Post-smoothing:**

$$B : L^2(\mu) \to L^2(\mu), \quad B = \overline{GV}.$$

- $GV \subset (GV)^{**} = B = -A^*$.
- $B$ is a Hilbert-Schmidt integral operator with

$$Bf = -\int_\Omega k'(\cdot, \omega) f(\omega) \, d\mu(\omega).$$

# Compactification schemes for the Koopman generator

**Given:**

Positive-definite, $C^1$ kernel $k : \Omega \times \Omega \to \mathbb{R}$.

Integral operators $K : L^2(\mu) \to \mathcal{K}$, $G = K^* K$.

**Skew-adjoint compactification on the RKHS:**

$$W : \mathcal{K} \to \mathcal{K}, \quad W = KVK^*.$$

- $W$ is a skew-adjoint, Hilbert-Schmidt operator on $\mathcal{K}$ satisfying

$$Wf = -\int_\Omega k'(\omega, \cdot) f(\omega) \, d\mu(\omega).$$

# Compactification schemes for the Koopman generator

**Given:**

Positive-definite, $C^1$ kernel $k : \Omega \times \Omega \to \mathbb{R}$.

Integral operators $K : L^2(\mu) \to \mathcal{K}$, $G = K^* K$.

**Skew-adjoint compactification on $L^2(\mu)$:**

$$\tilde{V} : L^2(\mu) \to L^2(\mu), \quad \tilde{V} = G^{1/2} V G^{1/2}.$$

- $K = \mathcal{U} G^{1/2}$ (polar decomposition).
- $\tilde{V}$ is a skew-adjoint, Hilbert-Schmidt operator on $L^2(\mu)$ related to $W$ by

$$\tilde{V} = \mathcal{U}^* W \mathcal{U}.$$

# Eigenvalues and eigenfunctions

**Proposition 4.41.**

*Let $k : \Omega \times \Omega \to \mathbb{R}$ be a $C^1$, $L^2$-universal, $\mu$-Markov ergodic kernel.*

1. *There exists an orthonormal basis $\tilde{z}_0, \tilde{z}_1, \dots,$ of $L^2(\mu)$ consisting of eigenfunctions of $\tilde{V}$,*

$$\tilde{V} z_j = i\alpha_j \tilde{z}_j, \quad \alpha_j \in \mathbb{R}.$$

2. *In the above, $i\alpha_0 = 0$ is a simple eigenvalue corresponding to the constant eigenfunction $\tilde{z}_0 = 1$.*
3. *$\tilde{V}$ has an associated purely atomic PVM $\tilde{E} : \mathfrak{B}(\mathbb{R}) \to B(L^2(\mu))$ such that*

$$\tilde{E}(S) = \sum_{j:\alpha_j \in S} \langle \tilde{z}_j, \cdot \rangle_{L^2(\mu)} \tilde{z}_j, \quad \tilde{V} = \int_{\mathbb{R}} i\alpha \, d\tilde{E}(\alpha).$$

# Strong resolvent convergence

**Definition 4.42.**

1. A one-parameter family of operators $A_\tau : D(A_\tau) \to H$, $\tau > 0$, on a Hilbert space $H$ is said to converge to a skew-adjoint operator $A : D(A) \to H$ in strong resolvent sense if for every $\rho \in \mathbb{C} \setminus \{i\mathbb{R}\}$ in the resolvent set of $A$ the resolvents $(A_\tau - \rho)^{-1}$ converge to $(A - \rho)^{-1}$ strongly.
2. The family $A_\tau$ is said to be p2-continuous if it is uniformly bounded and $\tau \mapsto \|p(A_\tau)\|$ is continuous for every degree-2 polynomial $p$.
3. If $A_\tau$ is skew-adjoint, $A_\tau$ is said to converge to $A$ in strong dynamical sense if for every $t \in \mathbb{R}$, $e^{tA_\tau}$ converges to $e^{tA}$ strongly.

# Strong resolvent convergence

**Theorem 4.43.**
*With the notation of Definition 4.42, suppose that $A_\tau$ is skew-adjoint. Then:*

1. *Strong resolvent convergence is equivalent to strong dynamical convergence.*
2. *A sufficient condition for strong resolvent convergence $A_\tau \to A$ is that $A_\tau$ converges to $A$ strongly in a core, i.e., a subspace $C \subseteq D(A)$ such that $\overline{A|_C} = A$.*
3. *The domain $D(A^2)$ is a core for $A$.*

# Strong resolvent convergence

**Theorem 4.44.**
*Let $A_\tau : D(A_\tau) \to H$ be a one-parameter family of skew-adjoint operators that converges to a skew-adjoint operator $A : D(A) \to H$ in strong resolvent sense. Let $E_\tau : \mathfrak{B}(R) \to B(H)$ and $E : \mathfrak{B}(R) \to B(H)$ be the PVMs associated with $A_\tau$ and $A$, respectively.*

1. *For every bounded, Borel-measurable set $\Omega \subset R$ such that $E(\partial\Omega) = 0$, $E_\tau(\Omega)$ converges strongly to $E(\Omega)$.*
2. *For every bounded, continuous function $Z : i\mathbb{R} \to \mathbb{C}$, $Z(A_\tau)$ converges strongly to $Z(A)$.*
3. *If the operators $A_\tau$ are compact, then for every element $i\alpha \in i\mathbb{R}$ of the spectrum of $A$ there exists a one-parameter family $i\alpha_\tau$ of eigenvalues of $A_\tau$ such that $\lim_{\tau\to 0} \alpha_\tau = \alpha$. Moreover, if $A_\tau$ is p2-continuous, the curve $\tau \mapsto \alpha_\tau$ is continuous.*

# Spectral convergence of the compactified generators

**Theorem 4.45.**
*Let* $\{G_\tau\}_{\tau \geq 0}$ *be a strongly continuous, ergodic semigroup of Markov operators on* $L^2(\mu)$ *such that for every* $\tau > 0$*,*

$$G_\tau f = \int_\Omega k_\tau(\cdot, \omega) f(\omega) \, d\mu(\omega),$$

*where* $k_\tau : \Omega \times \Omega \to \mathbb{R}$ *is a* $C^1$*,* $L^2$*-universal, positive-definite kernel. Then, Theorem 4.44 holds for the compactified generators*

$$\tilde{V}_\tau = G_\tau^{1/2} V G_\tau^{1/2}.$$

# Construction of the semigroup $G_\tau$

1. Start from an $L^2$-universal, $C^1$ kernel $\kappa : \Omega \times \Omega \to \mathbb{R}$.
2. Normalize $\kappa$ to an $L^2$-universal, $C^1$, bistochastic Markov kernel $p : \Omega \times \Omega \to \mathbb{R}$ (Coifman & Hirn '13). Let $P : L^2(\mu) \to L^2(\mu)$ be the associated integral operator.
3. Define the Laplace-like operator $\Delta = (I - P)^{-1}$.
4. Define $G_\tau = e^{-\tau \Delta}$.

# Dirichlet energy

$$P\phi_j = \lambda_j \phi_j, \quad \lambda_j > 0, \quad \langle \phi_i, \phi_j \rangle_{L^2(\mu)} = \delta_{ij}$$

$$G_\tau \phi_j = \lambda_{j,\tau} \phi_j, \quad \lambda_{j,\tau} = e^{-\tau \eta_j}, \quad \eta_j = 1 - \frac{1}{\lambda_j}.$$

- $\mathcal{H}$: RKHS associated with $p$.
- $f \in L^2(\mu)$ has a representative in $\mathcal{H}$ iff

$$\tilde{\mathcal{D}}(f) := \sum_{j=0}^{\infty} \frac{|\langle \phi_j, f \rangle_{L^2(\mu)}|^2}{\lambda_j} < \infty.$$

- For every such (nonzero) $f$, we define the Dirichlet energy

$$\mathcal{D}(f) = \frac{\tilde{\mathcal{D}}(f)}{\|f\|^2_{L^2(\mu)}} - 1.$$

# Coherent observables

$$W_\tau = K_\tau V K_\tau^*$$

$$W_\tau \zeta_{j,\tau} = i\omega_{j,\tau}\zeta_{j,\tau}, \quad z_{j,\tau} = \frac{K_\tau^* \zeta_{j,\tau}}{\|K_\tau^* \zeta_{j,\tau}\|_{L^2(\mu)}}.$$

**Proposition 4.46.**

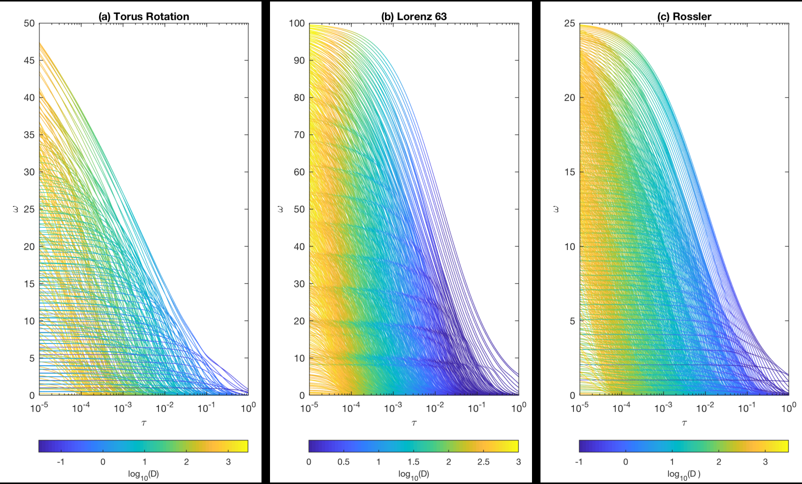
*There exists a continuous function $R(\epsilon, \tau)$ that diverges as $\tau \to 0$ for every $\epsilon > 0$ such that*

$$\|U^t z_{j,\tau} - e^{i\omega_{j,\tau}} z_{j,\tau}\|_{L^2(\mu)} < \epsilon, \quad |t| \leq T(\epsilon, \tau) := \frac{R(\epsilon, \tau)}{\sqrt{\mathcal{D}(z_{j,\tau}) + 1}}.$$

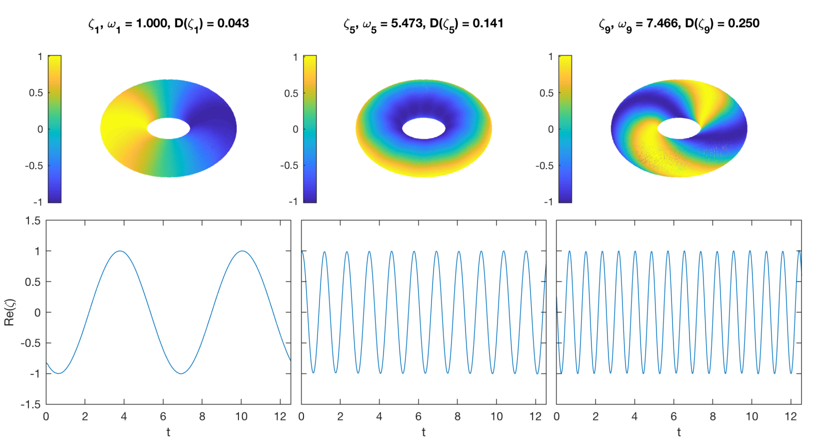
*Moreover:*

1. *If* $\lim_{\tau\to 0} \omega_{j,\tau} =: \omega_j$ *exists and* $T(\epsilon, \tau)$ *diverges as* $\tau \to 0$ *for every* $\epsilon > 0$*, then* $i\omega$ *is an element of the spectrum of* $\tilde{V}$*.*
2. *If* $\lim_{\tau\to\omega}$ *exists and* $\mathcal{D}(z_{j,\tau})$ *is bounded as* $\tau \to 0$*, then* $i\omega$ *is an eigenvalue of* $V$*. Moreover,* $z_{j,\tau}$ *converges to the eigenspace of* $V$ *corresponding to* $i\omega$*.*

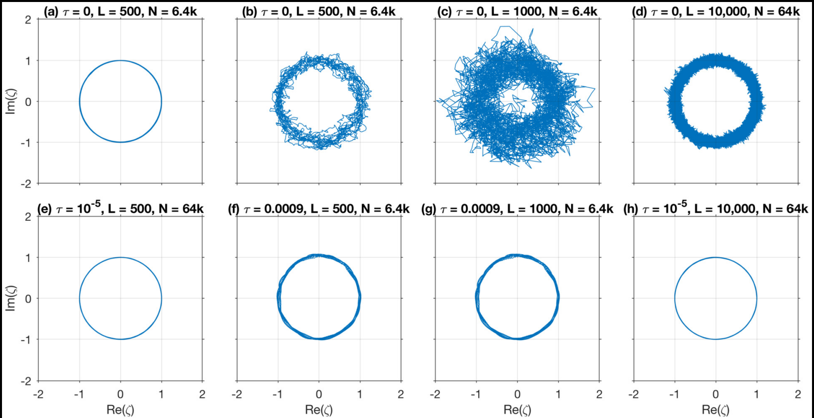
# Numerical examples



# Torus rotation—eigenfunctions of $W_\tau$

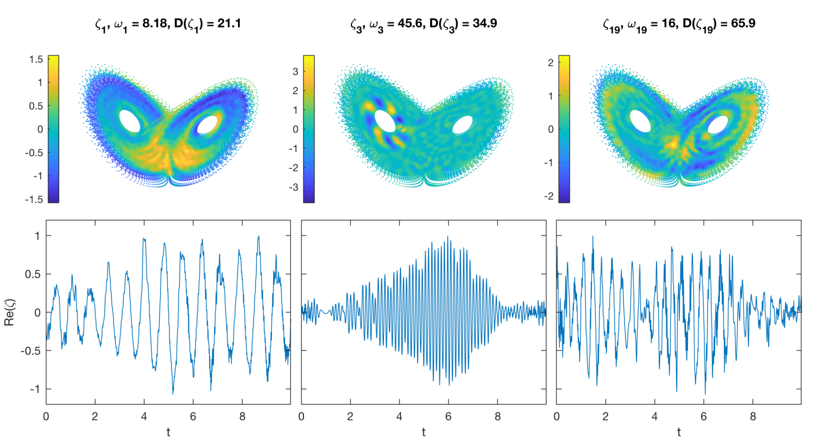


# Torus rotation

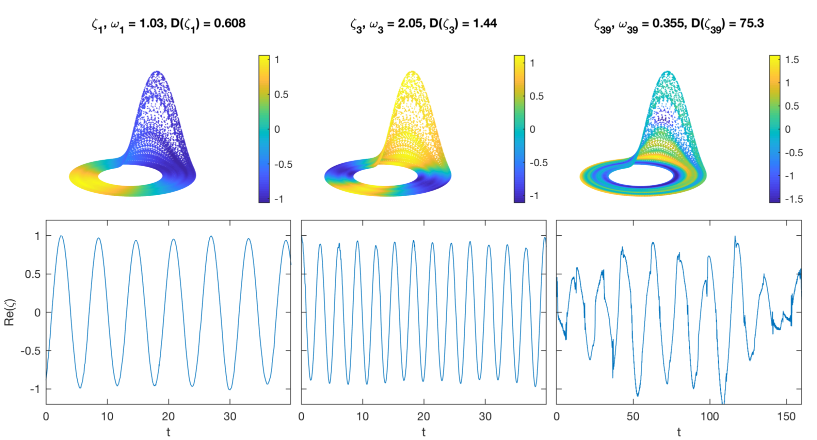



Due to the density of the spectrum in the imaginary line, regularization is important, even for a system with pure point spectrum.

# L63 system—eigenfunctions of $W_\tau$



# Rössler system—eigenfunctions of $W_\tau$



# Further reading


[1] F. Chatelin, *Spectral Approximation of Linear Operators*, ser. Classics in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics, 2011.

[2] S. Das and D. Giannakis, "Delay-coordinate maps and the spectra of Koopman operators," *J. Stat. Phys.*, vol. 175, no. 6, pp. 1107–1145, 2019. DOI: 10.1007/s10955-019-02272-w.

[3] S. Das, D. Giannakis, and J. Slawinska, "Reproducing kernel Hilbert space quantification of unitary evolution groups," *Appl. Comput. Harmon. Anal.*, vol. 54, pp. 75–136, 2021. DOI: 10.1016/j.acha.2021.02.004.

[4] D. Giannakis, "Delay-coordinate maps, coherence, and approximate spectra of evolution operators," *Res. Math. Sci.*, vol. 8, p. 8, 2021. DOI: 10.1007/s40687-020-00239-y.

[5] C. R. de Oliveira, *Intermediate Spectral Theory and Quantum Dynamics*, ser. Progress in Mathematical Physics. Basel: Birkhäuser, 2009, vol. 54.

# Further reading

[6] M. Rédei and C. Werndl, “On the history of the isomorphism problem of dynamical systems with special regard to von Neumann’s contribution,” *Arch. Hist. Exact Sci.*, vol. 66, pp. 71–93, 2012. DOI: 10.1007/s00407-011-0089-y.