CHAPTER III TRANSFER OPERATORS

III.1 STUDYING DYNAMICAL SYSTEMS WITH DENSITIES

3.1 IN PRACTICE

Q: How do we get μ, if only $T : X \to X$ is known?

Trajectory simulation:

Empirical measure: $\mu_N := \frac{1}{N} \sum_{k=0}^{N-1} \delta_{T^k x} \longrightarrow \mu$

weak convergence towards an ergodic measure (cf. ergodic decomposition & BET)

Q: How fast is this convergence?

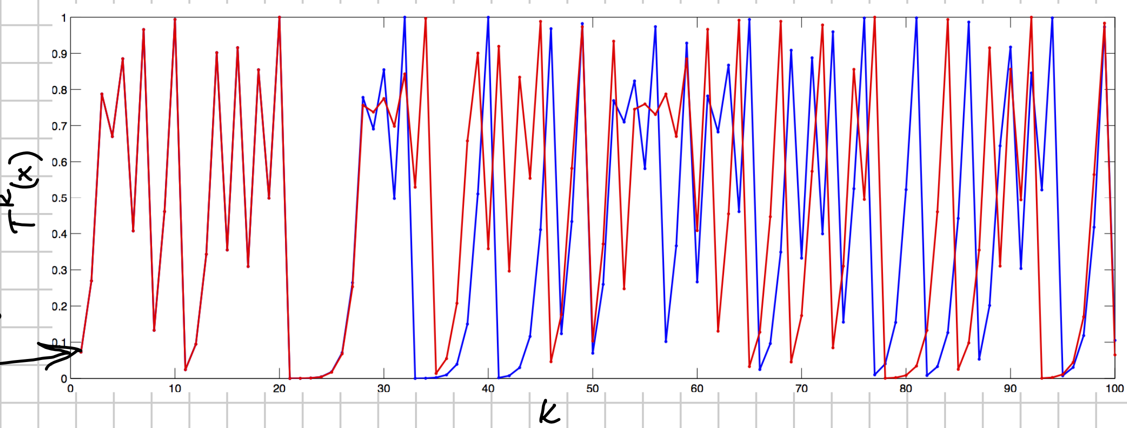
↳ What is the dynamical behavior on subdominant time-scales?

3.2 LONG TRAJECTORIES VS DENSITIES

- Round-off errors accumulate ⇝ faith in the results? (cf. angle doubling exercise)
- Large condition number ⇒ close initial conditions quickly diverge ("unpredictability", "chaos")

$T : [0,1] \to [0,1]$

$x \mapsto 4x(1-x)$



Initial conditions 10^{-10} apart

- Gibbs' insight: For chaotic systems it's easier to predict the evolution of densities of a large collection of initial conditions, than individual trajectories.

Recall from exercises: Sensitivity of trajectories not reflected if histograms of different trajectories are plotted.

$\rightsquigarrow$ promising approach: consider dynamics on ensemble of states

$\shortparallel$ many states described by a distribution

Random variable : $Y \longmapsto T(Y)$

$\shortparallel$ one state with prob. distribution

$?$ $?$

$$L^1 \ni f_0 \longmapsto f_1 \in L^1$$

transfer operator (linear, bounded, see below)

Comparison:

1) Non-linear, (possibly) low-dimensional dynamics vs linear ∞-dim. operator

2) One long trajectory vs many (∞) short trajectories

In applications, trajectories and operators are often used hand-in-hand:

(i) Molecular dynamics: $T: X \to X$, where $\dim(X) \gg 1$, hence it is unthinkable to approximate the full operator

$\rightsquigarrow$ Long simulations are used to set up a discretized operator

(ii) Data-driven approaches: Sometimes the dynamics isn't known fully, just through some (few) realizations of the system.

$\rightsquigarrow$ Set-up under uncertainty / smoothness assumptions on densities

3.3 FROBENIUS-PERRON OPERATOR

Def A: $(X, \mathcal{B}, \mu)$ prob. space, $T: X \to X$ measurable map. T is called non-singular if for any $A \in \mathcal{B}$ $\mu(A) = 0$ implies that $\mu(T^{-1}A) = 0$ too; or equivalently: $\mu \circ T^{-1} \ll \mu$

In words: one cannot create mass out of nothing by taking pre-images; or: positive measure cannot disappear under forward iteration.

Example:

1) $T:[0,1]\to[0,1]$, $T(x)=1/2$ is singular wrt Lebesgue
is non-singular wrt $\delta_{1/2}$

2) $T:[0,1]\to[0,1]$, $T(x)=4x(1-x)$ is non-singular wrt Lebesgue

How does the dynamics propagate (prob.) densities?

A distribution of infinitely many initial conditions $\Leftrightarrow$ one initial random variable with the same density. We work with the second.

$Y\sim f_0\in L^1(X,\mu)$, i.e. $\mathbb{P}[Y\in A]=\int_A f_0\,d\mu \quad \forall A\in\mathcal{B}$

$T(Y)\sim f_1=?$

$$\mathbb{P}[T(Y)\in A]=\mathbb{P}[Y\in T^{-1}(A)]=\int_{T^{-1}(A)} f_0\,d\mu \overset{!}{=} \int_A f_1\,d\mu$$

Observe: $\nu:\mathcal{B}\to[0,1]$ with $\nu(A)=\int_{T^{-1}(A)} f_0\,d\mu$ is a prob. measure,

and if T is non-singular, then $\nu\ll\mu$

$\xrightarrow[\text{Nikodym}]{\text{Radon-}} \exists_1\ 0\le f_1\in L^1(X,\mu)$ s.t. $\nu(A)=\int_A f_1\,d\mu$

Def B: Let T be a non-singular transformation on a prob. space $(X,\mathcal{B},\mu)$. The Frobenius-Perron operator (FPO) $P: L^1(X,\mu)\to L^1(X,\mu)$ is the unique operator satisfying

$$\int_A Pf\,d\mu=\int_{T^{-1}(A)} f\,d\mu \qquad \forall A\in\mathcal{B},\ f\in L^1.$$

(Note: uniqueness follows from the Radon-Nikodym thm, applied to the positive & negative parts, f^+ and f^-, respectively)

Example: Circle doubling: $T: S^1 \to S^1$, $T(x) = 2x \mod 1$

$I \subseteq S^1$ interval $\Rightarrow$ $T^{-1}(I) = \frac{1}{2}I \,\dot\cup\, \{\frac{1}{2}I + \frac{1}{2}\}$

$$\int_{T^{-1}I} f(x)dx = \int_{\frac{1}{2}I} f(x)dx + \int_{\{\frac{1}{2}I+\frac{1}{2}\}} f(x)dx$$

subst: $x := y/2$ $\qquad$ $x := y/2 + 1/2$

$$= \frac{1}{2}\int_I f(y/2)\,dy + \frac{1}{2}\int_I f(y/2 + 1/2)\,dy$$

$$= \int_I \underbrace{\frac{1}{2}\left(f(y/2) + f(y/2+1/2)\right)}_{} dy \overset{!}{=} \int_I Pf(y)\,dy$$

$$\Rightarrow \qquad = Pf$$

3.4 PROPERTIES OF THE FPO

PropA: For a non-singular transformation $T: X \to X$, the FPO satisfies

(a) $P(\alpha f + g) = \alpha Pf + Pg \quad \forall f, g \in L^1, \alpha \in \mathbb{C}$ (linearity)

(b) $f \geq 0 \Rightarrow Pf \geq 0$ (positivity)

(c) $\int Pf\,d\mu = \int f\,d\mu$ (integral preserving)

(d) $\|Pf\|_{L^1} \leq \|f\|_{L^1}$ (contraction)

(e) If $S: X \to X$ is another non-sing. map, then

$$P_{T\circ S} = P_T \circ P_S$$

Remark: An operator satisfying (a), (b), (d) is called a Markov operator

Proof:

(a) For $A \in \mathcal{B}$:

$$\int_A P(\alpha f + g)\,d\mu = \int_{T^{-1}A} \alpha f + g\,d\mu = \alpha\int_{T^{-1}A} f\,d\mu + \int_{T^{-1}A} g\,d\mu =$$

$$= \alpha\int_A Pf\,d\mu + \int_A Pg\,d\mu \quad \Rightarrow \text{claim}$$

(b) For $A \in \mathcal{B}$, if $f \geq 0$:

$$\int_A Pf\,d\mu = \int_{T^{-1}A} f\,d\mu \geq 0.$$ Since A arbitrary $\Rightarrow Pf \geq 0$ a.e.

(c) With $A = X$:

$$\int_X Pf \, d\mu = \int_{\underbrace{T^{-1}X}_{=X}} f \, d\mu = \int_X f \, d\mu$$

(d) Let $f = f^+ - f^-$, both $f^+, f^- \geq 0$. Then

$$\|Pf\|_{L^1} = \int |Pf| \, d\mu = \int |Pf^+ - Pf^-| \, d\mu \leq \int |Pf^+| + |Pf^-| \, d\mu$$

$$\overset{(b)}{=} \int Pf^+ + Pf^- \, d\mu \overset{(c)}{=} \int f^+ + f^- \, d\mu = \int |f| \, d\mu = \|f\|_{L^1}$$

(e) $T \circ S$ non-singular: $(T \circ S)^{-1}(A) = S^{-1} \circ T^{-1}(A)$

$$\mu(A) = 0 \underset{T \text{ n-s.}}{\Rightarrow} \mu(T^{-1}A) = 0 \underset{S \text{ n-s.}}{\Rightarrow} \mu(S^{-1} \circ T^{-1}A) = 0$$

$\Rightarrow P_{T \circ S}$ well-defined

For $A \in \mathcal{B}$, $f \in L^1$

$$\int_A P_{T \circ S} f \, d\mu = \int_{S^{-1} \circ T^{-1}(A)} f \, d\mu = \int_{T^{-1}A} P_S f \, d\mu = \int_A P_T(P_S f) \, d\mu \quad \blacksquare$$

Prop B: If $(X, \mathcal{B}, \mu, T)$ is a ppt, then

$$Pf \circ T = \mathbb{E}(f \mid T^{-1}\mathcal{B}) \quad \text{a.e.}$$

Proof: $Pf \circ T$ is clearly $T^{-1}\mathcal{B}$-measurable. We have for $A = T^{-1}B \in T^{-1}\mathcal{B}$

$$\int_A Pf \circ T \, d\mu = \int_{T^{-1}B} Pf \circ T \, d\mu \underset{\mu \text{ inv.}}{=} \int_B Pf \, d\mu = \int_{T^{-1}B} f \, d\mu$$

$$= \int_A f \, d\mu \quad \blacksquare$$

Corollary A: If $X \subseteq \mathbb{R}^d$ open and $T: X \to X$ is a Lebesgue-preserving homeomorphism (T invertible and both T and T^{-1} are continuous), then $Pf = f \circ T^{-1}$ a.e.

Proof: T, T^{-1} continuous $\Rightarrow T^{-1}\mathcal{B} = \mathcal{B}$ (Borel) $\Rightarrow \mathbb{E}(f \mid T^{-1}\mathcal{B}) = f$, and Prop B yields the claim $\blacksquare$

<u>Corollary B</u>: If $(X, \mathcal{B}, \mu, T)$ is a ppt, then P is a contraction on $L^p(X,\mu)$ for every $1 \le p \le \infty$.

<u>Proof</u>: For $1 \le p < \infty$:

$$\|Pf\|_{L^p}^p = \int |Pf|^p \, d\mu \overset{\mu \text{ inv.}}{=} \int |Pf \circ T|^p \, d\mu$$

$$\overset{\text{Prop B}}{=} \int \left| \mathbb{E}(f \mid T^{-1}\mathcal{B}) \right|^p d\mu \le \int \mathbb{E}(|f|^p \mid T^{-1}\mathcal{B}) \, d\mu$$

$x \mapsto |x|^p$ convex on $\mathbb{R}_+$ & Exercise 1 (Sheet 7)

$$= \int |f|^p \, d\mu = \|f\|_{L^p}^p$$

For $p = \infty$

$$\|Pf\|_{L^\infty} = \|Pf \circ T\|_{L^\infty} = \|\mathbb{E}(f \mid T^{-1}\mathcal{B})\|_{L^\infty} \le \|f\|_{L^\infty}$$

μ inv $\Rightarrow \mu(T(X)) = \mu(X) = 1$ Ex 1 (Sheet 7)

3.5 ERGODICITY AND MIXING

<u>Def</u>: The operator $\mathcal{U}: L^\infty(X,\mu) \to L^\infty(X,\mu)$, $f \mapsto f \circ T$ is called the Koopman operator associated with T.

<u>Prop A</u>: The Koopman and Frobenius-Perron operators are adjoint, i.e. $\forall f \in L^1, g \in L^\infty$

$$\int Pf \cdot g \, d\mu = \int f \cdot \mathcal{U}g \, d\mu,$$

we also write $\langle Pf, g \rangle = \langle f, \mathcal{U}g \rangle$

<u>Proof</u>: For $f \in L^1$, and A m'ble:

$$\int f \, \mathcal{U}\chi_A \, d\mu = \int f \cdot \chi_A \circ T \, d\mu = \int f \cdot \chi_{T^{-1}A} \, d\mu = \int_{T^{-1}A} f \, d\mu$$

$$= \int_A Pf \, d\mu = \int Pf \cdot \chi_A \, d\mu$$

Since characteristic functions span L^∞, the claim follows ∎

Let $\mathcal{D} = \mathcal{D}(X,\mathcal{B},\mu) = \{ f \in L^1(X,\mathcal{B},\mu) \mid f \geq 0, \|f\|_{L^1} = 1 \}$ denote the set of densities.

Prop B: Let $(X,\mathcal{B},\mu,T)$ be a non-sing. ppt. Then

(a) T is ergodic if and only if

$$\frac{1}{n}\sum_{k=0}^{n-1} \langle P^k f, g\rangle \xrightarrow[n\to\infty]{} \langle \mathbb{1}, g\rangle = \int g \, d\mu \qquad \forall f\in\mathcal{D}, g\in L^\infty$$

(b) T is (strongly) mixing if and only if

$$\langle P^n f, g\rangle \xrightarrow[n\to\infty]{} \langle \mathbb{1}, g\rangle \qquad \forall f\in\mathcal{D}, g\in L^\infty$$

Proof: Recall characterizations 2.2 and 1.13 of ergodicity and mixing, respectively:

$$\text{ergodicity} \iff \frac{1}{n}\sum_{k}^{n-1} \mu(A\cap T^{-k}B) \xrightarrow[n\to\infty]{} \mu(A)\mu(B) \qquad (1)$$

$$\text{mixing} \iff \mu(A\cap T^{-n}B) \xrightarrow[n\to\infty]{} \mu(A)\mu(B) \qquad (2)$$

These are equivalent with

$$\frac{1}{n}\sum_{k}^{n-1} \langle f, U^k g\rangle \xrightarrow[n\to\infty]{} \int f\,d\mu \cdot \int g\,d\mu \qquad (1,1)$$

$$\langle f, U^n g\rangle \xrightarrow[n\to\infty]{} \int f\,d\mu \cdot \int g\,d\mu \qquad (2,1)$$

if one takes $f = \chi_A$, $g = \chi_B$, and hence with (1,1) resp. (2,1) holding $\forall f\in L^1, g\in L^\infty$ (characteristic functions span L^1, L^∞).

Use the adjoint property (Prop A) and restrict to $f\in\mathcal{D}$ to obtain the claim. $\blacksquare$

3.6 ABSOLUTELY CONTINUOUS INVARIANT MEASURES AND THE FPO

Recall : $\nu \ll \mu \;:\Longleftrightarrow\; \mu(A)=0 \Rightarrow \nu(A)=0$

Prop: Let $(X,\mathcal{B},\mu)$ be a prob. space and $T: X\to X$ non-singular. Let $f^* \in D(X,\mathcal{B},\mu)$, and $\nu = f^*\cdot\mu$ (meaning $\frac{d\nu}{d\mu}=f^*$). Then

$$Pf^* = f^* \quad \Longleftrightarrow \quad \nu \text{ is } T\text{-invariant, i.e. } \nu=\nu\circ T^{-1}$$

Proof: For any $A\in\mathcal{B}$:

$$\nu(A)=\nu(T^{-1}A) \Longleftrightarrow \int_A f^*\,d\mu = \int_{T^{-1}A} f^*\,d\mu = \int_A Pf^*\,d\mu \qquad \blacksquare$$

Why is this interesting?

Recall $T:[0,1]\to[0,1]$, $T(x)=\sqrt{x}$, for which δ_0 is an ergodic measure, but not one that we "observe", since $T^n x \longrightarrow 1$ for all x, except $x=0$.

What is wrong with that?

In our perception, it is "natural" that significant sets have nonzero Lebesgue measure, and $\{0\}$ doesn't.

Or: δ_0 is not absolutely continuous wrt Lebesgue

If μ represents a measure which we consider to be "natural", and wrt this we find $Pf^*=f^*$, then BET gives (suppose ν ergodic)

$$\frac{1}{n}\sum_{k}^{n-1} h\circ T^k(x) \xrightarrow[n\to\infty]{} \int h\,d\nu = \int h f^*\,d\mu$$

for ν-a.e. $x\in X$, which means for a set of x with positive μ-meas.

$\Rightarrow$ we can observe ν !

Existence of f^*: No general results. For specific case: [GB, Chap. 7]

III.2 NUMERICAL APPROXIMATION OF TRANSFER OPERATORS

3.7 ABSTRACT SETTING

Q: How to approximate fixed points of P (or eigenfunctions, in general)?

$P: L^1 \to L^1$: linear operator

$V_m \subseteq L^1$: subspace with $\dim V_m = m \in \mathbb{N}$

$\pi_m: L^1 \to V_m$: linear projection, i.e. $\pi_m \circ \pi_m = \pi_m$

Solve eigenproblem for $P_m: V_m \to V_m$, $P_m = \pi_m \circ P$, instead!

Q: How to choose V_m / its basis ?

Examples:
- "hat functions" $\rightsquigarrow$ finite element methods (local support $\Rightarrow$ sparse stiffness mat.)
- polynomials of different order $\rightsquigarrow$ spectral methods (fast convergence for "smooth problems", global support of basis fcns $\Rightarrow$ full matrix)

3.8 ULAM'S METHOD (ULAM 1960)

Partition of X : $\mathcal{P}_m = \{B_1, \dots, B_m\}$, where
- $B_i \in \mathcal{B}$
- $m(B_i \cap B_j) = 0$
- $\dot{\bigcup}_{i=1}^{m} B_i = X$

usually rectangles, i.e. "boxes"

m = Lebesgue

Let $\chi_i := \chi_{B_i}$, and define projection

$$\pi_m f = \sum_{i=1}^{m} c_i \frac{\chi_i}{m(B_i)}, \quad c_i = \int_{B_i} f \, dm \quad \Rightarrow V_m = \operatorname{span}\{\chi_1, \dots, \chi_m\}$$

Discretized operator: $P_m := \pi_m P$

Matrix representation w.r.t. basis $\{\chi_1/m(B_1), \dots, \chi_m/m(B_m)\}$ also denoted by P_m:

For $c \in \mathbb{R}^m$ write $c \simeq f \in V_m$ iff $f = \sum_i c_i \frac{\chi_i}{m(B_i)}$

Find $P_n \in \mathbb{R}^{n\times n}$ s.t. $\pi_n P f = c^T P_n$

$$\iff \pi_n P\left(\frac{\chi_i}{m(B_i)}\right) = \sum_{j=1}^{n} \underline{P_{n,ij}}\, \frac{\chi_j}{m(B_j)} \quad \forall i = 1,\dots,n$$

$$\| $$

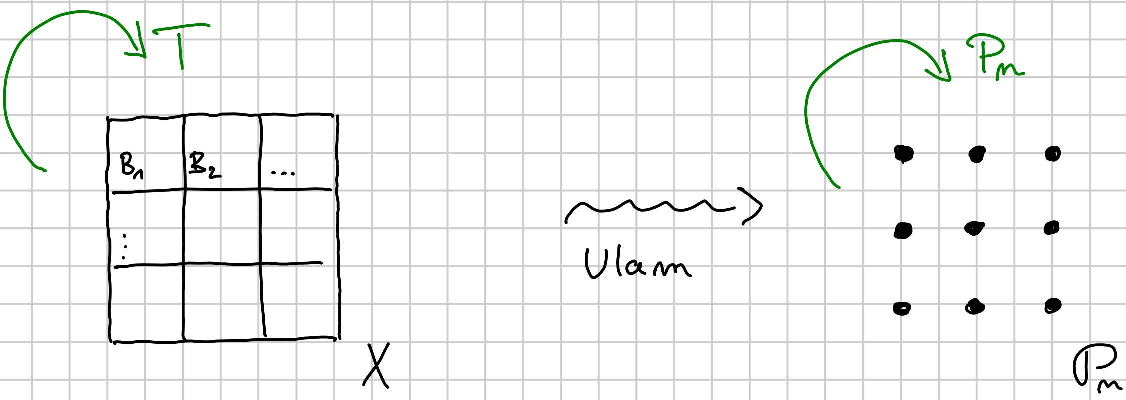
$$\sum_{j=1}^{n} \left(\underline{\int_{B_j} P \frac{\chi_i}{m(B_i)}\, dm} \right) \frac{\chi_j}{m(B_j)}$$

$$P_{n,ij} = \int_{B_j} P\left(\frac{\chi_i}{m(B_i)}\right) dm = \frac{1}{m(B_i)} \int_{T^{-1}(B_j)} \chi_i \, dm = \boxed{\frac{m(B_i \cap T^{-1} B_j)}{m(B_i)}}$$

3.9 PROBABILISTIC INTERPRETATION

$$P_{n,ij} = \frac{m(B_i \cap T^{-1} B_j)}{m(B_i)} \implies P_{n,ij} = \mathbb{P}\left[T(x) \in B_j \,\middle|\, x \sim \mathrm{unif}(B_i) \right]$$

"probability that a uniformly sampled state from B_i will jump into B_j"



Markov chain on the partition elements

P_n stochastic matrix $\overset{1.18}{\Longrightarrow}$ $\exists c^*$: stationary probability vector

$\Longrightarrow f = \sum_i c_i^* \frac{\chi_i}{m(B_i)}$ stationary density of P_n

Prop: P_n always has an invariant density. Even if P doesn't.

3.10 COMPUTATIONAL ASPECTS

Recall $P_{m,ij} = \frac{1}{m(B_i)} \int \underbrace{\chi_{B_i} \cdot \chi_{T^{-1}B_j}} dm$

discontinuous functions, high order quadrature doesn't help

$\curvearrowright$ Monte-Carlo sampling:

Sample uniformly from B_i : $x_1, x_2, \dots, x_N$

Approximate $P_{m,ij} \approx \hat{P}_{m,ij} = \frac{1}{N} \sum_{k=1}^{N} \chi_j(Tx_k)$

Properties:

- $\hat{P}_m \to P_m$ as $N^{-1/2}$
- curse of dimension (if $\dim X \gg 1$, it is very expensive (exponentially in $\dim X$) to resolve every coordinate dimension)
- Ulam's method (P_m) preserves properties of the Markov operator P: linearity, positivity, integral-preserving (cf. 3.4)
- $\hat{P}_m$ preserves the same properties
- Up to numerical errors: If $P_{m,ij} = 0$, then $\hat{P}_{m,ij} = 0$
- $B_i \cap T^{-1}B_j \neq \emptyset \iff TB_i \cap B_j \neq \emptyset$
  If T Lipschitz continuous, and boxes are "local" (i.e. $\operatorname{diam} B_i$ small), then P_m and $\hat{P}_m$ are sparse
- Computation of $\hat{P}_m$ parallelizable

Convergence: [Li / 1976: Finite Approximation for the Frobenius-Perron operator. A solution to Ulam's conjecture]

For a restricted class of systems: $Pf^* = f^*$, and $P_m f_m = f_m$. Then

$$f_m \xrightarrow[L^1]{} f^* \quad (m \to \infty)$$

3.11 REMARKS

1) If there is a subsequence $\{f_{m_i}\}$ and a prob. measure μ s.t. $\forall g \in C_b^0(X)$

$$\int f_{m_i}\, g \xrightarrow[i\to\infty]{} \int g\, d\mu \text{ , then } \mu \text{ is an invariant measure}$$

[Froyland 1996]

Weaker, but less restrictive than Li's result.

2) P_m^T is a discretization of the Koopman operator:

$$(P_m^T)_{ij} = P_{m,ji} = \frac{m(B_j \cap T^{-1}B_i)}{m(B_j)} = \frac{1}{m(B_j)} \int_{B_j} \chi_{T^{-1}B_i}\, dm$$

$$= \frac{1}{m(B_j)} \int_{B_j} \chi_i \circ T\, dm = \frac{1}{m(B_j)} \int_{B_j} \mathcal{U}\chi_i\, dm = \mathbb{E}\left[\mathcal{U}\chi_i(x) \,\middle|\, x \sim \mathrm{unif}(B_j)\right],$$

or, equivalently, if $\mathbb{R}^n \ni c = f \in L^\infty$, then $P_m c = \mathbb{E}\left[\mathcal{U}f \,\middle|\, \sigma(\mathcal{P}_m)\right]$

3) In the last 20 years Ulam's method (and other discretizations) have been used to approximate further eigenfunctions of the FPO (associated with eigenvals $|\lambda| < 1$)

$\rightsquigarrow$ almost-invariant- / persistent- / metastable sets

Dynamical meaning of $Pf = \lambda f$:

- $\lambda = 1$: dominant behavior for $t \to \infty$ (cf. BET)
- $|\lambda| < 1$: dominant behavior on time-scales $t < \infty$

Applications:
- atmospheric science
- molecular dynamics
- oceanography

3.12 GENERALIZED GALERKIN METHODS

a)
- $\mathcal{H}$: Hilbert space with inner product $\langle\cdot,\cdot\rangle$
- $L: \mathcal{H} \circlearrowleft$ bounded linear operator
- $\mathrm{span}(\varphi_1, \dots, \varphi_k) =: \mathcal{V}$

Galerkin projection: $L_k : \mathcal{V} \circlearrowleft$ is the unique lin. operator with

$$\langle \varphi_j, L\varphi_i \rangle = \langle \varphi_j, L_k \varphi_i \rangle \qquad \forall i,j = 1, \dots, k$$

If $\Pi_k : \mathcal{H} \to \mathcal{V}$ denotes the $\langle\cdot,\cdot\rangle$-orthogonal proj., then $L_k = \Pi_k L$.

Exercise: Ulam's method is a Galerkin proj. on L^2_μ, if μ invariant.

b) If L is not living on a Hilbert space, it is advantageous to differentiate between basis and test functions.

- $L : \mathcal{Y} \circlearrowleft$
- $\mathrm{span}\{\varphi_i\} = \mathcal{V} \subseteq \mathcal{Y}$
- $\mathrm{span}\{\varphi_i^*\} = \mathcal{W} \subseteq \mathcal{Y}^*$ (dual space)

Petrov-Galerkin projection: $L_k : \mathcal{V} \circlearrowleft$ is the unique lin. op. with

$$\langle \varphi_j^*, L\varphi_i \rangle = \langle \varphi_j^*, L_k \varphi_i \rangle \qquad \forall i,j = 1, \dots, k$$

where $\langle \varphi_j^*, f \rangle = \varphi_j^*(f)$ "duality pairing"

c) What if we choose $\ell = \dim \mathcal{W} > \dim \mathcal{V} = k$?

Over-determined Petrov-Galerkin, set up, e.g., as least squares problem:

$$\sum_{j=1}^{\ell} \sum_{i=1}^{k} \langle \varphi_j^*, L\varphi_i - L_k \varphi_i \rangle^2 = \min!$$

3.13 EXTENDED DYNAMIC MODE DECOMPOSITION (EDMD)

[Williams, Kevrekidis, Rowley 2015]

We are only given a fixed set of dynamical data:

$X = [x_1 \cdots x_m]$, $Y = [y_1 \cdots y_m]$ with $y_i = T(x_i)$, called snapshots.

Define

- $\Psi = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_k \end{bmatrix} : \underbrace{\mathcal{X}}_{\text{state space}} \to \mathbb{R}^k$
- $\Psi_X = [\Psi(x_1) \cdots \Psi(x_m)] \in \mathbb{R}^{k \times m}$
- $\Psi_Y = [\Psi(y_1) \cdots \Psi(y_m)] = [\mathcal{U}\Psi(x_1) \cdots \mathcal{U}\Psi(x_m)]$

$\mathcal{U}$: Koopman operator

Point-evaluations $\rightsquigarrow$ set up a Petrov-Galerkin method of collocation type, i.e. $\psi_j^* = \delta_{x_j}$ (Dirac delta)

By overloading notation, let $U_k \in \mathbb{R}^{k \times k}$, such that for $f = \sum_{i=1}^{k} c_i \psi_i \in V$ we identify $U_k c \in \mathbb{R}^k$ with $U_k f \in V$.

Over-determined Petrov-Galerkin gives

$$\sum_{\ell=1}^{m} \sum_{i=1}^{k} \langle \delta_{x_\ell}, \mathcal{U}\psi_i - U_k \psi_i \rangle^2 = \min!$$

$$U_k \psi_i = U_k \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \underbrace{U_{k,:i}}_{i^{th}\text{ column}}$$

(i^{th} entry)

$$\Leftrightarrow \sum_{i=1}^{k} \sum_{\ell=1}^{m} \Big(\psi_i(y_\ell) - \sum_{j=1}^{k} U_{k,ji} \psi_j(x_\ell) \Big)^2 = \min! \quad (*)$$

$$\Leftrightarrow \sum_{i=1}^{k} \| \Psi_{Y,i:} - (U_k^T)_{i:} \Psi_X \|_2^2 = \min!$$

$$\Leftrightarrow \sum_{i=1}^{k} \underbrace{\| \Psi_{Y,i:}^T - \Psi_X^T U_{k,:i} \|_2^2}_{\text{LSP for } U_{k,:i}} = \min!$$

decoupled subproblems for $i = 1, \dots, k$

Since $\ell > k$, solution given by normal equation:

$$\mathcal{U}_{k,:i} = \left(\Psi_X \Psi_X^T\right)^{-1} \Psi_X \left(\Psi_{Y,i:}\right)^T$$

$\rightsquigarrow$ all columns:

$$\mathcal{U}_k = \left(\Psi_X \Psi_X^T\right)^{-1} \Psi_X \Psi_Y^T \iff \boxed{\mathcal{U}_k^T = \Psi_Y \Psi_X^+}$$

NB: $\circledast \iff \|\Psi_Y - \mathcal{U}_k^T \Psi_X\|_F^2 = \min!$ Moore-Penrose pseudoinverse, $\Psi_X \Psi_X^+ = I$

<u>Remark</u>: $\mathcal{X} = \mathbb{R}^d$, $\psi_i(x) = x^i$ (i^{th} coordinate, i.e. $k = d$) gives the so-called Dynamic Mode Decomposition (DMD). Schmid 2008

3.14 CONVERGENCE TO A GALERKIN METHOD

$$\mathcal{U}_k = \left(\Psi_X \Psi_X^T\right)^{-1} \Psi_X \Psi_Y^T = \underbrace{\left(\frac{1}{m} \Psi_X \Psi_X^T\right)^{-1}}_{=: G_k} \underbrace{\left(\frac{1}{m} \Psi_X \Psi_Y^T\right)}_{=: A_k}, \quad \text{i.e. } G_k \mathcal{U}_k = A_k$$

Let:

(a) $x_1, x_2, \dots$ be the trajectory of a μ-ergodic process (μ prob. measure)

(b) $\psi_i \in L^\infty_\mu$, thus $\psi_i, \mathcal{U}\psi_i \in L^p_\mu$ for any $1 \le p \le \infty$

$$\Longrightarrow A_{k,ij} = \frac{1}{m} \sum_{\ell=1}^m \psi_i(x_\ell)\psi_j(y_\ell) = \frac{1}{m} \sum_{\ell=1}^m \psi_i(x_\ell)\,\mathcal{U}\psi_j(x_\ell)$$

$\psi_i, \mathcal{U}\psi_j \in L^2_\mu$, i.e. BET is applicable

$$\xrightarrow[m\to\infty]{\text{BET}} \int \psi_i(x)\,\mathcal{U}\psi_j(x)\,d\mu(x) = \langle \psi_i^*, \mathcal{U}\psi_j \rangle_\mu$$

$$G_{k,ij} = \frac{1}{m} \sum_{\ell=1}^m \psi_i(x_\ell)\psi_j(x_\ell) \xrightarrow[m\to\infty]{\text{BET}} \langle \psi_i^*, \psi_j \rangle_\mu$$

with $\psi_i^*(\cdot) = \int \psi_i \cdot d\mu$ $\Longrightarrow$

<u>Thm</u>:

$\mathcal{U}_k = G_k^{-1} A_k$ converges as $m \to \infty$ almost surely to the (Petrov-) Galerkin projection of $\mathcal{U}: L^\infty_\mu \circlearrowleft$ wrt $V = \operatorname{span}\{\psi_i\}$, $W = \operatorname{span}\{\int \psi_i \cdot d\mu\}$.

<u>Remarks</u>:

- Often $y_k = x_{k+1} = T(x_k)$ (especially widely used in MD), and then μ is automatically the ergodic measure of T.

- G_k might not be invertible for finitely many snapshots ⇝ use G_k^+ instead. BET & continuity of determinant ⇒ G_k inv'ble if m suff. large
- Spectral convergence of EDMD: [Korda, Mezić, arXiv: 1703.04680, 2017]

3.15 EDMD FOR THE FPO

Intuition: $\mathcal{U}_k^T = A_k G_k^{-1}$, with $A_k \xrightarrow[m\to\infty]{} \langle \psi_i, \mathcal{U}\psi_j \rangle_\mu = \langle \mathcal{P}\psi_i, \psi_j \rangle_\mu$

FPO wrt μ

⇝ $P_k^T = A_k^T G_k^{-1}$ approximates Galerkin projection of the FPO.

<u>Cor.</u> (of Thm 3.14): Assume that T is non-singular w.r.t. μ (Def 3.3A). Then, $P_k = G_k^{-1} A_k^T$ converges as $m \to \infty$ almost surely to the (Petrov-) Galerkin projection of $\mathcal{P} : L^1_\mu \circlearrowleft$ wrt $V = \operatorname{span}\{\psi_i\}$, $W = \operatorname{span}\{\int \psi_i \cdot d\mu\}$.

<u>Note</u>: P_k is in general not an over-determined Petrov-Galerkin method of collocation type for $\mathcal{P}$, because we have no access to $\mathcal{P}\psi_i(x_\ell)$.
Due to the same reason, $P_k \neq \mathcal{U}_k^*$ on V, in general. Duality holds only in the ∞-snapshot limit.

<u>Remark</u>: The considerations of 3.14 & 3.15 can be extended to non-deterministic dynamics:
[Klus, Koltai, Schütte. JCD 2016], [Klus et al, arXiv: 1703.10112, 2017]