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:

\[
\mu_N : = \frac{1}{N} \sum_{\xi=0}^{N-1} \mathcal{T}_{\xi} \xrightarrow{\text{weak convergence}} \mu
\]

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

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 \( \Rightarrow \) 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.

A promising approach: consider dynamics on ensemble of states

Random variable: \( Y \rightarrow T(Y) \) "many states described by a distribution"

\( Y \) state with prob. distribution \( L^1 = \mathbb{P}_Y \rightarrow f_Y \in L^1 \) (transfer operator, linear, bounded, see below)

Comparison:

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

2) One long trajectory vs many (\( \infty \)) short trajectories

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

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

\( \leadsto \) 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.

\( \leadsto \) Set-up under uncertainty / smoothness assumptions on densities

3.3 **FROBENIUS-PERRON OPERATOR**

**Definition**: \( (X, \mathcal{B}, \mu) \) prob. space, \( T: X \rightarrow X \) 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) \( T: [0,1] \to [0,1] \), \( T(x) = \frac{1}{2} \) is singular wrt Lebesgue is non-singular wrt \( \delta_{\frac{1}{2}} \)

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 \( \iff \) one initial random variable with the same density. We work with the second.

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

\( T(Y) \sim f_1 = ? \)

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

Observe: \( \nu: B \to [0,1] \) with \( \nu(A) = \int_{T^{-1}(A)} f_0 \, d\mu \) is a prob. measure, and if \( T \) is non-singular, then \( \nu \ll \mu \)

Random \( \Rightarrow \) \( \exists f_1 \in L^1(X,\mu) \) s.t. \( \nu(A) = \int_A f_1 \, d\mu \)

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

\[ \int_A Pf_1 \, d\mu = \int_{T^{-1}(A)} f_1 \, d\mu \quad \forall A \in \mathcal{B}, f \in L^1 \]

(Note: uniqueness follows from the Radon-Nikodym thm, applied to the positive & negative parts \( f^+ \) and \( f^- \) respectively.)
Example: Circle doubling: $T: S^1 \to S^1$, $T(x) = 2x \mod 1$

$I \subset S^1$ interval $\Rightarrow T^{-1}(I) = \frac{1}{2}I \cup \left[ \frac{1}{2}, \frac{1}{2} + \frac{1}{2} \right]$

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

Subst.: $x = \frac{y}{2}$

\[
= \frac{1}{2} \int_{I} f\left(\frac{y}{2}\right) \, dy + \frac{1}{2} \int_{I} f\left(\frac{y}{2} + \frac{1}{2}\right) \, dy
\]

\[
= \int_{I} \frac{1}{2} \left( f\left(\frac{y}{2}\right) + f\left(\frac{y}{2} + \frac{1}{2}\right) \right) \, dy = \int_{I} Pf(y) \, dy
\]

\[
\Rightarrow Pf \leq Pf
\]

3.4 PROPERTIES OF THE FPO

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

(a) $P(\lambda f + g) = \lambda Pf + Pg \quad \forall f, g \in L^1, \lambda \in \mathbb{R}$ (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-negative 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 $A \subset B$:

\[
\int_A Pf \, d\mu = \int_{T^{-1}(A)} f \, d\mu = \lambda \int_{T^{-1}(A)} f \, d\mu = \int_{T^{-1}(A)} f \, d\mu + \int_{T^{-1}(A)} g \, d\mu
\]

\[
= \lambda \int_A Pf \, d\mu + \int_APg \, d\mu \quad \Rightarrow \text{claim}
\]

(b) For $A \subset B$, if $f \geq 0$:

\[
\int_A Pf \, d\mu = \int_{T^{-1}(A)} f \, d\mu \geq 0. \quad \text{Since } A \text{ arbitrary } \Rightarrow Pf \geq 0 \text{ a.e.}
\]
(c) With $A = X$:

$$\int_X Pf \, d\nu = \int_X f \, d\mu = \int_X f \, d\mu$$

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

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

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

(e) To $S$ non-singular:

$$(T^s)^{-1}(A) = T^{-1} \circ S^{-1}(A)$$

$$\mu(A) = 0 \implies \mu(T^{-1} A) = 0 \implies \mu(S^{-1} T^{-1} A) = 0$$

$\implies P_{T^s}$ well-defined

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

$$\int_A P_{T^s} f \, d\mu = \int_{S^{-1} T^{-1} A} f \, d\mu = \int_{T^{-1} A} P_s f \, d\mu = \int_A P_s f \, d\mu$$

**Prop B:** If $(X, \beta, \mu, T)$ is a ppt, then

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

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

$$\int_A Pf \circ T \, d\mu = \int_{T^{-1} \beta} Pf \circ T \, d\mu = \int \beta Pf \, d\mu = \int T^{-1} \beta f \, d\mu$$

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

**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:** $T$ continuous $\implies T^{-1} \beta = \beta$ (Borel) $\implies E(f \mid T^{-1} \beta) = f$ and Prop B yields the claim.
Corollary: If \((X, \mathcal{B}, (\mu_t)_{t \in \mathbb{R}})\) is a ppt, then \(P\) is a contraction on \(L^p(X, \mu)\) for every \(1 \leq p < \infty\).

Proof: For \(1 \leq p < \infty\):
\[
\begin{align*}
\|Pf\|_p^p &= \int |Pf|^p \, d\mu = \int |Pf \circ T|^p \, d\mu \\
\leq & \int E(|f| \circ T^p) \, d\mu \\
= & \int |f|^p \, d\mu = \|f\|_p^p
\end{align*}
\]

For \(p = \infty\):
\[
\|Pf\|_{L^\infty} = \|Pf \circ T\|_{L^\infty} = \|E(|f| \circ T^\infty)\|_{L^\infty} \leq \|f\|_{L^\infty}
\]

3.5 Ergodicity and Mixing

Definition: The operator \(U: L^\infty(X, \mu) \to L^\infty(X, \mu), f \mapsto f \circ T\) is called the Koopman operator associated with \(T\).

Property A: The Koopman and Frobenius-Perron operators are adjoint, i.e., for \(f \in L^1, g \in L^\infty\):
\[
\int Pf \cdot g \, d\mu = \int f \cdot Ug \, d\mu,
\]
we also write \(\langle Pf, g \rangle = \langle f, Ug \rangle\).

Remark: For \(f \in L^1\) and \(A\) a m'b, \(X_A\):
\[
\int f \, d\mu = \int f \cdot X_A \, d\mu = \int f \cdot X_{\mu^{-1} A} \, d\mu = \int \mu_t^{-1} f \, d\mu
\]
\[
= \int Pf \, d\mu = \int Pf \cdot X_A \, d\mu
\]
Since characteristic functions span \(L^\infty\), the claim follows.
Let $D = D(X, B, \mu) = \{ f \in L^1(X, B, \mu) \mid f \geq 0, \|f\|_1 = 1 \}$ denote the set of densities.

**Prop B:** Let $(X, B, \mu, T)$ be a man-ring, ppt. Then

(a) $T$ is ergodic if and only if

$$\frac{1}{m} \sum_{k=0}^{m-1} \langle P^k f g \rangle \underset{m \to \infty}{\longrightarrow} \langle \Pi g \rangle = \int g \, d\mu \quad \forall f \in D, g \in L^\infty$$

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

$$\langle P^k f g \rangle \underset{m \to \infty}{\longrightarrow} \langle \Pi g \rangle \quad \forall f \in D, g \in L^\infty$$

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

- **Ergodicity:**
  $$\text{ergodicity } \Rightarrow \frac{1}{m} \sum_{k=0}^{m-1} \mu(A \cap T^{-k} B) \underset{m \to \infty}{\longrightarrow} \mu(A) \mu(B) \quad (1)$$

- **Mixing:**
  $$\text{mixing } \Rightarrow \mu(A \cap T^{-m} B) \underset{m \to \infty}{\longrightarrow} \mu(A) \mu(B) \quad (2)$$

These are equivalent with

$$\frac{1}{m} \sum_{k=0}^{m-1} \langle f^k \Pi g \rangle \underset{m \to \infty}{\longrightarrow} \int f \, d\mu \cdot \int g \, d\mu \quad (5.1)$$

$$\langle f^k \Pi g \rangle \underset{m \to \infty}{\longrightarrow} \int f \, d\mu \cdot \int g \, d\mu \quad (6.1)$$

if one takes $f \chi_A, g \chi_B$, and hence with (5.1) resp. (6.1) holding $\forall f \in L^1, g \in L^\infty$ (characteristic functions span $L^1(L^\infty)$).

Use the adjoint property (Prop A) and restrict to $f \in D$ to obtain the claim.
Recall: \( \langle \mu, f \rangle = 0 \Rightarrow \mu(\{x\}) = 0 \Rightarrow \nu(A) = 0 \)

Prop: Let \((X, B, \mu)\) be a prob. space and \(T: X \to X\) non-singular. Let \(f^* \in D(X, B, \mu)\), and \(\nu = f^* \mu\) (meaning \(\frac{d\nu}{d\mu} = f^*\)). Then \(Pf^* = f^* \iff \nu\text{ is } T\text{-invariant, i.e. } \nu = \nu \circ T^{-1}\)

\[ \nu(A) = \nu(T^{-1} A) \iff \int_A f^* d\mu = \int_{T^{-1} A} f^* d\mu = \int_A Pf^* d\mu \]

Why is this interesting?

Recall \(T: [0,1] \to [0,1], T(x) = \sqrt{x}\), for which \(\lambda_0\) is an ergodic measure, but not one that we "observe," since \(T^n x \to 1\) 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: \(\lambda_0\) is not absolutely continuous wrt Lebesgue.

If \(\mu\) represents a measure which we consider to be "natural" and wrt this we find \(Pf^* = f^*\), then B.E.T. gives (suppose \(\nu\) ergodic)

\[ \frac{1}{n} \sum_{k=0}^{n-1} h \circ T^k (x) \to \int h d\nu = \int h f^* d\mu \]

for \(\nu\)-a.e. \(x \in X\), which means for a set of \(x\) with positive \(\mu\)-measure, \(h = f^*\) we can observe \(\nu\)!

Existence of \(f^*\): No general results. For specific case: \([Ge, Ch. 7]\)
3.7. ABSTRACT SETTING

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

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

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

$\Pi_m: L^1 \rightarrow V_m$: linear projection, i.e. $\Pi_m \circ \Pi_m = \Pi_m$

Solve eigenproblem for $P_m: V_m \rightarrow V_m$, $P_m = \Pi_m \circ P \circ \Pi_m$, instead!

Q: How to choose $V_m$/its basis?

Examples: 

• "hat functions" $\Rightarrow$ finite element methods
  (local support $\Rightarrow$ sparse stiffness mat.)

• polynomials of different order $\Rightarrow$ spectral methods
  (fast convergence for "smooth problems", global support
  of basis fans $\Rightarrow$ full matrix)

3.8. ULAM'S METHOD (ULAM 1960)

Partition of $X$: $P_m = \{B_1, \ldots, B_m\}$, where

• $B_i \in \mathcal{B}$

• $m(B_i \cap B_j) = 0$

• $\bigcup_{i=1}^m B_i = X$

Let $x_i := x_{B_i}$, and define projection

$$\Pi_m f = \sum_{i=1}^m c_i \frac{x_i}{m(B_i)}, \quad c_i = \int_{B_i} f \, dm \quad \Rightarrow \quad V_m = \text{span} \{x_1, \ldots, x_m\}$$

Discretized operator: $P_m := \Pi_m P$

Matrix representation w.r.t. basis $\{x_i/m(B_i)\}_{i=1}^m$ also denoted by $P_m$

For $x \in \mathbb{R}^m$, write $c = x \in V_m$ iff $f = \sum_{i=1}^m c_i \frac{x_i}{m(B_i)}$
Find \( P_m \in \mathbb{R}^{m \times m} \) s.t. \( \pi m \mathcal{P} f = c^T P_m \)

\[\iff \quad \pi_m \mathcal{P} \left( \frac{X_i}{m(B_i)} \right) = \sum_{j=1}^{m} P_{m,ij} \frac{X_i}{m(B_j)} \quad \forall i = 1, \ldots, m\]

\[= \sum_{j=1}^{m} P_{m,ij} \frac{X_i}{m(B_j)} \]

\[P_{m,ij} = \frac{1}{m(B_i)} \int_{B_j} \mathcal{P} \left( \frac{X_i}{m(B_i)} \right) \, dm\]

\[= \frac{1}{m(B_i)} \int_{B_j} \mathcal{P} \left( \frac{X_i}{m(B_i)} \right) \, dm\]

\[= \frac{m \left( B_i \cap T^{-1} B_j \right)}{m(B_i)}\]

### 3.9 Probabilistic Interpretation

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

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

\[\begin{array}{ccc}
B_1 & B_2 & \ldots \\
\vdots & & \vdots \\
X & & \\
\end{array}\]

\[\xrightarrow{\text{Markov chain on the partition elements}}\]

\[\xrightarrow{\text{Ulam}}\]

\[P_m \text{ stochastic matrix} \implies \exists c^* : \text{stationary probability vector}\]

\[f = \sum_{i} c_i^* \frac{X_i}{m(B_i)} \text{ stationary density of } P_m\]

Prop: \( P_m \) always has an invariant density. Even if \( P \) doesn't.
Recall $P_{m,ij} = \frac{1}{m(B_i)} \int \mathcal{X}_{B_i} : \mathcal{X}_T B_j \, dm$

discontinuous functions, high order quadrature doesn't help.

Monte-Carlo sampling:

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

Approximate $P_{m,ij} \approx \hat{P}_{m,ij} = \frac{1}{N} \sum_{k=1}^{N} \mathcal{X}_{x_k}(T B_j)$

Properties:

- $\hat{P}_m \rightarrow P_m$ 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 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 $\hat{P}_{m,ij} = 0$
- $B_i \cap T B_j \neq \emptyset \iff T B_i \cap B_j \neq \emptyset$

If $T$ Lipschitz continuous, and boxes are "local" (i.e. diam $B_i$ small), then $P_m$ and $\hat{P}_m$ are sparse.
- Computation of $\hat{P}_m$ parallelizable

Convergence: \cite{Li1976:approximation}

A solution to Ulam's conjecture

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

$\begin{align*}
f_m \xrightarrow{L^1} f^* & \quad (m \rightarrow \infty) \\
\end{align*}
1) If there is a subsequence \( \{ f_{n_i} \} \) and a prob. measure \( \mu \) s.t. \( \forall g \in \mathcal{C}_b(X) \)

\[
\int f_{n_i} \, g \, d\mu_i \to \int g \, d\mu,
\]

then \( \mu \) is an invariant measure

[Froyland 1996]

Weaker, but less restrictive than Li's result.

2) \( P_m \) is a discretization of the Koopman operator:

\[
(P_m)_{ij} = P_m(g_i) = \frac{m(B_j \cap T^{-1}B_i)}{m(B_j)} = \frac{1}{m(B_j)} \int_{B_j} X_{T^{-1}B_i} \, d\mu
\]

\[
= \frac{1}{m(B_j)} \int_{B_j} X_i \circ T \, d\mu = \frac{1}{m(B_j)} \int X_i \, d\mu = E\left[ X_i(x) \mid x \sim \text{uni}(B_j) \right],
\]

or equivalently, if \( \mathbb{R}^n \ni c = f \in \mathcal{L}^\infty \), then \( P_m c = E\left[ f \mid \mathcal{E}(P_m) \right] \)

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

\[ \lambda \approx \text{almost-invariant - / persistent - / metastable sets} \]

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
a) \( H \): Hilbert space with inner product \( \langle \cdot, \cdot \rangle \)
- \( L: H \to H \) bounded linear operator
- \( \text{span} \{ \psi_1, \ldots, \psi_k \} = V \)

**Galerkin projection:** \( L_k: V \to V \) is the unique lin. op. with
\[
\langle \psi_i, L_k \psi_i \rangle = \langle \phi_i, L_k \psi_i \rangle \quad \forall i, j = 1, \ldots, k
\]

If \( \tilde{P}_k: H \to V \) denotes the \( \langle \cdot, \cdot \rangle \)-orthogonal proj., then \( L_k = \tilde{P}_k L \).

*Exercise:* Ulam's method is a Galerkin proj. on \( L^2 \mu \), if \( \mu \) invariant.

b) If \( L \) is not living on a Hilbert space, it is advantageous to differentiate between basis and test functions.
- \( L: Y \to Y \)
- \( \text{span} \{ \psi_i \} = V \subseteq Y \)
- \( \text{span} \{ \psi_i \} = W \subseteq Y^* \) (dual space)

**Petrov-Galerkin projection:** \( L_k: V \to V \) is the unique lin. op. with
\[
\langle \psi_i^*, L_k \psi_i \rangle = \langle \phi_i^*, L_k \psi_i \rangle \quad \forall i, j = 1, \ldots, k
\]
where \( \langle \psi_i^*, f \rangle = \psi_i^* (f) \) "duality pairing"

What if we choose \( k = \dim W > \dim V = k \)?

**Over-determined Petrov-Galerkin**, set up, e.g., as least squares problem:
\[
\sum_{j=1}^{k} \sum_{i=1}^{k} \langle \psi_i^*, L_k \psi_i - L_k \psi_i \rangle^2 = \text{min}!
\]
We are only given a fixed set of dynamical data: 
\[ X = [x_1 \cdots x_m], \quad Y = [y_1 \cdots y_m] \] with \( y_i = T(x_i) \), called snapshots.

Define
\[ \Psi = \begin{bmatrix} \Psi_x \\ \Psi_y \end{bmatrix} : \mathcal{X} \to \mathbb{R}^k \]

\[ \Psi_x = [\Psi(x_1) \cdots \Psi(x_m)] \in \mathbb{R}^{k \times m} \]
\[ \Psi_y = [\Psi(y_1) \cdots \Psi(y_m)] = [U \Psi(x_1) \cdots U \Psi(x_m)] \quad U : Koopman \text{ operator} \]

Point evaluations set up a Petrov-Galerkin method of collocation type 1 i.e. \( \psi_i \) = \( \delta_{i,j} \) (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
\[ \sum_{i=1}^m \sum_{j=1}^k \left< \delta_{x_j}, U_k \psi_i - U_k \psi_i \right>^2 = \min \]

\[ \sum_{i=1}^m \sum_{j=1}^k \left( \psi_i \psi_j(x_k) - \sum_{j=1}^k U_k \psi_i (x_k) \right)^2 = \min \]

\[ \sum_{i=1}^k \left\| \Psi_y \psi_i - (U_k^T) \Psi_x \right\|^2 = \min \]

\[ \sum_{i=1}^k \left\| \Psi_y \psi_i - U_k^T \Psi_x \right\|^2 = \min \]

LSP for \( U_k : i \)

Since \( k > k_1 \) solutions given by normal equation.
\[ U_{k,i} = \left( \Psi_x \Psi_x^T \right)^{-1} \Psi_x \Psi_y^T \]

all columns:

\[ U_k = \left( \Psi_x \Psi_x^T \right)^{-1} \Psi_x \Psi_y^T \iff U_k^T \Psi_y \Psi_x^+ \]

NB: \( \iff ||\Psi_y - U_k \Psi_x||_F^2 = \min \)

Moore-Penrose pseudo-inverse, \( \Psi_x \Psi_x^+ = I \)

Remark: \( X = \mathbb{R}^d \) \( \Psi_i(x) = x^i \) (i-th coordinate, i.e. \( k = d \)) gives the so-called Dynamic Mode Decomposition (DMD); Schmid 2008

### 3.14 Convergence to a Galerkin Method

\[ U_k = \left( \Psi_x \Psi_x^T \right)^{-1} \Psi_x \Psi_y^T = \left( \frac{1}{m} \Psi_x \Psi_x^T \right)^{-1} \left( \frac{1}{m} \Psi_x \Psi_y^T \right) \]

i.e. \( G_k U_k = A_k \)

Let:

(a) \( x_1, x_2, \ldots \) be the trajectory of a \( \mu \)-ergodic process (\( \mu \) prob. measure)

(b) \( \Psi_i \in L^\infty \), thus \( \Psi_i, U \Psi_i \in L^p \) for any \( 1 \leq p \leq \infty \)

\[ \Rightarrow A_k \Psi_i = \frac{1}{m} \sum_{l=1}^{m} \Psi_i(x_l) \Psi_j(x_l) = \frac{1}{m} \sum_{l=1}^{m} \Psi_i(x_l) U \Psi_j(x_l) \]

if applicable

\[ \sum_{k=1}^{m} \Psi_i(x) U \Psi_j(x) \, d\mu(x) \Rightarrow \left\langle \Psi_i, U \Psi_j \right\rangle_\mu \]

\[ G_k \Psi_i = \frac{1}{m} \sum_{l=1}^{m} \Psi_i(x_l) \Psi_j(x_l) \overset{\text{BET}}{\longrightarrow} \left\langle \Psi_i, U \Psi_j \right\rangle_\mu \]

Thus:

\( U_k = G_k^{-1} A_k \) converges as \( m \to \infty \) almost surely to the (Petrov-)

Galerkin projection of \( U : L^\infty \mu \to V = \text{span} \{ \Psi_i \} \) on \( \text{span} \{ \int x \Psi_i \, d\mu \} \).

Remarks:

- Often \( y_k = x_{k+r} = T(x_k) \) (especially widely used in MD),
  and then \( \mu \) is automatically the ergodic measure of \( T \).
\* $G_k$ might not be invertible for finitely many snapshots. Use $G_k^+$ instead. BEF & Continuity of determinant $\Rightarrow G_k$ invertible if $m$ sufficiently large.


### 3.15 EDMD FOR THE PPO

**Intuition:** $U_k^T = A_k G_k^{-1}$ with $A_k \xrightarrow{m\to\infty} \langle \Psi_i, \mu \rangle = \langle \Phi_k, \Psi_j \rangle$

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

**Cor. (of Thm 3.14):** Assume that $T$ is nonsingular w.r.t. $\mu$ (Def 3.3.1).

Then, $P_k = G_k^{-1} A_k^T$ converges as $m \to \infty$ almost surely to the (Petrov-)

Galerkin projection of $P$: $L^2(\mu) \text{ wrt } V = \text{span } \{ \Psi_i \}$, $V = \text{span } \{ \int \Psi_i \, d\mu \}$.

**Note:** $P_k$ is in general not an over-determined Petrov-Galerkin method of collocation type for $P$, because we have no access to $\Phi_k(x_e)$.

Due to the same reason, $P_k \neq U_k^*$ on $V$, in general. Duality holds only in the $\infty$-snapshot limit.

**Remark:** The considerations of 3.14 & 3.15 can be extended to non-deterministic dynamics: