

Markov State Modelling (MSM)

- analysis of data sets with orders.
- simplest e.g. one MD trajectories
- other exs of ordered datasets:
 - ① genome → if we reshuffle the genome, nothing meaningful will be coded.
 - ② clinical recordings → EEG trace, time series over all the info is coded in the sequencing the signal we observe.
Also true for fMRIs.
 - ③ text :- reshuffling orders of text makes it gibberish, so Order matters.
- we lose information if we reshuffle the data in time-ordered series.
- we can use ML techniques that are relevant for non-time-ordered data for time-series data.

Dynamic Systems:

→ In a dynamic system,

(x) = vars. of the system

→ The data points that we have labelled
so far as $\underline{x^i}$, i is now $\underline{x^t}$

$x^i \rightarrow x^t$ ($t \equiv t_{N-1}$ label)

→ In general, x is a multi-dimensional vector.

→ (t) in x^t plays same role as (i) in x^i

$$\rightarrow \boxed{x^{t+1} = f(x^t)}$$

(Be consider only
markovian
processes now)

$f \rightarrow$ can be deterministic, i.e.,
 x^{t+1} is an explicit func² of
 x^t

→ can be stochastic, & typically
we have a deterministic part &
some random noise we can add
to the dynamics.

$$\Rightarrow \text{In } \mathbb{M}^D : X = (q, \dot{q}) \xrightarrow{(3N-2)} (3N-1)$$

$\hookrightarrow X \in 6N-1$ vector

If be we Hamilton's eq^{ns} of motion.
 (If is a deterministic func²)

However, if operates on a very high dimensional ($6N$) space, if is a highly nonlinear func² that even though it is deterministic, it behaves in a somewhat stochastic manner.

"KAM" theorem

Assumption: $\{X\}$ vector is large enough that the map is considered

"Markovian"

→ For a non-Markovian process, -

$$x^{t+1} = g(x^t, x^{t-1}, x^{t-2})$$

we map (x) to a state vector (\bar{x})

which includes $(\underline{x^{t-1}, x^{t-2}})$

$$\therefore \bar{x}^t = (x^t, x^{t-1}, x^{t-2})$$

Now, we can transfer this to a

Markovian process; -

$$\boxed{\bar{x}^{t+1} = g(\bar{x}^t)}$$

⇒ So, we can sort of convert a non-Markovian process to a Markovian process, (Nambu's eq \rightleftharpoons → Hamilton's eq \rightleftharpoons)

⇒ We generally assume the Markovian approximation in our derivations here; it's reflected well in real life datasets.

⇒ Let's take a simple Markov Map :-

$$x^{t+1} = f(x^t, \xi^t)$$

→ many of the topics we explore can also be applied to strictly deterministic maps, in the case where transition probabilities

are 5-functional, i.e., we can only go from one point to another.

→ So, without loss of generality, we assume

(f) as a stochastic map

→ ξ^t = random variable independently generated at each time-step

→ Let's assume ξ^t = Gaussian process

→ Another process we can write like this is MC (Monte-Carlo). MC can be markovian ($m_i m_j$) & new state depends on prev state + 2 random #s (displacement of x , acceptance prob.)

→ For simplicity we work with MC / MD
with a stochastic thermostat

⇒ In Monte-Carlo (MC) or MD :-

→ f is chosen so that (x^t) can
be thought of as harvested from a
given prob. density.

→ In MC & Metropolis-Hastings is assigned
so that (x^t) is from a
specific prob. density

→ ∴ $x^t \sim p_{eq}(x)$

$$p_{eq} \equiv \text{eqb}^{\text{--}} \text{ prob. distrib}^{\text{--}}$$

$$\cdot = \frac{1}{T} \exp\left(-\frac{V(x)}{T}\right)$$

= Canonical prob.
distrib[?] (as in eg.)

\Rightarrow Ignoring the time-ordered aspect:

- (1) we can do PCA/RPCA on X
- (2) we can compute the intrinsic dim of our data points
- (3) we can compute the density of our data points & find the density peaks clusters

the common assumption of all of these approaches is

\rightarrow statistical independence of our data

$$\text{i.e., } \underline{X^t} \perp \underline{\underline{X^{t-1}}}$$

$\perp \perp \rightarrow$ notation for stat. indep.

Red.-r Poisson Distrib^b \rightarrow data is harvested from a Poisson Distrib^b $\rightarrow Z^N N_{ID \text{ etc.}}$

→ For $m < \infty$ we could collect data

points in "steps". That is if we look with $\sim N$ particles, we attempt $\sim 3^N$ moves, so that we are reasonably sure that our data pts are statistically independent.

→ The simplest way of ensuring statistical independence in a dynamical system is

"Decimation",
↓

Q) Now do we verify if we have decimated our sys enough?

→ Empirically → Verify that results are same if we decimate sys by factor of 10 or 20, if this is the case, we trust our results.

→ If we get $ID = 3$, $dt = 1 \text{ ps}$
 $ID = 5$, $dt = 2 \text{ ps}$

BAD → we are not happy.

- \Rightarrow For mD trajectories if we don't
 generate at all & instead take every config. as generated. What happens?
- \rightarrow x is = very high dim vector
- \rightarrow So, it is very unlikely that we find the exact same config. "be as in a very high dim space with chaotic tendencies."
- \rightarrow If we compute the ID without determination of an mD traj, we always get $\boxed{ID = 1}$ $\{\because mD \in \text{deterministic}\}$
- \rightarrow ID for Brownian motion $\neq 1$
 \therefore or back \Rightarrow Gaussian noise.
- \rightarrow For mD , $ID = 1$, \therefore the process is deterministic, If we take every step of the MD sim = rigorous.

→ TL;DR: we can use LLL techniques from before as long as we are very careful about fulfilling the criteria of statistical independence of the variables,

Time Correlation Structure of the Data:

→ Idea: we want to derive a low-dimensional description that can reproduce non-trivial time correlation structure of the data.

⇒ What does non-trivial time-correl. of data mean?

(i) In dynamic systems, our dynamics in many cases of interest are trapped in "attractors".



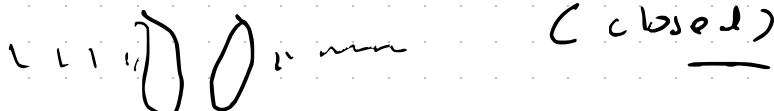
chaotic sys. not can

jump b/w \Rightarrow 2 completely diff orbits

\rightarrow even in det. dyn. systems the system jumps to another orbit (gray) + at some point can go back to the previous orbit.

\rightarrow The molecular equivalent is transitions b/w \Rightarrow metastable states,

(\odot) a metastable state is a conformation of a molecule/protein whose short-time dynamics appears to not move & just oscillate in a single β -min. For ex., a membrane channel that is closed



\rightarrow only rarely the channel performs a transition to an open state - e.g. (\odot) - Equivalently to transition in dynamical state attractors (also same in 1st order phase transitions)

- The techniques we will introduce will allow us to compute these time - cons. \approx func
- That is, we can compute the probability of transitions $b^{tj} =$ of different attractors + $b^{ti} =$ different metastable states.

→ We want integer labels of probabilities of each metastable state.

- Let's start from a long trajectory (x^t)
- Our goal: map x^t to a finite set of metastable states (with integer labels)

i.e.: $x^t \rightarrow C^t$
 $\begin{array}{c} \text{finite} \\ \text{set of} \\ \text{metastable} \\ \text{states} \end{array}$ $\begin{array}{c} C^t \\ \text{(integer labels)} \\ (C^t \rightarrow \text{individual} \\ \text{metastable} \\ \text{states}) \end{array}$

∴ We go from: $\mathbb{R}^D \rightarrow \mathbb{N}$
 $\begin{array}{c} \text{(subset of} \\ \text{natural nos.)} \end{array}$

(e.g. $x \in \mathbb{R}^n$ from say, \rightarrow \mathbb{R}^n -d vector
to say, 1D metastable states,

Markov - State Modelling (MSM)

→ clustering technique, other instead of looking at prob. distribⁿ of data, one instead looks at time corr² of data.

→ For the Lorenz attractor, we get nonsense results if we do density based clustering: "the 2 orbits touch, so there is no red separation, but often, there are 2 diff. kinetic states."

→ i.e. MSM is a clustering technique to reproduce time-corr² str., in the dynamic system (X^t) with a dynamics on the integers.

→ The clusters defined by MSM are called (Markov States → MS).

→ For eg, for the Lorenz attractor :-

black orbit $\rightarrow 1\ 1\ 1\ 1\dots$

grey orbit $\rightarrow 2\ 2\ 2\ 2\dots$

in overall dynamics :

1 1 1 1 1 1 2 2 2 2 2 2 1 1 1 1 1

↓

(We go back to
block orbit)

→ This segment allows us to compute quantities relevant for mesoscopic analysis of our system.

↳ (i) time to wait in state (1) before performing transition to state (2)

(ii) reproduce prob distib² of this waiting time \rightarrow weight in MD.

→ A transition matrix with only 1 state event
Should be extremely noisy.

→ MSM is not enhanced sampling technique, but is typically used for analysis a -posteriori.

→ We want to map \mathbb{R}^m to very high-dim.

$\text{spec}(\mathbb{R} \rightarrow \mathbb{R}) \not\cong \mathbb{Z}$ for genomes etc.)

to a low-dim. space of integers (tree-based
representable
state clusters)

→ $x^t \rightarrow k$ -means Clustering, for \approx very
large K , say $N = 10,000$

→  for every large K , we actually end up partitioning our data points into cells. The larger K the smaller the cells we obtain, so all pts. belonging to a single cell are very, very similar to each other.

→ i.e. we choose (K) large enough, such that data points x^t belonging to a cluster are very similar.

- We can't guarantee "very similar" to it
 is a degree of arbitrariness that we introduce.
- These "clusters" after K-means clustering
 are called "microstates" (NOT the final
clusters that we want)
- ↓
 e.g., they contain configurations that
 are microscopically indistinguishable.

$$\therefore \boxed{x^t \rightarrow m^t}$$

Microstates diff. from
 c^t , which are the macrostate)

- $m^t \rightarrow$ integer space : (This is particularly useful when x^t is non-integers
 otherwise exists in \mathbb{R}^d .)
 convenient basis for
 representing dynamics but

otherwise exists in $\underline{\mathbb{R}^d}$.

We are exploring the power of
 clustering here & finally we want a convenient
 basis for representing our traj. in integer space
 that is large enough that we can be sure about faithfully
 representing our dynamics.

\Rightarrow In on Lorenz sp, - (for eg.)



→ the microstates just represent sections along the orbit

→ they have NOTHING to do with ID of the system

→ They are just a bag.

\Rightarrow We characterize the system with the probability, (P_m) of observing our system in a microstate (m) at any (t) .

\Rightarrow Forgetting (t) we can simply count how many times we observe each microstate during our whole trajectory.

$$P_m \sim \# \text{ of times we observe } (m) \text{ in all the trajectory}$$

→ We start from :-

→ stochastic noise

$$\therefore X^{t+1} = f(X^t, g_t)$$

(# microstates)

(this happens under certain assumptions)

$$P_m^{t+1} = \sum_{m'=1}^N k_{m'm} P_{m'}^t$$

→ we also assume
it is "linear"
dynamics

we assume

we can

write the

dynamics in

terms of prob.

of observing y_i in microstate m)

$N = \text{no. of clusters from K-means clustering}$

$= \text{no. of microstates}$

$\rightarrow k_{m'm} = \text{matrix } (N \times N) \rightarrow \text{"rate" matrix}$

$\rightarrow P_m^t = \text{vector with } N \text{ components}$

- ① We want to see what are the cond's
under which we can go from dynamics
in $\{X\}$ to dynamics in a probability
vector in a "not-so-large" space.

$$\Rightarrow x^{t+1} = f(x^t, \xi^t) \xrightarrow{\substack{\text{random var} \\ \text{original dynamics}}} \xrightarrow{\substack{\text{dynamic map} \\ \text{(dyn map)}}}$$

→ This means if we know $p^t(x)$ we can compute $p^{t+1}(x)$, $\xrightarrow{\substack{\downarrow \\ \text{prob of observing } x \text{ at time } t'}}$

$$\therefore p^t(x) \rightarrow p^{t+1}(x)$$

→ Now do we compute this prob.?

→ with no assumptions & without loss of generality we can write it as:-

$$p^{t+1}(x) = \int dx' p^{t+1}(x') \underbrace{p(x_{t+1} | x'_t, t)}_{(\text{cond prob})}$$

→ ∵ $x^{t+1} = f(x^t, \xi^t)$ is a Markov process, & ∵ $p^{t+1}(x)$ doesn't depend explicitly on time we can rewrite it as $\underline{\underline{p^t(x)}}$

$$\underbrace{P(x_{t+1} | x_t, t)}_{\downarrow} = P(x_{t+1} | x_t)$$

transition prob. can be considered as time indep. i.e., it is same if we compute it b/w $t \rightarrow t+1$, $t+1 \rightarrow t+2$, \vdots , $t \rightarrow \infty$ etc.

\Rightarrow This is the time invariant condition
prob. of observing the sys. in 1 config. at $t=1$ & another at $t=\infty$. This is called the "transition Kernel"

$$\therefore P(x_{t+1} | x_t, t) = P(x_{t+1} | x_t) = K(x \rightarrow x')$$

func. $\overset{\leftarrow}{\text{b/w}} \mathbb{R}^D \rightarrow \mathbb{R}^D$, so it lies in a very high dim space.

$\rightarrow K$ can be written explicitly for a certain dynamics (Langevin / Master eqn etc.) for small enough timestep

in continuous space :- \rightarrow (linear)

$$\Rightarrow P^{t+1}(x') = \int dx' K(x' \rightarrow x) P^t(x')$$

\rightarrow given that our dynamics is a time-dep. Markov Process this eq² can be formally associated with our dynamics,

\Rightarrow Now let's discretize this continuous eq² :-

$$\therefore P_m^{t+1} = \sum_{m'} K_{m' \rightarrow m} P_{m'}^t$$

this is a linear transformation; the

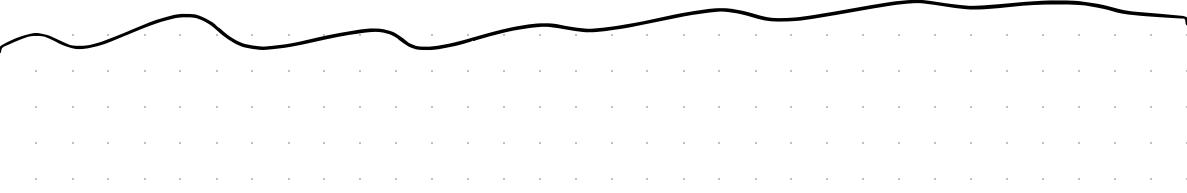
K is a non-symm. matrix
is it a conditional prob.
where the cond? is on
the first index.

interpretation we derived this from it is of linear.

\Rightarrow We can think of this transformation as linear
ONLY if our partition is fine/small enough.

This is controlled by choosing "N" during the K-Mans clustering.

→ We can also choose the microstates in a different / smarter way, but better-choice approach is to make N larger.



⇒ Properties of $K_{m \rightarrow m}$ (stochastic matrix) :-

(1) $K_{m \rightarrow m}$ is the conditional prob. of observing the system in microstate "m" given that the system was in microstate "m'" at time $(t-1)$

e.g., if traj/segment is $\textcircled{1} \textcircled{1} \textcircled{3} \Rightarrow \textcircled{1} \textcircled{3} \dots \textcircled{5} \textcircled{5} \textcircled{2}$

$$K_{1 \rightarrow 3} \sim \frac{2}{\text{# times } 1 \rightarrow 3} = \frac{(\# 1 \rightarrow 3)}{(\# times sp. is 1)}$$

approximately not exactly so we estimate & there might be other factors it.

(2) $K_{m' \rightarrow m} \geq 0$, always: it is
a conditional prob.

positive
defined

at least 10
events,
 $\Rightarrow K_{m' \rightarrow m} > 0$

(normalized)

$$\sum_m K_{m' \rightarrow m} = 1 \quad \text{if } m'$$

i.e., it is normalized
with respect to its 2nd index (m)
i.e., index to the right of the arrow.

\Rightarrow if $K_{m' \rightarrow m}$ satisfies these 2 criteria
it is called a
"stochastic Matrix"

\Rightarrow Now if $(J) \rightarrow \text{map} \rightarrow$ is explicitly
time-dep., we can still do this same
formalism, the only change is that ' K '
becomes time-dependent, $(K^t_{m' \rightarrow m})$.
However, $K^t_{m' \rightarrow m}$ still remains a stochastic
matrix.

- \Rightarrow The problem will be in estimating (K^t) due to the time-dependence, but the theoretical derivation remains valid, if $K_{m' \rightarrow m}$ is \equiv cond⁻¹ prob., & it is \equiv a stochastic matrix.
- \Rightarrow The manner of estimating it, however, becomes tricky once time dependence is introduced, & is problem-dependent.

→ macrostates should be considered a basis on which we represent our prob. density which in principle lives in \mathbb{R}^D .

→ $K_{m' \rightarrow m} \equiv$ cond⁻¹ prob. of observing sys. in microstate ' m' ' at time - t , given that sys. is in microstate ' m ' at time = 0

→ " this is time-indep., we estimate it by averaging over all the trajectories.

→ The rate matrix ($K_{m' \rightarrow m}$)

satisfies both the properties mentioned.

→ We investigate properties of stochastic matrices that allow us to perform a dim. reduction on our system. This implies that instead of describing our system's dynamics in

a basis of say 1000 microstates we can (under certain cond^{ns}) describe our system in 3-4 Markov/metastable states

→ Let's rewrite properties in a way that we can refer to them repeatedly in a convenient manner in future derivations:

(P1)

$$K_{m' \rightarrow m} \geq 0 \quad \forall m, m'$$

(P2)

$$\sum_m K_{m' \rightarrow m} = 1, \quad \forall m'$$

→ Let's rewrite $\underline{P_2}$ in matrix form :

① $K_{n' \times n} = \text{non-symmetric matrix}$

② $\underline{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ (vector with value = 1 in column format)

So $\underline{P_2}$ can be rewritten as

$$\rightarrow K \underline{1} = \underline{1}$$

⇒ ∵ $\underline{1}$ is a ^{right} eigenvector of K

→ ∵ is $\underline{P_2}$ we sum over 2nd index (n)

& ∵ K is non symmetric ∴ left & right

eigenvectors are diff.

→ $\underline{1}$ is a right eigec. with one eigenval = 1

P.B. If column of K sums to 1, then $\underline{1}$ is left eigec.
if row sums to 1, then $\underline{1}$ is right eigec.

$\rightarrow \text{TLDR} \rightarrow \textcircled{1}$ is a right eigenv. of K

\textcircled{ii} any stochastic matrix has
at least one eigenvalue = 1

$\textcircled{1}$ K is a matrix of the non \neq 's, which

is square & non-sym.

(left & right
eigvals are
same)

eigvecs (left & right)
are diff.

$\textcircled{2}$ But eigvals are the same, \therefore the secular eq $=0$ is the same. So, left & right eigvals are the same.

$\textcircled{3}$ eigval = 1 is associated to right eigvec, but there's also a left eigvec with eigval = 1

$\textcircled{4}$ let's write the left eigvec eq $=0$ associated to

$$\therefore \sum_{m'} K_{m' \rightarrow m} v_{m'} = v_m \quad \lambda = 1 \text{ :-}$$

(left eigval eq $=0$ of K , associated to eigval = 1)

→ we sum over first index, i.e., integrate from where we start.

→ Comparing this to eq^{c} from before (EQ-DYN)

$$P_m^{(t+1)} = \sum_{m'} K_{m' \rightarrow m} P_{m'}^{(t)}$$

We can see that left eigenvectors associated to eigenvalue = 1 is a stationary sol² of my dynamics

⇒ ∵ left eigenvectors associated to $\lambda = 1$ is a stationary sol² to

(EQ-DYN),

$$\therefore \vartheta = P^{\text{eq}}$$

($\text{eq}^{\text{c}} = \text{prob distrib}^{\text{?}}$)

N.B.: In (EQ-DYN) K plays the role of time-propagator in a linear eq^{n} .

\rightarrow ; There must be at least ONE stationary

Solution.

Q1: Is this stationary sol² stable?

"YES"

What does stable mean? $(P_m^{ev} + \delta_m)$ $\xrightarrow{\text{perturbation}}$

Imagine we generate an initial config at P^{ev} & then add a small random perturbation. There are 2 possibilities:-

i) perturbation \uparrow , i.e., sol² is stationary but NOT stable

OR ii) dynamics goes back to P^{ev} i.e. sol² is stable.

Q2: Is the stationary sol² unique?

in general the answer is "No". But there are condⁿs under which this sol² is unique. In many dynamic systems we work with these condⁿs are often satisfied.

→ Q2 is strongly related to topic of metastability & ergodicity breaking.

→ Q1 → "YES" i.e., dynamics described by stochastic matrices has a stationary set that is at least locally stable. They are globally stable if π^* is unique, & locally if π^* is non-unique.

Q Let's answer Q1 mathematically :-

To answer this, we need to prove the "Gersh-Gorin" theorem :-

→ All the eigenvalues of a stochastic matrix are ≤ 1 .

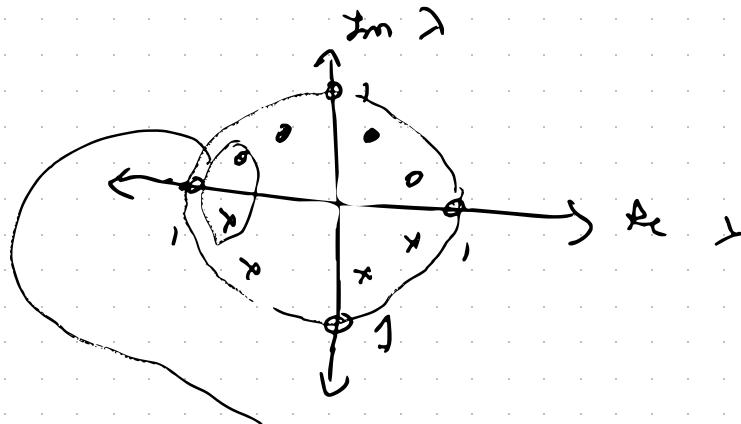
That is,

$$|\lambda_m| \leq 1$$

note the "modulus"
∴ $\operatorname{Re}(\lambda)$ might be
-ve ∵ λ is complex no.

→ remember that K is non-symm. &
eigenvls of non-symm. matrices are complex
numbers, but they will all fall within

The unit circle



→ M2 notes the eigenvalues always come
in conjugate pairs. The qualitative reason
for this is that \therefore the prob is real
conj. pairs are needed to remove
the complex part when calculating the prob,

⇒ Proof of Gershgorin Theorem :-

$$\sum_{m'} K_{m' \rightarrow m} v_{m'} = \lambda^L v_m$$

$$m' \underbrace{\qquad\qquad\qquad}_{\downarrow}$$

(left eigvec eq^L)
 \therefore summation over
 st index of K)

$(N \times N)$ matrix

↓
 so, it will have $N - \text{sd}^{2^L}$.

→ we add ' λ^L ' as superscript & it labels
 the different eigvals & eigvecs.

→ so we rewrite as :-

$$\sum_{m'} K_{m' \rightarrow m} v_m^L = \lambda^L v_m^L$$

→ And right eigvec eq^L is :-

$$\sum_{m'} K_{m \rightarrow m'} v_{m'}^L = \lambda^L v_m^L$$

(summation over 2^{2^L} index)

\rightarrow Take the norm of both sides (of right
eig vec eq $\hat{=}$)

$$|x^\alpha| |v_m^\alpha| = \left| \sum_{m'} K_{m \rightarrow m'} v_{m'}^\alpha \right|$$

\rightarrow Now we use (triangle inequality):

$$|a+b| \leq |a| + |b| \quad \text{for 2 vectors } (a) \text{ & } (b)$$

$$\therefore \left| \sum_{m'} K_{m \rightarrow m'} v_{m'}^\alpha \right|$$

$$\leq \sum_{m'} |K_{m \rightarrow m'} v_{m'}^\alpha|$$

$\rightarrow \because K_{m \rightarrow m'} \geq 0$ we can bring it out from the norm

$$\therefore \left| \sum_{m'} K_{m \rightarrow m'} \hat{v}_{m'}^{\alpha} \right| = |\lambda^{\alpha}| |v_m^{\alpha}|$$

$$|\lambda^{\alpha}| \leq \sum_{m'} |K_{m \rightarrow m'} v_{m'}^{\alpha}|$$

$$|\lambda^{\alpha}| \leq \sum_{m'} K_{m \rightarrow m'} |v_{m'}^{\alpha}|$$

$$\Rightarrow \therefore |\lambda^{\alpha}| |v_m^{\alpha}| \leq \sum_{m'} K_{m \rightarrow m'} |v_{m'}^{\alpha}|$$

→ Let's now define $\{\hat{v}^{\alpha}\}$ → a +ve numbers

S. Poch

$$\hat{v}^{\alpha} = \max_{m'} |v_{m'}^{\alpha}|$$

(max possible value of v^{α})

max modulus of v^{α} w.r.t. m'

$$\therefore |\lambda^{\alpha}| |v_m^{\alpha}| \leq \sum_{m'} K_{m \rightarrow m'} \hat{v}^{\alpha}$$

$\therefore \hat{v}^\alpha$ is indep. of m' \therefore we can bring it out of the summation :-

$$\therefore |\lambda^\alpha| |v_{m'}^\alpha| \leq \hat{v}^\alpha \sum_m k_{m \rightarrow m'} = 1$$

This is why
it's important to
start from right
eigenv. eq²

(due to P2)

↓
summation over 2nd
index)

$$\therefore |\lambda^\alpha| |v_{m'}^\alpha| \leq \hat{v}^\alpha \quad , \quad \text{if } m$$

no assumption
made

i. This must also be true for (\hat{m}) , where

$$\hat{m} = \arg \max_{m'} |v_{m'}^\alpha|$$

That is \hat{m} is the m' that corresponds to max value of $|v_m^x|$ that is

$$\hat{v}^x = |v_{\hat{m}}^x|$$

$\rightarrow \therefore$ The inequality must also hold for $m = \hat{m}$, \because it holds for all m .

$$\therefore |x^x| |v_{m=\hat{m}}^x| \leq \hat{v}^x$$

|

$$|x^x| \hat{v}^x \leq \hat{v}^x$$

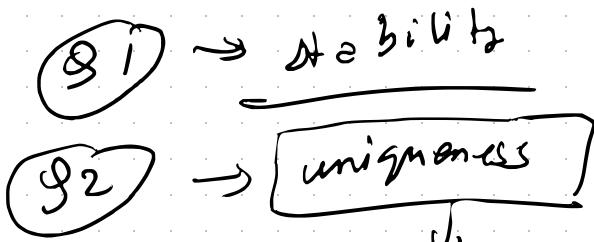
|

$$\Rightarrow \therefore |x^x| \leq 1 \quad (\text{Q.E.D})$$

Remember :-

to prove Gershgorin theorem, we start with RIGHT eig vec eq^T.

→ defⁿ's recap :-



→ addressed by "Perron-Frobenius"

Theorem,

→ To get some insight into the uniqueness of the solⁿ, let's start from a case where the solⁿ is NOT unique.

→ As we saw before :-

$$K \underline{1} = \underline{1} \quad , \quad \sum_{m'} K_{m \rightarrow m'} = 1, \forall m$$

→ Let's define 2 weird unique vectors.

$$\underline{1}_H = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \underline{1}_L = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}$$

That is, for $N=8$, we have 2 vectors

with "high (H)" values = 1

& "low (L)" values = 1

→ We can imagine a situation where $\mathbb{1}_n$ & $\mathbb{1}_L$ are separately eigvecs. of

K_0

$$\rightarrow \therefore K \mathbb{1}_n = \mathbb{1}_H \quad \left. \begin{array}{l} \text{is this case,} \\ \text{we would have} \\ 2 \text{ eigvecs with} \\ \lambda = 1 \end{array} \right\}$$
$$\text{and } K \mathbb{1}_L = \mathbb{1}_L$$

i.e. $\lambda_1 = \lambda_n = 1$

$$\lambda_2 = \lambda_L = 1$$

⇒ But to satisfy these 2 cond \Rightarrow
simultaneously, how must K look like?

$\rightarrow \underline{K}$ must be "block-diagonal"

\rightarrow That is over 2 eq \Rightarrow can be valid only if

$$K = \begin{pmatrix} K_n & 0 \\ 0 & K_L \end{pmatrix}$$

\rightarrow To see the reasoning, let's write the first component of fair eq \Rightarrow :-

i) $\sum_{m'=1}^{\infty} K_{n \rightarrow m'} = 1$ for (\underline{I}_n)

We truncate sum to Σ , i.e.
lower half of I_n is 0's

But also ii) $\sum_{m'=1}^{N} K_{n \rightarrow m'} = 1$
(by def \Rightarrow)

$$\text{And } \kappa_{n \rightarrow m} \geq 0 \quad \forall n'$$

\therefore to satisfy both (i) & (ii),

$$\kappa_{n \rightarrow m'} = 0, \quad \forall n' \neq m$$

\therefore κ must be of block-diagonal form

\Rightarrow That is, we show that we can have more than one eigenvalue = 1, but then $\underline{\kappa}$ must be "block-diagonal".

\Rightarrow T_{ij} DR :- if $\underline{\kappa}$ is block-diagonal then we have \geq number of eigenvalues = 1, which is equal to no. of blocks.

○ Recap of 32 :-

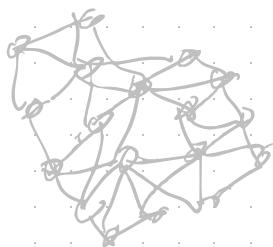
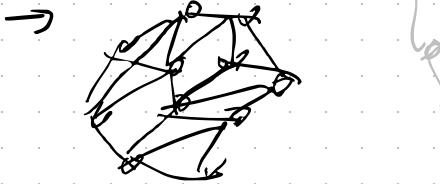
$\rightarrow X^1 = 1$ is unique if it is impossible to bring K into a block diagonal form by a permutation of indexes
i.e., relabelling my microstates.

Corollary :- If $K_{m' \rightarrow m}$ general is non-negative and strictly positive then $(X^1 = 1)$ is unique
this is transition from any microstate to any other is $\neq 0$, even with $m =$ very small probability, ($K_{m' \rightarrow m} > 0$)
(which is)

→ Physically each ~~pure~~ right eigenvector is associated to a left eigenvector that gives us a stationary set? of our state matrix.

∴ There can be > 1 stat. (P^{ev})

① What does block-diagonal K mean physically?



- say we have our microstates (as shown above)
- & we draw a line for microstates connected by a non-zero element of K .
- Then for a block-diag K , there will be no lines connecting black & gray microstates.

- That is there will be 2 subsets of
connected metastates, but no connections
 b/w \Rightarrow these 2 subsets,
 → So there is no way of jumping from
 one region to another.
 → So our sys. will have 2 prob.
 measures, one for left & one for right
 parts.
 → So we can separate our rate eq^{ns} into
 2 totally indep. sub-blocks, & proceed
 as before + $\begin{pmatrix} p^{(r)}_M & p^{(r)}_L \\ p^{(l)}_M & p^{(l)}_L \end{pmatrix}$ as
 2 ordinary stationary prob. distrib^{ns},
 → That is, if our sys. is separated in blocks
 we will be in an extreme region of
metastability ; it is completely
 impossible to pass from one region of phase
 space to another.

$\rightarrow \therefore$ Extreme instability is detected
by seeing if our system has more
than one eigenvalue = 1.



① Let's rewrite some stuff (Final lecture) :-

⇒ Dynamic eq^r in vector of probabilities

defined on small sets of points very similar to each other (called "microstates")

$$\rightarrow P_\alpha \leftarrow \frac{n}{\downarrow} \text{ microstates}$$

(n ~ 10^3 - 10^5 etc.)

$\rightarrow r_\alpha = \text{func}^c$ of time

$$\Rightarrow P_\alpha(t+1) = \underbrace{\sum_{\beta} \underbrace{P_{\beta \rightarrow \alpha}}_{\text{transition prob. from } \beta} \underbrace{P_\beta(t)}_{\text{existing in state } \beta}}_{\text{prob.}}$$

time lag b to =

2 subsegment observations $\equiv (x^2 = 1)$

→ For MD traj $\tau = 1 \text{ ps} - 1 \text{ ns}$

→ For finance $\tau = 1 \text{ ms}$ etc.

→ So without loss of generality, we can replace $(+t)$ by $(+\tau)$

EQ. 1

$$\therefore P_\alpha(t + \tau) = \sum_{\beta} K_{\beta \rightarrow \alpha} P_\beta(t)$$

$$\rightarrow K_{\beta \rightarrow \alpha} \geq 0 \quad (\text{non-symm sign or matrix})$$

$$\rightarrow \sum_{\beta} K_{\alpha \rightarrow \beta} = 1 \quad (\text{complex eigvals})$$

→ 2 eigen val eq^{2s} left & right:

$$\therefore \sum_{\alpha} K_{\alpha \rightarrow \beta} u_{\alpha}^{\gamma} = \lambda u_{\beta}^{\gamma}$$

(sum over first index) (left)

($\gamma \rightarrow$ labels diff eigvals & eigvecs)

$$R \therefore \sum_{\alpha} \kappa_{\beta \rightarrow \alpha} v_{\alpha}^{\gamma} = v_{\beta}^{\gamma}$$

(sum over 2nd index)

\Rightarrow In previous lecture we said that left & right eigenvs are for sure the same \because they are sol^{ns} of the same secular eqn.

\Rightarrow Moreover, in general left & right eigenvcs of a square matrix are mutually orthogonal

Therefore,

$$\sum_{\alpha} u_{\alpha}^{\gamma} v_{\alpha}^{\delta'} = \delta^{\gamma \delta'}$$

\therefore scalar part of left & right eigvec = 1

ONLY \therefore they correspond to the same eigval.

Eg. 2

\Rightarrow Other properties proved so far :-

① $\|\lambda^y\| \leq 1$ \rightarrow γ^y

(Gershgorin th.)

② At least one eigenvalue = 1

$$\left[\begin{matrix} \gamma^1 & = & 1 \\ \downarrow & & \end{matrix} \right]$$

(by convention,
considered the first
eig val)

+ the corresponding right eig vec :-

$$\boxed{\gamma^1 = 1} \rightarrow \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \rightarrow \text{vector of}$$

(implicit normalization
cond? of rate matrix)

+ the left eigenvector is P_α^{eqv}

stationary sol²
to our dynamic eqv?

③ If $K_{\alpha \rightarrow \beta} > 0$, & (α, β)

OR if K is irreducible

then we have a single eigenv = 1

i.e. $|\lambda^*| < 1$, & $\gamma > 1$

N.B.; ' K ' corresponds to "transition kernel" in the theory of Markov processes

⇒ We want to find an explicit sol² to $EQ. (1) \rightarrow$ master eq²

→ we take the same approach as in the case of Schrödinger's eq².

$$\therefore P_p(t) = \sum_{\gamma} c^*(t) u_{\beta}^{\gamma}$$

Ansatz sol² for Eq. 1

"u" is left eigenv,
"v" is right

\rightarrow we substitute this into Eq. (1) :-

\downarrow we obtain :-

$$\begin{aligned} & \underbrace{\sum_{\gamma} c^{\gamma}(t+\tau) u_2^{\gamma}}_{\gamma} \xrightarrow{\quad} p_2(t+\tau) \\ & = \sum_{\beta} \lambda_{\beta \rightarrow \alpha} \underbrace{\sum_{\gamma'} c^{\gamma'}(t) u_{\beta}^{\gamma'}}_{\downarrow} \quad \left(\text{summing over } \beta \text{ the left eigenv, } u_{\beta}^{\gamma'} \right) \left(p_{\beta}(t) \right) \\ & = \sum_{\gamma'} c^{\gamma'}(t) \lambda' u_2^{\gamma'} \end{aligned}$$

\rightarrow multiply both sides by (v_2^{γ}) &

$\underline{\text{sum over } \gamma : -} \quad (\text{right eigenv})$

$\underline{(\text{we want to exploit the mutual orthogonality of } g_u \text{ & } v)}$

$$\begin{aligned}
 & \sum_{\alpha} \sum_{\gamma} c^{\gamma}(t+\tau) u_{\alpha}^{\gamma} v_{\alpha}^{\gamma} \\
 \Rightarrow & \sum_{\alpha} \sum_{\gamma} c^{\gamma}(t) \lambda^{\gamma} u_{\alpha}^{\gamma} v_{\alpha}^{\gamma} \\
 & = (\sum_{\gamma} \lambda^{\gamma}) \rightarrow \boxed{\text{by Eq. 2}} \\
 & \quad \text{(summing over } \alpha \text{)}
 \end{aligned}$$

if then we also sum over (γ)
 & get

$$c^{\gamma}(t+\tau) = c^{\gamma}(t) \lambda^{\gamma}$$



\therefore We can compute coefficients of
 expansion of prob. at time = $(t+\tau)$
 if we know prob. at time = t , by
 multiplying term by the corresponding eigval.

$$\rightarrow \therefore C^Y(t) = C^Y(t-\tau) \lambda^\tau$$

$$= C^Y(t-2\tau) (\lambda^\tau)^2$$

we starte

this eq^o

till we

reach t=0

\therefore it's \approx

D.E. & has
an initial
cond²:

$$P_\alpha(0) = P_\alpha$$

(by iteration $\rightarrow \frac{t}{\tau}$ # of times)

$$= C^Y(0) (\lambda^\tau)^{t/\tau}$$

Remember : I.C. :

$$P_\alpha(0) = P_\alpha$$

Given I.C.

so, now we need to find
 $C^Y(0)$ & then we are happy

\therefore we can get coefficients
at any time if we start here

$$\rightarrow \text{I.r.C.} + P_\alpha(0) = \tilde{P}_\alpha$$

$$= \sum_\gamma C^\gamma(0) u_\alpha^\gamma$$

$$\therefore C^\gamma(0) = \sum_\lambda v_\lambda^\gamma \tilde{P}_\alpha$$

(\because this is just a projection at $t=0$)

\Rightarrow Now, let's denote $P_p(t)$ with these eq^{ns} :-

$$\therefore P_p(t) = \sum_\gamma C^\gamma(0)(\lambda^\gamma) u_\beta^\gamma$$

$$= \sum_\gamma \left(\sum_\lambda v_\lambda^\gamma \tilde{P}_\alpha \right) (\lambda^\gamma) u_\beta^\gamma$$

Implication : if in our line-ordered dataset we can estimate $K_{\alpha \rightarrow \beta}$, we can also predict $P_\alpha(t)$ [prob. at time t], for any initial condⁿ, \tilde{P}_α

\Rightarrow This is an exact set² :- if we choose I.C. that sp. is in microstate 3 we can predict what sp. will be in any time = t, only if we can estimate eigvals & eigvecs of $\hat{U}(t)$

For Schrodinger eq², there can ONLY be determined for very simple systems but in our case, we can do it for any (t)

\Rightarrow Let's see what happens for $\gamma = 1$:-

$$\gamma = 1, u^2 = P^{\text{ev}} \quad \rightarrow \quad U^1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \vdots \\ 1 \end{pmatrix} = \underline{1L}$$

$$\therefore C(0) = \sum_{\alpha} \tilde{P}_{\alpha} \cdot \underline{1L} = 1 \quad \rightarrow \tilde{P}$$

\downarrow
sum over a vector of probabilities
(normalization) $= 1$

Moreover $\lambda^1 = 1$

$$\therefore \left(\sum_{\alpha} v_{\alpha}^r \cdot p_{\alpha} \right) = 1$$

$$(\lambda^r)^{t/\kappa} = 1$$

$$v_p^r = p^{er}$$

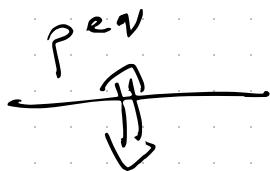
$$\therefore p_{\alpha}(t) = p^{er} + \sum_{r>1} c^r(0) (\lambda^r)^{t/\kappa} v_{\alpha}^r$$

Let's assume (3) holds i.e;

$\lambda_{\alpha \rightarrow p}$ is strictly <ve or
irreducible & $\therefore |\lambda| < 1$
& $r > 1$

Thus $|\lambda^r| < 1$ & $\lim_{t \rightarrow \infty} (\lambda^r)^{t/\kappa} = 0$

\Rightarrow This implies , if condition (3) holds
then: (3) for every possible \tilde{P}
 $(i.e., \tilde{P} \text{ initial cond})$
 the dynamics converges to



TNS justifies usage of MC / MD
 ∵ we sample a very high dim distrib =
s. that canonical distrib is
 start distrib for our system.

Also proves \downarrow \mathcal{Q}_1 → i.e., our
 solution is indeed "stable"

\rightarrow This is NOT true in general,
ONLY true if property (3) of strictly
 the $K_{\alpha > \beta}$ holds,

→ it's very easy to violate (3),
just by having a sys where we
break ergodicity & it is block
diagonal / reducible.

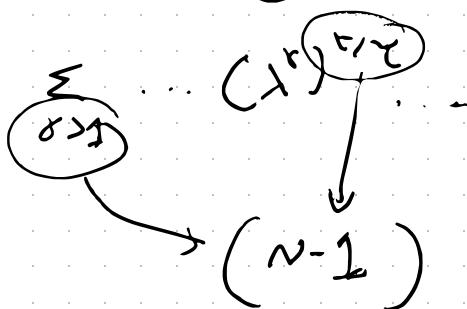
⇒ This technique allows us to perform
dim red^c by clustering

◻ Dimensional Reduction based on
time evolution properties :-

→ similar procedure to PCA / kPCA etc.

→ our system has $(N-1)$ relaxation

times



→ do all $(N-1)$ times really matter?

$\rightarrow \gamma^* \equiv \text{complex } \#$, $\therefore T_{\alpha \beta} \gamma^*$ is
non-symm.
symm. matrix.

$$\rightarrow (\gamma^*)^{t/\tau} = (|\lambda^*| e^{i\phi^*})^{t/\tau}$$

$$= |\lambda^*|^{t/\tau} e^{i\phi^* \frac{t}{\tau}}$$

NOT relevant for us, \because
we only want to

study relaxation to eq b $\frac{m}{\tau}$ & mat?

controlled by

$$|\lambda^*|^{t/\tau}$$

oscillating term

$$= \cos\left(\phi^* \frac{t}{\tau}\right)$$

$$+ i \sin\left(\phi^* \frac{t}{\tau}\right)$$

as said before (γ^r)
also comes with its
conj., so summing, we
only have real part.

$$\rightarrow [\lambda^\gamma]^t \chi = \exp(\log |\lambda^\gamma|^t \chi)$$

$\boxed{\tau = \text{time} - \text{lag}}$

$$= \exp\left(\frac{t}{\tau} \log |\lambda^\gamma|\right)$$

$$\rightarrow \frac{\tau}{\log |\lambda^\gamma|} = \{\text{time dimension}\}$$

$$\hookrightarrow \because |\lambda^\gamma| < 1, \log |\lambda^\gamma| = -ve$$

$$\rightarrow \underbrace{\therefore \text{we define}}_{T} \quad \tau^\gamma = \left(-\frac{\tau}{\log |\lambda^\gamma|} \right) > 0$$

relaxation time associated
to mode γ

$\therefore [|\lambda^\gamma|^t \chi] = \exp\left(-\frac{t}{\tau^\gamma}\right)$

$\boxed{\frac{\gamma > 1}{\downarrow}} \\ \text{remember}$

⇒ Role played by eigenvalues in PCA
is played here by these relaxation times.

⇒ if we have $\underline{N^2}$ microstates
we have $\underline{N-1}$ relaxation times

⇒ some of them will be equal ∵ they correspond to a conjugate pair, & so they count as ∞ .

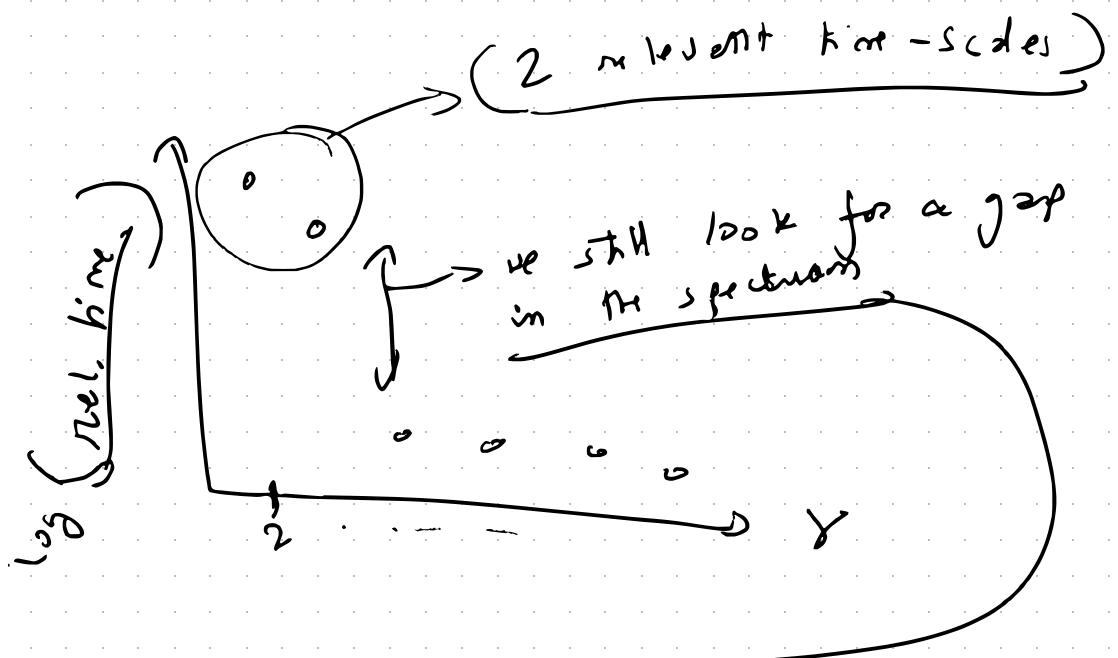
⇒ so we plot these relaxation times
in decreasing order i-

→ we start from $\gamma = \underline{2}$ ∵

$$\gamma = 1 \rightarrow \lambda' = \underline{1}$$

$$\therefore \log |\lambda'| = \underline{0} + \gamma' = \frac{1}{0} = \infty$$

infinite relaxation time corresponds to eg. ∞ distrib?



Unlike in P(A) finding gap in spectrum

of MSM is quite common, even in real world data. This is because number of relaxation times corresponds to $n - n_{\text{metastable states}}$ of the system + 1

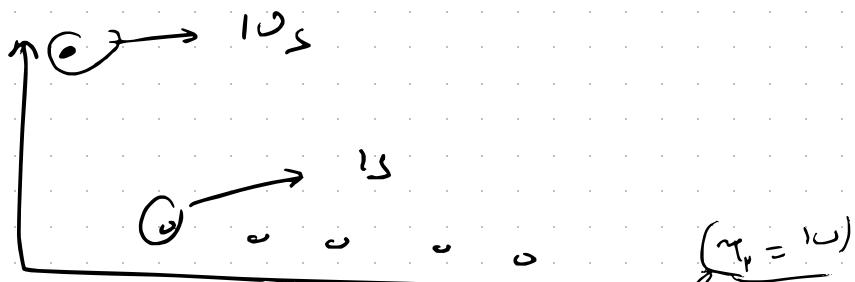
- many systems of relevance have metastable states & prob. densities that are not unimodal.
- N.B. → logarithmic scale is imp. in the spectrum.

→ A lot of physics can be described using relatively few degrees of freedom.

→ This is possible ∵ this gap exists

& practical physical systems are governed by few relevant timescales

→ Imagine if spectrum is like :-



$$\therefore P(t) = P^{eq} + c^1 e^{-\frac{t}{\tau_1}} u^1 + c^2 e^{-\frac{t}{\tau_2}} u^2 + \dots$$

\rightarrow if we are in range $t \sim (3-2\omega)$,

second term is already $(\sim e^{-3})$
 \downarrow
 (~ 0)

first term $\sim e^{-\frac{3}{10}} \sim 1$)

\rightarrow i.e., even if we have a gap of only

2 orders of magnitude we can
satisfactorily describe the dynamics
by truncating & considering only
the first relaxation time, error

is quite small

$$\therefore P_{\alpha}(t) \approx P_{\alpha}^{\text{ev}} + \left(\frac{u}{\omega} \right)^2 e^{-t/\tau_i}$$

(left eigenv)

→ ONLY γ^1 matters,

$$P_\alpha(t) = P_\alpha^{eq} + c e^{-\gamma^1 t}, \quad u_\alpha^1$$

must be orthogonal to v_α^0

$$\sum_\alpha u_\alpha^1 v_\alpha^0 = 0$$

$$\rightarrow \because v_\alpha^0 = 1$$

$$\therefore \sum_\alpha u_\alpha^1 = 0$$

true for all $\gamma > 1$

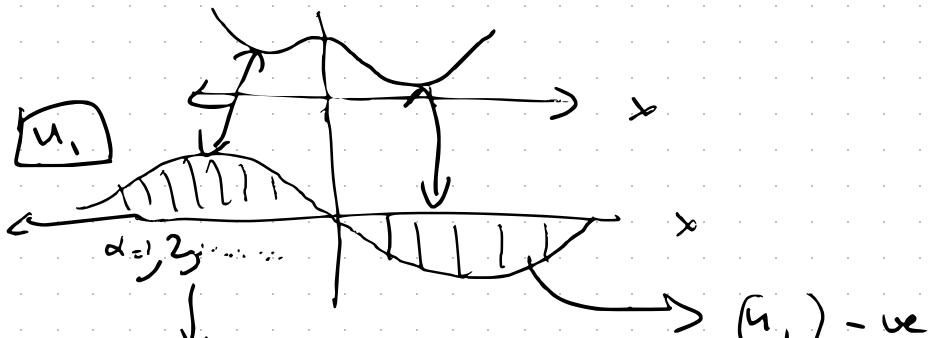
$$\therefore \sum_\alpha u_\alpha^\gamma = 0, \text{ if } \gamma \geq 1$$

for all possible $\gamma > 1$, eigenvectors (left) must sum to 0. (another weird property of stochastic matrices)

→ In general, the left eigenvectors of a stochastic matrix always sum to 0, except for eglo^m prob distrib⁼

⇒ This dynamics describes relaxation to 0 with (e^{-t/τ_i}) , for some part of my microstates associated to the (u')
some of " " " -ve (u')

⇒ Es: Langevin dynamics with 2 minima



(u') +ve belongs
to one minimum

belongs to
other minimum,

$\rightarrow \tau_{Lj} DR$ is the lowest relaxation time
 is associated to a transition $b \propto 2$
 subset of microstates, one in which
 (u_α') is +ve & other in which
 (u_α') is -ve .

$$\begin{aligned}
 A &: \left\{ \alpha ; \text{ such that } u_\alpha' < 0 \right\} \\
 B &: \left\{ \alpha ; \text{ such that } u_\alpha' > 0 \right\}
 \end{aligned}$$

These subsets are called "Markov States"

\rightarrow clusters of microstates (clusters) which
 have b common distⁿ of relaxing to
 eq b^m prob distribⁿ.

→ A markov state is a set of date points or configurations which relaxes to equilibrium in the same manner

i.e., whatever I.C. is, microstates belonging to state ' A ' always relaxes to erg A from the values of microstates belonging to state ' B ', always relaxes to erg B from -ve values.

∴ We performed in one dim. reduc 2 :

$R^o \rightarrow "n"$ → 1-2 Markov states
 10^{3-5} microstates

→ Instead of looking at prob. of a microstate, we look at prob. of the markov states.

$$\therefore P_A(t) = \sum_{\alpha \in A} P_\alpha(t)$$

$$+ P_B(t) = \sum_{\alpha \in B} P_\alpha(t)$$

$$\Rightarrow P_A(t) = P_A^{eq} + e^{-t/\tau_1} \Delta_A$$

$$\boxed{\Delta_A = \sum_{\alpha \in A} c' u_\alpha'}$$

(const.)

prob. of all
m-states $\in A$
relaxing to P^{eq}

\Rightarrow So we can describe the dynamics of our sys. as a jump from 1 m.s. to another.

+ it's prob. is determined by the leading relaxation time.

\Rightarrow if we have 2 relevant

relaxation times i -
 $u_\alpha^1 = \{1, -1, -1, +1\}$
 $u_\alpha^2 = \{-1, +1, +1, +1\}$

then eg. $B = [u_1^1, u_1^2, u_2^1, u_2^2]$

$A = \{u_\alpha^1 > 0, u_\alpha^2 > 0\}$ eg. J

$B = \{u_\alpha^1 > 0, u_\alpha^2 < 0\}$ how
the
signs
convention
of
eigenv
work,

$C = \{u_\alpha^1 < 0, u_\alpha^2 > 0\}$

$D = \{u_\alpha^1 < 0, u_\alpha^2 < 0\}$

\rightarrow 4 markov states

\rightarrow if there are $\leq d$ relevant relaxations

times then there are (2^d)
 $\therefore 2^d$ possible sign conventions

Markov States.

\rightarrow even if there are $\leq m.s.$ in practice (3)
is commonly observed 1 has very low
probability.

→ Key assumption made :-

discretized form of

continuous eq² $P^{t+1} = \int \dots$

$$\rightarrow P^{t+1} = \sum \dots -$$

4 this requires us to choose an appropriate number of microstates which must be large enough.

→ Remember that

We need to verify our results & there are techniques to ensure that we can trust our results.