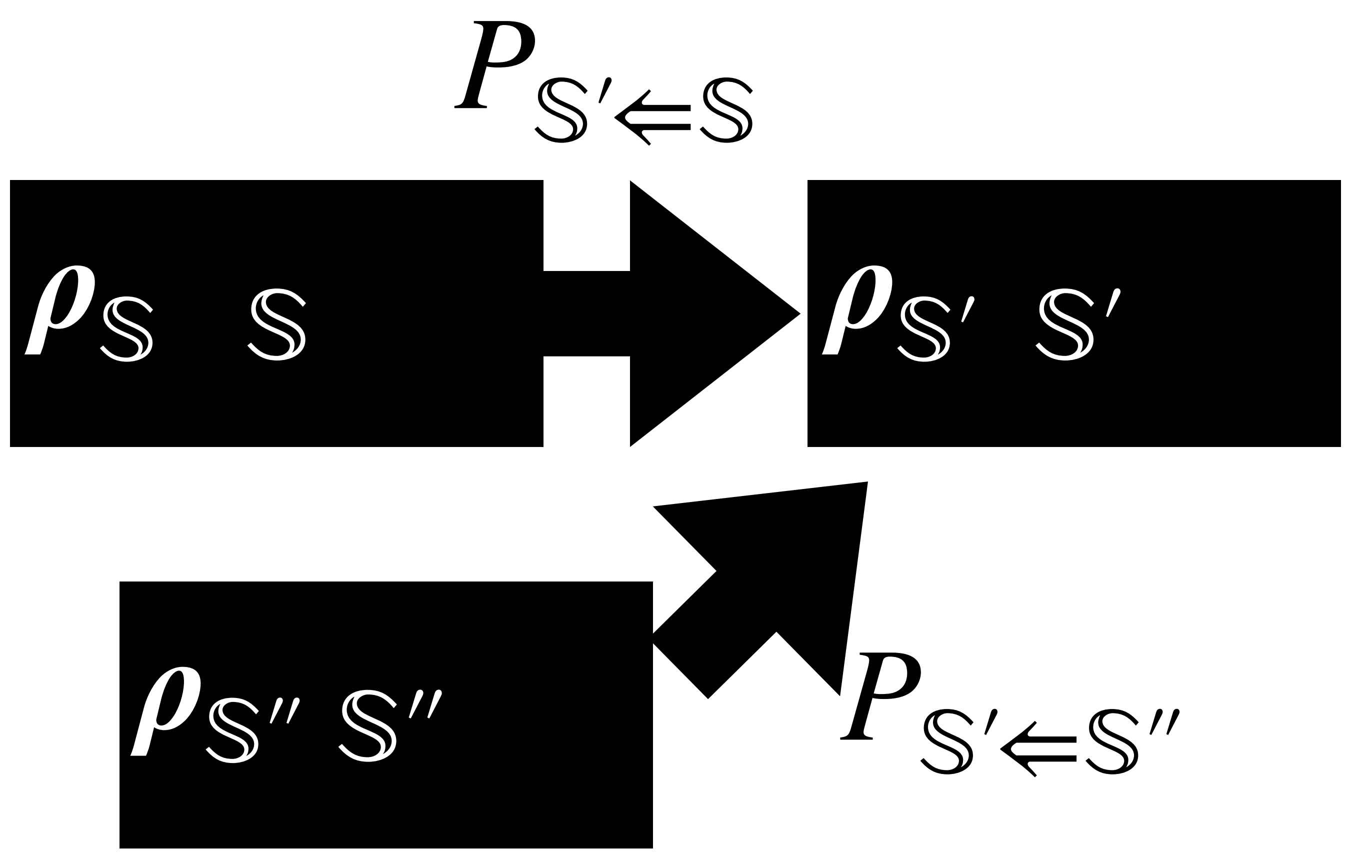
Markov chain theory in MD

Yongle Li

Based on Sethna, *Statistical Mechanics*

Setup

- A finite set of states $\{\alpha\}$
- Without memory
- Example: $N-$state Ising model, has 2^N states $\{s_i\}$
- Transfer probability: $P_{\mathbb{S}'\Leftarrow\mathbb{S}}$
- When at n_{th} step, the probability of the system sits at the α state is $\boldsymbol{\rho}_\alpha(n)$, then the next step, the probability of the system sits at the β state is:



$$\sum_\alpha P_{\beta\alpha}\boldsymbol{\rho}_\alpha(n) = \boldsymbol{\rho}_\beta(n+1) \qquad P_{\beta\alpha} \equiv P_{\beta\Leftarrow\alpha}$$

$$\sum_{\alpha} P_{\beta\alpha}\boldsymbol{\rho}_{\alpha}(n) = \boldsymbol{\rho}_{\beta}(n+1)$$

Expression with matrix element

$$P \cdot \boldsymbol{\rho}(n) = \boldsymbol{\rho}(n+1)$$

Expression with matrix form

What property needed of P?

1. Positivity $\quad 0 \leq P_{\beta\alpha} \leq 1$
2. Conservation of probability $\quad \sum_{\beta} P_{\beta\alpha} = 1$
3. Typically, it's not symmetric $\quad P_{\alpha\beta} \neq P_{\beta\alpha}$

These are the conditions for further usage!

Actually P here is *regular* or *positive*: with any $n \in \mathbb{N}$, $(P^n)_{\alpha\beta} > 0$

4. P here can be *diagonalizable*.[a] For each eigenvalue, there is at least one right eigenvector and one left eigenvector.

Right eigenvector of an eigenvalue:

$$P \cdot \boldsymbol{\rho}^{\lambda} = \lambda \boldsymbol{\rho}^{\lambda}$$

Left eigenvector of an eigenvalue:

$$\boldsymbol{\sigma}^{\lambda T} \cdot P = \lambda \boldsymbol{\sigma}^{\lambda T}$$

Counter example of not diagonalizable:
http://people.math.harvard.edu/~knill/teaching/math19b_2011/handouts/lecture33.pdf

a: https://www.math.wustl.edu/~freiwald/309markov.pdf

When equilibrium reached (1)

1. $P \cdot \boldsymbol{\rho}^* = \boldsymbol{\rho}^*$

Theorem 1: The matrix P has at least one right eigenvector $\boldsymbol{\rho}^*$ with eigenvalue one.

Proof: P has a left eigenvector $\boldsymbol{\sigma}^*$ with eigenvalue one — the vector all of whose components are one:

$$\boldsymbol{\sigma}^{*T} = (1,1,\cdots,1):\ (\boldsymbol{\sigma}^{*T} \cdot P)_\alpha = \sum_\beta \sigma^*_\beta P_{\beta\alpha} = \sum_\beta P_{\beta\alpha} = 1 = (\boldsymbol{\sigma}^{*T})_\alpha$$

Hence P must have an eigenvalue equal to one, and hence it must also have a right eigenvector with eigenvalue one.

Property 4

When equilibrium reached (2)

Theorem 2: Any right eigenvector $\boldsymbol{\rho}^\lambda$ with eigenvalue λ different from one must have components that sum to zero.

Proof: $\boldsymbol{\rho}^\lambda$ is a right eigenvector, $P \cdot \boldsymbol{\rho}^\lambda = \lambda \boldsymbol{\rho}^\lambda$.

Hence: $\lambda \sum_\beta \rho_\beta^\lambda = \sum_\beta (\lambda \rho_\beta^\lambda) = \sum_\beta \left(\sum_\alpha P_{\beta\alpha} \rho_\alpha^\lambda \right) = \sum_\alpha \left(\sum_\beta P_{\beta\alpha} \right) \rho_\alpha^\lambda = \sum_\alpha \rho_\alpha^\lambda$

So, there are only two probabilities: $\lambda = 1$, or $\sum_\alpha \rho_\alpha^\lambda = 0$

Ergodic Markov Chain

A finite-state Markov chain is ergodic if it does not have cycles and it is *irreducible*: that is, one can get from every state α to every other state β in a finite sequence of moves.

Theorem 3: (*Perron-Frobenius theorem*) Let A be a matrix with all non-negative matrix elements such that A^n has all positive elements. Then A has a positive eigenvalue λ_0, of multiplicity one, whose corresponding right and left eigenvectors have all positive components. Furthermore any other eigenvalue λ of A must be smaller, $|\lambda| < \lambda_0$.

Proof: http://people.math.harvard.edu/~knill/teaching/math19b_2011/handouts/lecture34.pdf

Corollary 1

With theorem 2 and theorem 3, we can see the largest eigenvalue of P is 1.

(since from theorem 2, all the eigenvectors with $\lambda^i < 1$ *have negative elements [* $\sum_{\alpha} \rho_{\alpha}^{\lambda_i} = 0$*], and from theorem 3,* λ_0 *with its eigenvector having all positive elements)*

Corollary 2

An ergodic Markov chain has a unique time-independent probability distribution ρ^*.

Corollary 3

If reached Detailed Balance

One can find a complete set of right eigenvectors for P.

Proof:

$$Q_{\alpha\beta} = P_{\alpha\beta}\sqrt{\frac{\rho^*_\beta}{\rho^*_\alpha}} = P_{\alpha\beta}\sqrt{\frac{\rho^*_\beta}{\rho^*_\beta}}\sqrt{\frac{\rho^*_\beta}{\rho^*_\alpha}}$$

$$= P_{\alpha\beta}\rho^*_\beta\frac{1}{\sqrt{\rho^*_\alpha\rho^*_\beta}}$$

Detailed Balance

$P_{\alpha\beta}\rho^*_\beta = P_{\beta\alpha}\rho^*_\alpha$

$$= P_{\beta\alpha}\rho^*_\alpha\frac{1}{\sqrt{\rho^*_\alpha\rho^*_\beta}} = P_{\beta\alpha}\sqrt{\frac{\rho^*_\alpha}{\rho^*_\beta}} = Q_{\beta\alpha}$$

Every symmetric matrix has a complete basis of eigenvectors.[a]

$$Q \cdot \boldsymbol{\tau}^\lambda = \lambda\boldsymbol{\tau}^\lambda$$

a: UC Berkeley EECS 223
Stochastic Systems: Estimation and Control
https://people.eecs.berkeley.edu/~ananth/223Spr07/ee223spr07_lec8.pdf

Proof of Corollary 3 (Cont.)

$$\sum_{\alpha} P_{\beta\alpha}\tau^{\lambda}_{\alpha}\sqrt{\rho^{*}_{\alpha}} = \sum_{\alpha}\left(Q_{\beta\alpha}\sqrt{\frac{\rho^{*}_{\beta}}{\rho^{*}_{\alpha}}}\right)\left(\tau^{\lambda}_{\alpha}\sqrt{\rho^{*}_{\alpha}}\right)$$

$$= \sum_{\alpha}\left(Q_{\beta\alpha}\tau^{\lambda}_{\alpha}\sqrt{\rho^{*}_{\beta}}\right) = \lambda\left(\tau^{\lambda}_{\beta}\sqrt{\rho^{*}_{\beta}}\right)$$

$$\Rightarrow \quad \rho^{\lambda}_{\alpha} = \tau^{\lambda}_{\alpha}\sqrt{\rho^{*}_{\alpha}}$$

So, $\{\rho^{\lambda}\}$ also form a complete set.

Detailed balance

1. If there is some probability distribution $\boldsymbol{\rho}^*$ satisfying:

   $$P_{\alpha\beta}\rho^*_\beta = P_{\beta\alpha}\rho^*_\alpha$$

   for each state α and β.

2. Detailed balance ⇔ global balance.

   Proof:

$$\sum_\beta P_{\alpha\beta}\rho^*_\beta = \sum_\beta P_{\beta\alpha}\rho^*_\alpha$$

$$\sum_\beta P_{\alpha\beta}\rho^*_\beta = \rho^*_\alpha \sum_\beta P_{\beta\alpha}$$

$$\sum_\beta P_{\alpha\beta}\rho^*_\beta = \rho^*_\alpha 1$$

$$\Rightarrow \boldsymbol{P}\cdot\boldsymbol{\rho}^* = \boldsymbol{\rho}^*$$

Reverse derivation needs:

1) the chain *irreducible* *Proof omitted. Can be left as homework!*

or

2) with initial distribution is π and the process is *reversible*.

Irreducible:

One can get from every state α to every other state β in a finite sequence of moves.

Reversible:

A reverse chain $\{Y_k\},\ Y_k = X_{n-k}$ is also with detailed balance.

Proof of reversible induce the detailed balance:

$$P(Y_{k=i} | Y_{k+1} = j) = P(X_{n-k=i} | X_{n-k+1=j})$$ Conditional probability[a]

$$= \frac{P(X_{n-k=i}, X_{n-k+1=j})}{P(X_{n-k+1=j})}$$ Joint probability

$$= \frac{P_{ji}\pi_i}{\pi_j}$$

$$P(X_{k=i} | X_{k-1=j}) = P_{ij}$$

So if the chain is reversible and with equilibrium distribution, there is detailed balance.

a: https://setosa.io/conditional/

Main theorem

Theorem 4: A discrete dynamical system with a finite number of states can be guaranteed to converge to an equilibrium distribution ρ^* if the computer algorithm:

1. is Markovian (has no memory)
2. is ergodic (can reach everywhere and is acyclic)
3. satisfies detailed balance.

Proof: detailed balance $\Leftrightarrow P$ has a complete set of eigenvectors $\boldsymbol{\rho}^{\lambda}$. Since our algorithm is ergodic there is only one right eigenvector $\boldsymbol{\rho}^{1}$ with eigenvalue one, which we can chose to be the stationary distribution $\boldsymbol{\rho}^{*}$; all the other eigenvalues λ have $|\lambda| < 1$. Decompose the initial condition $\boldsymbol{\rho}(0) = a_1\boldsymbol{\rho}^{*} + \sum_{|\lambda|<1} a_\lambda \boldsymbol{\rho}^{\lambda}$. Then:

$$
\begin{aligned}
\boldsymbol{\rho}(n) &= P^n \cdot \boldsymbol{\rho}(0) \\
&= a_1 P^n \boldsymbol{\rho}^{*} + \sum_{|\lambda_i|<1} a_\lambda P^n \boldsymbol{\rho}^{\lambda} \\
&= a_1 \boldsymbol{\rho}^{*} + \sum_{|\lambda<1|} a_\lambda \lambda^n \boldsymbol{\rho}^{\lambda}
\end{aligned}
$$

$$\lim_{n\to\infty} \boldsymbol{\rho}(n) = \boldsymbol{\rho}^{*}$$