An introduction to

Stochastic processes and applications

7CCMCS04

G. Sicuro

Abstract. These lecture notes are largely based on a previous version prepared by A. Annibale, for the 7CCMCS04 – *Dynamical analysis of complex systems* module at King's College London, part of the *Complex Systems Modelling* master up to 2021. The notes have been reviewed, rearranged and re-edited by myself in 2022.

Gabriele Sicuro
London, March 6, 2022

Contents

CHAPTER 1

Discrete-time Markov processes

ABSTRACT. In this first chapter we will introduce the main subject of the module, namely stochastic processes. We will focus then on a specific kind of process, discrete-time Markov chains. Markov chains are prototypical Markovian stochastic processes that allow isolating some crucial properties of Markovian processes. Our analysis will prepare the study of Markov processes in continuous time, developed in the next chapter. As a possible reference, see Chapter 6 of
GRIMMET and STIRZAKER, *Probability and random processes*, Oxford University Press, 2020.
A monograph dedicated to Markov chains is
NORRIS, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, 1998.

1. Stochastic processes: fundamental definitions

LECTURE 1

A stochastic process (SP) can be informally defined as a collection of random variables $(\mathsf{X}_t)_t$ taking value in some set Ω, called sample space. The elements in the collection depend on a label $t \in \mathcal{T}$, which we can think plays the role of "time". We will assume that $\mathcal{T} \subseteq \mathbb{R}$. The set Ω can be a *discrete* set (for example, X_t can represent the number of people in line at a post office at time t) or *continuous* (e.g., X_t might be how much time we have to wait to arrive at the post office at time t). For simplicity, we will start considering the case in which $\Omega \subseteq \mathbb{R}$, but it is easy to generalise the definitions below to higher dimensional cases, e.g., $\Omega \subseteq \mathbb{R}^d$ with $d > 1$. We will denote X_t the stochastic process and $x(t)$ an actual realisation of it.

The most basic classification of stochastic processes usually refers to the set $\mathcal{T}$ of possible values of the "time" variables. We distinguish (see Fig. 1)

- *discrete-time stochastic processes*, if $\mathcal{T}$ is a finite or discrete set (for example, $t \in \mathbb{N}$);
- *continuous-time stochastic processes*, if $\mathcal{T}$ is an interval (or the union of intervals) of $\subseteq \mathbb{R}$.

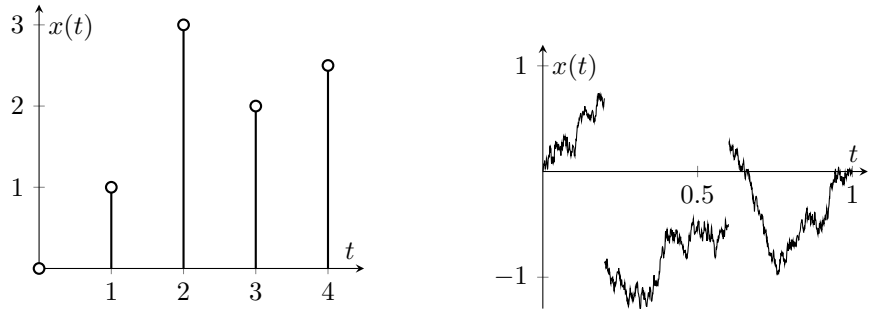


FIGURE 1. Example of a graph of discrete-time (left) and continuous-time (right) stochastic process. Observe that a continuous-time stochastic process is not necessarily continuous.

The time-dependence of the random variable X_t naturally suggests introducing, in the case of discrete valued SPs, the *joint probability*

$$\mathbb{P}_n[\mathsf{X}_{t_1} = x_1; \dots; \mathsf{X}_{t_n} = x_n] \equiv \mathbb{P}_n[\{\mathsf{X}_{t_i} = x_i\}_{i=1}^n] \tag{1.1}$$

that the SP X takes value $x_i \in \Omega$ at time t_i for $i = 1, \dots, n$. These joint probabilities satisfy

(1) $\mathbb{P}_n[\{\mathsf{X}_{t_i} = x_i\}_i] \geq 0$;
(2) $\sum_{x_j \in \Omega} \mathbb{P}_n[\{\mathsf{X}_{t_i} = x_i\}_i] = \mathbb{P}_{n-1}[\{\mathsf{X}_{t_i} = x_i\}_{i \neq j}]$;
(3) $\sum_{\{x_i\}_i} \mathbb{P}_n[\{\mathsf{X}_{t_i} = x_i\}_i] = 1$.

Similarly, in the case of continuous-values SP, we can introduce a *joint probability density*

$$p_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n) \equiv p_n(\{(x_i, t_i)\}_{i=1}^n) \tag{1.2}$$

defined so that

$$p_n(\{(x_i, t_i)\}_{i=1}^n) \prod_{i=1}^n \mathrm{d}\, x_i = \mathbb{P}[\{\mathsf{X}_{t_i} \in (x_i, x_i + \mathrm{d}\, x_i)\}_{i=1}^n]. \tag{1.3}$$

Such densities clearly satisfy

(1) $p_n(\{(x_i, t_i)\}_i) \geq 0$;
(2) $\int p_n(\{(x_i, t_i)\}_i)\, \mathrm{d}\, x_j = p_{n-1}(\{(x_i, t_i)\}_{i \neq j})$;
(3) $\int p_n(\{(x_i, t_i)\}_i) \prod_{i=1}^n \mathrm{d}\, x_i = 1$.

Averages of quantities involving an SP will in general depend on one or more values of t. Let us focus on continuous-valued SPs: similar formulas can be obtained for discrete-valued SPs just by replacing probability densities with probabilities and $\int \mathrm{d}\, x \to \sum_x$. So for example, the expectation of a function $f \colon \Omega \to \mathbb{R}$ of X_t is in general time-dependent as

$$\mathbb{E}[f(\mathsf{X}_t)] := \int_\Omega f(x) p_1(x,t) \mathrm{d}x \quad \text{and in particular} \quad \mathbb{E}[\mathsf{X}_t] := \int_\Omega x p_1(x,t) \mathrm{d}x. \tag{1.4}$$

More generally, we define the nth moment of the SP X_t as

$$\mathbb{E}\left[\prod_{i=1}^n \mathsf{X}_{t_i}\right] = \left[\prod_{i=1}^n \int_\Omega \mathrm{d}\, x_i\, x_i\right] p_n(\{(x_i, t_i)\}_i) \tag{1.5}$$

The $n = 2$ case is particularly relevant and called *autocorrelation function*:

$$C(t_1, t_2) = \mathbb{E}[\mathsf{X}_{t_1} \mathsf{X}_{t_2}] = \iint_{\Omega \times \Omega} \mathrm{d}\, x_1\, \mathrm{d}\, x_2\, x_1 x_2\, p_2(x_1, t_1; x_2, t_2) \tag{1.6}$$

The prefix *auto* is often used to distinguish from the case in which the correlation is computed between two different SPs, e.g., considering $\mathbb{E}[\mathsf{X}_{t_1} \mathsf{Y}_{t_2}]$ with Y different SP. If X_{t_1} and X_{t_2} are independent, then $\mathbb{E}[\mathsf{X}_{t_1} \mathsf{X}_{t_2}] = \mathbb{E}[\mathsf{X}_{t_1}] \mathbb{E}[\mathsf{X}_{t_2}]$, hence a good way to measure correlations between X_{t_1} and X_{t_2} is the *connected correlator* which subtract this contribution,

$$\langle\!\langle \mathsf{X}_{t_1} \mathsf{X}_{t_2} \rangle\!\rangle := \mathbb{E}[\mathsf{X}_{t_1} \mathsf{X}_{t_2}] - \mathbb{E}[\mathsf{X}_{t_1}] \mathbb{E}[\mathsf{X}_{t_2}] \tag{1.7}$$

and is zero by construction when X_{t_1} and X_{t_2} are independent. A positive value of the connected correlator means that when X_{t_1} is above (below) its average value, X_{t_2} tends to be above (below) its average as well. Conversely, X_{t_1} and X_{t_2} are negatively correlated if they tend to have opposite trends. For $t_1 = t_2$ the connected correlator reduces to the time-dependent variance $\langle\!\langle \mathsf{X}_t^2 \rangle\!\rangle =: \sigma_{\mathsf{X}}^2$.

Generating functions. — We can proceed generalising the concept of generating function and characteristic function. Let us start with a reminder.

If X is a random variable, then its generating function is defined as

$$G_{\mathsf{X}}(s) \coloneqq \mathbb{E}[s^{\mathsf{X}}]. \tag{1.8}$$

This function has interesting properties. For example, if X and Y are independent random variables, $\mathsf{Z} = \mathsf{X} + \mathsf{Y}$ has

$$G_{\mathsf{Z}}(s) \coloneqq \mathbb{E}[s^{\mathsf{X}+\mathsf{Y}}] = \mathbb{E}[s^{\mathsf{X}}]\mathbb{E}[s^{\mathsf{Y}}] = G_{\mathsf{X}}(s)G_{\mathsf{Y}}(s). \tag{1.9}$$

A special type of generating function is the *moment generating function*

$$M_{\mathsf{X}}(k) \coloneqq G_{\mathsf{X}}(\mathrm{e}^{u}) = \mathbb{E}[\mathrm{e}^{u\mathsf{X}}], \tag{1.10}$$

which is such that $\partial_u^n M_{\mathsf{X}}(u)|_{u=0} = \mathbb{E}[\mathsf{X}^n]$ and allows therefore to easily compute all moments of X (hence the name). The function M_{X} does not always exist, and it is therefore convenient to consider, instead, the *characteristic function*

$$\varphi_{\mathsf{X}}(u) \coloneqq G_{\mathsf{X}}(\mathrm{e}^{iu}) = \mathbb{E}[\mathrm{e}^{iu\mathsf{X}}], \tag{1.11}$$

which is instead always well-defined. Also in this case, the (possibly infinite) moments are obtained by derivation $\partial_u^n \varphi_{\mathsf{X}}(u)|_{u=0} = i^n \mathbb{E}[\mathsf{X}^n]$. Finally, the *cumulant generating function* is defined by taking the logarithm of φ_{X},

$$K(u) \coloneqq \ln \varphi_{\mathsf{X}}(u) = \ln \mathbb{E}[\mathrm{e}^{iu\mathsf{X}}]. \tag{1.12}$$

The quantities

$$\langle\!\langle \mathsf{X}^n \rangle\!\rangle \coloneqq i^{-n} \partial_u^n K(u)|_{u=0} \tag{1.13}$$

are called *cumulants* of X.

As an SP depends on time, to generalise the concept of characteristic generating function we will replace the scalar u with test function $u(t)$ and the product $u\mathsf{X}$ with $\int_{\mathcal{T}} u(t)\mathsf{X}_t \,\mathrm{d}\, t$. The characteristic generating function will become a *characteristic generating functional*.

The characteristic generating *functional* is defined as the following functional of $u(t)$

$$\varphi_{\mathsf{X}}[u] = \mathbb{E}\left[\mathrm{e}^{i\int_{\mathcal{T}} u(t)\mathsf{X}_t \,\mathrm{d}\, t}\right]. \tag{1.14}$$

The notation $\varphi_{\mathsf{X}}[u]$ emphasizes that φ_{X} depends on the *whole function* u. If we expand in powers of u, we get

$$\varphi_{\mathsf{X}}[u] = \sum_{n=0}^{\infty} \frac{i^n}{n!} \left[\prod_{i=1}^{n} \int u(t_i) \,\mathrm{d}\, t_i\right] \mathbb{E}\left[\prod_{i=1}^{n} \mathsf{X}_{t_i}\right] \tag{1.15}$$

Each moment of X_t can be therefore recovered as coefficient such expansion taking a *functional derivative*

$$\mathbb{E}[\mathsf{X}_{t_1} \cdots \mathsf{X}_{t_k}] = \frac{\delta^k \varphi_{\mathsf{X}}[u]}{\delta u(t_1) \cdots \delta u(t_k)}\Big|_{u=0}. \tag{1.16}$$

Similarly, we can define a *cumulant generating functional* as

$$K_{\mathsf{X}}[u] \coloneqq \ln \varphi_{\mathsf{X}}[u] = \sum_{n=0}^{\infty} \frac{i^n}{n!} \left[\prod_{i=1}^{n} \int u(t_i) \,\mathrm{d}\, t_i\right] \left\langle\!\!\left\langle \prod_{i=1}^{n} \mathsf{X}_{t_i} \right\rangle\!\!\right\rangle \tag{1.17}$$

It is easy to see that the connected autocorrelation function is indeed the cumulant $\langle\!\langle \mathsf{X}_t \mathsf{X}_{t'} \rangle\!\rangle$ appearing in this expansion, so that

$$\langle\!\langle \mathsf{X}_t \mathsf{X}_{t'} \rangle\!\rangle = -\frac{\delta^2 K_{\mathsf{X}}[u]}{\delta u(t)\, \delta u(t')}. \tag{1.18}$$

1.1. Stationary processes. Knowing all p_n of some SP X_t amounts to know the process itself. Depending on the properties of the different p_n, we can attempt a classification of SPs. A particular important family is one of *stationary processes*. A SP is *stationary* if for all p_n we have

$$p_n(\{(x_i, t_i + \tau)\}_{i=1}^n) = p_n(\{(x_i, t_i)\}_{i=1}^n) \tag{1.19}$$

i.e., shifting *all the times* does not change the joint probability densities. We say then that X_t and $\mathsf{X}_{t+\tau}$ have the same distribution, writing $\mathsf{X}_t \overset{\mathrm{d}}{\sim} \mathsf{X}_{t+\tau}$. It follows in particular that $p_1(x,t) \equiv p_1(x)$, and therefore $\mathbb{E}[\mathsf{X}_t] \equiv \mathbb{E}[\mathsf{X}]$ is time-independent. Similarly,

$$p_2(x_1, t_1; x_2, t_2) = p_2(x_1, 0; x_2, t_2 - t_1), \tag{1.20}$$

i.e., it depends on the time difference $t_2 - t_1$ only, so that

$$C(t + \tau, t) = \mathbb{E}[\mathsf{X}_{t+\tau}\mathsf{X}_t] \equiv C(\tau) = C(-\tau), \tag{1.21}$$

i.e., the autocorrelation is expressed in terms of a symmetric one-variable function $C(\tau) = C(-\tau)$: the important quantity is the *time distance* between the two observations $\mathsf{X}_{t+\tau}$ and X_t, not which one took place first or when they took place. If there exists a value τ_C such that $C(\tau) \simeq 0$ for $\tau > \tau_C$, then τ_c is called *autocorrelation time of the stationary SP*. We briefly mention here that, if $\mathbf{X}_t = (\mathsf{X}_t^\mu)_{\mu=1}^d$ is a stationary, multicomponent SP, the autocorrelation function is replaced with the correlation matrix $C_{\mu\nu}(\tau) = \mathbb{E}[\mathsf{X}_t^\mu \mathsf{X}_{t+\tau}^\nu]$, where the diagonal elements represent autocorrelations and off-diagonal elements are cross-correlations. In this case $C_{\mu\nu}(\tau) = C_{\nu\mu}(-\tau)$.

Suppose that we have a SP X_t. Its *energy* is defined as

$$\mathsf{E} \coloneqq \int_{-\infty}^{+\infty} \mathsf{X}_t^2 \,\mathrm{d}\, t \tag{1.22}$$

If we define the Fourier transform of the process

$$\hat{\mathsf{X}}_\omega \coloneqq \int_{-\infty}^{+\infty} \mathsf{X}_t \,\mathrm{e}^{-i\omega t} \,\mathrm{d}\, t \tag{1.23}$$

then Parseval's identity tells us that

$$\mathsf{E} \coloneqq \int_{-\infty}^{+\infty} \mathsf{X}_t^2 \,\mathrm{d}\, t = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{\mathsf{X}}_\omega|^2 \,\mathrm{d}\, \omega. \tag{1.24}$$

The quantity $|\hat{\mathsf{X}}_\omega|^2$ has the meaning of *energy density* in frequency, corresponding to the interval $(\omega, \omega + \mathrm{d}\,\omega)$. However, it is hardly possible to work on the entire line, so a more accessible quantity is

$$\hat{\mathsf{X}}_\omega^{(T)} \coloneqq \int_{-T}^{+T} \mathsf{X}_t \,\mathrm{e}^{-i\omega t} \,\mathrm{d}\, t \quad \text{and therefore} \quad |\hat{\mathsf{X}}_\omega^{(T)}|^2 \quad \text{in place of} \quad |\hat{\mathsf{X}}_\omega|^2. \tag{1.25}$$

Assume now that X_t is stationary. Its (rescaled) expected value is called *power spectrum*, and is defined as

$$S(\omega) \coloneqq \lim_{T\to 0} \frac{\mathbb{E}[|\hat{\mathsf{X}}_\omega^{(T)}|^2]}{2T} = \lim_{T\to 0} \frac{1}{2T} \int_{-T}^{+T} \mathrm{d}\, t \int_{-T}^{+T} \mathrm{d}\, t' \,\mathbb{E}[\mathsf{X}_t \mathsf{X}_{t'}] \,\mathrm{e}^{-i(t-t')\omega} = \int_{-\infty}^{+\infty} C(\tau) \,\mathrm{e}^{i\tau\omega} \,\mathrm{d}\, \omega. \tag{1.26}$$

The relation between $S(\omega)$ and $C(\tau)$ is the so-called *Wiener–Khinchin theorem*.

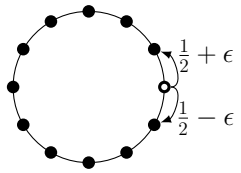
Equilibrium processes. — Equilibrium processes are stationary processes that also satisfy *dynamical reversibility*, i.e.,

$$\mathsf{X}_t \overset{\mathrm{d}}{\sim} \mathsf{X}_{T-t} \quad \forall T \in \mathbb{R}^+, \tag{1.27}$$

where $\overset{\mathrm{d}}{\sim}$ stresses equality in distribution. In other words, a forward trajectory is as likely as a backward one through the same values of the SP.

LEMMA 1.1. *All equilibrium processes are stationary.*

PROOF. It is enough to observe that $\mathsf{X}_t \overset{\mathrm{d}}{\sim} \mathsf{X}_{T-t}$ but also, for any τ, $\mathsf{X}_{t+\tau} \overset{\mathrm{d}}{\sim} \mathsf{X}_{-t}$ (it is enough to choose $T = \tau$ in the definition above). Therefore $\mathsf{X}_t \overset{\mathrm{d}}{\sim} \mathsf{X}_{t+\tau}$ for any τ. □



Note that the converse of the previous lemma is not true. To appreciate the difference between equilibrium and stationarity consider a system of one particle hopping between the adjacent sites of a lattice ring with N sites, $\{1, \dots, N\}$, *starting at random in one of the N sites at $t = 0$*. Its position at time t will be our SP X_t. The particle hops clockwise with probability $1/2 - \epsilon$ and counterclockwise with probability $1/2 + \epsilon$. The hopping probability does not depend on time t and it is evident that the starting time of the experiment is not relevant: the process is stationary. However, $\mathbb{P}[\mathsf{X}_t = i; \mathsf{X}_{t+1} = i+1] \neq \mathbb{P}[\mathsf{X}_t = i+1; \mathsf{X}_{t+1} = i]$: the process is not an equilibrium process unless $\epsilon = 0$ and, if for example $\epsilon > 0$, the particle will *mostly* flow in the counterclockwise direction.

Ergodic processes. — In practical applications, averaging over all possible realisations of an SP X, is simply not possible. If $x(t)$ is the outcome of an experiment corresponding to the SP X_t, this would require indeed to run the experiment a large number of times to average over all $x(t)$. Instead, the *time average* of the single instance $x(t)$ of X_t can be more easily accessible. For an important class of stationary processes, i.e., *ergodic processes*, such information (if available for long enough times) is enough to recover the averages over the measure of X_t. To be more precise, suppose that we compute

$$\overline{x(t)} := \lim_{T\to+\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\mathrm{d}t, \qquad \overline{x(t)x(t+\tau)} := \lim_{T\to+\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)\mathrm{d}t. \tag{1.28}$$

Then, for an ergodic process, the following equalities hold:

$$\overline{x(t)} = \mathbb{E}[\mathsf{X}_t], \qquad \overline{x(t)x(t+\tau)} = \mathbb{E}[\mathsf{X}_t\mathsf{X}_{t+\tau}] = C(\tau). \tag{1.29}$$

In other words, in ergodic processes, the ensemble average is equivalent to the time average over an arbitrarily large time T, which is then allowed to become infinite. Intuitively, this suggests that in ergodic processes if we partition a trajectory $x(t)$ in time intervals that are *long enough*, the obtained "pieces" have the same properties of different sampling of the SP X_t, see Fig. 2 for an example.

2. Finite Markov chains

LECTURE 2

2.1. Markov processes. Suppose that the SP X_t takes values in a finite or countable space Ω. We might assume that we know the value it takes in our observation at some times $t_1, \dots, t_s$,

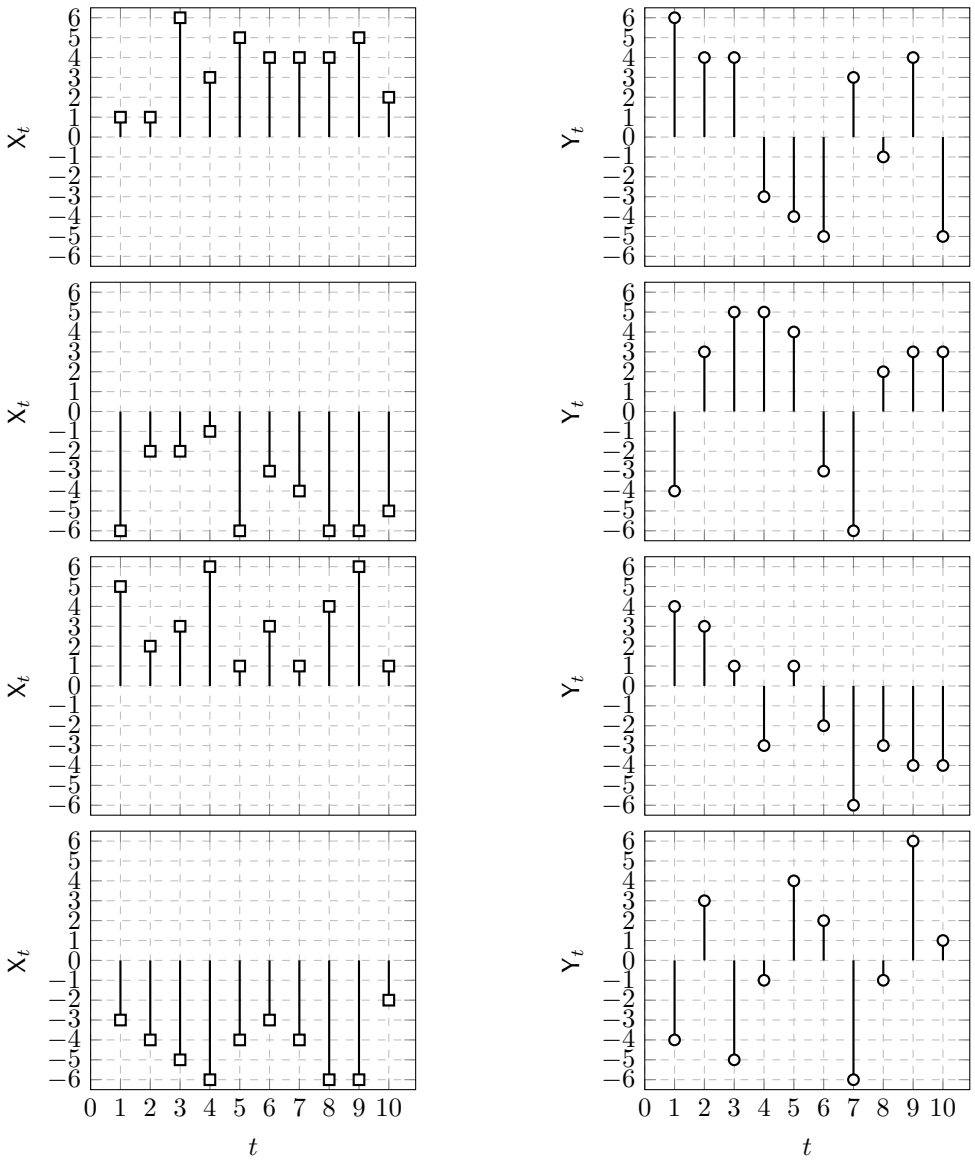


FIGURE 2. We represent here some realisations of two stationary processes, X_t and Y_t, in discrete time $t \in \mathbb{N}$. Both these processes are constructed from two other *independent and stationary* processes: a fair-coin toss, $\mathsf{C}_t \in \{-1, 1\}$, where we associated 1 to *head* and -1 to *tail*, and a fair-dice toss $\mathsf{D}_t \in \{1, 2, \dots, 6\}$. The process X_t is obtained as $\mathsf{X}_t = \mathsf{C}_1\mathsf{D}_t$: we toss a coin at the very beginning, and we assign a sign, depending on the outcome, to a sequence of dice toss. The process Y_t instead is obtained by tossing the coin at each dice toss to decide, at each step, the sign, so $\mathsf{Y}_t = \mathsf{C}_t\mathsf{D}_t$. It is intuitive that X_t is *not* ergodic. For example, $\mathbb{E}[\mathsf{X}_t] = \mathbb{E}[\mathsf{D}_t]\mathbb{E}[\mathsf{C}_t] = 0$, whilst on a single trajectory $\overline{\mathsf{X}_t} = \pm 7/2$, with sign depending on the realisation. On the other hand, Y_t is ergodic: for example, $\mathbb{E}[\mathsf{Y}_t] = \overline{\mathsf{Y}_t} = 0$.

let us say that $\mathsf{X}_{t_i} = x_i$ for $i = 1, \dots, s$. Using Bayes theorem, we can write the *conditional probability* that the SP will take values $\mathsf{X}_{t_i} = x_i$ for $i = s+1, \dots, n$ as

$$\mathbb{P}_{n-s|s}[\{\mathsf{X}_{t_i} = x_i\}_{i=s+1}^{n} | \{\mathsf{X}_{t_i} = x_i\}_{i=1}^{s}] = \frac{\mathbb{P}_n[\{\mathsf{X}_{t_i} = x_i\}_{i=1}^{n}]}{\mathbb{P}_s[\{\mathsf{X}_{t_i} = x_i\}_{i=1}^{s}]}. \tag{1.30}$$

Directly from Bayes theorem, we have that, for any time t' we have in particular

$$\mathbb{P}_1[\mathsf{X}_t = x] = \sum_{y \in \Omega} \mathbb{P}_{1|1}[\mathsf{X}_t = x | \mathsf{X}_{t'} = y] \mathbb{P}_1[\mathsf{X}_{t'} = y]. \tag{1.31}$$

Assuming now $t_1 < t_2 < \cdots < t_n$, a *Markov process* satisfies the following *Markov property*

$$\mathbb{P}_{1|n-1}[\mathsf{X}_{t_n} = x_n | \{\mathsf{X}_{t_i} = x_i\}_{i=1}^{n-1}] = \mathbb{P}_{1|1}[\mathsf{X}_{t_n} = x_n | \mathsf{X}_{t_{n-1}} = x_{n-1}]. \tag{1.32}$$

The Markov property expresses that for a Markov process, the conditioned probability of transition depends only on the last known value of X and not on the previous history of the system. In other words, if the present state of the system is known, we can determine the probability of any future state without reference to the past. We may say that Markov processes have "short" memory, although the attribute "memoryless" is often used.

Interestingly, in Markovian processes only $\mathbb{P}_{1|1}[\mathsf{X}_t = x | \mathsf{X}_{t'} = y]$ and $\mathbb{P}_1[\mathsf{X}_t = x]$ matters: this can be seen observing that, for any $n \geq 2$, $\mathbb{P}_n[\{\mathsf{X}_{t_i} = x_i\}_{i=1}^{n}]$ can be written in terms of this two quantities. For example, for $n = 3$, assuming $t_1 < t_2 < t_3$, we have

$$\begin{aligned} \mathbb{P}_3[\mathsf{X}_{t_1}{=}x_1; \mathsf{X}_{t_2}{=}x_2; \mathsf{X}_{t_3}{=}x_3] &= \mathbb{P}_{1|2}[\mathsf{X}_{t_3}{=}x_3 | \mathsf{X}_{t_1}{=}x_1; \mathsf{X}_{t_2}{=}x_2] \mathbb{P}_2[\mathsf{X}_{t_1}{=}x_1; \mathsf{X}_{t_2}{=}x_2] \\ &= \mathbb{P}_{1|1}[\mathsf{X}_{t_3}{=}x_3 | \mathsf{X}_{t_2}{=}x_2] \mathbb{P}_{1|1}[\mathsf{X}_{t_2}{=}x_2 | \mathsf{X}_{t_1}{=}x_1] \mathbb{P}_1[\mathsf{X}_{t_1}{=}x_1], \end{aligned} \tag{1.33}$$

and, for general n, assuming $t_1 < t_2 < \cdots < t_n$,

$$\mathbb{P}_n[\{\mathsf{X}_{t_i} = x_i\}_{i=1}^{n}] = \left(\prod_{i=2}^{n} \mathbb{P}_{1|1}[\mathsf{X}_{t_i} = x_i | \mathsf{X}_{t_{i-1}} = x_{i-1}] \right) \mathbb{P}_1[\mathsf{X}_{t_1} = x_1]. \tag{1.34}$$

The relation (1.33) can be useful to obtain an important equation for Markov processes. Let us integrate it over x_2 and divide both sides by $\mathbb{P}_1[\mathsf{X}_{t_1} = x_1]$

$$\mathbb{P}_{1|1}[\mathsf{X}_{t_3}{=}x_3 | \mathsf{X}_{t_1}{=}x_1] = \sum_{x_2 \in \Omega} \mathbb{P}_{1|1}[\mathsf{X}_{t_3}{=}x_3 | \mathsf{X}_{t_2}{=}x_2] \mathbb{P}_{1|1}[\mathsf{X}_{t_2}{=}x_2 | \mathsf{X}_{t_1}{=}x_1], \qquad t_1{<}t_2{<}t_3. \tag{1.35}$$

This equation is called *Chapman–Kolmogorov* equation and states that the probability of a path $x_1 \to x_2 \to x_3$ factorises in the product of the probabilities of the individual transitions $x_1 \to x_2$ and $x_2 \to x_3$.

A Markov process is *stationary* if $\mathbb{P}_1[\mathsf{X}_t = x]$ is actually independent on time and $\mathbb{P}_{1|1}[\mathsf{X}_{t_2} = x_2 | \mathsf{X}_{t_1} = x_1]$ depends only on the time interval $t_2 - t_1$, i.e.

$$\mathbb{P}_{1|1}[\mathsf{X}_{t_2} = x_2 | \mathsf{X}_{t_1} = x_1] = \mathbb{P}_{1|1}[\mathsf{X}_{t_2 - t_1} = x_2 | \mathsf{X}_0 = x_1]. \tag{1.36}$$

If $\mathbb{P}_1$ depends on time but (1.36) holds, the Markov process is said to be *homogeneous*.

✎ Explain why equation (1.42) does not hold for non-Markovian processes.

🔍 Similarly to what we have discussed for the case in which Ω is countable or finite, we can repeat our discussion for the case in which Ω has the cardinality of the continuum. The main difference is that probabilities will be replaced by *probability densities* and, for example, Bayes theorem takes the form

$$p_{n|s}(\{(x_i, t_i)\}_{i=s+1}^{n} | \{(x_i, t_i)\}_{i=1}^{s}) = \frac{p_n(\{(x_i, t_i)\}_{i=1}^{n})}{p_s(\{(x_i, t_i)\}_{i=1}^{s})}. \tag{1.37}$$

The Markov property takes then the form

$$p_{1|n-1}(x_n,t_n|\{(x_i,t_i)\}_{i=1}^{n-1}) = p_{1|1}(x_n,t_n|x_{n-1},t_{n-1}). \tag{1.38}$$

We have, of course, that $p_{1|1}(x_2,t_2|x_1,t_1) \geq 0$; moreover $\int_\Omega p_{1|1}(x_2,t_2|x_1,t_1)\,\mathrm{d}\,x_2 = 1$, expressing the fact that the system has to go somewhere with probability 1, and finally, directly from Bayes theorem,

$$p_1(x_2,t_2) = \int_\Omega p_{1|1}(x_2,t_2|x_1,t_1)p(x_1,t_1)\,\mathrm{d}\,x_1. \tag{1.39}$$

For a Markov process, the joint probabilities for $n \geq 3$ are all expressed in terms of p_1 and $p_{1|1}$. For example, for $n = 3$ we have

$$\begin{aligned} p_3(x_1,t_1;x_2,t_2;x_3,t_3) &= p_{1|2}(x_3,t_3|x_1,t_1;x_2,t_2)p_2(x_1,t_1;x_2,t_2) \\ &= p_{1|1}(x_3,t_3|x_2,t_2)p_2(x_2,t_2|x_1,t_1)p_1(x_1,t_1). \end{aligned} \tag{1.40}$$

For general n the general crucial result holds

$$p_n(\{(x_i,t_i)\}_{i=s+1}^{n}) = \prod_{i=2}^{n} p_{1|1}(x_i,t_i|x_{i-1},t_{i-1})\,p_1(x_1,t_1). \tag{1.41}$$

Taking relation (1.40), integrating it over x_2 and dividing both sides by $p_1(x_1,t_1)$ we obtain the CK equation

$$p_{1|1}(x_3,t_3|x_1,t_1) = \int \mathrm{d}x_2 p_{1|1}(x_3,t_3|x_2,t_2)p_{1|1}(x_2,t_2|x_1,t_1), \qquad t_1<t_2<t_3. \tag{1.42}$$

2.2. Finite Markov chains. One of the simplest examples of a Markov process is that of a Markov chain. A *discrete time Markov chain* is a discrete-time Markov process on some *countable or finite* space Ω. In this Section, we will limit for simplicity to Markov chains with in which Ω is a finite set, called *finite Markov chains*. This is not essential but removes distracting technical complications. Let us assume in particular that that X can take on values in $\Omega = \{1,2,\ldots,N\}$ and time is measured in units such that t takes values $t \in \mathbb{N}$. Then

$$\mathbb{P}_1[\mathsf{X}_n = i] = \sum_{j=1}^{N} \mathbb{P}_{1|1}[\mathsf{X}_n = i|\mathsf{X}_{n-1} = j]\mathbb{P}_1[\mathsf{X}_{n-1} = j] \tag{1.43}$$

The theory of Markov chains is highly developed for *homogeneous* chains and we shall mostly be concerned with these. Defining then

$$Q_{ij} \coloneqq \mathbb{P}_{1|1}[\mathsf{X}_n = i|\mathsf{X}_{n-1} = j] = \mathbb{P}_{1|1}[\mathsf{X}_1 = i|\mathsf{X}_0 = j] \tag{1.44}$$

this is n-independent. Since the system has to move to some state from any state j we have, for all j $\sum_{i=1}^{N} Q_{ij} = 1$. We can construct a *transition matrix* $\boldsymbol{Q}$ of *transition probabilities*

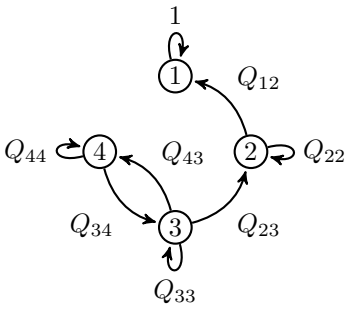
$$\boldsymbol{Q} = \begin{pmatrix} Q_{11} & Q_{12} & \cdots \\ Q_{21} & Q_{22} & \cdots \\ \vdots & \vdots & \vdots \end{pmatrix}. \tag{1.45}$$

Each matrix can have a graphical representation, in which, if each state i is associated with a node, a *directed* edge $i \to j$ is associated with Q_{ij}. For example, the following matrix with $N = 4$

$$\boldsymbol{Q} = \begin{pmatrix} 1 & Q_{12} & 0 & 0 \\ 0 & Q_{22} & Q_{23} & 0 \\ 0 & 0 & Q_{33} & Q_{34} \\ 0 & 0 & Q_{43} & Q_{44} \end{pmatrix}$$

can be associated with the graph

(1.46)



The matrix $\boldsymbol{Q}$ is a *stochastic matrix*: its elements are non-negative by construction and all its columns sum to one. In our convention, Q_{ij} is the probability of a transition from state j (column suffix) to i (row suffix).

The problem of finding the probability that after n steps the system is in a given state can be reduced to calculating entries in the n-th power of the transition matrix $\boldsymbol{Q}$. Let us introduce the column vector of state occupation probabilities at time t,

$$|\boldsymbol{P}(t)\rangle \equiv \begin{pmatrix} P_1(t) \\ \vdots \\ P_N(t) \end{pmatrix} \qquad \text{where } P_i(t) \coloneqq \mathbb{P}[\mathsf{X}_t = i]. \tag{1.47}$$

We can write the evolution of this vector for one step from $t_1 = t - 1$ to $t_2 = t$ as

$$|\boldsymbol{P}(t)\rangle = \boldsymbol{Q}|\boldsymbol{P}(t-1)\rangle \Leftrightarrow P_i(t) = \sum_j Q_{ij} P_j(t-1). \tag{1.48}$$

Iterating the very same equation,

$$|\boldsymbol{P}(t)\rangle = \boldsymbol{Q}^t|\boldsymbol{P}(0)\rangle, \tag{1.49}$$

with $\boldsymbol{Q}^0 \equiv \boldsymbol{I}$ being the identity matrix. Consequently, the process can be described completely by its *initial probability vector* $|\boldsymbol{P}(0)\rangle$ and the transition matrix $\boldsymbol{Q}$. For example, the probability of the system taking a specific *path* of states is

$$\mathbb{P}_n[\{\mathsf{X}_{t_i} = j_i\}_{i=1}^n] = \left(\prod_{k=2}^{n} Q_{j_k, j_{k-1}} \right) P_{j_1}(0). \tag{1.50}$$

If the system starts in any given state j, then $P_i(0) = \delta_{ij}$ and $P_k(t) = (\boldsymbol{Q}^t)_{kj}$. Finally, the simple fact that $\boldsymbol{Q}^{t+t'} = \boldsymbol{Q}^t \boldsymbol{Q}^{t'}$ expresses the Chapman–Kolmogorov equation for homogeneous Markov chain.

What we have said refers to the homogeneous case. In the non-homogeneous case the transition probability $\mathbb{P}_{1|1}[\mathsf{X}_t = i|\mathsf{X}_{t'} = j]$, $t > t'$, will depend on both t and t' and not just on $t - t'$. In this case, we write

$$Q_{ij}(t', t) = \mathbb{P}_{1|1}[\mathsf{X}_{t'} = i|\mathsf{X}_t = j]. \tag{1.51}$$

In particular, the jump probabilities $Q_{ij}(t+1, t)$ will depend on the time t. The Chapman-Kolmogorov equation reads in this case

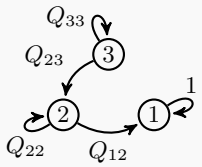
$$\boldsymbol{Q}(t, r) = \boldsymbol{Q}(t, s)\boldsymbol{Q}(s, r) \quad r < s < t. \tag{1.52}$$

2.3. Definitions based on the accessibility of states. Depending on the accessibility of the states in Ω in a Markov chain, different kinds of definitions can be given and are used in the literature. Looking at the diagram of a Markov chain, such as the one in (1.46), we can identify paths, closed sets, and absorbing states.

We say that there is a path $i \to j$ if $\exists n \geq 0$ such that $(\boldsymbol{Q}^n)_{ij} > 0$. On the other hand, a path does not exist if $\nexists n \geq 0$ such that $(\boldsymbol{Q}^n)_{ij} > 0$. Two states i and j are *communicating*, there exists a path $i \to j$ and a path $j \to i$, i.e., given sufficient time we can always get from i to j and from j to i: we write in this case $i \leftrightarrow j$.

EXAMPLE In the following simple Markov chain

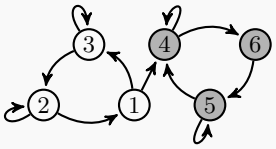
(1.53)



there is a path $3 \to 1$ but there is not a path $1 \to 3$: so 1 and 3 are not communicating.

A subset $\Omega_0 \subset \Omega$ is *closed* if $\forall i \in \Omega_0$ and $\forall j \in \Omega \setminus \Omega_0$, $i \not\to j$, i.e., starting from a state inside Ω_0 the system cannot ever reach any state outside Ω_0. In other words, $(\boldsymbol{Q}^n)_{ij} = 0$ for all $n \geq 0$ if $i \in X_0$ and $j \notin \Omega_0$. If Ω_0 contains a single element, such element is said to be an *absorbing state*. If, for each pair $i, j \in \Omega_0$, closed set of states, we have $i \leftrightarrow j$, then Ω_0 is a *closed irreducible subset* (or closed communicating class): every element inside Ω_0 can be accessible from any other, and the system cannot go out of Ω_0 once inside.

EXAMPLE In a chain of the form



the gray nodes belong to a closed irreducible subset of Ω: if you can start from a gray node, you cannot reach any white one, but you can get to any other gray node.

We can measure the *periodicity* of each state i, defined as

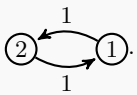
$$\tau_i := \gcd\{t : (\boldsymbol{Q}^t)_{ii} > 0\}. \tag{1.54}$$

If $\tau_i > 1$, then i is *periodic*, otherwise it is *aperiodic*. A periodic Markov chain is a chain with a periodic state, otherwise, it is said to be *aperiodic*.

EXAMPLE Consider for example the chain

(1.55)

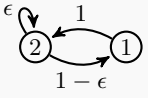
$$\boldsymbol{Q} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \Leftrightarrow$$



The state 1 in the chain (1.55), for example, has $\tau_1 = 2$. Similarly $\tau_2 = 2$. But if we consider for example, for $\epsilon \in (0, 1)$,

(1.56)

$$\boldsymbol{Q} = \begin{pmatrix} 0 & 1-\epsilon \\ 1 & \epsilon \end{pmatrix} \Leftrightarrow$$



then $\tau_1 = 1$: indeed, for any $t \geq 2$, the sequence of states $1 \to \overbrace{2 \to \cdots \to 2}^{t-1 \text{ times in } 2} \to 1$ has finite probability to happen (to be precise, it has probability $\epsilon^{t-2}(1-\epsilon)$), and the greatest common divisor of *all* $t \geq 2$ is 1: the state is aperiodic.

Periodic systems are usually characterised by the presence of eigenvalues $\lambda = -1$.There is an interesting property of periodicity, stated by the following theorem.

THEOREM 2.1. *Given two states i and j of a Markov chain, if communicating they have the same period, i.e.,*

$$i \leftrightarrow j \Rightarrow \tau_i = \tau_j. \tag{1.57}$$

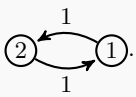
PROOF. If $i \leftrightarrow j$, then there exists $t > 0$ such that $(\boldsymbol{Q}^t)_{ij} > 0$ and a $t' > 0$ such that $(\boldsymbol{Q}^{t'})_{ji} > 0$. Now

$$\begin{aligned} \mathbb{P}_{1|1}[\mathsf{X}_{t+t'+t_0} = i | \mathsf{X}_0 = i] = \\ = \sum_{k,k' \in \Omega} \mathbb{P}_{1|1}[\mathsf{X}_{t+t'+t_0} = i | \mathsf{X}_{t'+t_0} = k] \mathbb{P}_{1|1}[\mathsf{X}_{t'+t_0} = k | \mathsf{X}_{t'} = k'] \mathbb{P}_{1|1}[\mathsf{X}_{t'} = k' | \mathsf{X}_0 = i] \\ \geq \mathbb{P}_{1|1}[\mathsf{X}_{t+t'+t_0} = i | \mathsf{X}_{t'+t_0} = j] \mathbb{P}_{1|1}[\mathsf{X}_{t'+t_0} = j | \mathsf{X}_{t'} = j] \mathbb{P}_{1|1}[\mathsf{X}_{t'} = j | \mathsf{X}_0 = i] = (\boldsymbol{Q}^t)_{ij} (\boldsymbol{Q}^{t_0})_{jj} (\boldsymbol{Q}^{t'})_{ji}. \end{aligned} \tag{1.58}$$

This equation must hold for any t_0 such that $(\boldsymbol{Q}^{t_0})_{jj} > 0$. In particular, it holds for $t_0 = 0$ and $t_0 = \tau_j$, the smallest possible time the system can take to come back to j after leaving j. But then, for $t_0 = 0$, $t + t'$ must be divisible by τ_i. By picking $t_0 = \tau_j$, the smallest possible nonzero value we can use, this fact implies that also τ_j must be divisible by τ_i. If we repeat the exact same operation starting and ending in j, we find that τ_i must be divisible by τ_j. So $\tau_i = \tau_j$. □

Reducible and irreducible Markov chains. — In an irreducible Markov chain, $\forall\, i, j \in \Omega$, $\exists t_{ij} > 0 \colon (\boldsymbol{Q}^t_{ij})_{ij} > 0$. This means that *all pairs of states are communicating*: one can go from any state in Ω to any other state in Ω in a finite number of steps. This also means, by Theorem 2.1, that all states have the same period τ. However, there is not necessarily a time that makes *all* the entries of the transition matrix positive (as in ergodic chains), meaning that some states may only be reached at certain times. Irreducibility is a *weaker* condition than regularity: all regular Markov chains are irreducible while not all irreducible Markov chains are regular. Non-irreducible Markov chains are called *reducible*.

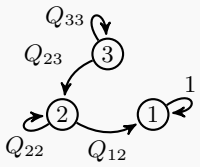
EXAMPLE An example of an irreducible chain which is not regular is given by a transition matrix of the type in (1.55),

$$\boldsymbol{Q} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \Leftrightarrow \text{  } \tag{1.59}$$

In this case $\boldsymbol{Q}^{2m+1} = \boldsymbol{Q}$ and $\boldsymbol{Q}^{2m} = \boldsymbol{I}$. However there is not n for which $\boldsymbol{Q}^n$ has all entries non-zero: the chain is *not* regular, as we will see below.

As an example of a reducible chain, consider

$$\boldsymbol{Q} = \begin{pmatrix} 1 & Q_{12} & 0 \\ 0 & Q_{22} & Q_{23} \\ 0 & 0 & Q_{33} \end{pmatrix} \Leftrightarrow \tag{1.60}$$



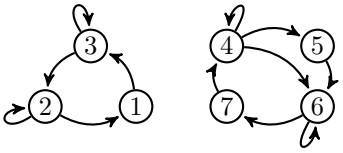
It is reducible as, if the system starts in 1, it cannot escape from it. Note that the state 1 is absorbing. A Markov chain has an absorbing state if one or more columns of the transition matrix contain all zeros except for the diagonal element, which must be one.

Reducible chains are also matrices in the form

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2 \end{pmatrix} \tag{1.61}$$

For example, this corresponds to a structure like

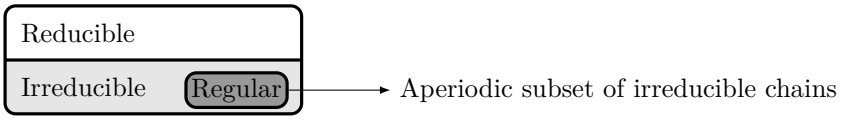
(1.62)



Let $\boldsymbol{Q}_1$ being $s \times s$ and $\boldsymbol{Q}_2$ being $(N-s) \times (N-s)$. This chain has two sets of dynamically separated states, Ω_1 and Ω_2. Hence, it will have two steady-states, $|\boldsymbol{\Pi}_1\rangle = (\Pi_1, \ldots, \Pi_s, 0, \ldots, 0)^\top$ and $|\boldsymbol{\Pi}_2\rangle = (0, \ldots, 0, \Pi_{s+1}, \ldots, \Pi_N)^\top$. Which one is reached, depends on whether the system is initialised in Ω_1 or Ω_2, corresponding to two components of the graph. As the name "reducible" suggests, this chain can, in fact, be reduced to two different Markov chains, with transition matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, respectively.

EXAMPLE The chain considered in (1.55) is an example of a periodic chain, as the system can only return to state 1 at even times $Q^n_{11} > 0$ for $n = 2, 4, \ldots$.

Regular Markov chain. — In a regular Markov chain $\exists t > 0 \colon (\boldsymbol{Q}^t)_{ij} > 0 \;\forall\; i, j \in \Omega$. Regular Markov chains are sometimes said to be *ergodic*. Note that if this property holds for some $t \geq 0$, it will hold for any $t' \geq t$. This property means that after a finite number of iterations, there is a non-zero probability for the system to be in *any* state, irrespective of the initial state i. This means that regular Markov chains are irreducible *and* aperiodic, i.e., they are the proper subset of irreducible chains with $\tau = 1$.



LECTURE 3

2.4. Steady-state. We define *stationary distribution* (sometimes for brevity *steady state*) of the Markov chain (1.48) a vector $|\boldsymbol{\Pi}\rangle$ of probabilities that does not evolve under the action of $\boldsymbol{Q}$, i.e.,

$$|\boldsymbol{\Pi}\rangle = \boldsymbol{Q}|\boldsymbol{\Pi}\rangle \quad \Pi_i \geq 0 \;\forall i \quad \sum_i \Pi_i = \langle \boldsymbol{1}|\boldsymbol{\Pi}\rangle = 1. \tag{1.63}$$

In other words, $\boldsymbol{\Pi}$ is the right eigenvector of the stochastic matrix $\boldsymbol{Q}$, associated with the eigenvalue $\lambda = 1$ with (non-negative) entries normalised to one and represents a time-independent

solution of the Markov chain equation (1.48). It is very important to note that the fact that all the column sums of a stochastic matrix are unity can be expressed in matrix notation by

$$\langle \mathbf{1}|\boldsymbol{Q} = \langle \mathbf{1}|, \tag{1.64}$$

with $\langle \mathbf{1}|$ is the row vector with all entries equal to 1. This means that $\langle \mathbf{1}|$ is always a left eigenvector of $\boldsymbol{Q}$, and $\lambda = 1$ is always an eigenvalue.

This also implies that a steady-state always exists: if a unit eigenvalue exists then the corresponding right eigenvector is the steady-state. However, the steady state may not be unique and convergence to it may not be guaranteed. We will see in later sections the conditions under which conditions a finite Markov chain converges to a unique steady-state.

For later reference, we note that multiplying each side by $\boldsymbol{Q}$ from right, we have $\langle \mathbf{1}|\boldsymbol{Q}^2 = \langle \mathbf{1}|\boldsymbol{Q} = \langle \mathbf{1}|$ and in general $\langle \mathbf{1}|\boldsymbol{Q}^n = \langle \mathbf{1}|$, so $\boldsymbol{Q}^n$ is also a stochastic matrix for each integral n.

Probability current. — The net *probability current* $J_{i\to j}$ from any state $i \in \Omega$ to any state $j \in \Omega$ in the stationary state $|\boldsymbol{\Pi}\rangle$ of a Markov chain is defined as

$$J_{j\to i} = Q_{ij}\Pi_j - Q_{ji}\Pi_i. \tag{1.65}$$

The current is by definition anti-symmetric under permutation of i and j, $J_{i\to j} = J_{j\to i}$. Since Π_i is the probability of the system to be in state i, and Q_{ji} the likelihood that it subsequently moves from i to j, $Q_{ji}\Pi_i$ is the probability that we observe the system moving from i to j. With multiple copies of our system evolving with the same Markov chain, it would be proportional to the *number* of observed moves from i to j. Thus $J_{i\to j}$ represents the net balance of observed transitions between i and j in the stationary state; hence the term 'current'. If $J_{i\to j} > 0$ there are more transitions $i \to j$ than $j \to i$; if $J_{i\to j} < 0$ there are more transitions $j \to i$ than $i \to j$. Conservation of probability implies that the sum over all currents is always zero:

$$\sum_{ij} J_{i\to j} = \sum_{ij} \Big(Q_{ji}\Pi_i - Q_{ij}\Pi_j \Big) = \sum_i \Pi_i - \sum_j \Pi_j = 0. \tag{1.66}$$

Detailed balance. — Stationarity implies that $\sum_j J_{i\to j} = 0$ (no net flow of probability outgoing from i). If we have that

$$J_{i\to j} = 0 \Longleftrightarrow Q_{ji}\Pi_i = Q_{ij}\Pi_j \qquad \forall i, j, \tag{1.67}$$

we say that there is *detailed balance*. This is a very strong condition. Detailed balance implies stationarity: if $\boldsymbol{Q}$ and $|\boldsymbol{\Pi}\rangle$ are in detailed balance, then $|\boldsymbol{\Pi}\rangle$ is the steady-state probability. Indeed

$$\sum_j Q_{ij}\Pi_j = \sum_j Q_{ji}\Pi_i = \Pi_i. \tag{1.68}$$

However, detailed balance is a stronger condition than stationarity: it implies, in addition to stationarity, dynamical *reversibility*, i.e., equilibrium. If X is a Markov chain with transition matrix $\boldsymbol{Q}$ and steady-state $|\boldsymbol{\Pi}\rangle$, detailed balance makes the probability of a forward path $i_0 \to i_1 \to ... \to i_N$ at stationarity, equal to the probability of the backward path $i_N \to ... \to i_1 \to i_0$

$$\begin{aligned} &\mathbb{P}[\mathsf{X}_0 = i_N; \mathsf{X}_1 = i_{N-1}; \dots; \mathsf{X}_n = i_0] \\ &\quad = Q_{i_0 i_1} \dots Q_{i_{N-1} i_N} \Pi_{i_N} = Q_{i_0 i_1} \dots Q_{i_{N-2}, i_{N-1}} \Pi_{i_{N-1}} Q_{i_N i_{N-1}} \\ &\qquad = \dots = \Pi_{i_0} Q_{i_0 i_1} \dots Q_{i_{N-1} i_N} = \mathbb{P}[\mathsf{X}_0 = i_0; \mathsf{X}_1 = i_1; \dots; \mathsf{X}_n = i_N]. \end{aligned} \tag{1.69}$$

If $N = 2$, i.e., X can take only value $\{1, 2\}$, and since $J_{i\to i} = 0$ by definition, the condition for stationarity, $\sum_j J_{i\to j} = 0\ \forall i$, and the one for detailed balance $J_{i\to j} = 0\ \forall\ i \neq j$, coincide, hence any 2-state Markov chain allowing a stationary state automatically satisfies detailed balance.

Detail balance is very important in physical simulations. Markov chains used to model closed physical many-particle systems with noise are usually of the detailed balance type, as a result of the invariance of Newton's laws of motion under time reversal $t \to -t$.

2.5. Eigenvalues and eigenvectors of stochastic matrices. We have seen that calculating the probability that the Markov chain is in a given state after n steps, normally involves calculating entries in the n-th power of the transition matrix. This is best done by using a spectral representation of the transition matrix, i.e., a decomposition of the matrix based on eigenvalues and eigenvectors. For $|\boldsymbol{P}(t)\rangle$ to remain well-defined for $t \to \infty$, it is vital that the eigenvalues are sufficiently small.

The matrix $\boldsymbol{Q}$ is in general not a symmetric matrix, therefore the right and left eigenvectors will be different and complex-valued. The left and the right eigenvector problems are

$$\boldsymbol{Q}|\boldsymbol{\psi}^a\rangle = \lambda_a|\boldsymbol{\psi}^a\rangle \tag{1.70}$$

$$\langle\boldsymbol{\phi}^b|\boldsymbol{Q} = \lambda_b\langle\boldsymbol{\phi}^b| \tag{1.71}$$

with $|\boldsymbol{\psi}^a\rangle, \langle\boldsymbol{\phi}^b| \in \mathbb{C}^N \setminus \{\boldsymbol{0}\}$ and $\lambda_\mu, \lambda_\nu \in \mathbb{C}$. Much can be extracted from the two defining properties $Q_{ij} \geq 0\ \forall i,j$ and $\sum_i Q_{ij} = 1\ \forall j$ alone. Note that eigenvectors $|\boldsymbol{\psi}^a\rangle, \langle\boldsymbol{\phi}^b|$ of $\boldsymbol{Q}$ need not be probabilities in the sense of $|\boldsymbol{P}\rangle$, as they could have negative or complex entries. A series of general facts can be stated, stemming directly from basic linear algebra.

LEMMA 2.2. *The sets of left- and right- eigenvalues of $\boldsymbol{Q}$ are identical, and, moreover, right and left eigenvectors are biorthogonal. Finally, it is possible to expand the matrix $\boldsymbol{Q}$ in terms of its left and right eigenvectors as*

$$\boldsymbol{Q} = \sum_a \lambda_a|\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a|. \tag{1.72}$$

PROOF. Equations (1.70) and (1.71) give $\det[\boldsymbol{Q} - \lambda_R\boldsymbol{I}] = 0$ for the right eigenvalues λ_R and $\det[\boldsymbol{Q}^\dagger - \lambda_L\boldsymbol{I}] = \det[(\boldsymbol{Q} - \lambda_L\boldsymbol{I})^\dagger] = \det[\boldsymbol{Q} - \lambda_L\boldsymbol{I}] = 0$ for the left eigenvalues λ_L respectively. These are the same equation for any $\boldsymbol{Q}$, involving a polynomial of degree N, and have therefore the same set of N solutions $\{\lambda_a\}_{a=1}^N$.
Multiplying (1.70) by $\langle\boldsymbol{\phi}^b|$ and (1.71) by $|\boldsymbol{\psi}^a\rangle$ and subtracting, we find $(\lambda_a - \lambda_b)\langle\boldsymbol{\phi}^b|\boldsymbol{\psi}^a\rangle = 0$, so $\langle\boldsymbol{\phi}^b|\boldsymbol{\psi}^a\rangle = 0$ if $\lambda_a \neq \lambda_b$. If $\langle\boldsymbol{\phi}^a|\boldsymbol{\psi}^a\rangle \neq 0$ (condition ensured when the set of eigenvectors of $\boldsymbol{Q}$ is *complete*), we can scale these vectors so that $\langle\boldsymbol{\phi}^b|\boldsymbol{\psi}^a\rangle = \delta_{ab}$ and (equivalently) $\sum_a |\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a| = \boldsymbol{I}$. From the fact that $\langle\boldsymbol{\phi}^a|\boldsymbol{Q} = \langle\boldsymbol{\phi}^a|\lambda_a$ if we tensor-multiply by $|\boldsymbol{\psi}^a\rangle$ and sum over a using the fact that $\sum_a |\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a| = \boldsymbol{I}$, we obtain the formula above. The result also implies that

$$\boldsymbol{Q}^n = \sum_a \lambda_a^n|\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a|. \tag{1.73}$$

□

LEMMA 2.3. *All eigenvalues λ of stochastic matrices $\boldsymbol{Q}$ obey*

$$|\lambda| \leq 1. \tag{1.74}$$

PROOF. This property is the first one which is specific to a stochastic matrix. Consider the right eigenvalue equation, $\lambda_a|\boldsymbol{\psi}^a\rangle = \boldsymbol{Q}|\boldsymbol{\psi}^a\rangle$. Then we can write

$$|\lambda_a|\sum_i |\psi_i^a| = \sum_i |\lambda_a\psi_i^a| = \sum_i \left|\sum_j Q_{ij}\psi_j^a\right| \leq \sum_i\sum_j |Q_{ij}\psi_j^a| = \sum_i\sum_j Q_{ij}|\psi_j^a| = \sum_j |\psi_j^a|.$$

Since $|\boldsymbol{\psi}^a\rangle$ is assumed to be nonzero, we know that $\sum_i |\psi_i^a| \neq 0$ and hence $|\lambda_a| \leq 1$. □

We will soon go come back to these results when discussing the spectral properties of irreducible Markov chains. From now on, we will order the eigenvalues in such a way that $\lambda_1 = 1$ and

$$1 = |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_N|.$$

As anticipated, the fact that $\langle \mathbf{1}|\boldsymbol{Q} = \langle \mathbf{1}|$ implies that $\langle \mathbf{1}|$ is always left eigenvector with eigenvalue $\lambda = 1$. Therefore, there must exist at least one right eigenvector with $\lambda = 1$. Due to the biorthonormality and the fact that $\langle \mathbf{1}|$ is left eigenvector with eigenvalue 1, right eigenvectors corresponding to eigenvalues $\lambda \neq 1$ of stochastic matrices $\boldsymbol{Q}$ obey

$$\langle \mathbf{1}|\boldsymbol{\psi}^a\rangle = 0. \tag{1.75}$$

2.6. Convergence to a stationary state. It is often very relevant to study the behaviour of the system after a large number of transitions $t \to \infty$ and to see if it retains some memory of its initial state, or if the asymptotic behavior is independent of it. Once again, to investigate this problem, we need to analyse the transition matrix $\boldsymbol{Q}$. A fundamental theorem, the *Perron–Frobenius theorem*, helps us in extracting the information we need. Let us state it in a slightly more general form, and then adapt it to our case.

THEOREM 2.4 (Perron–Frobenius). *Let $\boldsymbol{Q}$ be a stochastic matrix of an irreducible Markov chain with periodicity $\tau \geq 1$. Then*

- *the τ roots of the identity, $\lambda_k = \mathrm{e}^{2\pi i \frac{k-1}{\tau}}$, $k = 1, \ldots, \tau$, are eigenvalues of $\boldsymbol{Q}$, and in particular $\lambda_1 = 1$.*
- *the remaining eigenvalues $\lambda_{\tau+1}, \ldots, \lambda_N$ have $|\lambda_j| < 1$;*
- *there exists a unique vector $|\boldsymbol{\Pi}\rangle$ with strictly positive entries such that $\langle \mathbf{1}|\boldsymbol{\Pi}\rangle = 1$ and $\boldsymbol{Q}|\boldsymbol{\Pi}\rangle = |\boldsymbol{\Pi}\rangle$*

As for a *regular* Markov chain $\tau = 1$, this means that regular Markov chains have only one eigenvalue with $|\lambda_1| = 1$. We know the corresponding left eigenvector $\langle \boldsymbol{\phi}^1| = \langle \mathbf{1}|$, and the right eigenvector, normalised to one, is the steady state $|\boldsymbol{\psi}^1\rangle = |\boldsymbol{\Pi}\rangle$, that, by the Perron–Frobenius theorem, has all its components non-negative and is therefore, once normalised, a proper probability vector. Then (1.73) can be rewritten as

$$\boldsymbol{Q}^n = |\boldsymbol{\Pi}\rangle\langle \mathbf{1}| + \sum_{a>1} \lambda_a^n |\boldsymbol{\phi}^a\rangle\langle \boldsymbol{\psi}^a|. \tag{1.76}$$

Since $|\lambda_a| < 1 \; \forall \; a \neq 1$, in the limit $n \to \infty$ we obtain a unique value

$$\boldsymbol{Q}^n \xrightarrow{n\to+\infty} |\boldsymbol{\Pi}\rangle\langle \mathbf{1}| = \begin{pmatrix} \Pi_1 & \Pi_1 & \ldots & \Pi_1 \\ \Pi_2 & \Pi_2 & \ldots & \Pi_2 \\ \vdots & & & \\ \Pi_N & \Pi_N & \ldots & \Pi_N \end{pmatrix}. \tag{1.77}$$

It is easy to see that $|\boldsymbol{\Pi}\rangle\langle \mathbf{1}|\boldsymbol{Q} = |\boldsymbol{\Pi}\rangle\langle \mathbf{1}|$ and that, for any $|\boldsymbol{P}(0)\rangle$, $|\boldsymbol{\Pi}\rangle\langle \mathbf{1}|\boldsymbol{P}(0)\rangle = |\boldsymbol{\Pi}\rangle$ (because $\langle \mathbf{1}|\boldsymbol{P}(0)\rangle = 1$).

✎ Show that indeed $|\boldsymbol{\Pi}\rangle\langle \mathbf{1}|\boldsymbol{Q} = |\boldsymbol{\Pi}\rangle\langle \mathbf{1}|$.

In particular, under the hypotheses above, *the solution of the Markov chain converges to $|\boldsymbol{\Pi}\rangle$ independently from the initial state.* Indeed

$$\lim_{t\to+\infty} |\boldsymbol{P}(t)\rangle = \lim_{t\to+\infty} \boldsymbol{Q}^t |\boldsymbol{P}(0)\rangle = |\boldsymbol{\Pi}\rangle\langle \mathbf{1}|\boldsymbol{P}(0)\rangle = |\boldsymbol{\Pi}\rangle. \tag{1.78}$$

We can actually be even more precise. Indeed, using the spectral representation of $\boldsymbol{Q}^t$ one has

$$|\boldsymbol{P}(t)\rangle = \boldsymbol{Q}^t|\boldsymbol{P}(0)\rangle = |\boldsymbol{\Pi}\rangle + \sum_{a>1} \lambda_a^t |\psi^a\rangle\langle\phi^a|\boldsymbol{P}(0)\rangle. \tag{1.79}$$

For large t, the sum is dominated by the second largest (in absolute value) eigenvalue λ_2, so that

$$|P_i(t) - \Pi_i| \sim |\lambda_2|^t = \mathrm{e}^{-t \ln \frac{1}{|\lambda_2|}} = \mathrm{e}^{-\frac{t}{t_r}} .$$

Hence the *relaxation time* t_r is given by the second largest eigenvalue

$$t_r = -\frac{1}{\ln|\lambda_2|}$$

Note that for $\lambda_2 \to \pm 1$, $\tau \to \infty$. For a Markov chain to converge to a unique steady-state, it has to be irreducible *and* aperiodic. This defines a regular chain.

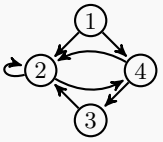
EXAMPLE We consider an interesting and important application of Markov chains to random walks on graphs. There are many processes that can be modeled in this way. A remarkable example is `PageRank`, the algorithm used by Google Search to rank websites in their search engine results.

A graph (or network) is defined by N 'nodes', labeled by $i = 1, \ldots, N$ and a set of 'links' specified by a connectivity matrix, also called adjacency matrix

$$\boldsymbol{A} = \begin{pmatrix} A_{11} & \cdots & A_{1N} \\ \vdots & \ddots & \vdots \\ A_{N1} & \cdots & A_{NN} \end{pmatrix} \tag{1.80}$$

such that $A_{ij} = 1$ if there is a link between i and j, and zero otherwise. For example,

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \Leftrightarrow$$



If links are undirected, the adjacency matrix is symmetric and $\boldsymbol{A}^\top = \boldsymbol{A}$. From now on, we will consider undirected graphs, without loops, i.e., $A_{ii} = 0\ \forall i$. We define the degree of a node i as $k_i = \sum_j A_{ij}$, the number of the 'neighbours' of i, i.e., nodes linked to i. The degree sequence $|\boldsymbol{k}\rangle = (k_1, k_2, \ldots, k_N)$ specifies the degrees of all the nodes in a graph.

We now consider a random walker, who takes steps along the links of a non-directed network $\boldsymbol{A}$, at discrete times $t = 1, 2, \ldots$, according to the following probabilistic rule: at each time step, the walker moves from the site j it currently occupies, to a new site i selected uniformly at random among the neighbours of j. We denote $P_i(t)$ as the probability to find the walker on the site i at time t. Translating the dynamical rule into equations, we have

$$P_i(t+1) = \sum_j \frac{A_{ij}}{k_j} P_j(t)$$

which describes a Markov chain. If $\boldsymbol{D}$ is the diagonal matrix having $D_{ij} = k_i^{-1}\delta_{ij}$, the transition matrix of the chain is

$$\boldsymbol{Q} = \boldsymbol{A}\boldsymbol{D}^{-1}.$$

It is easy to see that this is indeed a stochastic matrix, i.e. $\sum_i Q_{ij} = 1$ for all j. The node probability vector $|\boldsymbol{P}(t)\rangle = (P_1(t), \ldots, P_N(t))^\top$ then evolves according to the probabilistic rule $|\boldsymbol{P}(t+1)\rangle = \boldsymbol{Q}|\boldsymbol{P}(t)\rangle$ for Markov chains. In addition to the adjacency matrix, which defines the graph connectivity, one can define another important matrix, the so-called Laplacian

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$$

The stationary probability vector $|\boldsymbol{\Pi}\rangle$ can be expressed in terms of the eigenvectors of the Laplacian matrix $\boldsymbol{L}$ associated to zero eigenvalues. From its defining equation $|\boldsymbol{\Pi}\rangle = \boldsymbol{Q}|\boldsymbol{\Pi}\rangle = \boldsymbol{A}\boldsymbol{D}^{-1}|\boldsymbol{\Pi}\rangle$, setting

$|\boldsymbol{x}\rangle = \boldsymbol{D}^{-1}|\boldsymbol{\Pi}\rangle$, we have

$$\boldsymbol{D}|\boldsymbol{x}\rangle = \boldsymbol{A}|\boldsymbol{x}\rangle \Leftrightarrow (\boldsymbol{D}-\boldsymbol{A})|\boldsymbol{x}\rangle = \boldsymbol{L}|\boldsymbol{x}\rangle = 0.$$

It is easy to show that $|\mathbf{1}\rangle$ is always an eigenvector of the Laplacian associated to eigenvalue zero. One can also show that the number of zero eigenvalues of the Laplacian is equal to the number of disconnected components in the graph. Hence, if the graph is connected, there exists only one zero eigenvalue. In this case, the stationary distribution is unique and equal to

$$|\boldsymbol{\Pi}\rangle = c\boldsymbol{D}|\mathbf{1}\rangle$$

where c is a normalization constant. Therefore, at stationarity, the random walker visits each site with probability

$$\Pi_i = \frac{k_i}{\sum_j k_j}$$

i.e., nodes with higher degrees k_i are visited more frequently.

2.7. Accounting for memory. Markov processes are "memoryless", as for these systems, the future trajectory depends only on the current configuration and not on the past trajectory. This may seem a big limitation, as many systems may retain a longer memory. However, Markov chains can also be used to model systems with longer, but finite memory: the general strategy to do so is to augment sufficiently the dimension of the state-space. To illustrate this, consider the following example.

EXAMPLE Assume that London's weather can be modeled in terms of a stochastic variable X taking values in the set $\Omega = \{\text{Sun}, \text{Rain}\}$ and that X is updated daily according to the probabilistic rules:

(i): Tomorrow's weather is like today with probability 0.5

(ii): If the weather has stayed the same for two days in a row, it will change with probability 0.7.

It is clear that for this system (using the abbreviation S for Sun and R for Rain)

$$\mathbb{P}_{1|2}[\mathsf{X}_n = \mathrm{S}|\mathsf{X}_{n-1} = \mathrm{S}; \mathsf{X}_{n-2} = \mathrm{R}] \neq \mathbb{P}_{1|2}[\mathsf{X}_n = \mathrm{S}|\mathsf{X}_{n-1} = \mathrm{S}; \mathsf{X}_{n-2} = \mathrm{S}]$$

so that the Markovian property is apparently lost. However, one can retrieve it by duplicating the dimensionality of the stochastic variable

$$\mathsf{X} \to \mathbf{X} = (\mathsf{Y}, \mathsf{Z})$$

so that one component, e.g. Y, accounts for the current configuration, and the other, Z, for the configuration at the earlier time-step (in this example, the day before). In this way, we have enlarged the state-space from

$$\Omega = \{\mathrm{S}, \mathrm{R}\} \to \Omega' = \Omega \times \Omega.$$

In the enlarged state-space, the dynamics is Markovian, as given the system's configuration of today and yesterday, the future configuration is independent of the past.

This example illustrates that an initially non-Markovian process, where transition probabilities depend on how long the current state has been on, can be modeled as a Markovian process by augmenting sufficiently the state space. Obviously, this strategy does not work for processes that are more drastically non-Markovian such as self-avoiding random walks, in which a particle moving on a lattice cannot visit a site where it has already been. For these processes, the whole history is relevant, hence the state-space would have to be made larger and larger as time elapses, which becomes unfeasible at some point.

3. Markov chain MonteCarlo algorithms

So far we have been concerned with the problem of finding the steady-state probability $|\boldsymbol{\Pi}\rangle$ of a given Markov chain with transition matrix $\boldsymbol{Q}$ and studying how the system reaches such

a steady state. Here we address the inverse problem of constructing a Markov chain which is ensured to converge to a given probability distribution $|\mathbf{\Pi}\rangle$. This is can be very useful when we wish to sample configurations x from a finite set Ω, with a given probability Π_x. As an example, assume we wish to calculate the average of some observable A defined on the set Ω

$$\mathbb{E}[A(\mathsf{X})] = \sum_{x\in\Omega} \Pi_x A(x).$$

A naïve way of proceeding would be generating a large number of configurations x distributed according to Π_x, calculating A_x in each of them, and performing the arithmetic average of their values. The approach would lead to a naïve *MonteCarlo algorithm*:

(1) generate x and compute Π_x;
(2) generate a random number $z \in [0, 1]$;
(3) if $z < \Pi_x$ accept x and calculate $A(x)$ and store it;
(4) go back to Step 1.

The cycle is repeated a number of times, until some convergence criterion is met, and the average $\mathbb{E}[A(\mathsf{X})]$ is then estimated taking the mean of the accepted values $A(x)$. However, when x lives in a very large space, Π_x is typically very small and the acceptance rate generating x might be very low. In other words, one would spend most of the time generating random numbers and computing Π_x for configurations that are eventually rejected.

A more efficient way of proceeding is to build a dynamic process that allows sampling mostly the important configurations with occasional visits to unimportant ones. The idea behind *Markov chains MonteCarlo* (MCMC) strategies is to construct a Markov chain designed to converge to the targeted distribution $|\mathbf{\Pi}\rangle$, so that, when evolved to equilibrium, it will visit configurations x with probability Π_x. This is done by defining a *regular* Markov chain $\boldsymbol{Q}$ which satisfies *detailed balance* with $|\mathbf{\Pi}\rangle$.

In practice, one generates, at each iteration, a move from the current configuration x to a new configuration x', drawn from a distribution of moves $\eta(x \to x')$, and then execute the move with probability $a_{x'\,x}$ (with probability $1 - a_{x'\,x}$ the move is rejected and the system stays in x). This leads to an *aperiodic* Markov chain with transition matrix $\boldsymbol{Q}$, in which

$$Q_{x'\,x} = a_{x'x}\eta_{x'x}, \qquad x' \neq x \tag{1.81}$$

and by consequence $Q_{x\,x} = 1 - \sum_{x'\neq x} a_{x'x}\eta_{x'x}$.

The two functions η and a have to be chosen in such a way that the Markov chain with transition matrix $\boldsymbol{Q}$ has $|\mathbf{\Pi}\rangle$ as a steady state. We can impose in particular *detail balance*

$$Q_{xx'}\Pi_{x'} = Q_{x'x}\Pi_x$$

for all $x \neq x'$, so that $|\mathbf{\Pi}\rangle$ is an equilibrium distribution for the Markov chain. A standard choice for η is simply

$$\eta_{x'x} = \frac{1}{N_x} \quad \forall\ x'$$

where N_x is the size of the subset of Ω accessible from x. In other words, if we are in the configuration x, we propose randomly one configuration x' amongst the accessible ones: the acceptance or not of this configuration will depend on a: but how to choose a? The detailed balance condition gives

$$a_{x'x}\eta_{x'x}\Pi_x = a_{xx'}\eta_{xx'}\Pi_{x'}.$$

There are several possible choices of a that may satisfy this equation. One aims for the one allowing for the highest acceptance rate, given that rejections are wasteful. Starting with the

observation that acceptance probabilities are bounded by one, $0 \leq a_{x'x} \leq 1$, then observes that

$$a_{x'x} = \frac{\eta_{xx'}\Pi_{x'}}{\eta_{x'x}\Pi_x} a_{xx'} \leq \frac{\eta_{xx'}\Pi_{x'}}{\eta_{x'x}\Pi_x}. \tag{1.82}$$

Two choices are usually adopted

Metropolis–Hastings rule: In the Metropolis–Hastings algorithm, one tries to saturate the bound above, taking into account that a cannot exceed 1. So we adopt

$$a_{x'\,x} = \min\left\{1, \frac{\eta_{xx'}\Pi_{x'}}{\eta_{x'x}\Pi_x}\right\}.$$

Glauber rule: In the Glauber approach one takes

$$a_{x'\,x} = \frac{\eta_{xx'}\Pi_{x'}}{\eta_{x'x}\Pi_x + \eta_{xx'}\Pi_{x'}}.$$

Such a choice is convenient in many physical settings, as we will discuss below.

In many practical situations, $\eta_{x'x}$ does not depend on the current configuration x nor from the destination x', i.e., all configurations have the same "mobility", and we can drop the dependence from it in the equations for the acceptance rates so that the two choices mentioned above becomes

$$a_{x'\,x} = \begin{cases} \min\left\{1, \frac{\Pi_{x'}}{\Pi_x}\right\} & \text{(Metropolis)}, \\ \frac{\Pi_{x'}}{\Pi_x + \Pi_{x'}} & \text{(Glauber)}. \end{cases} \tag{1.83}$$

The MCMC algorithm can be implemented therefore by the following pseudo-code:

(1) initialize the system in x;
(2) draw a move $x \to x'$ from $\eta_{x'\,x'}$;
(3) calculate $a_{x'\,x}$;
(4) generate a random number $z \in [0, 1]$;
(5) if $z < a_{x'\,x}$ accept configuration x', and calculate $A(x')$;
(6) go to Step 2 and iterate.

As in the naïve algorithm, the cycle is repeated until a certain convergence criterion is satisfied. When the configuration space Ω is very big, the MCMC has a first advantage with respect to the naïve algorithm: the acceptance rate of the move $a_{x'\,x}$ is the *ratio* of two small numbers, as opposed to $\Pi_{x'}$ (which is a small number), and can be therefore finite. Another very important advantage appears when dealing with simulations in statistical mechanics.

Suppose that we are considering a thermodynamical system such that the steady-state is in Gibbs–Boltzmann form,

$$\Pi_x = \frac{\mathrm{e}^{-\beta H(x)}}{Z} \qquad Z \coloneqq \sum_{x\in\Omega} \mathrm{e}^{-\beta H(x)} \tag{1.84}$$

where β is the inverse temperature of our system and $H(x)$ is its Hamiltonian function, depending on the configuration x. Usually, β and the function $H(x)$ are the input of the analysis. The computation of Z, however, might be a formidable task for high-dimensional systems, where the size of the space Ω can be huge: this complication adds to the ones described for the adoption of the naïve MonteCarlo algorithm.

If we adopt instead a MCMC, we can use

$$a_{x'x} = \begin{cases} \min\left\{1, \mathrm{e}^{-\beta\Delta H}\right\} & \text{(Metropolis)}, \\ \frac{1}{1+\mathrm{e}^{-\beta\Delta H}} = \frac{1}{2}\left(1 - \tanh\frac{\beta\Delta H}{2}\right) & \text{(Glauber)}. \end{cases} \qquad \Delta H = H(x') - H(x). \tag{1.85}$$

At this point, we can immediately see one first major advantage of the MCMC method: *the computation of Z is not necessary to specify the acceptance rates a.* This is a really important simplification that makes MCMC fundamental tools for numerical statistical mechanics. Moreover, the above equations show that Metropolis prescription moves the system downhill, on the energy surface $H(x)$, with certainty, and uphill, with a probability that is smaller the larger the energy of the new configuration with respect to the current one. Glauber prescription moves the system both downhill and uphill with a probability which favours visiting or remaining in configurations with lower energy. Both prescriptions ensure that the algorithm does not get stuck in local minima, as there is a non-zero probability to go uphill.

LECTURE 4

4. Hitting times and hitting probabilities

As opposed to the asymptotic investigations above, one may be interested in the *short* time behaviour, for example, the first time a system reaches a given state (e.g., the first time a gambler goes bankruptcy). In this section we will consider a setting slightly more general than before, assuming at first that Ω is a possibly infinite, but countable, set. A quantity that has attracted large attention in Markov models is the *Mean First Passage Time* (MFPT), called sometimes *mean hitting time*. The time that it takes for the system to go from i to j is a random variable,

$$\mathsf{T}_{ji} \coloneqq \min\{t \geq 1 \colon \mathsf{X}_t = j \text{ and } \mathsf{X}_0 = i\}, \tag{1.86}$$

with the understanding that if j is *never* reached from i, then $\mathsf{T}_{ji} = +\infty$. This random variable is distributed with distribution

$$f_{ji}(t) = \mathbb{P}[\mathsf{T}_{ji} = t] \tag{1.87}$$

By definition, $f_{ji}(0) = 0$. In the following, for the sake of brevity, we will use the notation

$$q_{ji}(t) = \mathbb{P}_{1|1}[\mathsf{X}_t = j | \mathsf{X}_0 = i] \tag{1.88}$$

for the probability of getting to j starting from i in t steps (also, by definition $q_{ji}(0) = \delta_{ij}$). Note that, if the chain is on a finite number of states, $q_{ij}(t)$ is just (a power of) the stochastic matrix $\boldsymbol{Q}$, $q_{ji}(t) = (\boldsymbol{Q}^t)_{ji}$. Also, $q_{ji}(t)$ has not to be confused with $f_{ji}(t)$ which is the probability of getting to j *for the first time* starting from i in t steps.

The mean first passage time, or mean hitting time, from a state i to j is the *expectation* of such time

$$t_{ji} \coloneqq \mathbb{E}[\mathsf{T}_{ji}] = \sum_{t=1}^{\infty} t f_{ij}(t) + \infty \times f_{ji}(+\infty). \tag{1.89}$$

With the last term, I wanted to stress that $f_{ji}(+\infty)$ might be non-zero, and in that case $\mathbb{E}[\mathsf{T}_{ji}] = +\infty$. However, even if $f(+\infty) = +\infty$ it is of course still possible that $t_{ji} = +\infty$. If in particular $t_{ii} = +\infty$, the state i is said to be *null*; otherwise it is said to be *positive*. The probability that the chain ever "hits" j at some point for the first time starting from i is

$$f_{ij} \coloneqq \mathbb{P}[\mathsf{T}_{ji} < +\infty] = \sum_{t=1}^{\infty} f_{ji}(t). \tag{1.90}$$

A special role plays the case $i = j$. States can be classified depending on their *return time* properties. On the basis of the value of f_{ii}, we can give a classification of the states $i \in \Omega$. We say that a state i is *recurrent* if the probability of coming back to it at some finite time is 1,

$$f_{ii} = 1. \tag{1.91}$$

which means that there is some time at which the system, which started in i, goes back to i for sure. This also means that, if V_{ii} is the number of visits of i of the system after starting from i, then $\mathbb{P}[\mathsf{V}_{ii} = +\infty] = 1$. If otherwise there is a finite probability of never coming back to i, i.e.,

$f_{ii} < 1$ or equivalently $f_{ii}(+\infty) > 0$, then the state i is *transient*. But we can sort out a criterion for this classification that can be very useful.

THEOREM 4.1. *For each state $i \in \Omega$ in a Markov chain*

$$\sum_{t=0}^{\infty} q_{ii}(t) = \mathbb{E}[\mathsf{V}_{ii}] = \begin{cases} +\infty & \text{if } i \text{ is recurrent,} \\ \frac{1}{1-f_{ii}} & \text{if } i \text{ is transient.} \end{cases} \tag{1.92}$$

PROOF. First of all, let us observe

$$\sum_{t=0}^{\infty} q_{ii}(t) = \sum_{t=0}^{\infty} \mathbb{E}[\mathbb{I}(\mathsf{X}_t = i) | \mathsf{X}_0 = i] = \mathbb{E}\left[\sum_{t=0}^{\infty} \mathbb{I}(\mathsf{X}_t = i) \Big| \mathsf{X}_0 = i \right] = \mathbb{E}[\mathsf{V}_{ii}]. \tag{1.93}$$

To compute the expectation, let us look at the probability of different values $\mathbb{P}[\mathsf{V}_{ii} \geq k]$ for different values of k. First of all $\mathbb{P}[\mathsf{V}_{ii} \geq 1] = f_{ii}$, as the probability of one visit or more is equal to the probability of coming back to i in finite time at some point at least once. Similarly,

$$\mathbb{P}[\mathsf{V}_{ii} \geq 1] = f_{ii} \tag{1.94}$$

$$\mathbb{P}[\mathsf{V}_{ii} \geq 2] = f_{ii}^2 \tag{1.95}$$

$$\vdots \qquad \vdots \tag{1.96}$$

$$\mathbb{P}[\mathsf{V}_{ii} \geq k] = f_{ii}^k. \tag{1.97}$$

Then $\mathbb{P}[\mathsf{V}_{ii} = +\infty] = \lim_{k \to +\infty} \mathbb{P}[\mathsf{V}_{ii} \geq k] = \lim_{k \to +\infty} f_{ii}^k$. This limit can be only 1, if $f_{ii} = 1$ (i is recurrent), or 0, if $0 < f_{ii} < 1$ (i is transient). To complete the proof then

$$\sum_{t=0}^{\infty} q_{ii}(t) = \mathbb{E}[\mathsf{V}_{ii}] = \sum_{k=0}^{\infty} \mathbb{P}[\mathsf{V}_{ii} \geq k] = \sum_{k=0}^{\infty} f_{ii}^k = \begin{cases} +\infty & \text{if } i \text{ is recurrent,} \\ \frac{1}{1-f_{ii}} & \text{if } i \text{ is transient.} \end{cases} \tag{1.98}$$

□

The following theorem says that we can group states in "classes" containing only recurrent or only transient states: it implies that, in a closed irreducible subset, all states "share" the same property of being recurrent or transient.

THEOREM 4.2. *Suppose that, in a Markov chain, $i \leftrightarrow j$. If i is recurrent, then j is recurrent. On the other hand, if i is transient, then j is transient.*

PROOF. The fact that i and j are intercommunicating means that $\exists t_1 > 0$ and $\exists t_3 > 0$ such that $q_{ji}(t_1) > 0$ and $q_{ij}(t_3) > 0$. We can write now

$$q_{ii}(t_1 + t_2 + t_3) \geq q_{ij}(t_3) q_{jj}(t_2) q_{ji}(t_1). \tag{1.99}$$

Now, let us sum over t_2, obtaining $\sum_{t_2} q_{ii}(t_1 + t_2 + t_3) \geq q_{ji}(t_1) q_{ij}(t_3) \sum_{t_2} q_{jj}(t_2)$. If i is transient, then the term on the left is finite, and so the one on the right. Similarly, we can prove that if j is recurrent, so it is i. □

If the class is finite-sized, i.e., contains a finite number of states (as we usually assume in this chapter), however, then it *cannot be* transient.

THEOREM 4.3. *Every finite closed irreducible subset Ω_0 is recurrent.*

PROOF. Let us imagine starting from some state $j \in \Omega_0$: it must exist a state $i \in \Omega_0$ that is visited an infinite number of times because we run the chain for infinite time on a finite set. We can

write

$$\mathbb{P}[\mathsf{V}_{ij} = +\infty] = \mathbb{P}[\mathsf{V}_{ii} = +\infty | \mathsf{T}_{ij} < +\infty]\mathbb{P}[\mathsf{T}_{ij} < +\infty] \equiv \mathbb{P}[\mathsf{V}_{ii} = +\infty]\mathbb{P}[\mathsf{T}_{ij} < +\infty].$$

As $\mathbb{P}[\mathsf{V}_{ij} = +\infty] > 0$ and $\mathbb{P}[\mathsf{T}_{ij} < +\infty]$, it follows that $\mathbb{P}[\mathsf{V}_{ii} = +\infty] \neq 0$ and the only possibility is that $\mathbb{P}[\mathsf{V}_{ii} = +\infty] = 1$, i.e., i is recurrent, and therefore the entire Ω_0 is recurrent. □

In particular, observing that irreducible finite chains are made of a single closed irreducible subset, the following result is implied.

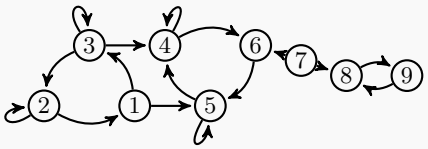
COROLLARY 4.4. *An irreducible finite Markov chain is recurrent.*

The theorem above implies that, if we consider a finite Markov chain, we can isolate the recurrent classes and the transient classes, so that the matrix $\boldsymbol{Q}$ can be rearranged in the form

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{\mathcal{T}\mathcal{T}} & \mathbf{0} \\ \boldsymbol{Q}_{\mathcal{R}\mathcal{T}} & \boldsymbol{Q}_{\mathcal{R}\mathcal{R}} \end{pmatrix} \tag{1.100}$$

where $\mathcal{R}$ is the set of recurrent states, $\mathcal{T}$ is the set of transient states: note that the submatrix $\boldsymbol{Q}_{\mathcal{R}\mathcal{R}}$ is a legit stochastic matrix itself.

EXAMPLE Consider for example the finite Markov chain

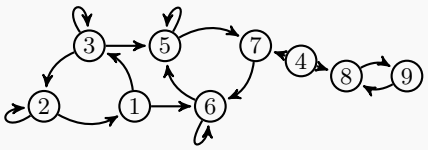


We can see that 1, 2, 3, and 7 are transient because they can be left for good, whilst all the others are recurrent. The matrix $\boldsymbol{Q}$ representing this chain has the form

$$\boldsymbol{Q} = \begin{pmatrix} 0 & Q_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & Q_{22} & Q_{23} & 0 & 0 & 0 & 0 & 0 & 0 \\ Q_{31} & 0 & Q_{33} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & Q_{43} & Q_{44} & Q_{45} & 0 & 0 & 0 & 0 \\ Q_{51} & 0 & 0 & 0 & Q_{55} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & Q_{64} & 0 & 0 & Q_{67} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & Q_{87} & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{1.101}$$

Let us now "rename" the nodes, calling, from 1 to $\hat{4}$ the transient nodes, and from 5 to 9 the recurrent ones, using for example the following "dictionary"

$$1 \mapsto 1 \quad 2 \mapsto 2 \quad 3 \mapsto 3 \quad 4 \mapsto 5 \quad 5 \mapsto 6 \quad 6 \mapsto 7 \quad 7 \mapsto 4 \quad 8 \mapsto 8 \quad 9 \mapsto 9.$$



This relabeling does not change at all the chain. It is just a matter of naming the nodes. If we write the associated transition matrix exhibits precisely the structure of Eq.(1.100):

$$\boldsymbol{Q} = \left(\begin{array}{cccc|ccccc} 0 & Q_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & Q_{22} & Q_{23} & 0 & 0 & 0 & 0 & 0 & 0 \\ Q_{31} & 0 & Q_{33} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & Q_{53} & 0 & Q_{55} & Q_{56} & 0 & 0 & 0 \\ Q_{61} & 0 & 0 & 0 & 0 & Q_{66} & 1 & 0 & 0 \\ 0 & 0 & 0 & Q_{74} & Q_{75} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & Q_{84} & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right). \tag{1.102}$$

4.1. Mean hitting times in finite Markov chains. Suppose that we want to study the mean time t_{ji} to go from i to j in a *finite* Markov chain. Starting from state i, the system can either move to j directly (i.e., in one step), with probability Q_{ji} or transition to some other state k with probability Q_{ki} (in one time step) and then move to j taking *on average* t_{jk} steps so that j is reached in, on average, $t_{jk}+1$ steps in total. This is expressed mathematically as

$$\begin{aligned} t_{ji} = \mathbb{E}[\mathsf{T}_{ji}] = \sum_{t=1}^{\infty} t f_{ji}(t) = f_{ji}(1) + \sum_{t=2}^{\infty} t f_{ji}(t) = Q_{ji} + \sum_{t=1}^{\infty}(t+1)\sum_{k\neq j} f_{jk}(t) Q_{ki} \\ = Q_{ji} + \sum_{k\neq j}(t_{jk}+1)Q_{ki} = 1 + \sum_{k\neq j} t_{jk} Q_{ki} \\ = 1 + \sum_{k}(t_{jk} - t_{jj}\delta_{kj})Q_{ki}. \end{aligned} \tag{1.103}$$

In the equation above, I denoted by $\sum_{t=1}^{\lceil\infty\rceil} t f_{ji}(t) =: \sum_{t=1}^{\lceil\infty\rceil} t f_{ji}(t) + \infty \times f_{ji}(\infty)$: this means that we take into account the possibility that some of these t_{ji} are actually infinite, therefore making the entire expression diverge. This set of equations can be solved numerically. Alternatively, one can express the MFPTs in terms of the eigenvalues and eigenvectors of the transition matrix (provided they form a complete orthonormal set). Eq. (1.103) can be written in vector notation introducing the two matrices

$$\boldsymbol{T} = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1N} \\ t_{21} & t_{22} & \dots & t_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ t_{N1} & t_{N2} & \dots & t_{NN} \end{pmatrix} \qquad \underbrace{\boldsymbol{T}^{(r)} = \begin{pmatrix} t_{11} & 0 & \dots & 0 \\ 0 & t_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t_{NN} \end{pmatrix}}_{\text{matrix of mean recurrence times}} \tag{1.104}$$

so that we have the formula

$$\boldsymbol{T} = |\mathbf{1}\rangle\langle\mathbf{1}| + (\boldsymbol{T} - \boldsymbol{T}^{(r)})\boldsymbol{Q}. \tag{1.105}$$

In general, the matrix $\boldsymbol{Q}$ has the structure in Eq. (1.100). This means that we can write the equation above as

$$\begin{pmatrix} \boldsymbol{T}_{\mathcal{T}\mathcal{T}} & \boldsymbol{T}_{\mathcal{T}\mathcal{R}} \\ \boldsymbol{T}_{\mathcal{R}\mathcal{T}} & \boldsymbol{T}_{\mathcal{R}\mathcal{R}} \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{T}_{\mathcal{T}\mathcal{T}} - \boldsymbol{T}^{(r)}_{\mathcal{T}\mathcal{T}} & \boldsymbol{T}_{\mathcal{T}\mathcal{R}} \\ \boldsymbol{T}_{\mathcal{R}\mathcal{T}} & \boldsymbol{T}_{\mathcal{R}\mathcal{R}} - \boldsymbol{T}^{(r)}_{\mathcal{R}\mathcal{R}} \end{pmatrix} \begin{pmatrix} \boldsymbol{Q}_{\mathcal{T}\mathcal{T}} & \mathbf{0} \\ \boldsymbol{Q}_{\mathcal{R}\mathcal{T}} & \boldsymbol{Q}_{\mathcal{R}\mathcal{R}} \end{pmatrix} \tag{1.106}$$

where $\mathbf{1}$ represents here a matrix of ones. In particular, this equation tells us that the array of times to get from the transient states to the recurrent states is given by

$$\boldsymbol{T}_{\mathcal{R}\mathcal{T}} = \mathbf{1} + \boldsymbol{T}_{\mathcal{R}\mathcal{T}}\boldsymbol{Q}_{\mathcal{T}\mathcal{T}} + (\boldsymbol{T}_{\mathcal{R}\mathcal{R}} - \boldsymbol{T}^{(r)}_{\mathcal{R}\mathcal{R}})\boldsymbol{Q}_{\mathcal{R}\mathcal{T}} \tag{1.107}$$

that can be solved by looking at the information coming from the recurrent part

$$\boldsymbol{T}_{\mathcal{R}\mathcal{R}} = \mathbf{1} + (\boldsymbol{T}_{\mathcal{R}\mathcal{R}} - \boldsymbol{T}^{(r)}_{\mathcal{R}\mathcal{R}})\boldsymbol{Q}_{\mathcal{R}\mathcal{R}}. \tag{1.108}$$

Note that from Eq. (1.106)

$$\boldsymbol{T}_{\mathcal{TR}} = \mathbf{1} + \boldsymbol{T}_{\mathcal{TR}}\boldsymbol{Q}_{\mathcal{RR}}, \tag{1.109}$$

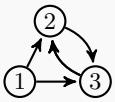
which means that $\boldsymbol{T}_{\mathcal{TR}}$ satisfies

$$\boldsymbol{T}_{\mathcal{TR}}(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{RR}}) = \mathbf{1}. \tag{1.110}$$

This equation has no finite solution because the matrix $\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{RR}}$ has at least one zero eigenvalue and cannot be inverted. This makes sense indeed: the expected time to go from a *recurrent* to a *transient* state is infinite!

EXAMPLE Let us consider as an example the Markov chain below

(1.111)



having transient state 1 and recurrent states 2 and 3, and associated to

$$\boldsymbol{Q} = \begin{pmatrix} \overbrace{0}^{\boldsymbol{Q}_{\mathcal{TT}}} & \overbrace{\begin{matrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{matrix}}^{\boldsymbol{Q}_{\mathcal{TR}}} \\ \underbrace{\begin{matrix} 1/2 \\ 1/2 \end{matrix}}_{\boldsymbol{Q}_{\mathcal{RT}}} & \underbrace{\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}}_{\boldsymbol{Q}_{\mathcal{RR}}} \end{pmatrix}. \tag{1.112}$$

We have that, using Eq. (1.108)

$$\boldsymbol{T}_{\mathcal{RR}} = \begin{pmatrix} t_{22} & t_{23} \\ t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & t_{23} \\ t_{32} & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 + t_{23} & 1 \\ 1 & 1 + t_{32} \end{pmatrix}. \tag{1.113}$$

This implies immediately

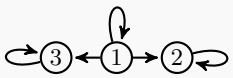
$$\boldsymbol{T}_{\mathcal{RR}} = \begin{pmatrix} t_{22} & t_{23} \\ t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \tag{1.114}$$

which was maybe expected. Using this in Eq. (1.107),

$$\boldsymbol{T}_{\mathcal{RT}} = \begin{pmatrix} t_{21} \\ t_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} t_{21} \\ t_{31} \end{pmatrix} \times 0 + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3/2 \\ 3/2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}. \tag{1.115}$$

EXAMPLE As a second example, consider the toy-sh Markov chain

(1.116)



having again transient state 1 and recurrent states 2 and 3, and associated to

$$\boldsymbol{Q} = \begin{pmatrix} \overbrace{1/3}^{\boldsymbol{Q}_{\mathcal{TT}}} & \overbrace{\begin{matrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{matrix}}^{\boldsymbol{Q}_{\mathcal{TR}}} \\ \underbrace{\begin{matrix} 1/3 \\ 1/3 \end{matrix}}_{\boldsymbol{Q}_{\mathcal{RT}}} & \underbrace{\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}}_{\boldsymbol{Q}_{\mathcal{RR}}} \end{pmatrix}. \tag{1.117}$$

Using Eq. (1.108)

$$\boldsymbol{T}_{\mathcal{RR}} = \begin{pmatrix} t_{22} & t_{23} \\ t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & t_{23} \\ t_{32} & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 + t_{23} \\ 1 + t_{23} & 1 \end{pmatrix}. \tag{1.118}$$

We obtain then that $t_{22} = t_{33} = 1$, but we also obtain the equations $t_{23} = t_{23} + 1$ and $t_{32} = t_{32} + 1$, which make sense only taking $t_{23} = +\infty$ and $t_{32} = +\infty$, which indeed is the case: it is impossible to access 3 from 2 and vice versa. Using this in Eq. (1.107), we get something that might appear as not

satisfactory,

$$\boldsymbol{T}_{\mathcal{RT}} = \begin{pmatrix} t_{21} \\ t_{31} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} t_{21} \\ t_{31} \end{pmatrix} \frac{1}{3} + \begin{pmatrix} 0 & +\infty \\ +\infty & 0 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix} = \begin{pmatrix} +\infty \\ +\infty \end{pmatrix}, \tag{1.119}$$

but however makes sense: in computing t_{21}, there is a finite probability that the system goes to 3 first, and gets stuck there forever and never reach 2.

Absorption time. — A very special, but important, case is the case in which the recurrent states are r absorbing states (see for example the chain in Eq. (1.116)). In this case $\boldsymbol{Q}_{\mathcal{RR}} = \boldsymbol{I}$, identity matrix of size r. If $r > 1$, the time to get absorbed in one specific state can be infinite, simply because of the possible absorption in *another* state: this is what happens, for example, in the example on the chain (1.116). We can, however, ask another question, i.e., *what is the average time t_i^a to get absorbed in some state starting from the transient state i?* These times can be collected in a vector $\langle \boldsymbol{T}^a|$ of size $N - r$ whose ith component is t_i^a. To compute the elements of this vector we can follow the spirit of the derivation of Eq. (1.103)

$$\begin{aligned} t_i^a &= \sum_{j\in\mathcal{R}} Q_{ji} + \sum_{j\in\mathcal{T}} (t_j^a + 1) Q_{ji} \\ &= 1 + \sum_{j\in\mathcal{T}} t_j^a Q_{ji}. \end{aligned} \tag{1.120}$$

The equation expresses the fact that we can get to an absorbing state starting from i either in one step or moving to another transient j and then from it to the absorbing states. We can rewrite this expression in the vectorial form

$$\langle \boldsymbol{T}^a| = \langle \boldsymbol{1}| + \langle \boldsymbol{T}^a|\boldsymbol{Q}_{\mathcal{TT}} \Rightarrow \langle \boldsymbol{T}^a| = \langle \boldsymbol{1}|(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{TT}})^{-1}. \tag{1.121}$$

In this case, we can also write an equation for the probability of hitting (i.e., arriving for the first time) one recurrent state j starting from a transient state i. We can indeed write

$$a_{ji} = Q_{ji} + \sum_{k\in\mathcal{T}} a_{ji} Q_{ki}. \tag{1.122}$$

The quantity a_{ji} is nothing but f_{ji} used before: we adopt a different letter to stress that we are in a situation in which all the recurrent states are absorbing states, so f_{ji} is the probability that, starting from i, the system will be "absorbed" in j. The first term takes into account a direct jump $j \to i$. The sum, instead, runs only on *transient* states: if the system gets to one of the recurrent ones, it gets absorbed there and cannot go to j. Defining now $\boldsymbol{A} \in \mathbb{R}^{r\times(N-r)}$ as the matrix of probability that the system is absorbed in j starting from i, it satisfies

$$\boldsymbol{A} = \boldsymbol{Q}_{\mathcal{RT}} + \boldsymbol{A}\boldsymbol{Q}_{\mathcal{TT}} \Rightarrow \boldsymbol{A} = \boldsymbol{Q}_{\mathcal{RT}}(\boldsymbol{I} - \boldsymbol{Q}_{\mathcal{TT}})^{-1}. \tag{1.123}$$

The case of regular chains. — Assuming now that our finite Markov chain is regular: this implies that *all states are recurrent* and there is a unique vector $|\boldsymbol{\Pi}\rangle$ that corresponds to a steady state probability. We can multiply both terms in (1.105) by $|\boldsymbol{\Pi}\rangle$, vector with positive entries, and use the fact that $\boldsymbol{Q}|\boldsymbol{\Pi}\rangle = |\boldsymbol{\Pi}\rangle$,

$$\boldsymbol{T}|\boldsymbol{\Pi}\rangle = |\boldsymbol{1}\rangle + (\boldsymbol{T} - \boldsymbol{T}^{(r)})|\boldsymbol{\Pi}\rangle \Leftrightarrow \boldsymbol{T}^{(r)}|\boldsymbol{\Pi}\rangle = |\boldsymbol{1}\rangle. \tag{1.124}$$

This means that the recurrence times t_{ii} satisfy

$$t_{ii}\Pi_i = 1 \Rightarrow t_{ii} = \frac{1}{\Pi_i} \tag{1.125}$$

In other words, what we proved is the following.

LEMMA 4.5 (Kač Lemma). *For a regular Markov chain, the element i of the steady-state distribution $|\boldsymbol{\Pi}\rangle$ is given by the reciprocal of the corresponding mean recurrence time,*

$$\Pi_i = \frac{1}{t_{ii}}. \tag{1.126}$$

This also implies that $\sum_i \frac{1}{t_{ii}} = 1$.

We have obtained therefore the expression of $\boldsymbol{T}^{(r)}$, which consists of these quantities only. The time t_{ii} can be interpreted as the expected number of time steps it takes to the system to first hit state i, *after* its release from state i itself. To obtain the remaining quantities, we start from the same equation (1.105)

$$\boldsymbol{T}(\boldsymbol{I}-\boldsymbol{Q}) = |\mathbf{1}\rangle\langle\mathbf{1}| - \boldsymbol{T}^{(r)}\boldsymbol{Q}$$

But applying both sides to $|\boldsymbol{\psi}^a\rangle$, $a \neq 1$,

$$\boldsymbol{T}(\boldsymbol{I}-\boldsymbol{Q})|\boldsymbol{\psi}^a\rangle = |\mathbf{1}\rangle\langle\mathbf{1}|\boldsymbol{\psi}^a\rangle - \boldsymbol{T}^{(r)}\boldsymbol{Q}|\boldsymbol{\psi}^a\rangle \Leftrightarrow (1-\lambda_a)\boldsymbol{T}|\boldsymbol{\psi}^a\rangle = -\lambda_a \boldsymbol{T}^{(r)}|\boldsymbol{\psi}^a\rangle$$

Tensor-multiplying by $\langle\boldsymbol{\phi}^a|$ and summing over a

$$\boldsymbol{T}\sum_{a>1}|\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a| = -\sum_{a>1}\frac{\lambda_a}{1-\lambda_a}\boldsymbol{T}^{(r)}|\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a|$$

If we use now $\sum_{a>1}|\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a| = \boldsymbol{I} - |\boldsymbol{\Pi}\rangle\langle\mathbf{1}|$ we obtain

$$\boldsymbol{T} = \boldsymbol{T}|\boldsymbol{\Pi}\rangle\langle\mathbf{1}| - \sum_{a>1}\frac{\lambda_a}{1-\lambda_a}\boldsymbol{T}^{(r)}|\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a|$$

The missing piece is the matrix $\boldsymbol{T}|\boldsymbol{\Pi}\rangle\langle\mathbf{1}|$. To deal with it we can observe that, switching to components, the previous equation read

$$t_{ji} = \sum_k t_{ik}\Pi_k - \sum_{a>1}\frac{\lambda_a}{1-\lambda_a}\frac{1}{\Pi_j}\psi_j^a\phi_i^a \Rightarrow \sum_k t_{ik}\Pi_k = t_{ji} + \sum_{a>1}\frac{\lambda_a}{1-\lambda_a}\frac{\psi_j^a\phi_i^a}{\Pi_j} \tag{1.127}$$

For $i = j$

$$\sum_k t_{jk}\Pi_k = \frac{1}{\Pi_j} + \sum_{a>1}\frac{\lambda_a}{1-\lambda_a}\frac{\psi_j^a\phi_j^a}{\Pi_j} \tag{1.128}$$

We can now collect all our pieces and write finally

$$t_{ji} = \frac{1}{\Pi_j} + \sum_{a>1}\frac{\lambda_a}{1-\lambda_a}\frac{\psi_j^a\phi_j^a - \psi_j^a\phi_i^a}{\Pi_j}. \tag{1.129}$$

The Kemeny constant. — A "typical time" of the dynamics can be obtained averaging t_{ji} over the possible destinations j, sampling them with probability Π_j,

$$\sum_j \Pi_j t_{ji} = N + \sum_{a>1}\frac{\lambda_a}{1-\lambda_a}\sum_j \psi_j^a(\phi_j^a - \phi_i^a) = N + \sum_{a>1}\frac{\lambda_a}{1-\lambda_a} = 1 + \sum_{a>1}\frac{1}{1-\lambda_a}, \tag{1.130}$$

where we have used $\sum_j \psi_j^a = \langle\mathbf{1}|\boldsymbol{\psi}^a\rangle = \delta_{a,1}$ and the orthonormality of eigenvectors. This also means that, if we exclude the recurrence time t_{ii} from the sum,

$$\sum_{j\neq i}\Pi_j t_{ji} = \sum_{a>1}\frac{1}{1-\lambda_a} \equiv \zeta. \tag{1.131}$$

The quantity ζ is known as the Kemeny constant and it has attracted a large interest, over the years, since its introduction in 1960 by Kemeny and Snell. It represents the sum of all relaxation

timescales in a Markov chain. Surprisingly, the Kemeny constant is independent of the starting state i. This intriguing constancy has been the subject of several studies.

CHAPTER 2

Continuous-time Markov processes

ABSTRACT. In the previous chapter, we have considered Markovian processes which involve transitions between states in a finite set Ω at discrete times. In this chapter we look at Markov processes evolving in continuous time, in either discrete or continuous space Ω. Continuous-time processes require the introduction of *transition rates*, i.e., transition probabilities over an infinitesimal time interval, as opposed to transition probability in one time-step described by $\boldsymbol{Q}$ in the case of Markov chains. The *master equation* is the fundamental tool to describe such processes and in particular the probability of finding the system in a given state. Such probability changes until the system reaches a final equilibrium steady state in which transitions cannot alter the probability distribution[1]. Once again, the Markovian property will be crucial to derive the master equation, which will be the starting point of many analyses that we will perform during the module.

1. The master equation

LECTURE 5

1.1. Derivation. Suppose that we have a stochastic process X_t taking values in the finite set $\Omega = \{1, \dots, N\}$. We can write down a Markov chain evolving at time steps $t = 0, \tau, 2\tau, \dots$. We start from the equation

$$P_j(t+\tau) \coloneqq \mathbb{P}_1[\mathsf{X}_{t+\tau} = j] = \sum_{i\in\Omega} \mathbb{P}_{1|1}[\mathsf{X}_{t+\tau} = j|\mathsf{X}_t = i]\mathbb{P}_1[\mathsf{X}_t = i] \tag{2.1}$$

that we can write in a vectorial form as

$$|\boldsymbol{P}(t+\tau)\rangle = \boldsymbol{Q}(\tau)|\boldsymbol{P}(t)\rangle \tag{2.2}$$

where $\boldsymbol{Q}(\tau)$ encloses the transition probabilities, between pairs of state, on a lag-time τ

$$Q_{ji}(\tau) = \mathbb{P}_{1|1}[\mathsf{X}_{t+\tau} = j|\mathsf{X}_t = i].$$

We assume here that the process is *homogeneous*. Note that $\boldsymbol{Q}(\tau)$, at given τ, is a stochastic matrix precisely of the same type of the ones discussed in the previous chapter: it has, therefore, the same properties, e.g., $\langle\mathbf{1}|\boldsymbol{Q}(\tau) = \langle\mathbf{1}|$. The *master equation* can be derived as the continuous time limit $\tau \to 0$ of Eq. (2.2). Indeed, let us subtract from both term $|\boldsymbol{P}(t)\rangle$ and dividing by τ

$$\frac{|\boldsymbol{P}(t+\tau)\rangle - |\boldsymbol{P}(t)\rangle}{\tau} = \frac{\boldsymbol{Q}(\tau) - \boldsymbol{I}}{\tau}|\boldsymbol{P}(t)\rangle. \tag{2.3}$$

Assuming that the matrix

$$\boldsymbol{W} \coloneqq \lim_{\tau\to 0} \frac{\boldsymbol{Q}(\tau) - \boldsymbol{I}}{\tau} \tag{2.4}$$

exists, we can write the master equation in the form

$$\frac{\mathrm{d}\,|\boldsymbol{P}(t)\rangle}{\mathrm{d}\,t} = \boldsymbol{W}|\boldsymbol{P}(t)\rangle, \tag{2.5}$$

where the matrix $\boldsymbol{W}$ is the matrix of the *transition rates*. Since $\langle\mathbf{1}|\boldsymbol{Q} = \langle\mathbf{1}|$ it follows that $\langle\mathbf{1}|\boldsymbol{W} = \lim_{\tau\to 0}\frac{1}{\tau}\langle\mathbf{1}|(\boldsymbol{Q}(\tau) - \boldsymbol{I}) = \langle\mathbf{0}|$, vector of zeros. I.e., the vector $\langle\mathbf{1}|$ is *left eigenvector* of

$\boldsymbol{W}$ with corresponding eigenvalue 0. This fact expresses the property

$$\sum_i W_{ij} = 0 \Leftrightarrow W_{ii} = -\sum_{j\neq i} W_{ji}. \tag{2.6}$$

Switching to components, this property allows us to rewrite the equation as follows

$$\frac{\mathrm{d}\, P_i(t)}{\mathrm{d}\, t} = \sum_{j\neq i} W_{ij}P_j(t) + W_{ii}P_i(t) = \sum_{j\neq i} [W_{ij}P_j(t) - W_{ji}P_i(t)] \tag{2.7}$$

This equation can be interpreted as follows: the first piece in parenthesis is the gain in probability due to the transitions from other states j to i; the second term is the loss due to transitions from i to other states j.

🔍 If Ω has the cardinality of the continuum, we can derive an analogous equation for the *probability density*,

$$\partial_t p(x,t) = \int [W(x|y)p(y,t) - W(y|x)p(x,t)]\, \mathrm{d}\, y. \tag{2.8}$$

1.2. Steady state. A stationary solution $|\boldsymbol{\Pi}\rangle$ is such that

$$\boldsymbol{W}|\boldsymbol{\Pi}\rangle = |\mathbf{0}\rangle \Leftrightarrow \sum_{j\neq i} W_{ij}\Pi_j = \sum_{j\neq i} W_{ji}\Pi_i. \tag{2.9}$$

In other words, $|\boldsymbol{\Pi}\rangle$ is the right eigenvector corresponding to eigenvalue 0. The equation above states that, at stationarity, the total number of transition per time into a state i equals on average the total number of transition out of it. One has *detailed balance* if each individual transition is balanced,

$$W_{ij}\Pi_j = W_{ji}\Pi_i. \tag{2.10}$$

When detailed balance holds, the stationary distribution is often called an *equilibrium distribution*[2]. In detailed balance regime, the entries stationary distribution vector can be obtained by iteration, i.e., expressing all the steady-state probabilities in terms of Π_1, using the fact that

$$\Pi_2 = \frac{W_{21}}{W_{12}}\Pi_1, \tag{2.11}$$

$$\Pi_3 = \frac{W_{32}}{W_{23}}\frac{W_{21}}{W_{12}}\Pi_1, \tag{2.12}$$

$$\vdots \quad \vdots \tag{2.13}$$

$$\Pi_N = \prod_{i=1}^{N-1} \frac{W_{i+1\,i}}{W_{i\,i+1}}\Pi_1 \tag{2.14}$$

and finally using the normalization condition $\sum_i \Pi_i = 1$ to determine Π_1.

[2]Note however that detailed balance is a necessary but not sufficient condition for thermodynamic equilibrium.

1.3. The propagator. As it happened with $\boldsymbol{Q}$ in discrete-time Markov chains, $\boldsymbol{W}$ is in general not symmetric, so it has left and right eigenvectors. We have already seen that $\langle \mathbf{1}$ is a left eigenvector corresponding to eigenvalue 0, whose corresponding right eigenvector is stationary (but it is not obvious that is a good probability vector). In general, we will define

$$\boldsymbol{W}|\boldsymbol{\psi}^a\rangle = \mu_a|\boldsymbol{\psi}^a\rangle, \tag{2.15}$$

$$\langle\boldsymbol{\phi}^b|\boldsymbol{W} = \mu_b\langle\boldsymbol{\phi}^b|, \tag{2.16}$$

with a spectrum $\{\mu_a\}_{a\in[N]}$. If they form a complete set (i.e., $\boldsymbol{W}$ is diagonalizable), we can write

$$\boldsymbol{W} = \sum_a \mu_a|\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a| \tag{2.17}$$

and the right eigenvectors can be used as a basis to expand any vector, and in particular any probability vector

$$|\boldsymbol{P}(t)\rangle = \sum_{a=1}^{N} \alpha_a(t)|\boldsymbol{\psi}^a\rangle, \tag{2.18}$$

where the time-dependent coefficients $\alpha_a(t) = \langle\boldsymbol{\phi}^a|\boldsymbol{P}(t)\rangle$ represent the projections of the probability vector on the left eigenvectors. Inserting into the master equations we have

$$\frac{\mathrm{d}\,|\boldsymbol{P}(t)\rangle}{\mathrm{d}\,t} = \sum_a \dot\alpha_a|\boldsymbol{\psi}^a\rangle = \sum_a \alpha_a\boldsymbol{W}|\boldsymbol{\psi}^a\rangle = \sum_a \mu_a\alpha_a|\boldsymbol{\psi}^a\rangle \tag{2.19}$$

Multiplying from left by $\langle\boldsymbol{\phi}^a|$ and using the bio-orthogonality of left and right eigenvectors $\langle\boldsymbol{\phi}^a|\boldsymbol{\psi}^b\rangle = \delta_{ab}$, we get

$$\dot\alpha_a(t) = \mu_a\alpha_a(t) \Rightarrow \alpha_a(t) = \alpha_a(0)\,\mathrm{e}^{\mu_a t} \tag{2.20}$$

so that

$$|\boldsymbol{P}(t)\rangle = \sum_{a=1}^{N} \alpha_a(0)\,\mathrm{e}^{\mu_a t}\,|\boldsymbol{\psi}^a\rangle \equiv \mathrm{e}^{t\boldsymbol{W}}\,|\boldsymbol{P}(0)\rangle. \tag{2.21}$$

The object $\mathrm{e}^{t\boldsymbol{W}}$ plays the same role played by $\boldsymbol{Q}$ in discrete-time Markov chains: it is the *propagator* that makes the probability vector evolve, so that, if we give a unite of time τ, a propagation in time of $t = \ell\tau$ is given applying the operator $\mathrm{e}^{\ell\tau\boldsymbol{W}} = (\mathrm{e}^{\tau\boldsymbol{W}})^\ell$. We can define therefore

$$\boldsymbol{Q}(\tau) \coloneqq \mathrm{e}^{\boldsymbol{W}\tau} \quad \text{so that} \quad Q_{ji}(\tau) = \mathbb{P}_{1|1}[\mathsf{X}_\tau = j|\mathsf{X}_0 = i]. \tag{2.22}$$

This definition is consistent with the one given for $\boldsymbol{W}$, being $\boldsymbol{Q}(\tau) = \boldsymbol{I} + \tau\boldsymbol{W} + o(\tau)$. Note that $Q_{ii}(t) = 1 + tW_{ii} + o(t) > 0$ for t small enough: in other words, the matrix $\boldsymbol{Q}(t)$ corresponds to an *aperiodic* chain. Also, we have

$$|\boldsymbol{P}(t)\rangle = \boldsymbol{Q}(t)|\boldsymbol{P}(0)\rangle = \boldsymbol{Q}(t-t')|\boldsymbol{P}(t')\rangle.$$

Finally, since $\boldsymbol{Q}(t) = \mathrm{e}^{\boldsymbol{W}t}$, one has

$$\frac{\mathrm{d}\,\boldsymbol{Q}(t)}{\mathrm{d}\,t} = \boldsymbol{W}\boldsymbol{Q}(t)$$

i.e., $\boldsymbol{Q}(t)$, as a matrix, solves the same equation as $|\boldsymbol{P}(t)\rangle$ with initial conditon $\boldsymbol{Q}(0) = \boldsymbol{I}$. This makes sense: rembember that $\boldsymbol{Q}$ is a matrix of *conditional probability*, and their evolution, in a Markov process, is the same as the one of full probabilities (the conditioning playing the role of an initial condition). The detailed balance condition in Eq. (2.10) translates to $Q_{ij}(t)\Pi_j = Q_{ji}(t)\Pi_i$,

as we would have written for a discrete-time chain. Finally $\boldsymbol{Q}(t)$ satisfies a Chapman–Kolmogorov equation,

$$\boldsymbol{Q}(t-t') = \boldsymbol{Q}(t-t'')\boldsymbol{Q}(t''-t'). \tag{2.23}$$

✎ Prove that if $W_{ij}\Pi_j = W_{ji}\Pi_i$, then $Q_{ij}(t)\Pi_j = Q_{ji}(t)\Pi_i$.

2. Convergence to equilibrium

As in the case of discrete-time Markov chains, the spectral properties of $\boldsymbol{W}$ are crucial to understand the long time behavior of the system. Let us assume that $\boldsymbol{W}$ has a complete set of (left and right) eigenvectors, and corresponding eigenvalues. If $|\boldsymbol{\psi}^a\rangle$ is eigenvector of $\boldsymbol{W}$ with eigenvalue μ_a, then, for any τ, it is also eigenvector of $\boldsymbol{Q}(\tau)$ with eigenvalue $\mathrm{e}^{\mu_a \tau}$. This means that we can write

$$\boldsymbol{Q}(\tau) = \sum_{a=1}^{N} \mathrm{e}^{\mu_a \tau}\, |\boldsymbol{\psi}^a\rangle\langle\boldsymbol{\phi}^a|. \tag{2.24}$$

Remembering the fundamental properties of $\boldsymbol{Q}$ and applying the Perron–Frobenius theorem to $\boldsymbol{Q}(\tau)$ at given τ exactly as we did for discrete-time Markov chains, we can give a series of statements. First of all, $\boldsymbol{Q}(\tau)$ has always an eigenvalue equal to 1, so $\boldsymbol{W}$ has an eigenvalue $\mu_1 = 0$, as we already noted. We also know that the corresponding left eigenvector is $\langle\mathbf{1}|$, and the corresponding right eigenvector is a probability vector $|\boldsymbol{\Pi}\rangle$, such that $\boldsymbol{Q}(\tau)|\boldsymbol{\Pi}\rangle = |\boldsymbol{\Pi}\rangle$ and therefore $\boldsymbol{W}|\boldsymbol{\Pi}\rangle = |\mathbf{0}\rangle$.

If $\boldsymbol{Q}(\tau)$ is irreducible, being also aperiodic for small time τ, it is *regular*: this means there is *only one* eigenvalue λ_1 of $\boldsymbol{Q}(\tau)$ which has modulus equal to 1, and therefore only one eigenvalue of $\boldsymbol{W}$ such that $\mu_1 = 0$. Moreover, all other eigenvalues λ_a of $\boldsymbol{Q}(\tau)$ have $|\lambda_a| < 1$, meaning that $|\,\mathrm{e}^{\mu_a \tau}\,| < 1$, i.e., $\Re(\mu_a) < 0$. As a consequence, *in the evolution under an irreducible Markov process, a generic initial state converges to the steady state*,

$$|\boldsymbol{P}(t)\rangle = \mathrm{e}^{t\boldsymbol{W}}\,|\boldsymbol{P}(0)\rangle = \sum_{a=1}^{N} \alpha_a(0)\,\mathrm{e}^{\mu_a t}\,|\boldsymbol{\psi}^a\rangle \xrightarrow{t\to+\infty} |\boldsymbol{\Pi}\rangle. \tag{2.25}$$

🔍 Suppose that $\boldsymbol{W}$ satisfies detailed balance with its steady state, i.e., $W_{ij}\Pi_j = W_{ji}\Pi_i$. This implies

$$\frac{1}{\sqrt{\Pi_j}} W_{ji} \sqrt{\Pi_i} = \frac{1}{\sqrt{\Pi_i}} W_{ij} \sqrt{\Pi_j}$$

We can define then the matrix $\boldsymbol{M} = \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{1/2}$, where $D_{ij} = \delta_{ij}\Pi_i$. The equation $\boldsymbol{M}$ is by construction related to $\boldsymbol{W}$ by an orthogonal transformation, plus the equation above tells us that $\boldsymbol{M}$ is symmetric. This implies that $\boldsymbol{M}$ has the same eigenvalues as $\boldsymbol{W}$ and it has a complete set of orthonormal eigenvevctors[a] $|\boldsymbol{\chi}^a\rangle$, such that $\boldsymbol{M}|\boldsymbol{\chi}^a\rangle = \mu_a|\boldsymbol{\chi}^a\rangle$. So

$$\boldsymbol{M} = \sum_a \mu_a |\boldsymbol{\chi}^a\rangle\langle\boldsymbol{\chi}^a| \Leftrightarrow \boldsymbol{W} = \sum_a \mu_a \boldsymbol{D}^{1/2}|\boldsymbol{\chi}^a\rangle\langle\boldsymbol{\chi}^a|\boldsymbol{D}^{-1/2}. \tag{2.26}$$

The set

$$|\boldsymbol{\psi}^a\rangle = \boldsymbol{D}^{1/2}|\boldsymbol{\chi}^a\rangle \qquad \langle\boldsymbol{\phi}^a| = \langle\boldsymbol{\chi}^a|\boldsymbol{D}^{-1/2} \qquad a \in [N] \tag{2.27}$$

and form a complete, orthonormal set, as $\langle\boldsymbol{\phi}^b|\boldsymbol{\psi}^a\rangle = \langle\boldsymbol{\chi}^b|\boldsymbol{\chi}^a\rangle = \delta_{ab}$.

[a] This time, left and right eigenvectors are identical because the matrix is symmetric.

2.1. H-theorem. One of the main results concerning the evolution of $|\boldsymbol{P}(t)\rangle$ towards $|\boldsymbol{\Pi}\rangle$ is the so-called *H-theorem*, introduced by Ludwig Boltzmann in 1872. In its original formulation, Boltzmann H-theorem proves that a certain functional, called *entropy* and defined as

$$\mathcal{H}[\boldsymbol{P}(t)] = -\sum_i P_i(t) \ln P_i(t) \tag{2.28}$$

is a non-decreasing function of time in systems with symmetric transition rates $\boldsymbol{W} = \boldsymbol{W}^\top$, and that the time-dependent distribution converges to the uniform distribution, which maximizes the entropy.

We will give a precise statement below, but, before that, we can work in a slightly more general setting and prove that detail balance in the steady state is a sufficient (although not necessary) condition for any probability $|\boldsymbol{P}(t)\rangle$ to converge to a steady state. To prove this statement, we consider the Kullback–Leibler divergence between the steady state $|\boldsymbol{\Pi}\rangle$ and a different distribution $|\boldsymbol{P}(t)\rangle$. This is defined as

$$\mathrm{KL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi}) \coloneqq \sum_i P_i(t) \ln \frac{P_i(t)}{\Pi_i}. \tag{2.29}$$

This is a sort of "distance" between the two distributions and satisfies a series of properties. For example, it is always non-negative, and it is equal to zero if, and only if, the two distributions are the same, as proven in the following Lemma.

LEMMA 2.1 (Gibbs inequality). *The Kullback–Leibler divergence between two probability distributions $|\boldsymbol{P}\rangle$ and $|\boldsymbol{P}'\rangle$ is such that*

$$\mathrm{KL}(\boldsymbol{P}\|\boldsymbol{P}') \geq 0. \tag{2.30}$$

Moreover, $\mathrm{KL}(\boldsymbol{P}\|\boldsymbol{P}') = 0$ if and only if the the two distributions are identical, $|\boldsymbol{P}\rangle = |\boldsymbol{P}'\rangle$.

PROOF. Let us show prove the previous lemma. We start observing that, being $\ln x > x - 1$ for $0 < x < 1$, then

$$\mathrm{KL}(\boldsymbol{P}\|\boldsymbol{P}') = \sum_i P_i \ln \frac{P_i}{P_i'} = -\sum_i P_i \ln \frac{P_i'}{P_i} \geq -\sum_i P_i' + \sum_i P_i = 0. \tag{2.31}$$

It is evident that, if $|\boldsymbol{P}\rangle = |\boldsymbol{P}'\rangle$, then $\mathrm{KL}(\boldsymbol{P}\|\boldsymbol{P}') = 0$. On the other hand, in Eq. (2.31) equality holds if $P_i = P_i'$ for all i, so that $\ln x = x - 1$ exactly. □

Using $\mathrm{KL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi})$ as measure of distance between $|\boldsymbol{P}(t)\rangle$ and $|\boldsymbol{\Pi}\rangle$, we can show that this quantity decreases with the evolution of our system if we assume that $|\boldsymbol{\Pi}\rangle$ satisfies detailed balance, and in particular $|\boldsymbol{P}(t)\rangle$ evolves precisely towards $|\boldsymbol{\Pi}\rangle$.

PROPOSITION 2.2. *Let us consider a finite Markov chain having stationary state $|\boldsymbol{\Pi}\rangle$ and let us assume that such state satisfies detailed balance. Then the Kullback–Leibler divergence $\mathrm{KL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi})$ is a decreasing function in time. Moreover, for $t \to +\infty$, $|\boldsymbol{P}(t)\rangle \to |\boldsymbol{\Pi}\rangle$.*

PROOF. Let us start by showing that the Kullback–Leibler divergence between $|\boldsymbol{P}(t)\rangle$ and $|\boldsymbol{\Pi}\rangle$ decreases with the evolution of the system. Using the master equation we have

$$\begin{aligned}\frac{\mathrm{dKL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi})}{\mathrm{d}t}&=\sum_i\left(\ln\frac{P_i(t)}{\Pi_i}+1\right)\frac{\mathrm{d}P_i(t)}{\mathrm{d}t}\\&=\sum_i\ln\frac{P_i(t)}{\Pi_i}\sum_j(W_{ij}P_j(t)-W_{ji}P_i(t))\\&=\sum_i\ln\frac{P_i(t)}{\Pi_i}\sum_j\left(W_{ij}\Pi_j\frac{P_j(t)}{\Pi_j}-W_{ji}\Pi_i\frac{P_i(t)}{\Pi_i}\right)\\&=\sum_{ij}W_{ij}\Pi_j\ln\frac{P_i(t)}{\Pi_i}\left(\frac{P_j(t)}{\Pi_j}-\frac{P_i(t)}{\Pi_i}\right)\\&=\frac{1}{2}\sum_{ij}W_{ij}\Pi_j\left(\ln\frac{P_i(t)}{\Pi_i}-\ln\frac{P_j(t)}{\Pi_j}\right)\left(\frac{P_j(t)}{\Pi_j}-\frac{P_i(t)}{\Pi_i}\right)<0.\end{aligned}\tag{2.32}$$

Observe that in the last line the contribution $i=j$ are absent. On the other hand, for $i\neq j$ $W_{ij}\geq 0$ by definition. The last inequality derives from the general inequality $(x-y)(\ln x-\ln y)\geq 0$ for all $x,y>0$ with equality if only if $x=y$. It follows that, as $\mathrm{KL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi})$ is bounded from below, for $t\to+\infty$

$$\lim_{t\to+\infty}\frac{\mathrm{d}\,\mathrm{KL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi})}{\mathrm{d}\,t}=0\Leftrightarrow\frac{\lim_{t\to+\infty}P_i(t)}{\Pi_i}=\frac{\lim_{t\to+\infty}P_j(t)}{\Pi_j}\quad\forall i,j.\tag{2.33}$$

In other words, $\frac{1}{\Pi_i}\lim_{t\to+\infty}P_i(t)=\gamma$ is some constant that does not depend on i. But then $1=\sum_i\lim_{t\to+\infty}P_i(t)=\gamma\sum_i\Pi_i=\gamma$, which means that, $\forall i$, $P_i(t)\to\Pi_i$. □

We can now state and prove the core result of this subsection.

THEOREM 2.3 (H-theorem). *Let us consider a Markov process on N states and let us assume $\boldsymbol{W}=\boldsymbol{W}^\top$. Then its steady state is $|\boldsymbol{\Pi}\rangle=\frac{1}{N}|\mathbf{1}\rangle$ and, moreover,*

$$\frac{\mathrm{d}\,\mathcal{H}[\boldsymbol{P}(t)]}{\mathrm{d}\,t}\geq 0.\tag{2.34}$$

PROOF. Let us first prove that $\Pi_i=1/N$, i.e., $|\boldsymbol{\Pi}\rangle=\frac{1}{N}|\mathbf{1}\rangle$ is indeed a steady state. It is enough to observe that, as $\langle\mathbf{1}|\boldsymbol{W}=\langle\mathbf{0}|$, being $\boldsymbol{W}$ symmetric, $\frac{1}{N}\boldsymbol{W}|\mathbf{1}\rangle=(\langle\mathbf{1}|\boldsymbol{W})^\top=(\langle\mathbf{0}|)^\top=|\mathbf{0}\rangle$. To complete the proof, note that

$$\mathcal{H}[\boldsymbol{P}(t)]=-\mathrm{KL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi})-\sum_i P_i(t)\ln\Pi_i=-\mathrm{KL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi})-\ln\frac{1}{N}\sum_i P_i(t)=-\mathrm{KL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi})+\ln N.\tag{2.35}$$

Taking a time derivative and applying Proposition 2.2

$$\frac{\mathrm{d}\,\mathcal{H}[\boldsymbol{P}(t)]}{\mathrm{d}\,t}=-\frac{\mathrm{d}\,\mathrm{KL}(\boldsymbol{P}(t)\|\boldsymbol{\Pi})}{\mathrm{d}\,t}\geq 0.\tag{2.36}$$

□

2.2. Expectations and correlations. We can derive dynamical equations for the averages, directly from the master equation, without solving the latter. Let us consider for example the average $\mathbb{E}[f(\mathsf{X}_t)]=\sum_i f_iP_i(t)$ of a function f of a our process X_t. Then

$$\frac{\mathrm{d}\,\mathbb{E}[f(\mathsf{X}_t)]}{\mathrm{d}\,t}=\sum_i f_i\frac{\mathrm{d}\,P_i(t)}{\mathrm{d}\,t}=\sum_{ij}f_i[W_{ij}P_j(t)-W_{ji}P_i(t)]=\sum_{ij}(f_i-f_j)W_{ij}P_j(t).\tag{2.37}$$

Two-time averages can be easily rewritten in terms of matrix $\boldsymbol{W}$. For example, given two observables A_{X_t} and B_{X_t},

$$\mathbb{E}[A_{\mathsf{X}_{t+\tau}} B_{\mathsf{X}_t}] = \sum_{ij} \mathbb{P}_2[\mathsf{X}_{t+\tau} = i; \mathsf{X}_t = j] A_i B_j = \sum_{ij} \mathbb{P}_{1|1}[\mathsf{X}_{t+\tau} = i | \mathsf{X}_t = j] P_j(t) A_i B_j = \sum_{ij} Q_{ij}(\tau) P_j(t) A_i B_j. \tag{2.38}$$

In other words, a two-time average is weighted with the distribution of *dynamical paths* starting at time t and ending at time $t+\tau$. At stationarity, $\mathbb{E}[A_{\mathsf{X}_{t+\tau}} B_{\mathsf{X}_t}] = \sum_{ij} Q_{ij}(\tau)\Pi_j A_i B_j$ is a function of the time difference τ only, as we know. If detailed balance is satisfied, $C_{AB}(\tau) = \sum_{ij} Q_{ij}(\tau)\Pi_j A_i B_j = \sum_{ij} Q_{ji}(\tau)\Pi_i A_i B_j = C_{BA}(\tau)$. Hence, stationarity and detailed balance jointly lead to time-reversal symmetry $C_{AB}(\tau) = C_{BA}(\tau) = C_{AB}(-\tau)$ as expected.

A useful observable, when analysing stochastic trajectories X_t is the occupancy-number function, defined as $\theta_i(t) := \mathbb{I}(\mathsf{X}_t = i)$, a stochastic process itself, equal to 1 if $\mathsf{X}_t = i$ and zero otherwise. The occupancy-number correlation function $C_{ij}(t,t') := \mathbb{E}[\theta_i(t)\theta_j(t')]$ gives the probability to find the system is in i at time t when released in j at time t'. At stationarity, using the spectral representation of $\boldsymbol{W}$, one has

$$\mathbb{E}[\theta_i(t+\tau)\theta_j(t)] = Q_{ij}(\tau)\Pi_j = \sum_{a=1}^{N} \mathrm{e}^{\tau\mu_a}\, \psi_i^a \phi_j^a \Pi_j \tag{2.39}$$

The connected correlator is then

$$\langle\!\langle \theta_i(t+\tau)\theta_j(t) \rangle\!\rangle = Q_{ij}(\tau)\Pi_j - \Pi_i\Pi_j = \sum_{a=1}^{N} \mathrm{e}^{\tau\mu_a}\, \psi_i^a \phi_j^a \Pi_j - \Pi_i\Pi_j = \sum_{a=2}^{N} \mathrm{e}^{\tau\mu_a}\, \psi_i^a \phi_j^a \Pi_j \tag{2.40}$$

i.e., a superposition of $N-1$ exponential functions of time. In particular, for 2-state Markov models, it is a single exponential function $\langle\!\langle \theta_i(t+\tau)\theta_j(t) \rangle\!\rangle = \mathrm{e}^{\tau\mu_2} \langle\!\langle \theta_i(0)\theta_j(0) \rangle\!\rangle$.

3. Hitting times

Results on hitting times in continuous-time dynamics can be derived directly from what we obtained for Markov chains by setting the time step to τ in the discrete time dynamics, and taking the limit $\tau \to 0$ at the end. The eigenvalues of the matrix $\boldsymbol{Q}(\tau)$ are given by $\lambda_a(\tau) = \mathrm{e}^{\tau\mu_a}$, as we saw. For a Markov chain evolving in discrete time t_{ji} was the average number of steps to go from state i to state j. To get the *time* required to go from i to j we need to multiply both sides of (1.129) by τ, unit of time,

$$t_{ji}\tau = \frac{\tau}{\Pi_j}\left[1 + \sum_{a>1} \frac{\mathrm{e}^{\mu_a\tau}}{1-\mathrm{e}^{\mu_a\tau}} \psi_j^a(\phi_j^a - \phi_i^a)\right]. \tag{2.41}$$

Let us call $t_{ji}\tau = \hat{t}_{ji}$ the hitting time, and take the limit $\tau \to 0$. We obtain

$$\hat{t}_{ji} = -\frac{1}{\Pi_j} \sum_{a>1} \frac{\psi_j^a(\phi_j^a - \phi_i^a)}{\mu_a}. \tag{2.42}$$

In contrast to the discrete time result, we have

$$\hat{t}_{jj} = 0, \tag{2.43}$$

which is intuitively understood, as here there is no time step to wait to return to the state (and all states are aperiodic!).

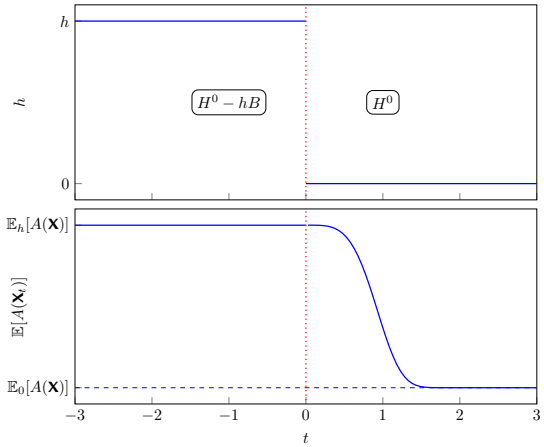


FIGURE 1. Pictorial representation of the perturbation discussed in Section 4: at $t = 0$ the parameter h is put to 0 and the expectation of the observable A start to relax towards a new equilibrium value.

Eq. (1.131) also can be treated in the $\tau \to 0$ limit to give us the Kemeny constant ζ in continuous time as

$$\sum_{j\neq i} \Pi_j \hat{t}_{jk} = \sum_{j\neq i} \Pi_j t_{jk} \tau = \sum_{a>1} \frac{\tau}{1 - \mathrm{e}^{\tau\mu_a}} \xrightarrow{\tau\to 0} -\sum_{a>1} \frac{1}{\mu_a} \equiv \zeta. \tag{2.44}$$

4. Linear response theory

LECTURE 6

In this section we examine the relaxation towards equilibrium of a system that has been brought out of equilibrium by an external perturbation. The main result is that for small deviations from equilibrium this relaxation is described by the equilibrium time correlation function. This is the so-called *Onsager regression law*. To do so, we will use a slightly different notation than before, assuming that our stochastic process $\mathbf{X}_t$ consists of a vectorial random quantity. We will denote $\boldsymbol{x}$ the values that $\mathbf{X}_t$ can take. This is just a reminder of the fact that the treatment includes the case in which the stochastic process involves many degrees of freedom.

We will assume that the system is described by a function, the *classical Hamiltonian*

$$H(\boldsymbol{x}) = H_0(\boldsymbol{x}) - hB(\boldsymbol{x}), \tag{2.45}$$

where $H_0(\boldsymbol{x})$ and $B(\boldsymbol{x})$ are some observables, i.e., functions of a value $\boldsymbol{x}$ that the stochastic process can take. We imagine that, since $t = -\infty$, the system is in contact with a infinitely large reservoire at temperature β^{-1}, and in equilibrium with it. Statistical mechanics tells us that the equilibrium distribution can be written as

$$\Pi_{\boldsymbol{x}} = \frac{\mathrm{e}^{-\beta H(\boldsymbol{x})}}{\sum_{\boldsymbol{x}} \mathrm{e}^{-\beta H(\boldsymbol{x})}} = \frac{\mathrm{e}^{-\beta H_0(\boldsymbol{x}) + \beta h B(\boldsymbol{x})}}{\sum_{\boldsymbol{x}} \mathrm{e}^{-\beta H_0(\boldsymbol{x}) + \beta h B(\boldsymbol{x})}}. \tag{2.46}$$

An observable $A(\mathbf{X})$ is averaged as $\mathbb{E}[A(\mathbf{X})] = \sum_{\boldsymbol{x}} A(\boldsymbol{x})\Pi_{\boldsymbol{x}} \equiv \mathbb{E}_h[A(\mathbf{X})]$, without any time dependence as we are at equilibrium (the reason of the subscript h will be clear in a moment). We imagine that this is the set-up for $t < 0$, then, at time $t = 0$, we perturb the Hamiltonian $H(\boldsymbol{x})$ by *turning off* the additional term in $B(\boldsymbol{x})$, i.e., by suddenly putting $h = 0$. This changes the Hamiltonian of the system to $H_0(\boldsymbol{x})$: the system, which was at equilibrium for $t < 0$, finds

itself *out of equilibrium* as the Hamiltonian has suddenly changed. What we expect is that, for $t \to +\infty$, the system will reach a new equilibrium, namely

$$\Pi^0_{\boldsymbol{x}} = \frac{\mathrm{e}^{-\beta H_0(\boldsymbol{x})}}{Z_0} \qquad Z_0 \coloneqq \sum_{\boldsymbol{x}} \mathrm{e}^{-\beta H_0(\boldsymbol{x})}. \tag{2.47}$$

so that, for example, the same observable $A(\mathbf{X})$ will have, for $t \to +\infty$, expectation $\mathbb{E}_0[A(\mathbf{X})]$ (we have added a subscript 0 to stress that it is the *equilibrium* expectation at $h = 0$). However, how $\mathbb{E}[A(\mathbf{X}_t)]$ will approach this new equilibrium value after the change of the parameter?

The starting point of our study is the equilibrium distribution $\Pi_{\boldsymbol{x}}$, which plays the role of *initial distribution* $|\boldsymbol{P}(0)\rangle \equiv |\boldsymbol{\Pi}\rangle$. For small h, we can Taylor-expand numerator and denominator of Eq. (2.46), to get

$$\Pi_{\boldsymbol{x}} \simeq \frac{\mathrm{e}^{-\beta H_0(\boldsymbol{x})}(1+\beta h B(\boldsymbol{x}))}{\sum_{\boldsymbol{x}} \mathrm{e}^{-\beta H_0(\boldsymbol{x})}(1+\beta h B(\boldsymbol{x}))} = \frac{\mathrm{e}^{-\beta H_0(\boldsymbol{x})}(1+\beta h B(\boldsymbol{x}))}{Z_0(1+\beta h \mathbb{E}_0[B(\mathbf{X})])} = \Pi^0_{\boldsymbol{x}}\Big(1+\beta h(B(\boldsymbol{x})-\mathbb{E}_0[B(\mathbf{X})])\Big). \tag{2.48}$$

To make expressions lighters, we will use in the following the notation

$$\delta B(\boldsymbol{x}) \coloneqq B(\boldsymbol{x}) - \mathbb{E}_0[B(\mathbf{X})]$$

so that

$$\Pi_{\boldsymbol{x}} \simeq \Pi^0_{\boldsymbol{x}}\Big(1 + \beta h\, \delta B(\boldsymbol{x})\Big). \tag{2.49}$$

Next, we calculate the expectation value of $A(\mathbf{X}_t)$ for $t > 0$, i.e., in the non-equilibrium regime:

$$\begin{aligned}
\mathbb{E}[A(\mathbf{X}_t)] &= \sum_{\boldsymbol{x}} A(\boldsymbol{x}) P_{\boldsymbol{x}}(t) \\
&= \sum_{\boldsymbol{x}} A(\boldsymbol{x}) \sum_{\boldsymbol{y}} \mathbb{P}^0_{1|1}[\mathbf{X}_t = \boldsymbol{x} | \mathbf{X}_0 = \boldsymbol{y}] \Pi_{\boldsymbol{y}} \\
&\simeq \sum_{\boldsymbol{x}} A(\boldsymbol{x}) \sum_{\boldsymbol{y}} \mathbb{P}^0_{1|1}[\mathbf{X}_t = \boldsymbol{x} | \mathbf{X}_0 = \boldsymbol{y}] \Pi^0_{\boldsymbol{y}} \Big(1 + \beta h\, \delta B(\boldsymbol{y})\Big) \\
&= \mathbb{E}_0[A(\mathbf{X})] + \beta h \Big(\mathbb{E}_0[A(\mathbf{X}_t)B(\mathbf{X}_0)] - \mathbb{E}_0[A(\mathbf{X})]\mathbb{E}_0[B(\mathbf{X})]\Big) \\
&= \mathbb{E}_0[A(\mathbf{X})] + \beta h \langle\!\langle A(\mathbf{X}_t) B(\mathbf{X}_0) \rangle\!\rangle_0
\end{aligned} \tag{2.50}$$

In the equations above, we have stressed, using $\mathbb{P}^0_{1|1}$, that the evolution takes place with $h = 0$ and therefore $|\boldsymbol{\Pi}^0\rangle$ is stationary with respect to it. Hence,

$$\mathbb{E}[\delta A(\mathbf{X}_t)] \coloneqq \mathbb{E}[A(\mathbf{X}_t)] - \mathbb{E}_0[A(\mathbf{X})] = \beta h \langle\!\langle A(\mathbf{X}_t) B(\mathbf{X}_0) \rangle\!\rangle_0. \tag{2.51}$$

A special case is obtained when $A = B$ and the equation takes the form

$$\mathbb{E}[\delta B(\mathbf{X}_t)] = \beta h \langle\!\langle B(\mathbf{X}_t) B(\mathbf{X}_0) \rangle\!\rangle_0. \tag{2.52}$$

Eq. (2.52) is normally referred to as Onsager regression law and involves an *out-of-equilibrium relaxation* induced by a perturbation (on the left) expressed in terms of a *fluctuation at equilibrium* (on the right). In other words, *the relaxation of macroscopic non-equilibrium perturbation is governed by the same laws as the regression of spontaneous microscopic fluctuations at equilibrium.*

We can expect $\mathbb{E}[\delta A(\mathbf{X}_t)]$ to depends linearly from h: for $h \to 0$, after all, $\mathbb{E}[\delta A(\mathbf{X}_t)] \to 0$. We can make therefore the following ansatz

$$\mathbb{E}[\delta A(\mathbf{X}_t)] = h \int_{-\infty}^{0} R_{AB}(t, t')\, \mathrm{d}\, t' \tag{2.53}$$

where we have introduced a *response function* $R_{AB}(t,t')$: this function measure how much the fact that h was on at time $t' < 0$ affects $\mathbb{E}[\delta A(\mathbf{X}_t)]$ at time $t > 0$. In particular, under the assumption of stationarity, we expect the precise times t and t' are not important and, instead, R_{AB} will depend on the *time interval* $\tau = t - t'$ between the perturbation and the effects we observe. Moreover, for reasons of causality we must have

$$R_{AB}(\tau) = 0 \qquad \text{if } \tau < 0. \tag{2.54}$$

Indeed, the fact that h was on at some time t' cannot affect $\mathbb{E}[\delta A(\mathbf{X}_t)]$ at some previous time $t < t'$. Assuming $t > 0$, we can rewrite then

$$\mathbb{E}[\delta A(\mathbf{X}_t)] = h \int_{-\infty}^{0} R_{AB}(t-t')\,\mathrm{d}\,t' = h \int_{t}^{\infty} R_{AB}(\tau)\,\mathrm{d}\,\tau. \tag{2.55}$$

Inserting this relation into (2.51) and differentiating with respect to time then gives

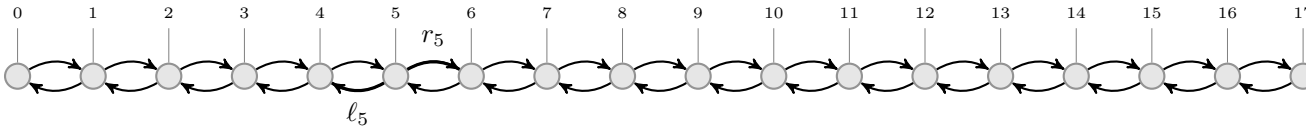
$$R_{AB}(t) = -\beta\theta(t)\frac{\mathrm{d}}{\mathrm{d}\,t}\langle\!\langle A(\mathbf{X}_t)B(\mathbf{X}_0)\rangle\!\rangle_0. \tag{2.56}$$

This relation is known as *fluctuation-dissipation theorem* (FDT). A special case of particular interest is the one for which the variable $A(\mathbf{X})$ is chosen to be equal to $B(\mathbf{X})$. In this case, if $\langle\!\langle B(\mathbf{X}_t)B(\mathbf{X}_0)\rangle\!\rangle_0 \equiv C_B(t)$ is the connected self-correlation of B, and dropping one subscript $R_{BB} \equiv R_B$, the FDT reads as

$$R_B(t) = -\beta\theta(t)\dot{C}_B(t). \tag{2.57}$$

5. Jump processes

The family of jump processes includes several processes of practical interest, such as Poisson processes, random walks, and linear birth-death processes. In a jump process, the set of states Ω is discrete and can be labeled by integer numbers, in such a way that the system can move only between adjacent states. A pictorial representation is the following,



In this chapter we will assume, unless otherwise specified, that $\Omega = \mathbb{Z}$ or $\Omega = \mathbb{N}_0$: in other words, we will assume that the space of possible configurations of our system is *infinite* but countable. Moreover, as we are assuming that there is an ordering between states (which is the ordering of $\mathbb{Z}$ of $\mathbb{N}_0$) and the system jumps only between neighbouring configurations, the relevant rates are, for each $k \in \Omega$, the rate of moving "right", $r_k \coloneqq W_{k+1\,k}$, and $\ell_k \coloneqq W_{k-1\,k}$ rate of moving left. Therefore, the master equation reads

$$\frac{\mathrm{d}\,P_k(t)}{\mathrm{d}\,t} = \ell_{k+1}P_{k+1}(t) + r_{k-1}P_{k-1}(t) - (\ell_k + r_k)P_k(t). \tag{2.58}$$

We can at this point specify different type of very relevant jump processes, depending on the values of ℓ_k and r_k.

5.1. The Poisson process. In Poisson processes $\Omega = \mathbb{N}_0$, so that $i = 0, 1, \dots$ with steps to the right only, occurring at random times with fixed rate. This means in particular

$$r_k \equiv \lambda, \qquad \ell_k = 0, \qquad \forall k. \tag{2.59}$$

Here $\lambda \in \mathbb{R}^+$ is constant. A typical Poisson process is the one concerning the probability that N_t independent events occurred at time t. These events could be for example customers arriving in a shop, or jobs arriving in a queuing system, or particles emitted from unstable nuclei. Usually, the number of events at time $t = 0$ is set to zero, i.e., $\mathsf{N}_0 = 0$ and therefore $P_k(0) = \delta_{k0}$. Given that steps occur only to the right, $P_k(t) = 0 \ \forall \ k < 0$ and $\forall \ t$. The master equation for Poisson process can be promptly written as

$$\dot{P}_k = \lambda(P_{k-1} - P_k), \tag{2.60}$$

and is most easily solved through use of the *generating function* method, a method we will used frequently in this Chapter. In this method we introduce an auxiliary function

$$F(z,t) \coloneqq \sum_{k=0}^{\infty} P_k(t) z^k \tag{2.61}$$

for $P_k(t) \coloneqq \mathbb{P}[\mathsf{N}_t = k]$. Note that, first of all, $F(1,t) = \sum_{k=0}^{\infty} P_k(t) = 1 \ \forall t$: this is always true and simply expresses the normalisation of probability. Various moments of the stochastic process N_t are obtained by taking derivatives of $F(z,t)$ with respect to z and allowing $z \to 1$. For example

$$\begin{aligned} \lim_{z\to 1} \frac{\partial F(z,t)}{\partial z} &= \sum_{k=0}^{\infty} k P_k(t) = \mathbb{E}[\mathsf{X}_t], \\ \lim_{z\to 1} \frac{\partial^2 F(z,t)}{\partial z^2} &= \sum_{k=0}^{\infty} k(k-1) P_k(t) = \mathbb{E}[\mathsf{X}_t^2] - \mathbb{E}[\mathsf{X}_t], \end{aligned} \tag{2.62}$$

and so on. From the master equations (2.60), we can obtain a single equation for the generating function $F(z,t)$, by multiplying (2.60) by z^k and summing over k,

$$\begin{aligned} \frac{\partial F(z,t)}{\partial t} = \sum_{k=0}^{\infty} z^k \dot{P}_k(t) &= \lambda \sum_{k=0}^{\infty} z^k (P_{k-1}(t) - P_k(t)) \\ &= \lambda z \sum_{k=1}^{\infty} z^{k-1} P_{k-1}(t) - \lambda \sum_{k=0}^{\infty} z^k P_k(t) \\ &= \lambda z \sum_{k=0}^{\infty} z^k P_k(t) - \lambda \sum_{k=0}^{\infty} z^k P_k(t) \\ &= \lambda(z-1) F(z,t) \end{aligned} \tag{2.63}$$

where we have used $P_{-1}(t) = 0$ and shifted the index $k \to k+1$ in the first sum (this does not affect the upper bound as it is infinite). The solution of the previous equation is

$$F(z,t) = F(z,0)\, \mathrm{e}^{\lambda(z-1)t} \tag{2.64}$$

and contains an arbitrary function $F(z,0) = \sum_{k=0}^{\infty} P_k(0) z^k$ which is determined from initial conditions. If we assume $P_k(0) = \delta_{k\,0}$, we get $F(z,0) = 1$. Having $F(z,t)$, we can find $P_k(t)$ by expanding $F(z,t) = \mathrm{e}^{\lambda(z-1)t}$ in powers of z

$$F(z,t) = \mathrm{e}^{\lambda(z-1)t} = \mathrm{e}^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda z t)^k}{k!} \to P_k(t) = \frac{(\lambda t)^k}{k!}\, \mathrm{e}^{-\lambda t} \tag{2.65}$$

which is a *Poissonian distribution* (hence the name of the process) with average $\mathbb{E}[\mathsf{X}_t] = \lambda t$.

Distribution of waiting times. — In a Poisson process, the random variable N_t increases by one with probability $\lambda\Delta t$ during the interval Δt, independently from the history of the process up to that moment: this is what $r_k \equiv \lambda$ means. A realisation of the process has the form in Fig. 2. Let us set $t = 0$ as the time origin. An event may or may not have occurred at $t = 0$. Any

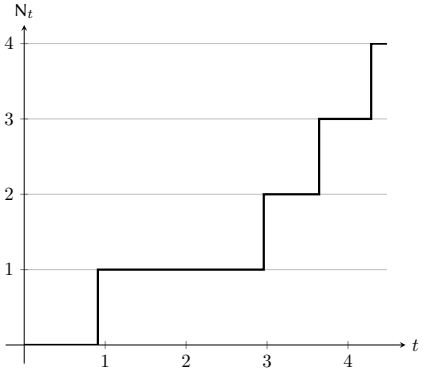


FIGURE 2. Sample path of a Poissonian process with $\lambda = 1$.

property of the process after $t = 0$ is independent of what happened at $t = 0$ or before. Let T be the time at which the *first* event occurs, e.g., the first customer arrives. This is a random variable that we can formally define as $\mathsf{T} = \min_{t\geq 0}\{t\colon \mathsf{N}_t = 1\}$. We will call its probability distribution $\Phi_{\mathsf{T}}(t) = \mathbb{P}[\mathsf{T} > t]$. In particular, $\Phi_{\mathsf{T}}(t)$ is the probability that T is larger than t. Then

$$\begin{aligned}\Phi_{\mathsf{T}}(t+\Delta t) &= \mathbb{P}[\mathsf{T} > t+\Delta t] = \mathbb{P}[\mathsf{T} > t,\ \mathsf{N}_{t+\Delta t} = \mathsf{N}_t]\\ &= \mathbb{P}[\mathsf{T} > t]\mathbb{P}[\mathsf{N}_{t+\Delta t} = \mathsf{N}_t] = \Phi_{\mathsf{T}}(t)\mathbb{P}[\mathsf{N}_{\Delta t} = \mathsf{N}_0]\\ &= \Phi_{\mathsf{T}}(t)P_0(\Delta t) = \Phi_{\mathsf{T}}(t)\,\mathrm{e}^{-\lambda\Delta t} = \Phi_{\mathsf{T}}(t)(1-\lambda\Delta t + o(\Delta t)),\end{aligned} \tag{2.66}$$

meaning that $\frac{\Phi_{\mathsf{T}}(t+\Delta t)-\Phi_{\mathsf{T}}(t)}{\Delta t} = -\lambda\Phi_{\mathsf{T}}(t)$, i.e., taking $\Delta t \to 0$, $\partial_t\Phi_{\mathsf{T}}(t) = -\lambda\Phi_{\mathsf{T}}(t)$ and finally

$$\Phi_{\mathsf{T}}(t) = \mathrm{e}^{-\lambda t}\,. \tag{2.67}$$

where we used the fact that $\Phi_{\mathsf{T}}(0) = 1$. The probability density $p(t) = -\partial_t\Phi_{\mathsf{T}}(t)$ is then an *exponential distribution*

$$p(t) = \lambda\,\mathrm{e}^{-\lambda}\,. \tag{2.68}$$

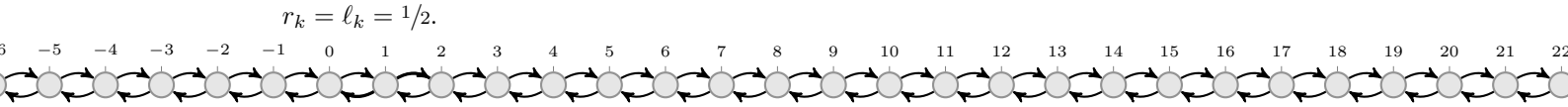
The typical time interval between two events is obtained as

$$\mathbb{E}[\mathsf{T}] = \int_0^\infty t\lambda\,\mathrm{e}^{-\lambda t}\,\mathrm{d}\,t = \frac{1}{\lambda} \tag{2.69}$$

as expected, being λ the transition rate (i.e., number of transitions per unit of time).

6. Continuous-time random walks on the line

6.1. The symmetric random walk. An unbiased one dimensional random walk process X_t can be pictured as a walker jumping on the integer positions, and has therefore $\Omega = \mathbb{Z}$. At any position k, the rate of going left and right is the same, and we can for example fix it as $r_k = \ell_k = 1/2$.



The master equation is, therefore,

$$\frac{\mathrm{d}\,P_k(t)}{\mathrm{d}\,t} = \frac{P_{k-1}(t) + P_{k+1}(t)}{2} - P_k(t). \tag{2.70}$$

As in the case of Poisson processes, a good strategy to solve this system is to write down an equation for the generating function

$$F(z,t) := \sum_{k=-\infty}^{\infty} P_k(t) z^k. \tag{2.71}$$

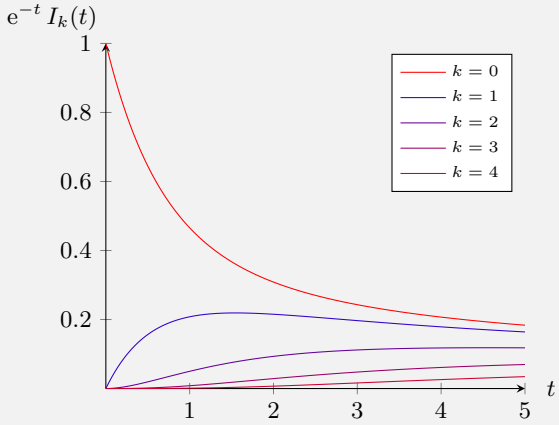
Note that this time the sum runs over *all integers*. It is a simple exercise to show that in this case $F(z,t)$ satisfies

$$\partial_t F(z,t) = \left(\frac{z+z^{-1}}{2} - 1\right) F(z,t) \Rightarrow F(z,t) = F(z,0)\,\mathrm{e}^{\frac{1}{2}(z+z^{-1}-2)t}. \tag{2.72}$$

The generating function that we obtained might look a little bit convoluted and not extremely easy to expand. But it turns out that is related to important special functions, namely the modified Bessel functions $I_k(t)$, as described below.

🔍 The modified Bessel functions of the first kind $I_k(t)$ are a family of functions satisfying the equations

$$\frac{\mathrm{d}\,I_k(t)}{\mathrm{d}\,t} = \frac{I_{k-1}(t) + I_{k+1}(t)}{2}, \quad \text{with } I_k(0) = \delta_{k,0} \text{ and } I_k(t) = I_{-k}(t). \tag{2.73}$$



A useful property for us of modified Bessel functions is that if we consider $k \gg 1$ and $t \gg 1$ such that $\frac{k^2}{t}$ is kept fixed, then

$$I_k(t) \simeq \frac{1}{\sqrt{2\pi t}}\,\mathrm{e}^{t-\frac{k^2}{2t}}. \tag{2.74}$$

Introducing the generating function

$$G(z,t) := \sum_{k=-\infty}^{\infty} I_k(t) z^k$$

it turns out that

$$G(z,t) = \mathrm{e}^{\frac{1}{2}(z+z^{-1})t}. \tag{2.75}$$

Using the results of the box above, we can write

$$\begin{aligned} F(z,t) &= \mathrm{e}^{-t}\, F(z,0)G(z,t) = \mathrm{e}^{-t}\, F(z,0) \sum_{k=-\infty}^{\infty} I_k(t) z^k = \mathrm{e}^{-t} \sum_n z^n P_n(0) \sum_k z^k I_k(t) \\ &= \mathrm{e}^{-t} \sum_k \left(\sum_n I_{k-n}(t) P_n(0) \right) z^k \end{aligned} \tag{2.76}$$

This means that

$$P_k(t) = \mathrm{e}^{-t} \sum_n I_{k-n}(t) P_n(0). \tag{2.77}$$

This is an interesting structure: it is telling us that the *propagator* to move from n to k in time t is $Q_{kn}(t) = \mathrm{e}^{-t}\, I_{k-n}(t)$. In particular, if $P_k(0) = \delta_{k\,0}$,

$$P_k(t) = \mathrm{e}^{-t}\, I_k(t). \tag{2.78}$$

For $t \gg 1$ and $k \gg 1$, so that k^2/t is fixed, the asymptotic expression of the modified Bessel function can give us information on $P_k(t)$ (under the assumption $\mathsf{X}_0 = 0$):

$$P_k(t) = \mathrm{e}^{-t}\, I_k(t) \simeq \frac{1}{\sqrt{2\pi t}}\, \mathrm{e}^{-\frac{k^2}{2t}}$$

which has Gaussian shape.

🔍 The case of a biased one-dimensional random walker can be treated in a similar way. In such setting the two rates are different, for example $r_n = \lambda$ and $\ell_n = \mu$, $\mu \neq \lambda$, so that the master equation is

$$\dot{P}_k(t) = \lambda P_{k-1}(t) + \mu P_{k+1}(t) - (\lambda + \mu) P_k(t). \tag{2.79}$$

This problem reduces to the one we have solved before, if we modify our definition of generating function to

$$\hat{F}(z,t) \equiv F\left(\sqrt{\frac{\mu}{\lambda}} z, t\right) = \sum_{n=-\infty}^{\infty} z^k P_k(t) \left(\frac{\mu}{\lambda}\right)^{\frac{k}{2}}. \tag{2.80}$$

Taking a time derivative of $\hat{F}(z,t)$ and plugging the master equation we find

$$\frac{\partial \hat{F}(z,t)}{\partial t} = \left[-(\lambda+\mu) + \sqrt{\lambda\mu} \left(z + \frac{1}{z} \right) \right] \hat{F}(z,t) \tag{2.81}$$

which is solved by

$$\hat{F}(z,t) = \hat{F}(z,0) \exp\left[-(\lambda+\mu)t + \sqrt{\lambda\mu}(z + 1/z)t \right]. \tag{2.82}$$

As you can see, the generating function of the modified Bessel functions $G(z, 2\sqrt{\mu\lambda}t)$ appears: repeating the arguments above, the expression can be expanded in series so that in the end

$$P_k(t) = \mathrm{e}^{-(\lambda+\mu)t} \sum_{n=-\infty}^{\infty} \left(\frac{\lambda}{\mu}\right)^{\frac{k-n}{2}} I_{k-n}(2\sqrt{\lambda\mu}t) P_n(0) \tag{2.83}$$

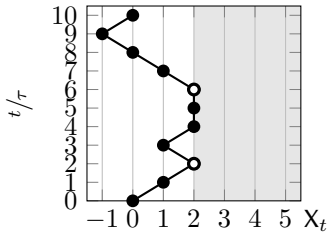
which indeed reduces to the expression previously found for $\lambda = \mu = 1/2$. The propagator, this time, is $Q_{kn}(t) = \mathrm{e}^{-(\lambda+\mu)t} \left(\frac{\lambda}{\mu}\right)^{\frac{k-n}{2}} I_{k-n}(2\sqrt{\lambda\mu}t)$.

Lecture 7

6.2. The effect of boundary conditions. So far we have considered the unbounded random walker, where $\Omega = \mathbb{Z}$ and our system can explore any possible integer position. We consider now the presence of boundaries that confine the walker in a certain region of the state space. We will distinguish three different cases: single boundary, double boundary, and periodic boundary.

Single reflecting boundary. — We start considering a single boundary, and we will focus on the *symmetric* random walk. We will assume that there is a "wall" at $k^\star$, so that the walker is confined in the region $k \leq k^\star$: for positions in the allowed regions and not on the boundaries, $W_{k\pm1\,k} = r_k = \ell_k = \mu$. We distinguish between two types of single-boundary settings, namely *reflecting* and *absorbing*. Let us start with the case of reflecting boundary.

To make the discussion simpler, let us put ourselves in the *discrete time* setting first, so that we work at time steps τ and $t = 0, \tau, 2\tau, 3\tau, \ldots$. In discrete time we have a Markov chain described by a stochastic matrix $\boldsymbol{Q}$. If there is no barrier, we fix $Q_{k\pm1\,k} = \mu\tau$ for all k, and by consequence $Q_{kk} = 1 - 2\mu\tau$; all other entries are zero (the walker can only jump between neighbour sites). If there is a barrier in $k^\star$, instead, $Q_{k\pm1\,k} = \mu\tau$ for $k < k^\star$, but $Q_{k^\star+1\,k^\star} = 0$ and $Q_{k^\star-1\,k^\star} = 2\mu\tau$. Also in this case $Q_{kk} = 1 - 2\mu\tau$ for all $k \leq k^\star$, whereas all other entries are zero. We want to estimate $P^{\mathrm{r}}_k(n\tau)$ with a barrier in $k^\star$ imaging that $\mathsf{X}_0 = 0$. This probability is the sum of all contributions of paths starting in the origin and arriving in k at time $t = n\tau$ compatibly with the presence of the barrier. For example, for $k = 0$, $k^\star = 2$ and $n = 10$, P^{r}_0 contains the contribution of the legit path below, which I will call $\mathfrak{X}$



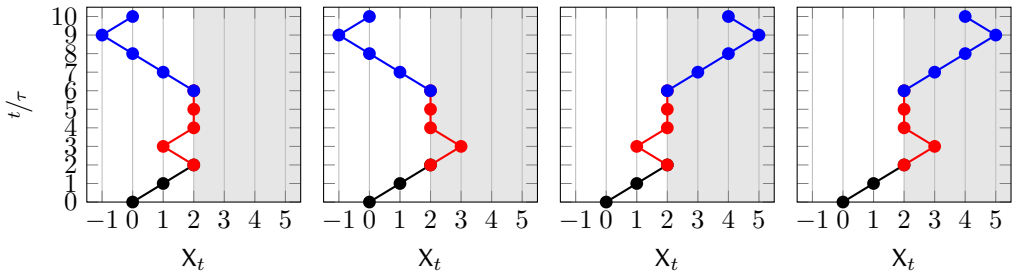
The probability of observing this path in presence of the barrier is

$$\mathbb{P}[\mathfrak{X}|\text{barrier}] = (2\mu\tau)^{n_c}(\mu\tau)^{n-n_s-n_c}(1-2\mu\tau)^{n_s} = 2^{n_c}(\mu\tau)^{n-n_s}(1-2\mu\tau)^{n_s}. \tag{2.84}$$

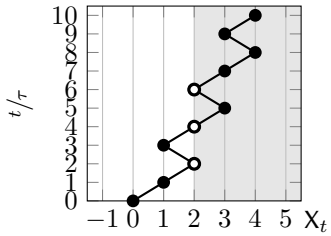
In the formula above I have denoted by n_c the number of times that the walker "take off from the barrier" (in white in the picture, where $n_c = 2$) and by n_s the number of times the walker stays where it is (in the picture $n_s = 2$, for example). Now, it is interesting to compute what is the probability of this path *in absence* of the barrier. This is

$$\mathbb{P}[\mathfrak{X}|\text{no barrier}] = (\mu\tau)^{n-n_s}(1-2\mu\tau)^{n_s}. \tag{2.85}$$

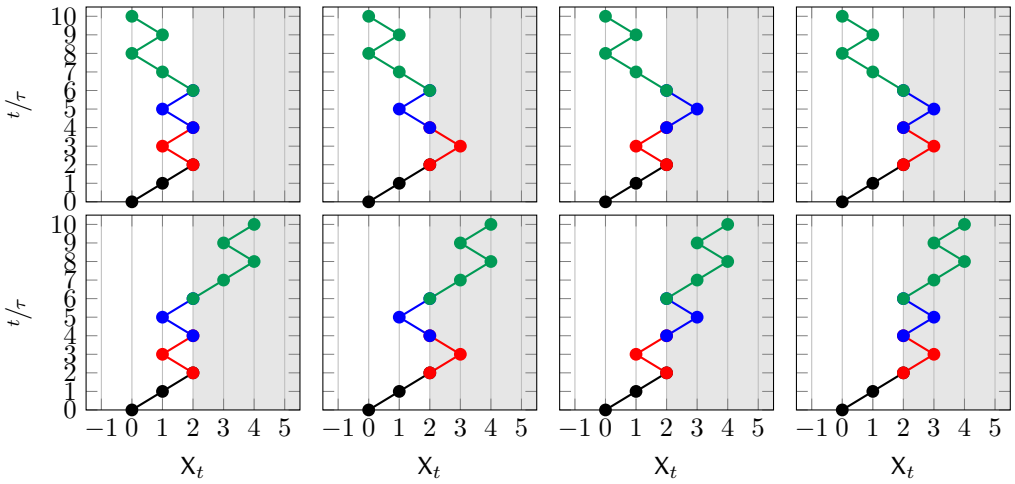
In other words $\mathbb{P}[\mathfrak{X}|\text{barrier}] = 2^{n_c}\mathbb{P}[\mathfrak{X}|\text{no barrier}]$. As in the example $n_c = 2$, the contribution to P^{r}_k of $\mathfrak{X}$ is $2^2 = 4$ times what is in P_k. Such contribution can be thought of as the sum of all 2^{n_c} paths you can obtain from the original legit path by "reflecting" the n_c portions in between "take-off points" and between the last collision and the destination. For example, in our path $\mathfrak{X}$ there are two portions, below in red and blue. You can obtain $2^2 = 4$ "sibling paths" reflecting none of the portions, the first, the second, or both. The resulting 4 paths have all the same weight in absence of the barrier $\mathbb{P}[\mathfrak{X}|\text{no barrier}] = (\mu\tau)^{n-n_s}(1-2\mu\tau)^{n_s}$:



Such sibling paths lead to k *or* to $2k^\star - k$, so that half of them are counted in P_k and half of them in $P_{2k^\star-k}$. As they are 2^{n_0}, summing their (equal) contribution automatically takes into account the combinatoric prefactor. This seems to suggest that to compute $P_k(t)$ we need to sum all contributions contained in $P_k(t)$ and all contributions contained in $P_{2k^\star-k}$, which is indeed the case. Every legit path corresponds to a family of paths counted in $P_k(t)$ and $P_{2k^\star-k}$, and vice versa every single path counted in $P_k(t)$ or $P_{2k^\star-k}(t)$ is a member of a unique family corresponding to a legit path. For example, for $k^\star = 2$ and $n = 10$, starting from the path



it has $n_c = 3$ (in white) and we can *uniquely* construct its family of $2^{n_c} = 8$ siblings in which one, and only one of them, lies in the $k < k^\star$ half-plane and correspond to a legit path in presence of a barrier:



As a result, we can write our law for $P_k^{\rm r}$ as

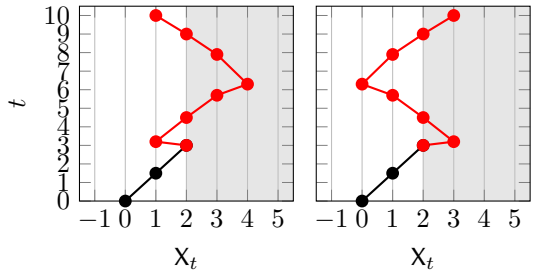
$$P_k^{\rm r}(t) = P_k(t) + P_{2k^\star-k}(t) \qquad t = n\tau, \quad n \in \mathbb{N}. \tag{2.86}$$

The relation holds for $\tau \to 0$ as well, so that we can use it in the case of continuous time random walks as well, with transition rates

$$\ell_k = r_k = \lim_{\tau\to 0} \frac{Q_{k\pm 1k}}{\tau} = \mu \text{ for } k<k^\star, \quad r_{k^\star} = \lim_{\tau\to 0} \frac{Q_{k^\star\pm 1k}}{\tau} = 0, \quad \ell_{k^\star} = \lim_{\tau\to 0} \frac{Q_{k^\star-1k}}{\tau} = 2\mu. \tag{2.87}$$

Single absorbing boundary. — In the case of an absorbing barrier, when the walker hits the wall, approaching from the left, it will be 'absorbed' by the wall and removed from the system. Computing the probability $P_k^{\rm a}(t)$ that the walker is in a location $k < k^\star$ in the presence of an absorbing boundary with similar arguments as before. First, we start observing that $P_k(t)$ contains all paths getting to k. Some of them never touch the boundary (and therefore have to be counted). Some of them instead touch it and have to be discarded from $P_k(t)$. The observation is that *each of these paths can be associated with a path going to* $2k^\star - k$ and, vice versa: it is

enough to reflect the entire original path after the first contact with the barrier. For example, the (forbidden) path below is put next to its reflected version.

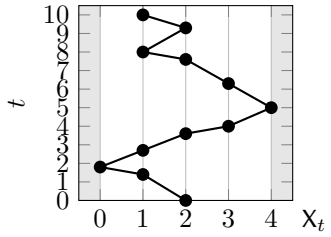


Again, the two paths "weights" are the same as we assumed that the rates to go left and right are equal. Therefore to get P_k^{a} we need to subtract the contributions of the forbidden paths which is equal to $P_{2k^\star-k}(t)$, i.e.,

$$P_k^{\mathrm{a}}(t) = P_k(t) - P_{2k^\star-k}(t). \tag{2.88}$$

Note that by construction $P_{k^\star}^{\mathrm{a}}(t) = 0$ for all t, meaning that the walker is removed from the system when it reaches the wall.

Double reflecting barrier. — Finally, let us consider the case of *two* reflecting boundaries, for example one at $k = 0$ and one at $k = K$. In this setting let us consider the generic case of *asymmetric* walker, so with different rates of going left and right.



In the presence of such barriers, $\ell_0 = r_{-1} = 0$, $\ell_{K+1} = r_K = 0$. Using generating functions and shifting indices becomes cumbersome, and gives undesired boundary terms. As the number of states, it is finite (because of the boundaries, $\Omega = \{0, 1, \dots, K\}$) we can write down the master equation in vector notation reads

$$\frac{\mathrm{d}\,|\boldsymbol{P}(t)\rangle}{\mathrm{d}\,t} = \boldsymbol{W}|\boldsymbol{P}(t)\rangle = \begin{pmatrix} -\lambda & \mu & 0 & \dots & \dots & 0 \\ \lambda & -(\lambda+\mu) & \mu & 0 & \dots & 0 \\ 0 & \lambda & -(\lambda+\mu) & \mu & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & 0 & \lambda & -\mu \end{pmatrix} |\boldsymbol{P}(t)\rangle$$

and finding the solution boils down to computing eigenvalues and eigenvectors of a tri-diagonal matrix. The stationary distribution in particular solves $\boldsymbol{W}|\boldsymbol{\Pi}\rangle = |\boldsymbol{0}\rangle$ that gives us the following set of equations

$$\lambda\Pi_k = \mu\Pi_{k+1} \Rightarrow \Pi_k = \frac{\lambda}{\mu}\Pi_{k-1} = \dots = \left(\frac{\lambda}{\mu}\right)^k \Pi_0 \tag{2.89}$$

Setting $\rho = {}^{\lambda}\!/_{\mu}$ and using the sum of a geometric series we have

$$1 = \sum_{k=0}^{N} \Pi_k = \Pi_0 \sum_{k=0}^{K} \rho^k = \Pi_0 \frac{1-\rho^{K+1}}{1-\rho} \Rightarrow \Pi_0 = \frac{1-\rho}{1-\rho^{K+1}}. \tag{2.90}$$

and from this we can compute any Π_k.

Queues. — In queueing theory, one has often to deal with, e.g., requests arriving on a server or customers entering in a shop at a certain rate λ, and orders being processed at a rate μ. This is often referred to as M/M/1 process (memoryless arrival, memoryless departure and 1 server) and it can be generalized to more servers. This problem sounds like the problem of the (asymmetric) random walk we considered above, so that each arriving customer makes N_t increase by one, and each "processed" operation makes N_t decrease by one. One may be interested, e.g., in the number N_t of customers waiting at any time t, or, more interestingly, to its average $\mathbb{E}[\mathsf{N}_t]$ when the process reaches stationarity. This require computing the stationary distribution, in the presence of a reflecting boundary at $k = 0$ (as the number of jobs cannot be negative). On the other hand, there is no upper bound on the number of jobs arriving, and one can take $K \to \infty$. Clearly, in this limit, a steady state only exists for $\rho < 1$, otherwise the queue grows indefinitely. Taking the limit $K \to \infty$ in (2.90) we have $\Pi_0 = 1 - \rho$ so the steady-state distribution is

$$\Pi_k = \rho^k \, (1 - \rho) \tag{2.91}$$

The average number of customers waiting at stationarity follows as

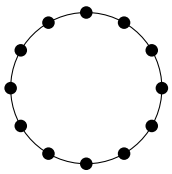
$$\mathbb{E}[\mathsf{N}_t] = (1 - \rho) \sum_{k=1}^{\infty} k \rho^k = (1 - \rho) \rho \frac{\mathrm{d}}{\mathrm{d}\, \rho} \sum_{k=0}^{\infty} \rho^k = \frac{\rho}{1 - \rho}. \tag{2.92}$$

Periodic boundary conditions. — Periodic boundary conditions consist in identifying the state $k = 0$ with the state $k = K$, so that the random walker can be thought of as jumping between the sites of a ring lattice, with sites $k = 1, 2 \ldots, K \equiv 0$. The master equation

$$\dot{P}_k(t) = \lambda P_{k-1}(t) + \mu P_{k+1}(t) - (\lambda + \mu) P_k(t) \tag{2.93}$$

holds for all $1 \leq k \leq K$, with the identification of $K \equiv 0$, and it can be solved again by a generating function formalism, with the generating function now taking the form of a *discrete Fourier transform*, which reflects the spacial periodicity of the system.

The steady-state is clearly given by $\Pi_k = {}^1\!/_K \;\forall\; k$, as inserting this expression in (2.93) makes both the RHS and LHS equal to zero. However, the detailed balance condition (2.89) is violated for any $\lambda \neq \mu$, i.e., $\rho \neq 1$, by the boundary condition $\Pi_K = \Pi_0$, as



$$\Pi_K = \left(\frac{\lambda}{\mu} \right)^K \Pi_0 = \rho^K \Pi_0 \neq \Pi_0. \tag{2.94}$$

In other words, for $\rho \neq 1$, the uniform distribution is not an equilibrium stationary state. Note how the steady-state solution for periodic boundaries was not valid for reflecting boundaries, as it would have violated the steady-state equations $\dot{P}_1 = \dot{P}_K = 0$ in that case.

Let us now calculate the time-dependent distribution of a random walker with periodic boundary conditions. The crucial tool will be the *discrete Fourier transform*, which will serve as the generating function to solve equations (2.93). The discrete Fourier transform of $P_k(t)$ is defined as

$$\hat{P}_u(t) \coloneqq \sum_{k=1}^{K} P_k(t) \, \mathrm{e}^{\frac{2\pi i k}{K} u}, \qquad u = 1, \ldots, K. \tag{2.95}$$

The discrete Fourier transform can be inverted using the formula

$$P_k(t) \coloneqq \frac{1}{K} \sum_{u=1}^{K} \hat{P}_u(t) \, \mathrm{e}^{-\frac{2\pi i k}{K} u}, \qquad k = 1, \ldots, K. \tag{2.96}$$

Proceeding similarly to the case of the unbounded random walker, we manipulate the set of master equations, multiplying both terms by $\mathrm{e}^{\frac{2\pi i k}{K}u}$ and summing over k,

$$\sum_{k=1}^{K} \dot{P}_k(t)\,\mathrm{e}^{\frac{2\pi i k}{K}u} = \sum_{k=1}^{K} \left[\lambda P_{k-1}(t) + \mu P_{k+1}(t) - (\lambda+\mu)P_k(t)\right]\mathrm{e}^{\frac{2\pi i k}{K}u}$$

in such a way to obtain a PDE for $\hat{P}_u(t)$,

$$\frac{\mathrm{d}\,\hat{P}_u(t)}{\mathrm{d}\,t} = \alpha_u \hat{P}_u(t), \qquad \alpha_u \coloneqq \lambda\left(\mathrm{e}^{-\frac{2\pi i u}{K}} - 1\right) + \mu\left(\mathrm{e}^{\frac{2\pi i u}{K}} - 1\right).$$

which is solved by

$$\hat{P}_u(t) = \mathrm{e}^{\alpha_k t}\,\hat{P}_u(0)$$

The time-dependent solution is obtained taking the inverse discrete Fourier transform,

$$P_k(t) = \frac{1}{K}\sum_{u=1}^{K} \hat{P}_u(t)\,\mathrm{e}^{-\frac{2\pi i k}{K}u} = \frac{1}{K}\sum_{u=1}^{K} \hat{P}_u(0)\,\mathrm{e}^{-\frac{2\pi i k}{K}u + \alpha_u t}. \tag{2.97}$$

Observe now that $\Re[\alpha_u] < 0\ \forall u \neq K$, and $\alpha_K = 0$. Singling out the contribution from $k = K$ in (2.97) we obtain

$$P_k(t) = \frac{1}{K}\hat{P}_K(0) + \frac{1}{K}\sum_{u=1}^{K-1} \hat{P}_u(0)\,\mathrm{e}^{-\frac{2\pi i k}{K}u + \alpha_u t}$$

Note now that, because of the normalisation of $|\boldsymbol{P}(0)\rangle$, $\hat{P}_K(0) = 1$ hence for large time $\lim_{t\to\infty} P_k(t) = {}^1\!/_K\ \forall k$ as expected. Each Fourier mode $u \neq K$ will decay exponentially as e^{-t/τ_u}, with a rate $\tau_u = |\Re[\alpha_u]|^{-1}$. For u finite and large K

$$\alpha_u = \lambda\left(\mathrm{e}^{-\frac{2\pi i u}{K}} - 1\right) + \mu\left(\mathrm{e}^{\frac{2\pi i u}{K}} - 1\right) = -(\lambda-\mu)\frac{2\pi i u}{K} - (\lambda+\mu)\frac{4\pi^2 u^2}{K^2} + o\left(\frac{1}{K^2}\right), \tag{2.98}$$

so that the first term is imaginary and gives raise to an oscillatory contribution in the time-dependent probability (due to a probability flux through the ring when $\lambda \neq \mu$), while the second term gives an exponential decay (to the equilibrium distribution). Hence the equilibration time can be read off as $\tau \sim K^2$. As we will see, this relation exhibits the typical dynamic exponents of a diffusion process. In practice, this means that the number of time-steps we would expect to run a simulation for, on a large ring, to reach equilibration, increases quadratically in the system size.

7. Birth–death processes

In this section we will study a final example of jump processes. *Birth–death processes* are processes appearing in the modeling of a variety of phenomena in which the focus is the dynamics of a population (of individuals, or chemical compounds) subject to some evolution laws. Consider for example a population of individuals or 'particles' • which evolves stochastically via offspring production and spontaneous death. Each particle may undergo one of the following 'reactions'

$$\bullet \to \begin{cases} \bullet + \bullet & \text{reproduction at rate } \lambda \text{ per particle,} \\ \varnothing & \text{death at rate } \mu \text{ per particle.} \end{cases} \tag{2.99}$$

Let N_t be the number of particles at time t. Since reactions occur stochastically, the evolution of N_t is random, and N_t is indeed a stochastic process. Clearly, N_t takes values in $\mathbb{N}_0 \coloneqq \{0, 1, 2, \ldots\}$. The transition rates from a state with n particles to a state with $n+1$ and $n-1$ particles are, respectively,

$$r_n = \lambda n, \qquad \ell_n = \mu n$$

Since both rates are proportional to n, $n = 0$ is an absorbing state, as there are no transitions out of it. This means in particular that detailed balance is not satisfied, as there is probability flow into the extinction state but not out of it. The dynamics will therefore be always a non-equilibrium process. Denoting $P_n(t) \coloneqq \mathbb{P}[\mathsf{N}_t = n]$, the master equation is given by

$$\dot{P}_n = \lambda(n-1)P_{n-1} + \mu(n+1)P_{n+1} - (\lambda+\mu)nP_n \tag{2.100}$$

and, precisely as we have done before for other processes, it can be solved via the generating function $F(z,t) = \sum_{n=0}^{\infty} z^n P_n(t)$. In this setting, we will be particularly interested in one specific probability, namely the *extinction probability* $P_0(t) = F(0,t)$, or the *survival probability* $1 - P_0(t) = 1 - F(0,t)$.

Getting a PDE for the generating function from the set of coupled master equations can be done following the same steps that we followed for the Poisson processes and for random walks. As usual, we will have boundary condition $F(1,t) = 1 \;\forall\; t$, due to the normalization of P_n. Proceeding in the standard way, we get

$$\begin{aligned} &\sum_{n\geq 0}\dot{P}_n z^n = \sum_{n\geq 0}[\lambda(n-1)P_{n-1} + \mu(n+1)P_{n+1} - (\lambda+\mu)nP_n]z^n \Longrightarrow \\ &\partial_t F(z,t) = \lambda z^2 \sum_{n\geq 1}(n-1)z^{n-2}P_{n-1} + \mu\sum_{n\geq 0}(n+1)z^n P_{n+1} - (\lambda+\mu)z\sum_{n\geq 0} n z^{n-1}P_n. \end{aligned} \tag{2.101}$$

Although the prefactors n and $n \pm 1$ do not allow to express the RHS in terms of F, they will allow to express the RHS in terms of its derivative, via the following manipulations

$$\begin{aligned} \partial_t F(x,t) &= \lambda z^2 \partial_z \sum_{n\geq 1} z^{n-1}P_{n-1} + \mu\partial_z \sum_{n\geq 0} z^{n+1}P_{n+1} - (\lambda+\mu)z\partial_z \sum_{n\geq 0} z^n P_n \\ &= \lambda z^2 \partial_z \sum_{n\geq 0} z^n P_n + \mu\partial_z \sum_{n\geq 0} z^n P_n - (\lambda+\mu)z\partial_z \sum_{n\geq 0} z^n P_n. \end{aligned} \tag{2.102}$$

We have shifted $n-1 \to n$ in the first term of the RHS, as the upper bound is infinite, and $n+1 \to n$ in the second term, which can therefore be written as $\sum_{n\geq 1} z^n P_n = \sum_{n\geq 0} z^n P_n - P_0$. We can finally obtain a closed equation for F

$$\partial_t F(z,t) = \mu(1-z)\,(1-\rho z)\,\partial_z F(z,t), \qquad \rho \coloneqq {}^{\lambda}\!/_{\mu}. \tag{2.103}$$

We solve for the initial condition $P_n(0) = \delta_{n,n_0}$ i.e., $F(z,0) = \sum_n z^n P_n(0) = z^{n_0}$, using the *method of characteristics*.

🔍 The method consists in finding a parametrization $z(s), t(s)$ of the variables of F such that

$$\frac{\mathrm{d}F}{\mathrm{d}s} = \frac{\partial F}{\partial t}\frac{\mathrm{d}t}{\mathrm{d}s} + \frac{\partial F}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}s}.$$

Comparing with (2.103) and equating the coefficients of the derivatives we obtain

$$\frac{\mathrm{d}t}{\mathrm{d}s} = 1, \tag{2.104}$$

$$\frac{\mathrm{d}z}{\mathrm{d}s} = -\mu(1-z)(1-\rho z), \tag{2.105}$$

$$\frac{\mathrm{d}F}{\mathrm{d}s} = 0. \tag{2.106}$$

Eliminating the s parameter, we obtain a system of two ODEs

$$\frac{\mathrm{d}F}{\mathrm{d}z} = 0, \qquad \frac{\mathrm{d}z}{\mathrm{d}t} = -\mu(1-z)(1-\rho z) \tag{2.107}$$

Separating variables and integrating by using partial fractions (in the case $\rho \neq 1$, i.e., $\lambda \neq \mu$ — the case $\rho = 1$ does not require partial fractions), we have

$$F = c_1 \quad \begin{cases} \dfrac{1-z}{1-\rho z} = c_2\, \mathrm{e}^{-\mu t(\rho-1)} & \rho \neq 1 \\ \dfrac{1}{1-z} = c_2 - \mu t & \rho = 1 \end{cases} \tag{2.108}$$

where c_1 and c_2 are constants of integration. From now on, we will have to distinguish the case $\rho \neq 1$ and $\rho = 1$. We know from the method of characteristics that $c_1 = \phi(c_2)$ for some function ϕ to be identified.

$$F = \begin{cases} \phi\left(\frac{1-z}{1-\rho z}\, \mathrm{e}^{\mu t(\rho-1)}\right) & \rho \neq 1 \\ \phi\left(\frac{1}{1-z} + \mu t\right) & \rho = 1 \end{cases} \tag{2.109}$$

where ϕ is, as yet, an unknown function, that we shall get from the initial conditions. Using $F(x,0) = x^{n_0}$,

$$\begin{cases} z^{n_0} = \phi\left(\frac{1-z}{1-\rho z}\right) \Rightarrow \phi(z) = \left(\frac{z-1}{\rho z-1}\right)^{n_0} & \rho \neq 1 \\ z^{n_0} = \phi\left(\frac{1}{1-z}\right) \Rightarrow \phi(z) = \left(1-\frac{1}{z}\right)^{n_0} & \rho = 1. \end{cases} \tag{2.110}$$

Inserting this in (2.109), we finally get

$$F(z,t) = \begin{cases} \left(\frac{(1-\rho z)\, \mathrm{e}^{\mu(1-\rho)t} + z - 1}{(1-\rho z)\, \mathrm{e}^{\mu(1-\rho)t} + 1/\rho (z-1)}\right)^{n_0} & \rho \neq 1 \\ \left(1 - \frac{1-z}{1+(1-z)\mu t}\right)^{n_0} & \rho = 1 \end{cases} \tag{2.111}$$

Using the obtained solution

$$F(z,t) = \begin{cases} \left(\frac{(1-\rho z)\, \mathrm{e}^{\mu(1-\rho)t} + z - 1}{(1-\rho z)\, \mathrm{e}^{\mu(1-\rho)t} + 1/\rho (z-1)}\right)^{n_0} & \rho \neq 1 \\ \left(1 - \frac{1-z}{1+(1-z)\mu t}\right)^{n_0} & \rho = 1 \end{cases} \tag{2.112}$$

the extinction probabilities $P_0(t) = F(0,t)$ follows as

$$P_0(t) = \begin{cases} \left(\frac{\mathrm{e}^{\mu(1-\rho)t} - 1}{\mathrm{e}^{\mu(1-\rho)t} - 1/\rho}\right)^{n_0} & \rho \neq 1, \\ \left(1 - \frac{1}{1+\mu t}\right)^{n_0} & \rho = 1. \end{cases} \tag{2.113}$$

For large time, we have

$$P_0(t) \sim \begin{cases} \rho^{-n_0} + \mathrm{e}^{\mu(\rho-1)t} & \rho > 1, \\ 1 + \mathrm{e}^{-\mu(1-\rho)t} & \rho < 1 \\ 1 - \frac{n_0}{\mu t} & \rho = 1. \end{cases} \tag{2.114}$$

showing that $P_0(\infty) = 1$ for $\rho \leq 1$, i.e. extinction is a *sure* event for $\lambda \leq \mu$, however, it is also *possible* for $\lambda > \mu$. We note that for $\mu \neq \lambda$ the steady-state is approached with an exponential rate, $\tau_c = |\mu - \lambda|^{-1}$. However, for $\rho \to 1$, $\tau_c \to \infty$ and the approach to the steady-state is much slower, i.e., power-law like. This is reminiscent of the so-called "critical slowing down' around a "critical point" (in this case, $\rho = 1$) in statistical physics, a phenomenon normally observed in second order phase transitions.

The asymptotic survival probability can be also computed as

$$\mathbb{P}[\mathsf{N}_\infty > 0] \coloneqq 1 - \lim_{t\to\infty} P_0(t) = \begin{cases} 1 - \rho^{-n_0} & \rho > 1, \\ 0 & \rho \leq 1. \end{cases}$$

🔍 It is worth noting that a "mean-field approximation" $\mathsf{N}_t \simeq \mathbb{E}[\mathsf{N}_t]$ would miss the possibility of extinction at $\lambda > \mu$. Such mean-field approximation is obtained writing an equation for $\mathbb{E}[\mathsf{N}_t]$, obtained by multiplying the master equation by n and summing over n. Unsurprisingly, the equation for the first moment closes due to the linearity of the rates

$$\partial_t \mathbb{E}[\mathsf{N}_t] = -\mu(1-\rho)\mathbb{E}[\mathsf{N}_t]. \tag{2.115}$$

The solution $\mathbb{E}[\mathsf{N}_t] = n_0\, \mathrm{e}^{-\mu(1-\rho)t}$ indicates a demographic explosion at $\mu < \lambda$, extinction at $\mu > \lambda$ and a constant population size n_0 for $\mu = \lambda$. However, fluctuations about the average are important and make the mean-field description inaccurate. In particular, ODEs like (2.115) do not capture well the behaviour of discrete systems, which evolve by discrete jumps, rather than infinitesimal variations, especially when the population size is small, and a stochastic jump can wipe out the entire population.

✎ Show that Eq. (2.115) can be obtained from the equation for $F(z,t)$ taking the derivative with respect to z and imposing $z = 1$.

7.1. Chemical reaction kinetics. Linear chemical reaction kinetics may also be regarded as linear birth-death processes. One such example is provided by the stochastic transitions of the molecules of a chemical specie between two conformational states, X to Y, upon interaction with another molecule A

$$\mathrm{A} + \mathrm{X} \underset{c_2}{\overset{c_1}{\rightleftharpoons}} \mathrm{A} + \mathrm{Y}.$$

If we denote N_t the number of molecules X at time t, N_A the number of molecules A (kept fixed) and N the number of molecules X and Y assumed to be constant, we can write the master equation for $P_n(t) = \mathbb{P}[\mathsf{N}_t = n]$ as

$$\dot{P}_n = c_2 N_\mathrm{A}(N-n+1)P_{n-1} + c_1 N_A(n+1)P_{n+1} - N_A(c_1 n + c_2(N-n))P_n. \tag{2.116}$$

This is linear and can be solved by applying the generating functions method. However, most of times master equations for chemical reaction kinetics are non-linear and cannot be solved exactly. We provide below an example.

⦿ Example For the chemical reactions

$$\mathrm{X} + \mathrm{X} \xrightarrow{c_1} \varnothing, \qquad \mathrm{X} \xrightarrow{c_2} \mathrm{X} + \mathrm{X}$$

the master equation for the number N_t of molecules X is

$$\dot{P}_n = c_1 \frac{(n+1)(n+2)}{2} P_{n+2} + c_2(n-1)P_{n-1} - \left(c_1 \frac{n(n-1)}{2} + c_2 n \right) P_n. \tag{2.117}$$

Due to non-linearities, the equation for the first moment does not close

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\mathbb{E}[\mathsf{N}_t] = (c_1 + c_2)\mathbb{E}[\mathsf{N}_t] - c_1 \mathbb{E}[\mathsf{N}_t^2]. \tag{2.118}$$

Closure can be achieved with a mean-field approximation $\mathbb{E}[\mathsf{N}_t^2] \simeq \mathbb{E}[\mathsf{N}_t]^2$,

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\mathbb{E}[\mathsf{N}_t] \approx (c_1 + c_2)\mathbb{E}[\mathsf{N}_t] - c_1 \mathbb{E}[\mathsf{N}_t]^2. \tag{2.119}$$

This predicts a stable fixed point $\mathbb{E}[\mathsf{N}_t] = 1 + {}^{c_2}\!/_{c_1}$, however, there is clearly a finite probability that the system goes extinct, that is not captured by the mean-field approximation.

🔍 The Gillespie algorithm When master equations cannot be solved analytically, one may wish to take recourse to numerical methods. We consider a system of K chemically reactive species X_i, $i = 1, \dots, K$, which can react with each other via M reactions $\mathcal{R}_\mu$, $\mu = 1, \dots, M$, each with its own rate c_μ. We denote with N_i the number of molecules of species X_i and with $\mathbf{N} = (\mathsf{N}_1, \dots, \mathsf{N}_K)$ the

state of the system. Furthermore, we indicate with $\boldsymbol{\nu}^\mu$ the change in $\mathbf{N}$ produced by reaction $\mathcal{R}_\mu$ (as given from the stechiometric coefficients), e.g. $\nu_i^\mu = -3$ means that, by effect of the reaction $\mathcal{R}_\mu$, the $\mathsf{N}_i \to \mathsf{N}_i - 3$. Finally, we denote with $h_\mu(\boldsymbol{n})$ the number of ways in which $\mathcal{R}_\mu$ can be realized in state $\boldsymbol{n}$, e.g., in the reaction $\mathrm{X}_1 + \mathrm{X}_1 \to \mathrm{X}_2$, there are $\frac{1}{2}n_1(n_1-1)$ ways in which the reaction can take place. In order to derive the chemical master equation, we introduce the so-called *propensity function* $W_\mu(\boldsymbol{n}) = c_\mu h_\mu(\boldsymbol{n})$, such that

$$W_\mu(\boldsymbol{n})\,\mathrm{d}\,t = \mathbb{P}\big[\mathcal{R}_\mu \text{ takes place in the interval } [t, t+\mathrm{d}\,t)\big|\mathbf{N}_t = \boldsymbol{n}\big] \tag{2.120}$$

The chemical master equation is then

$$\partial_t P_{\boldsymbol{n}}(t) = \sum_\mu \left[W_\mu(\boldsymbol{n}-\boldsymbol{\nu}^\mu)P_{\boldsymbol{n}-\boldsymbol{\nu}^\mu}(t) - W_\mu(\boldsymbol{n})P_{\boldsymbol{n}}(t)\right] \tag{2.121}$$

An efficient way to simulate such master equation is via the *Gillepsie algorithm*. The starting point of the algorithm is the definition of the so-called 'next-reaction density function'

$$\rho(\mathcal{R}_\mu, \tau|\boldsymbol{n}, t)\,\mathrm{d}\,t = \mathbb{P}\Big[\mathcal{R}_\mu \text{ happens in } [t+\tau, t+\tau+\mathrm{d}\,t)\Big|\mathbf{N}_t = \boldsymbol{n}\Big]. \tag{2.122}$$

It is reasonable to assume that τ and μ are conditionally independent. Hence, since the distribution of waiting times in memoryless processes is exponential, we can write

$$\begin{aligned}\rho(\mathcal{R}_\mu,\tau|\boldsymbol{n},t) &= \mathbb{P}\Big[\mathcal{R}_\mu \text{ happens}\Big|\mathbf{N}_t=\boldsymbol{n}\Big]\,\mathbb{P}\Big[\text{Something happens in } [t+\tau,t+\tau+\mathrm{d}t)\Big|\mathbf{N}_t=\boldsymbol{n}\Big] \\ &= \underbrace{\frac{W_\mu(\boldsymbol{n})}{W(\boldsymbol{n})}}_{\rho(\mathcal{R}_\mu|\boldsymbol{n})}\ \underbrace{W(\boldsymbol{n})\,\mathrm{e}^{-W(\boldsymbol{n})\tau}}_{\rho(\tau|\boldsymbol{n})}, \qquad W(\boldsymbol{n}) \coloneqq \sum_\mu W_\mu(\boldsymbol{n}).\end{aligned} \tag{2.123}$$

The Gillespie algorithm consists in iteratively sampling the next reaction time and type and executing it, as exemplified by the pseudocode below:

(1) Set $t = 0$. Initialize the system in $\boldsymbol{n} = \boldsymbol{n}_0$.
(2) Draw (τ, μ) from the distribution $\rho(\mathcal{R}_\mu, \tau|\boldsymbol{n}, t)$. Set $t = t + \tau$.
(3) Execute $\boldsymbol{n} \mapsto \boldsymbol{n} + \boldsymbol{\nu}^\mu$
(4) Go back to (2).

To realize step 2, one draws τ and μ independently, each from its distribution $\rho(\mathcal{R}_\mu|\boldsymbol{n})$ and $\rho(\tau|\boldsymbol{n})$ respectively. To efficiently sample τ from $\rho(\tau|\boldsymbol{n})$ we note that

$$p = \mathbb{P}[\text{Something happens in time} < \tau] = \int_0^\tau W(\boldsymbol{n})\,\mathrm{e}^{-W(\boldsymbol{n})t}\,\mathrm{d}\,t = 1 - \mathrm{e}^{-W(\boldsymbol{n})\tau}\,.$$

Rearranging,

$$\tau = \frac{1}{W(\boldsymbol{n})}\ln\frac{1}{1-p}$$

Hence, at each iteration of the algorithm, we can draw p uniformly at random in $(0,1)$ and set $\tau = \frac{1}{W(\boldsymbol{n})}\ln\frac{1}{1-p}$.

As a final remark, we note that the Gillespie algorithm does not solve the master equation, but it simulates a sample path $\boldsymbol{n}(t)$ of the Markov process, so one needs to run many statistically independent simulations (with the same $\boldsymbol{n}(0)$ and the same final t) to get a statistics, i.e. a distribution, for $\boldsymbol{n}(t)$. We also note that for large systems, the exponential rate $W(\boldsymbol{n})$ will typically get large, and the sampled time steps very small, so one may need to execute a large number of iterations to simulate the evolution of the system on a finite time span. This may get impractical for many particles and many reactions.

CHAPTER 3

Dynamics of spin systems

ABSTRACT. In this chapter, we will study the dynamics of systems with many interacting *binary units*. In other words, our stochastic variable will be a N-dimensional vector $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$. Each σ_i corresponds to a unit in a system with N components and takes value $\sigma_i = +1$ ('active') or $\sigma_i = -1$ ('inactive'). Many systems can be modeled in these terms, from financial systems (where agents can buy or sell financial products) to gene-regulatory networks (where genes can be expressed or not expressed), from neural networks (where neurons can either fire an electrical signal or stay quiescent) to epidemiological models (where individuals are either infected or susceptible). The units usually interact according to complex patterns[1]. In the realm of such models, the *Ising model* plays the role of prototype for many systems. In this model, the N units are organized on a graph, and are regarded as the atoms of some material. The unit σ_i indicates the orientation of the magnetic momentum (or 'spin') of an atom (for simplicity taken as pointing up or down). We will refer therefore to the binary units σ as *spins*.

1. Neural Networks

LECTURE 8

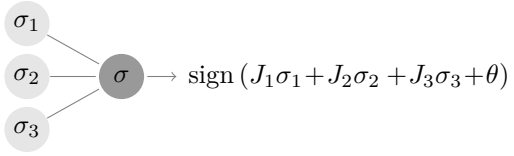
Let us consider a system of N neurons. Each neuron i is represented by a variable σ_i, such that $\sigma_i = +1$ if the neuron is firing, $\sigma_i = -1$ if the neuron is quiescent. We assume now that an "output neuron" σ collects the information about the neurons $\boldsymbol{\sigma} = (\sigma_i)_i$ according to this simple rule

$$\sigma = \operatorname{sign}\left(\sum_j J_j \sigma_j + \theta\right) = \operatorname{sign}\left(\langle \boldsymbol{J} | \boldsymbol{\sigma} \rangle + \theta\right). \tag{3.1}$$

This is a fundamental architecture called *perceptron* and has been introduced by McCulloch and Pitts to simulate the action of biological neurons. As you can see, the output σ, representing the state of the neuron itself, is influenced by a 'local field' resulting from the state of other neurons

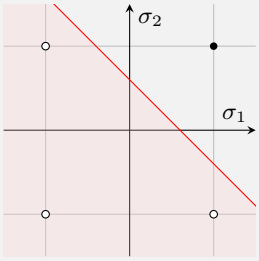
$$h(\boldsymbol{\sigma}) \coloneqq \sum_j J_j \sigma_j + \theta. \tag{3.2}$$

The quantities J_j measure the influence of the neuron j on the output neuron. If $J_j > 0$, the neuron j will 'encourage' the output neuron to take its same sign, otherwise will induce a disagreement. The quantity θ can be thought of as a 'bias' of the output neuron: if there were no other neurons, σ would just be the sign of θ. Pictorially, for $N = 3$,

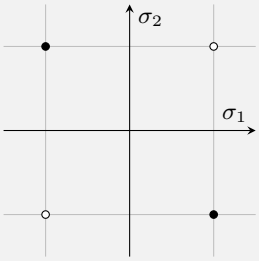


For $N = 2$, the perceptron can execute logic operations such as `AND`, `OR`, and `NOT` for appropriate choice of J_{0j}, $j = 1, 2$, and θ_0. In particular, it can execute any *linearly separable* operation.

🔍 The condition $J_1\sigma_1 + J_2\sigma_2 + \theta > 0$ defines a halfplane in the (σ_1, σ_2) plane:

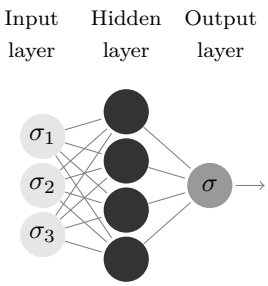


The border of this region is exactly given by the line $J_1\sigma_1 + J_2\sigma_2 + \theta = 0$, and we can arrange it in a way that the output is the one we desire. For example, suppose that we want to implement the `AND` function. This means that we want $\sigma = +1$ if $\sigma_1 = \sigma_2 = +1$, and $\sigma = -1$ otherwise. In the plot above, we represented this desired output coloring in black the dot corresponding to the input values giving output $+1$ using `AND`. Any straight line that separates the black dot from the white dots gives this kind of output. If we have instead the `XOR` function, i.e., $\sigma = -\sigma_1\sigma_2$, this is non-linear and corresponds to the plot



It is clear that this time we cannot separate white and black dots using a single straight line.

In order to execute operations that are not linearly separable, like the `XOR` logic gate, one needs at least one layer of neurons, sometimes called the "hidden" layer, between the input and the output neurons, i.e., an architecture of the following type



Each link in the picture corresponds to a coupling J to be carefully chosen so that the output corresponds to the desired one. Such networks used are normally 'feed-forward', i.e., the signal goes from the input layer toward the output layer moving from one layer to the next one. Deep neural networks are feed-forward networks with many layers. *Training* a neural network to execute a certain operation means to find a suitable set of couplings $\boldsymbol{J}$ and thresholds $\boldsymbol{\theta}$, such that the neuron in the output layer returns the right value associated to each input. When this occurs, we say that the network has 'learnt' to execute the task. Suppose for example that we want to train our network, let us call it $\phi_{\boldsymbol{J}}(\boldsymbol{\sigma}^\mu)$ (assume for simplicity that the thresholds are given and fixed), to reproduce some function $f(\boldsymbol{\sigma})$ which we do not know exactly. What we know is a set of n *examples*, $\boldsymbol{\sigma}^\mu$, that we know correspond each to a certain output $y^\mu = f(\boldsymbol{\sigma}^\mu)$. For

each of these examples, our network gives us the output $\phi_{\boldsymbol{J}}(\boldsymbol{\sigma}^\mu)$, so that the error over the set of n examples can be quantified as

$$\ell(\boldsymbol{J}) \coloneqq \frac{1}{n}\sum_{\mu=1}^{n}\left(y^\mu - \phi_{\boldsymbol{J}}(\boldsymbol{\sigma}^\mu)\right)^2 \tag{3.3}$$

We need to find $\boldsymbol{J}$ such that this error is as small as possible, i.e., we need to find $\boldsymbol{J}^\star = \arg\min_{\boldsymbol{J}} \ell(\boldsymbol{J})$. The simplest way is to take a gradient and impose $\nabla_{\boldsymbol{J}}\ell(\boldsymbol{J}) = 0$. Learning algorithms try to achieve such minima: observe that they generally converge to a *local* minimum of the error function, where the weights will in reproduce well the training dataset, but may generalise poorly (i.e., may have poor performances on unseen inputs). Theoretical and empirical studies of the predictive abilities and limitations of deep neural networks or in general neural networks with a huge number of parameters J constitutes nowadays an intense research field.

1.1. Neural networks as associative memories. Neural networks can be used to perform different types of tasks. One of them is storing and retrieving memories, and we will precisely study the dynamics of these networks to retrieve stored configurations. In order to use neural networks as *associative memories*, they must be able to store a multiplicity of patterns. Memorised patterns are thought of as stable attractors of the neural network dynamics: if the neural network is initialised in a configuration that is 'close' to one of the stored patterns, the network dynamics will converge to it and we will say that the network has 'retrieved' the pattern associated to the initial configuration.

If the neural network dynamics can be cast into a gradient-descent on an energy landscape, then memorised patterns will correspond to the local minima of the energy landscape. *Feedback loops* are a desired feature in these networks: they are a necessary ingredient to have many local minima in the energy landscape: if the spin i and j are connected, then the dynamic of i will be influenced by the dynamic of j and vice versa (at difference with feed-forward networks, used for learning, where each spin is influenced by the spins of the previous layers but not of the next). In a neural network used as associative memory, spins are on some graph $\mathcal{G}$ and they are subject to an evolution in time of the type

$$\sigma_i(t+\tau) = \operatorname{sign}\left[\sum_j J_{ij}\sigma_j(t) + \theta_i + \beta^{-1}\eta_i(t)\right] \equiv \operatorname{sign}\Big[h_i(\boldsymbol{\sigma}(t)) + \beta^{-1}\eta_i(t)\Big], \tag{3.4}$$

at *discrete* time steps $t = \tau,\ 2\tau\ldots$. The law above is the same used in (3.1) for the learning task. We have also introduced $\eta_i(t)$ that is a random noise, e.g., a stochastic process independent from the dynamics of the spin, with $\mathbb{E}[\eta_i(t)] = 0$, $\mathbb{E}[\eta_i^2(t)] = 1$ drawn from a symmetric distribution, i.e., $\mathbb{P}[\eta(t) \geq 0] = \mathbb{P}[\eta(t) \leq 0]$. One can interpret $\beta^{-1} > 0$ as a noise level, i.e., a measure of the intensity of the noise: for $\beta^{-1} = 0$ the dynamics is deterministic, while for $\beta^{-1} = \infty$ the dynamics is fully stochastic, ruled by the noise η. Due to the presence of the noise term, the update equations have probabilistic form. From (3.4), and using the symmetry of the distribution of η, we have

$$\begin{aligned}\mathbb{P}_{1|1}[\sigma_i(t+\tau)=1|\boldsymbol{\sigma}(t)=\boldsymbol{\sigma}]=\mathbb{P}[\eta(t)>-\beta h_i(\boldsymbol{\sigma})],\\ \mathbb{P}_{1|1}[\sigma_i(t+\tau)=-1|\boldsymbol{\sigma}(t)=\boldsymbol{\sigma}]=\mathbb{P}[\eta(t)<-\beta h_i(\boldsymbol{\sigma})].\end{aligned} \tag{3.5}$$

A natural choice for the noise distribution would be a Gaussian. In this case $\mathbb{P}[\eta > x] = \frac{1}{2}\mathrm{erf}(x)$. A qualitatively very similar choice, which makes calculations simpler, is the *Glauber choice* $\mathbb{P}[\eta > x] = \frac{1-\tanh(x)}{2}$ which leads to

$$\mathbb{P}_{1|1}[\sigma_i(t+\tau) = \sigma|\boldsymbol{\sigma}(t) = \boldsymbol{\sigma}] = \frac{1+\sigma\tanh\left(\beta h_i(\boldsymbol{\sigma})\right)}{2}$$

With this choice, the above expression can be rewritten

$$\mathbb{P}_{1|1}[\sigma_i(t+\tau)=\sigma|\boldsymbol{\sigma}(t)=\boldsymbol{\sigma}]=\frac{\mathrm{e}^{\beta\sigma h_i(\boldsymbol{\sigma})}}{2\cosh(\beta h_i(\boldsymbol{\sigma}))}=\frac{1}{1+\mathrm{e}^{-2\beta\sigma h_i(\boldsymbol{\sigma})}}$$

and the average value of a neuron state is then $\mathbb{E}[\sigma_i(t+\tau)]_{\boldsymbol{\sigma}(t)=\boldsymbol{\sigma}}=\tanh\beta h_i(\boldsymbol{\sigma})$. Now that we have this transition probability, we can make different choices for the evolution of the spin network.

1.2. Parallel dynamics. We say that the network evolves via *parallel dynamics* if all of the neurons in the network are updated at the same time (i.e., synchronously), at regular time intervals of duration τ. In this case, assuming the Glauber rule, we have

$$Q_{\boldsymbol{\sigma}'\boldsymbol{\sigma}}\coloneqq\mathbb{P}_{1|1}[\boldsymbol{\sigma}(t+\tau)=\boldsymbol{\sigma}'|\boldsymbol{\sigma}(t)=\boldsymbol{\sigma}]=\prod_{i=1}^{N}\frac{\mathrm{e}^{\beta\sigma_i' h_i(\boldsymbol{\sigma})}}{2\cosh(\beta h_i(\boldsymbol{\sigma}))}. \tag{3.6}$$

It is clear now that we can write the dynamics as a Markov chain for the probability $P_{\boldsymbol{\sigma}}(t)\coloneqq\mathbb{P}[\boldsymbol{\sigma}(t)=\boldsymbol{\sigma}]$,

$$P_{\boldsymbol{\sigma}'}(t+\tau)=\sum_{\boldsymbol{\sigma}}Q_{\boldsymbol{\sigma}'\boldsymbol{\sigma}}(\tau)P_{\boldsymbol{\sigma}'}(t)$$

with transition matrix $\boldsymbol{Q}(\tau)$. Each row/column of $\boldsymbol{Q}(\tau)$ corresponds to a configuration of the system, and, if there are N spins, there are 2^N possible configurations, i.e., $\boldsymbol{Q}(\tau)$ is $2^N\times 2^N$, that for large N is a huge number: even finding the stationary distribution is then a formidable task. However, a simplification arises when interactions are symmetric, i.e., $J_{ij}=J_{ji}\ \forall\ i,j$. In this case it is possible to show that $\boldsymbol{Q}(\tau)$ satisfies detailed balance with

$$\Pi_{\boldsymbol{\sigma}}=\frac{\mathrm{e}^{-\beta H(\boldsymbol{\sigma})}}{Z},\qquad Z\coloneqq\sum_{\boldsymbol{\sigma}}\mathrm{e}^{-\beta H(\boldsymbol{\sigma})},$$

where

$$H(\boldsymbol{\sigma})=-\sum_{i=1}^{N}\left[\theta_i\sigma_i+\beta^{-1}\ln\cosh(\beta h_i(\boldsymbol{\sigma}))\right]$$

is a 'pseudo-Hamiltonian'[2].

1.3. Sequential dynamics. Another possibility is to update the network as follows:

(1) at time t, select a spin i at random in $\{1,\dots,N\}$, i.e., with probability $1/N$;
(2) update it using a transition probability $\mathbb{P}_{1|1}[\sigma_i(t+\tau)=\sigma|\boldsymbol{\sigma}(t)=\boldsymbol{\sigma}]$;
(3) $t+\tau\mapsto t$ and go to step (1).

In this case we say that the network is updated via *sequential* (or asynchronous) dynamics: transitions only occur between configurations that differ by a single spin flip, e.g., $\boldsymbol{\sigma}=(\sigma_1,\dots,\sigma_i,\dots,\sigma_N)\to F_i\boldsymbol{\sigma}=(\sigma_1,\dots,-\sigma_i,\dots,\sigma_N)$, with F_i the i-spin flip operator. Defining now the probability to flip neuron i

$$Q_{F_i\boldsymbol{\sigma},\boldsymbol{\sigma}}(\tau)\coloneqq\frac{1}{N}\mathbb{P}_{1|1}[\sigma_i(t+\tau)=-\sigma_i|\boldsymbol{\sigma}(t)=\boldsymbol{\sigma}]=\frac{1-\sigma_i\tanh\beta h_i(\boldsymbol{\sigma})}{2N}\equiv\frac{W_i(\boldsymbol{\sigma})}{N}. \tag{3.7}$$

Let us take now the $\tau\to 0$ limit and write down a master equation for the process. We also consider $N\to+\infty$ and we combine these two limits choosing $\tau=1/N\to 0$. Let us introduce then our transition rates, using the definition

$$W_{F_i\boldsymbol{\sigma}\,\boldsymbol{\sigma}}=\lim_{N\to+\infty}\frac{Q_{F_i\boldsymbol{\sigma},\boldsymbol{\sigma}}(1/N)}{1/N}=W_i(\boldsymbol{\sigma}). \tag{3.8}$$

[2]The specification 'pseudo' is due to the fact that H retains a dependence on β.

The quantities $W_i(\boldsymbol{\sigma})$ are called *Glauber transition rates* and the master equation is obtained immediately as

$$\partial_t P_{\boldsymbol{\sigma}}(t) = \sum_{i=1}^{N} \Big[W_i(F_i\boldsymbol{\sigma}) P_{F_i\boldsymbol{\sigma}}(t) - W_i(\boldsymbol{\sigma}) P_{\boldsymbol{\sigma}}(t) \Big]. \tag{3.9}$$

1.3.1. *The equilibrium distribution.* We show that for $J_{ij} = J_{ji}$ and $J_{ii} = 0$, the Glauber transition rates satisfy detailed balance with $\Pi_{\boldsymbol{\sigma}} = \frac{1}{Z} \mathrm{e}^{-\beta H(\boldsymbol{\sigma})}$, where

$$H(\boldsymbol{\sigma}) = -\frac{1}{2} \sum_{i \neq k} J_{ik} \sigma_i \sigma_k - \sum_i \theta_i \sigma_i.$$

To prove detailed balance

$$\Pi_{\boldsymbol{\sigma}} W_i(\boldsymbol{\sigma}) = \Pi_{F_i\boldsymbol{\sigma}} W_i(F_i\boldsymbol{\sigma}) \Leftrightarrow \frac{W_i(\boldsymbol{\sigma})}{W_i(F_i\boldsymbol{\sigma})} = \mathrm{e}^{-\beta[H(F_i\boldsymbol{\sigma}) - H(\boldsymbol{\sigma})]}.$$

It is convenient to define $\hat{h}_i(\boldsymbol{\sigma}) \coloneqq \sum_k J_{ik}\sigma_k$ and write

$$H(\boldsymbol{\sigma}) = -\frac{1}{2} \sum_k \sigma_k \hat{h}_k(\boldsymbol{\sigma}) - \sum_k \theta_k \sigma_k.$$

We note that $\hat{h}_i(\boldsymbol{\sigma})$ does not depend on σ_i (because $J_{ii} = 0$) and therefore $\hat{h}_i(F_i\boldsymbol{\sigma}) = \hat{h}_i(\boldsymbol{\sigma})$ and $\hat{h}_k(F_i\boldsymbol{\sigma}) - \hat{h}_k(\boldsymbol{\sigma}) = -2J_{ki}\sigma_i$. Therefore

$$\begin{aligned} H(F_i\boldsymbol{\sigma}) - H(\boldsymbol{\sigma}) &= -\frac{1}{2} \sum_{k \neq i} \sigma_k (\hat{h}_k(F_i\boldsymbol{\sigma}) - \hat{h}_k(\boldsymbol{\sigma})) + \frac{1}{2} \sigma_i (\hat{h}_i(F_i\boldsymbol{\sigma}) + \hat{h}_i(\boldsymbol{\sigma})) + 2\theta_i\sigma_i \\ &= \sum_{k \neq i} J_{ki}\sigma_k\sigma_i + \sigma_i \hat{h}_i(\boldsymbol{\sigma}) + 2\theta_i\sigma_i = 2\sigma_i \left(\sum_{k \neq i} J_{ki} + \sigma_k \theta_i \right) = 2\sigma_i h_i(\boldsymbol{\sigma}). \end{aligned} \tag{3.10}$$

Using $\frac{1 \pm \tanh x}{2} = \frac{\mathrm{e}^{\pm x}}{2 \cosh x}$, we can write

$$\begin{aligned} \frac{\Pi_{F_i\boldsymbol{\sigma}}}{\Pi_{\boldsymbol{\sigma}}} &= \mathrm{e}^{-\beta[H(F_i\boldsymbol{\sigma}) - H(\boldsymbol{\sigma})]} = \mathrm{e}^{-2\beta\sigma_i h_i(\boldsymbol{\sigma})} \\ &= \frac{\mathrm{e}^{-\beta\sigma_i h_i(\boldsymbol{\sigma})}}{\mathrm{e}^{\beta\sigma_i h_i(\boldsymbol{\sigma})}} = \frac{1 - \tanh \beta\sigma_i h_i(\boldsymbol{\sigma})}{1 + \tanh \beta\sigma_i h_i(\boldsymbol{\sigma})} = \frac{1 - \sigma_i \tanh \beta h_i(\boldsymbol{\sigma})}{1 + \sigma_i \tanh \beta h_i(\boldsymbol{\sigma})} = \frac{W_i(\boldsymbol{\sigma})}{W_i(F_i\boldsymbol{\sigma})}. \end{aligned} \tag{3.11}$$

This proves that Glauber rates satisfy detailed balance with the Boltzmann distribution. Hence, for symmetric interactions and no self-interactions, we can determine the equilibrium distribution by appealing to the detailed balance condition, without solving the master equation. For asymmetric interactions, in contrast, such simplification does not arise, and one would have in principle to solve the master equation. However, due to the large dimensionality of the system, even computing its steady-state $\Pi_{\boldsymbol{\sigma}}$ is very hard. A strategy that is generally used with non-equilibrium systems, is to get, from the master equation, a closed set of equations for averages and fluctuations of a suitable set of observables, ideally providing a coarse-grained description of the system. This is not always possible, but it can normally be done for one-dimensional or fully connected systems, where the mean-field approximation becomes exact, and for particular structures of the interactions.

1.3.2. *Equations for the averages.* From the master equation

$$\partial_t P_{\boldsymbol{\sigma}}(t) = \sum_k \Big[W_k(F_k\boldsymbol{\sigma}) P_{F_k\boldsymbol{\sigma}}(t) - W_k(\boldsymbol{\sigma}) P_{\boldsymbol{\sigma}}(t) \Big] \tag{3.12}$$

one can derive equations for the averages of observables like $\mathbb{E}[A_{\boldsymbol{\sigma}}(t)] = \sum_{\boldsymbol{\sigma}} A_{\boldsymbol{\sigma}} P_{\boldsymbol{\sigma}}(t)$. We start with $\mathbb{E}[\sigma_i(t)] = \sum_{\boldsymbol{\sigma}} \sigma_i \, P_{\boldsymbol{\sigma}}(t)$:

$$\partial_t \mathbb{E}[\sigma_i(t)] = \sum_k \sum_{\boldsymbol{\sigma}} W_k(\boldsymbol{\sigma}) P_{\boldsymbol{\sigma}}(t) \big(F_k\sigma_i - \sigma_i \big) = -2\mathbb{E}[W_i(\boldsymbol{\sigma})\sigma_i] \tag{3.13}$$

which is intuitively equal to the change, in the neuron state, upon a single flip $-2\sigma_i$, times the flipping probability. Inserting the expression for the Glauber rate $W_i(\boldsymbol{\sigma}) = \frac{1}{2}[1 - \sigma_i \tanh \beta h_i(\boldsymbol{\sigma})]$, we have

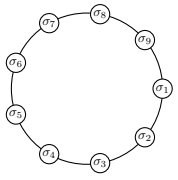
$$\partial_t \mathbb{E}[\sigma_i] = -\mathbb{E}[\sigma_i] + \mathbb{E}[\tanh \beta h_i(\boldsymbol{\sigma})] \tag{3.14}$$

The complication is that, while $h_i(\boldsymbol{\sigma}) = \sum_k J_{ik}\sigma_k + \theta_i$ is a linear function of the neuron states, $\tanh \beta h_i(\boldsymbol{\sigma})$ makes the equations non-linear, so the equations do not generally close. However, one may be able to close the equation within a *mean-field approximation*, where the effective field is replaced with its average $h_i(\boldsymbol{\sigma}) \to \mathbb{E}[h_i(\boldsymbol{\sigma})] = \sum_k J_{ik}\mathbb{E}[\sigma_k] + \theta_i$. This approximation ignores the fluctuations of the effective field about its average, and it becomes exact if the sum over k is over a large number of terms (so that the random fluctuations on average cancel out).

2. Dynamics in simple ferromagnets

In this Section we will specify the form of $h_i(\boldsymbol{\sigma})$ and derive the dynamics in the case of *ferromagnetic interactions* $J_{ij} \equiv J > 0$. We will consider two prototypical model, the *one-dimensional Ising model* and the *Curie–Weiss model*.

2.1. The one-dimensional Ising ferromagnet. The equation above are very generic and to attempt a more explicit computation we need to make some assumption on the structure of $h_i(\boldsymbol{\sigma})$. The simplest model one can consider is the one-dimensional Ising ferromagnet, where the spins are imagined to be on a circumference, and each spin interacts with two neighbours via couplings $J > 0$, which are identical for any pair of spins, so



$$h_i(\boldsymbol{\sigma}) = \frac{J}{2}\sigma_{i-1} + \frac{J}{2}\sigma_{i+1}.$$

Using $\tanh(\epsilon x) = \epsilon \tanh(x)$ for $\epsilon = 0, \pm 1$, and denoting $\gamma = \tanh(\beta J)$, this means that

$$\begin{aligned} W_i(\boldsymbol{\sigma}) &= \frac{1 - \sigma_i \tanh\left(\beta J \frac{\sigma_{i-1}+\sigma_{i+1}}{2}\right)}{2} \\ &= \frac{1}{2}\left(1 - \sigma_i \frac{\sigma_{i-1}+\sigma_{i+1}}{2} \tanh(\beta J)\right) \\ &= \frac{1}{2}\left(1 - \gamma\sigma_i \frac{\sigma_{i-1}+\sigma_{i+1}}{2}\right). \end{aligned} \tag{3.15}$$

Since interactions are symmetric the equilibrium distribution is of Boltzmann-type with Hamiltonian

$$H(\boldsymbol{\sigma}) = -\frac{J}{2}\sum_i \sigma_i(\sigma_{i-1} + \sigma_{i+1}) = -J\sum_i \sigma_i\sigma_{i+1}.$$

Let us now adapt Eq. (3.14) to the current case:

$$\partial_t \mathbb{E}[\sigma_i] = -\mathbb{E}[\sigma_i] + \gamma \frac{\mathbb{E}[\sigma_{i-1}] + \mathbb{E}[\sigma_{i+1}]}{2}. \tag{3.16}$$

If we call for brevity $m_i(t) \coloneqq \mathbb{E}[\sigma_i(t)]$ the *magnetisation* of the spin i, this means that we have obtained the set of equations

$$\partial_t m_i(t) = -m_i(t) + \gamma \frac{m_{i-1}(t) + m_{i+1}(t)}{2}, \quad i = 1, \dots, N. \tag{3.17}$$

From this equation we can extract many information about the dynamics of the system.

2.1.1. *Single spin magnetisation.* We have already met Eq. (3.17) when working on random walks. Solving for the initial condition $m_k(0) = \delta_{k\,0}$ (corresponding to an 'up' spin at the origin, in a sea of 'undecided' spins) we obtain the solution

$$m_k(t) = I_k(\gamma t)\, \mathrm{e}^{-t} \sim \mathrm{e}^{-(1-\gamma)t}\, \frac{\mathrm{e}^{-\frac{k^2}{2\gamma t}}}{\sqrt{2\pi\gamma t}}.$$

If β is finite (e.g., non-zero temperature), then $\gamma < 1$, hence all the local magnetizations decay exponentially to zero with rate $\tau_r = (1-\gamma)^{-1}$. Conversely, at $\beta = +\infty$ (zero temperature), $\gamma = 1$, and the decay is no longer exponential but power-law, indicating *critical slowing down*. At zero temperature, the local magnetization, depends on the space k and time t variables, through a special combination, i.e., according to the *scaling form*

$$m_k(t) \simeq \frac{\mathrm{e}^{-\frac{k^2}{2t}}}{\sqrt{2\pi t}} = \frac{1}{\sqrt{t}} f\left(\frac{k^2}{2t}\right).$$

This decays algebraically, as $t^{-1/2}$, for $k^2 \ll t$, while for $k^2 \gg t$ it decays exponentially. The scaling form of $m_k(t)$ suggest *self-similarity*, i.e. if one rescales distances (from the origin) k with time, the value of the scaling function f is unchanged. This is due to the fact that the spin at the origin is extending its influence on domains which grow with a lengthscale which is order $\sqrt{t}$. This is typical of *diffusion at interfaces*. In other words, the magnetisation "diffuse" in the system starting from the origin where we forced a spin to be up.

2.1.2. *Global magnetisation.* Let us now look at the global magnetization $m = \frac{1}{N}\sum_k m_k$. From the master equation (3.17), it is immediate to obtain (it is enough to sum both sides over i)

$$\partial_t m = -(1-\gamma)m \Longrightarrow m(t) = m(0)\, \mathrm{e}^{-(1-\gamma)t}$$

For $\beta^{-1} > 0$, $\gamma < 1$, so $m(t)$ decays exponentially fast to zero. For $\beta^{-1} = 0$, $\gamma = 1$, so $m(t) = m(0)$, i.e., the global magnetization is conserved. For the initial condition $m_k(0) = \delta_{k,0}$, and N large, $m(0) = 0$, hence $m(\infty) = 0$. This shows that a one-dimensional system does not spontaneously develop a global magnetization, so there is no phase transition to ferromagnetic order.

2.1.3. *Correlation functions.* To explore the system's behaviour further, we can look at the time behaviour of correlation functions, in particular, the equal-time correlator $C_{ij}(t) \coloneqq \mathbb{E}[\sigma_i(t)\sigma_j(t)]$. To write an equation for its evolution, we note that

$$\sigma_i(t + \mathrm{d}\,t)\sigma_j(t + \mathrm{d}\,t) = \begin{cases} -\sigma_i(t)\sigma_j(t) & \text{with prob } (W_i(\boldsymbol{\sigma}) + W_j(\boldsymbol{\sigma}))\, \mathrm{d}\,t, \\ \sigma_i(t)\sigma_j(t) & \text{with prob } 1 - (W_i(\boldsymbol{\sigma}) + W_j(\boldsymbol{\sigma}))\, \mathrm{d}\,t. \end{cases} \tag{3.18}$$

Here we used the fact that the probability that *only one spin of the pair* (i, j) *flips in the interval* $[t, t + \mathrm{d}\,t]$ is $W_i(\boldsymbol{\sigma})\, \mathrm{d}\,t(1 - W_j(\boldsymbol{\sigma})\, \mathrm{d}\,t) + W_j(\boldsymbol{\sigma})\, \mathrm{d}\,t(1 - W_i(\boldsymbol{\sigma})\, \mathrm{d}\,t) \simeq (W_i(\boldsymbol{\sigma}) + W_j(\boldsymbol{\sigma}))\, \mathrm{d}\,t$. It follows that

$$\partial_t C_{ij}(t) = -2\mathbb{E}[\sigma_i \sigma_j (W_i(\boldsymbol{\sigma}) + W_j(\boldsymbol{\sigma}))]$$

Substituting $W_i(\boldsymbol{\sigma}) = \frac{1}{2}\left(1 - \gamma\sigma_i \frac{\sigma_{i+1}+\sigma_{i-1}}{2}\right)$

$$\partial_t C_{ij} = -2C_{ij} + \gamma \mathbb{E}\left[\sigma_i \frac{\sigma_{j-1}+\sigma_{j+1}}{2} + \sigma_j \frac{\sigma_{i-1}+\sigma_{i+1}}{2}\right]. \tag{3.19}$$

Let us assume now, as initial condition, spins are initialised randomly, so that $C_{ij}(0) = \delta_{ij}$ for all pair i and j. Since the interactions are the same for any pair of spins, the system is expected to keep this homogeneity property, i.e., the correlation function is expected to depend on the distance between the spins, and not on their individual position, at all times, i.e., $C_{ij}(t) \equiv C_k(t)$ with $k = |i-j|$. Rewriting the equation above for C_k we get

$$\partial_t C_k = -2C_k + \gamma C_{k+1} + \gamma C_{k-1}.$$

This is nothing but (again) the equation for random walker, in the presence of the boundary condition $C_0(t) = 1 \ \forall\ t$: the equation can be solved by applying the method of images. Here, however, we limit to look at equilibrium, when the time derivative is zero. Substituting the ansatz $\lim_{t\to+\infty} C_k(t) = \eta^k$ in the equation and requiring that the RHS vanishes, we get

$$2 = \gamma\left(\eta + \frac{1}{\eta}\right) \Longrightarrow \eta = \frac{1 \pm \sqrt{1-\gamma^2}}{\gamma},$$

where the minus sign has to be selected for the solution to be physical as $0 \le C_k \le 1$. Some algebraic steps

$$\eta = \frac{1 - \frac{1}{\cosh(2\beta J)}}{\tanh(2\beta J)} = \frac{\cosh(2\beta J) - 1}{\sinh(2\beta J)} = \frac{2\sinh^2(\beta J)}{2\sinh(\beta J)\cosh(\beta J)} = \tanh(\beta J) \equiv \gamma$$

yield to

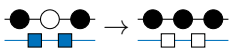
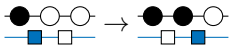
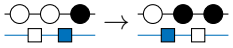
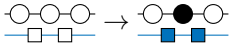
$$C_k^{\mathrm{eq}} := \lim_{t\to+\infty} C_k(t) = \tanh^k(\beta J) = \mathrm{e}^{-\frac{k}{\xi}}, \qquad \xi := -\frac{1}{\ln\gamma}.$$

In other words, at equilibrium $C_k^{\mathrm{eq}} = 0$ for $k \gg \xi$, while $C_k^{\mathrm{eq}} \simeq 1$ for $k \ll \xi$, hence ξ plays the role of a *correlation length*: spins at distances smaller than ξ are correlated while spins at larger distances than ξ are uncorrelated. In the limit $\beta^{-1} \to 0$, we have $\xi \to \infty$, hence one has a single domain of ferromagnetic order: $C_k^{\mathrm{eq}} = 1$, meaning that all spins are aligned.

2.1.4. *Domain walls.* To conclude, let us now look at the density ρ of domain walls, i.e., interfaces between neighbouring spins with opposite orientations, at equilibrium:

$$\rho = \frac{1}{N}\sum_k \mathbb{E}\left[\frac{1-\sigma_k\sigma_{k+1}}{2}\right] = \frac{1 - C_1^{\mathrm{eq}}}{2}$$

We know that at zero temperature $C_k^{\mathrm{eq}} = 1$, so $\rho = 0$; instead, for finite temperature, $C_1^{\mathrm{eq}} = \gamma$, so $\rho = \frac{1-\gamma}{2}$. In order to better understand the system's behaviour at zero temperature, we note that the dynamics can be mapped onto a process of *particle diffusion and annihilation*. At zero temperature (i.e., $\gamma = 1$) domain walls behave as particles diffusing and annihilating when they meet. This is understood by computing the rate $W_i(\boldsymbol{\sigma})$ for the following processes:

Annihilation		rate 1
Move right		rate $1/2$
Move left		rate $1/2$
Creation		rate 0

In the picture, the upper line represents the spin configuration (e.g., black is a $+1$ spin, white is a -1 spin) and the lower line represents the domain wall configuration (so a square is 'on' in between two spins of opposite sign). In particular, the last rule tells us that domains of

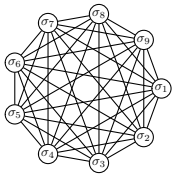
aligned spins cannot break up. In the long time limit, all the particles will have annihilated, so no domain wall will survive, and a single domain of ferromagnetic order will be found. How is this consistent with the finding $\lim_{t\to+\infty} m(t) = 0$? It must be noted that m is an average over all trajectories (ensemble average). The system does reach consensus in every single trajectory at zero temperature, i.e., the instantaneous magnetization $m(\boldsymbol{\sigma}) = N^{-1}\sum_i \sigma_i$ does converges to ± 1. However, half of the trajectories will converge to $m = 1$ and half to $m = -1$, so that the distribution of the thermodynamic magnetization at the end of the process will be $P(m) = \frac{1}{2}\delta_{m,1} + \frac{1}{2}\delta_{m,-1}$.

In the above analysis we had assumed the system initialised in a state with $m(0) = 0$, so $m(\infty) = 0$ follows from conservation of the magnetization at zero temperature. If the system is initialised in a different initial condition, with, say, K spins up and $N - K$ spins down, one would have

$$m(0) = \frac{K - (N-K)}{N} = 2\frac{K}{N} - 1.$$

Since the thermodynamic magnetization, at the end of the process, has to match the initial one, its distribution would have to be $P(m) = p\delta_{m,1} + (1-p)\delta_{m,-1}$ with $\lim_{t\to+\infty} m(\infty) = 2p - 1 \equiv 2\frac{K}{N} - 1 \Rightarrow p = \frac{K}{N}$, i.e., consensus is always achieved and the initial number of up-spins determines the likelihood of the consensus state with all spins up. As a final remark, we note that in this system the magnetization exhibits large fluctuations about the average behaviour $m = \mathbb{E}[m(\boldsymbol{\sigma})]$. This is typical of low-dimensional systems. Conversely, for large-dimensional systems, macroscopic observables fluctuate little about their average and the mean-field approximation becomes exact.

2.2. The Curie–Weiss model. The Curie–Weiss model is the *infinite range* version of the Ising model. In this case, every spin interact with all the others and



$$h_i(\boldsymbol{\sigma}) = \frac{J}{N}\sum_{j\neq i}\sigma_j.$$

Since interactions are symmetric the equilibrium distribution is of Boltzmann-type with Hamiltonian

$$H(\boldsymbol{\sigma}) = -\frac{J}{2}\sum_{i\neq j}\sigma_i\sigma_j = -\frac{NJ}{2}\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i\right)^2 + \frac{J}{2}.$$

One can write the local field as the sum of its average value plus fluctuations about the average

$$h_i(\boldsymbol{\sigma}) = \mathbb{E}[h_i(\boldsymbol{\sigma})] + \delta h_i(\boldsymbol{\sigma}) = Jm - \frac{\mathbb{E}[\sigma_i]}{N} + \delta h_i(\boldsymbol{\sigma}),$$

with $\mathbb{E}[\delta h_i(\boldsymbol{\sigma})] = 0$ and $m(t) = \frac{1}{N}\sum_i \mathbb{E}[\sigma_i(t)]$, for large N one has $h_i(\boldsymbol{\sigma}) \simeq Jm + \delta h_i(\boldsymbol{\sigma})$. Away from phase transitions, where correlations become important, spins fluctuate independently about their average value, so $\delta h_i(\boldsymbol{\sigma})$ is expected to be small, being a sum of many independent zero-average variables. Expanding Eq. (3.14) for small Δ_i we have

$$\begin{aligned}\partial_t\mathbb{E}[\sigma_i] &= -\mathbb{E}[\sigma_i] + \mathbb{E}[\tanh\beta[Jm + \delta h_i(\boldsymbol{\sigma})]]\\ &= -\mathbb{E}[\sigma_i] + \tanh\beta Jm + \beta(1-\tanh^2\beta Jm)\mathbb{E}[\delta h_i(\boldsymbol{\sigma})] + \dots\\ &\simeq -\mathbb{E}[\sigma_i] + \tanh\beta Jm\end{aligned} \tag{3.20}$$

The global magnetization is then found to evolve according to

$$\partial_t m = -m + \tanh J\beta m \equiv f(m) \tag{3.21}$$

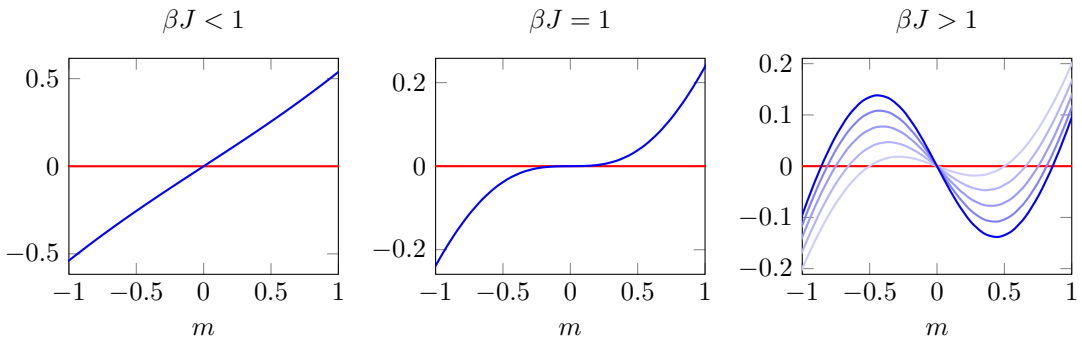


FIGURE 1. Plot $f(m) = -m + \tanh(\beta J m)$, with $J = 1$, for different values of βJ (in the right plot, darker color corresponds to higher value of βJ).

as expected from Eq. (3.14) within a mean-field approximation $\mathbb{E}[\tanh \beta h_i] \simeq \tanh \beta \mathbb{E}[h_i] \equiv \tanh J\beta m$. The steady state is found then imposing

$$m = \tanh \beta J m$$

Multiplying this by m, using that tanh is an odd function, and $|\tanh x| \leq |x|$, one has

$$m^2 = |m||\tanh \beta J m| \leq \beta J |m|^2$$

showing that for $\beta J < 1$ the only solution is $m = 0$. This can also be seen from the plot of $f(m)$ shown in Figure 1. One can see that for $\beta J \leq 1$, the function $f(m)$ has only one zero in $m = 0$, so this is the only (stable) steady state of the system, while for $\beta J > 1$, there are two additional zeros. We show below that the latter two solutions are stable, while the solution in zero becomes unstable for $\beta J > 1$.

This result also indicates the presence of a phase transition at $\beta J = 1$, where the steady state of the system m_{eq} changes from $m_{\text{eq}} = 0$, for $\beta J \leq 1$, to $m_{\text{eq}} = \pm m_T$ for $\beta J > 1$, $m_T > 0$, $\pm m_T$ being the new nonzero solution of the equation $f(m) = 0$. The solutions m_T depend on the temperature $T = \beta^{-1}$. To see this, one can solve the equation $f(m) = 0$ graphically, as in Fig. 1. For $T \to 0$, i.e., for $\beta \to +\infty$, $\tanh(\beta J m) \to \text{sign}(m)$, i.e., $m_T \to 1$.

Assuming that for $N \to 0$, $\frac{1}{N}\sum_i \sigma_i \to m$, the Hamiltonian of the system is an even function of m that can be written as $H = -\frac{JN}{2}m^2 + \frac{J}{2}$. The emergence of two stable solutions, $\pm m_T$, in a system described by a even Hamiltonian, is referred to as 'spontaneous symmetry breaking'. This refers to the fact that, although the Hamiltonian is symmetric under $m \to -m$, the system will select dynamically one of the two solutions, either m_T or $-m_T$, depending on the initial condition. In particular, for $m(0) = 0^+$ the system will converge to m_T, for $m(0) = 0^-$, the system will converge to $-m_T$. Hence, in contrast to the one-dimensional model, the evolution (3.21) does not conserve the magnetitzation of the system. A magnetization will spontaneously emerge in the system below the critical temperature $T = J$.

2.2.1. *Linear stability analysis.* We have seen that for $T > J$, $m_{\text{eq}} = 0$ is the only solution. To study the stability of this solution, we can linearise the dynamics about $m_{\text{eq}} = 0$. Calling $\hat{\beta} = \beta J$ for brevity, if $\hat{\beta} < 1$

$$\frac{\mathrm{d}\, m}{\mathrm{d}\, t} \simeq m(\hat{\beta} - 1) \Longrightarrow m(t) = m(0)\, \mathrm{e}^{-\frac{t}{\tau_r}} \quad \text{with} \quad \tau_r = \frac{1}{1 - \hat{\beta}} > 0$$

shows an exponential approach of the system to $m = 0$. For $\hat{\beta} \to 1^-$, the exponential rate diverges $\tau \to \infty$, showing critical slowing down. For $\hat{\beta} > 1$, $\tau < 0$, so $m \propto \mathrm{e}^{\frac{t}{|\tau|}}$, showing that $m = 0$ is unstable in this regime and the magnetisation grows.

We have shown before that, for $T < J$, two new solutions $m_{\mathrm{eq}} = \pm m_T$ appear. Linearising about m_T, by setting $m = m_T + \epsilon$, for ϵ small, we have

$$\begin{aligned} \frac{\mathrm{d}\,\epsilon}{\mathrm{d}\,t} &= \tanh(\hat{\beta}(m_T + \epsilon)) - m_T - \epsilon \\ &\simeq \tanh(\hat{\beta} m_T) + \hat{\beta}\epsilon[1 - \tanh^2(\hat{\beta} m_T)] - m_T - \epsilon \\ &= \epsilon[\hat{\beta}(1 - m_T^2) - 1] \Longrightarrow \\ \epsilon(t) &= \epsilon(0)\,\mathrm{e}^{-\frac{t}{\tau}}, \quad \tau_r \coloneqq \frac{1}{1 - \hat{\beta}(1 - m_T^2)}. \end{aligned} \tag{3.22}$$

The quantity τ_r is positive. To show this, we differentiate the steady state condition with respect to β

$$\begin{aligned} \partial_\beta m_T = \partial_\beta \tanh \beta J m_T = J(1 - \tanh^2 \beta J m)\,(m_T + \beta \partial_\beta m_T) \\ \equiv J(1 - m_T^2)\,(m_T + \beta \partial_\beta m_T) \end{aligned} \tag{3.23}$$

which can be rearranged as

$$\frac{1}{m_T}\frac{\partial m_T}{\partial \beta}\underbrace{(1 - \hat{\beta}(1 - m_T^2))}_{\tau_r^{-1}} = J(1 - m_T^2) \geq 0$$

The first factor is positive as m_T is an increasing function of T. This implies that the term in the round bracket on the LHS is positive as well, hence $\tau \geq 0$. It follows that the system approaches m_T exponentially fast.

We are left with the question of what happens at $T = J$, i.e., $\hat{\beta} = 1$, where $m_T \to 0$ and all our formulas fail. Setting $\hat{\beta} = 1$ in the equation for the magnetisation, and expanding about $m \simeq 0$,

$$\frac{\mathrm{d}\,m}{\mathrm{d}\,t} = \tanh m - m \simeq -\frac{m^3}{3} \Longrightarrow m = \sqrt{\frac{3}{2}}\frac{1}{\sqrt{t + \frac{3}{2m^2(0)}}} \sim \sqrt{\frac{3}{2t}},$$

which is a slow, power law decay towards $m = 0$.

Q In conclusion, we have three distinct behaviours in the Curie–Weiss model
***T* > *J*:** an exponential relaxation to $m = 0$ occurs;
***T* = *J*:** the system relaxes as $t^{-1/2}$ to $m = 0$;
***T* < *J*:** the system relaxes to one of the equilibrium ordered states, exponentially fast.
The order parameter m changes continuously from $m = 0$ (at $T > J$) to $m \neq 0$ (at $T < J$), hence we are in the presence of a second order phase transition.

Q The Ising model was originally introduced to explain ferromagnetic behaviour in magnetic systems, and it has soon become a paradigmatic model to understand phase transitions. These are classical subjects of statistical physics, which is concerned with the emergence of collective behaviour in large systems of simple units, which interact locally. For example, in lattice models, each atom interacts with only a few others (their neighbours in the crystalline structure), yet the system may exhibit long-range order, giving raise to non-trivial properties, like magnetism. As we saw, in the Ising model, neighbouring atoms i and j are assumed to interact via a coupling $J_{ij} > 0$ which favours alignment of their magnetic momenta.On the other hand, the temperature T tends to randomize their orientation, thus acting as an opposing force.

Ising solved the one-dimensional version of this model that we discussed above in 1925. As we saw, the one dimensional model does not not exhibit ferromagnetic behaviour: it was later understood that ferromagnetic transitions occur only for dimensionality above or equal to two. The solution to the two-dimensional model, due to Lars Onsager (1968 Chemistry Nobel Prize) appeared in 1944, and confirmed the emergence of a spontaneous magnetization for temperatures below a critical value. Since then, many numerical studies have been carried out at dimensionalities larger than two, however, an analytical solution of the three-dimensional model is still an open problem. Luckily, if we increase the dimensionality further to $d \geq 4$, the so-called mean-field approximation becomes exact, meaning that it gives the right critical temperature and exponents.

Not all systems have $J_{ij} \geq 0$ for all pairs i, j. In fact, most of systems are 'disordered', meaning that they have both positive and negative interactions (favouring alignment and opposite orientations, respectively). To understand the implications of this, consider the following setting:

Assume that $J_{12} > 0$, $J_{23} > 0$ but $J_{13} < 0$. In this case, σ_3 receives contrasting signal from spins σ_1 and σ_2 and, whatever configuration it takes, there will always be a link which will be 'frustrated'. It is easy to imagine how large loops involving positive and negative interactions in random order can make nontrivial the study of the 'optimal arrangement' of a spin systems, and induce many sub-optimal configurations, which frustrate some links (these normally correspond to local minima of an energy function $H(\boldsymbol{\sigma})$). This situation occurs often in real systems, where units may receive conflicting pieces of information, via different feedback loops. Neural circuits and neural networks tend to be of this type. The relevance of this kind of systems is behind the motivation for the assignment of the 2022 Physics Nobel Prize to Giorgio Parisi, who in 1979 solved the disordered version of the Curie–Weiss model, known as *Sherrington–Kirkpatrick model*.

3. Hebbian interactions

In the Curie–Weiss model, where all the interactions are the same, we could describe the system's behaviour in terms of a single order parameter, the global magnetization m. When the interactions are non-homogeneous, things are more complicated, however, for special structures of the interactions J_{ij} it may still be possible to obtain a coarse-grained description of the system in terms of a few observables, in closed form. This is the case for Hebbian interactions in neural networks.

3.1. Hebb's rule. We consider a neural network with N neurons, evolving according to the dynamics in Eq. (3.4). We wish to *store* P patterns in the network. A pattern is a vector $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \dots, \xi_N^\mu)$, $\mu = 1 \dots P$, with N binary entries $\xi_i^\mu \in \{-1, 1\}$. This pattern can represent, for example, a black and white picture, in which $\xi_i^\mu = +1$ means that the ith pixel is black, $\xi_i = -1$ means that the pixel is white.

We may wonder now if it is possible to make a suitable choice of the interactions J_{ik}, so that we get the neural network to store and retrieve patterns like $\boldsymbol{\xi}^\mu$. Let us start from the simplest case, i.e., let us assume that we want to store a single pattern $\boldsymbol{\xi} = (\xi_1, \dots, \xi_N)$, $\xi_i = \pm 1$. We say that the system "retrieves" the pattern, if, when initialised in a configuration $\boldsymbol{\sigma}(0) = \boldsymbol{\sigma}_0$, the network converges, under dynamical evolution, to configuration $\boldsymbol{\xi}$, i.e., $\lim_{t \to +\infty} \boldsymbol{\sigma}(t) = \boldsymbol{\xi}$. We will show that this can be achieved imposing *Hebbian interactions*, i.e.,

$$J_{ij} = \frac{1}{N} \xi_i \xi_j \quad \forall i \neq j, \qquad J_{ii} = 0. \tag{3.24}$$

Since the interactions are symmetric, we know from previous sections that a Glauber dynamics will converge to a Boltzmann equilibrium distribution described by the Hamiltonian

$$H(\boldsymbol{\sigma}) = -\frac{1}{2N}\sum_{i\neq j}\xi_i\xi_j\sigma_i\sigma_j - \sum_i \theta_i\sigma_i.$$

For $\theta_k = 0$, we can map the neural network to a ferromagnet with $J = 1$ by defining new 'spin' variables $\tau_i = \xi_i\sigma_i$, so that

$$H(\boldsymbol{\tau}) = -\frac{1}{2N}\sum_{i\neq j}\tau_i\tau_j.$$

This is a Curie–Weiss model and we know that a nonzero magnetisation $m \neq 0$ will develop in the system for $T < 1$, where the "magnetization" is now defined as $m = \frac{1}{N}\sum_i \tau_i \equiv \frac{1}{N}\sum_i \xi_i\sigma_i = \frac{1}{N}\langle\boldsymbol{\xi}|\boldsymbol{\sigma}\rangle$. But this quantity is the scalar product between the vector $\boldsymbol{\xi}$ representing the stored pattern and the system configuration. In other words, $m \neq 0$ means that our system will explore configurations $\boldsymbol{\sigma}$ which are aligned or anti-aligned with the pattern $\boldsymbol{\xi}$. The magnetisation m is also called the *overlap* between the system configuration and the stored pattern and it quantifies the system's retrieval of the pattern itself. In particular, we have seen that at zero temperature $m = \pm 1$, meaning that $\boldsymbol{\sigma} \to \pm\boldsymbol{\xi}$, where the $\pm$ sign will be selected from the initial condition. The network will thus have two attractors, one in $\boldsymbol{\xi}$ and one in its opposite, and will converge to one or the other depending on whether the initial configuration is sufficiently similar or dissimilar to the stored pattern. The fixed point $m = 0$, instead, corresponds to all non-sensical initial conditions (i.e. those which are neither similar nor dissimilar to the stored pattern) and is, as we know, unstable for $T < 1$.

3.2. The Hopfield model. In 1982, John Hopfield introduced a model for the study of associative neural networks that has soon played a central role in the field. Hopfield considered the problem of storing P patterns $\{\boldsymbol{\xi}^\mu\}_{\mu=1}^P$, but he assumed them to be random, i.e., for each μ and for each i, we randomly fix $\xi_i^\mu = 1$ or $\xi_i^\mu = -1$ with equal probability. He then considered a neural network with the following interactions

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^P \xi_i^\mu\xi_j^\mu \quad \forall i \neq j \qquad J_{ii} = 0. \tag{3.25}$$

This formula generalises the Hebb's rule above to the case of P patterns. To study this case, it is convenient to define P overlaps

$$m_\mu(\boldsymbol{\sigma}) = \frac{1}{N}\sum_{i=1}^P \sigma_i\xi_i^\mu = \frac{1}{N}\langle\boldsymbol{\xi}^\mu|\boldsymbol{\sigma}\rangle, \quad \mu = 1,\dots,P,$$

each quantifying the retrieval of one pattern. Again, interactions are symmetric and the system will converge to a Boltzmann distribution with Hamiltonian

$$H(\boldsymbol{\sigma}) = -\frac{1}{2N}\sum_{\mu=1}^P\sum_{i\neq j}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j = -\frac{N}{2}\sum_{\mu=1}^P m_\mu^2(\boldsymbol{\sigma}) + \frac{P}{2}.$$

Let us see how the system behaves in this case.

3.2.1. *Noiseless dynamics and storage capacity.* Let us start from the dynamics in Eq. (3.4) in the *noiseless regime*, i.e.,

$$\sigma_i(t+\tau) = \text{sign}\left(\sum_{j=1}^N J_{ik}\sigma_j(t)\right).$$

For the given prescription of the interactions, let us show that the stored patterns are fixed points of the above dynamics, i.e., if $\boldsymbol{\sigma}(t) = \boldsymbol{\xi}^\mu$, then $\boldsymbol{\sigma}(t+1) = \boldsymbol{\xi}^\mu$. Let us write this explicitly, assuming $\boldsymbol{\sigma}(t) = \boldsymbol{\xi}^\mu$:

$$\sigma_i(t{+}1){=}\operatorname{sign}\left(\frac{1}{N}\sum_{j=1}^{N}\sum_{\nu=1}^{P}\xi_i^\nu\xi_j^\nu\xi_j^\mu\right){=}\operatorname{sign}\Big(\xi_i^\mu{+}\underbrace{\frac{1}{N}\sum_{j}^{N}\sum_{\nu\neq\mu}^{P}\xi_i^\nu\xi_j^\nu\xi_j^\mu}_{z_i^\mu}\Big){=}\xi_i^\mu\operatorname{sign}(1{+}\xi_i^\mu z_i^\mu). \tag{3.26}$$

The quantity z_i^μ is a crosstalk term, denoting pattern interference: if z_i^μ is small, we will be able to correctly retrieve the i-th entry of the pattern μ. This is not something that can happen for any value of P and actually imposes some constraints on the number of patterns P that we can safely store, i.e., on the *storage capacity* of the system. Using that $\xi_i^\mu z_i^\mu$ is a normalised sum of NP variables equal to ±1, with equal likelihood, it follows that $\xi_i^\mu z_i^\mu$ is a Gaussian variable with zero mean and variance P/N for $NP \gg 1$. To have $\sigma_i(t) = \xi_i^\mu$ we need to have $1 + \xi_i^\mu z_i^\mu > 0$, otherwise we make a mistake. The probability of making an error on the single spin is then

$$P_{\text{error}} = \mathbb{P}[\xi_i^\mu z_i^\mu < -1] = \frac{1}{\sqrt{2\pi(P/N)}}\int_1^\infty \mathrm{e}^{-\frac{x^2}{2P/N}}\,\mathrm{d}\,x = \frac{1}{2}\left[1 - \operatorname{erf}\left(\sqrt{\frac{N}{2P}}\right)\right].$$

Setting a threshold $p^\star$ on the probability to perform an error-free recall of the pattern, we want

$$(1 - P_{\text{error}})^N > p^\star.$$

This condition gives an upper bound on the number of patterns we can store in the network. From the considerations above, we can expect strong crosstalk (and thus a decrease in the retrieval accuracy) for $P = \alpha N$. This is often referred to as the *saturated regime*.

Observe that, if $\boldsymbol{\xi}^\mu$ is stable, $-\boldsymbol{\xi}^\mu$ is also stable, and in general any odd mixture of stored pattern, e.g. $\boldsymbol{\xi}^{\mu_1} \pm \boldsymbol{\xi}^{\mu_2} \pm \boldsymbol{\xi}^{\mu_3}$, will be stable as well. This last feature may be undesired, as it leads to mixing memories, however, it can be shown that noise will help in destabilizing such mixtures.

3.2.2. *Glauber dynamics.* Let us now study the dynamics at finite temperature β of the magnetisation $m(t)$ in the Hopfield model. We start from the usual master equation for the expectation of $\sigma_i(t)$,

$$\partial_t\mathbb{E}[\sigma_i(t)] = -\mathbb{E}[\sigma_i(t)] + \mathbb{E}[\tanh\beta h_i(\boldsymbol{\sigma}(t))]. \tag{3.27}$$

We will aim to express the dynamics in coarse-grained form, in terms of the overlaps $m_\mu(\boldsymbol{\sigma})$, i.e., we aim to get a closed set of equations for their thermodynamic values

$$m_\mu(t) = \mathbb{E}[m_\mu(\boldsymbol{\sigma}(t))] = \frac{1}{N}\sum_i \xi_i^\mu\mathbb{E}[\sigma_i(t)].$$

For large N, we can express the local field as

$$\begin{aligned} h_i(\boldsymbol{\sigma}) &= \frac{1}{N}\sum_{j\neq i}J_{ij}\sigma_j = \sum_{\mu=1}^{P}\xi_i^\mu\frac{1}{N}\sum_{j\neq i}\xi_j^\mu\sigma_j \\ &= \sum_{\mu=1}^{P}\xi_i^\mu\left(m_\mu(\boldsymbol{\sigma}) - \frac{1}{N}\xi_i^\mu\sigma_i\right) \simeq \sum_{\mu=1}^{P}\xi_i^\mu m_\mu(\boldsymbol{\sigma}). \end{aligned} \tag{3.28}$$

Multiplying (3.27) by ξ_i^μ, summing over i, and dividing by N we obtain

$$\partial_t m_\mu = -m_\mu + \frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu \mathbb{E}\left[\tanh\left(\beta\sum_{\nu=1}^{P}\xi_i^\nu m_\nu(\boldsymbol{\sigma})\right)\right]. \tag{3.29}$$

We next do a mean-field approximation, i.e., we assume fluctuations about thermodynamic averages small $m_\nu(\boldsymbol{\sigma}) \simeq m_\nu$, allowing us to drop the average $\mathbb{E}[\bullet]$. Defining $\boldsymbol{\xi}_i = (\xi_i^1, \dots, \xi_i^P)$, vector of the patterns in i, and $\boldsymbol{m} = (m_1, \dots, m_P)$, we get

$$\begin{aligned}
\partial_t m_\mu &= -m_\mu + \frac{1}{N}\sum_{i=1}^{N}\xi^\mu \tanh\left(\beta\sum_{\nu=1}^{P}\xi_i^\nu m_\nu\right)\\
&= -m_\mu + \frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu \tanh\left(\beta\langle\boldsymbol{\xi}_i|\boldsymbol{m}\rangle\right)\\
&= -m_\mu + \frac{1}{N}\sum_{i=1}^{N}\sum_{\boldsymbol{\xi}}\delta_{\boldsymbol{\xi},\boldsymbol{\xi}_i}\xi_i^\mu \tanh\left(\beta\langle\boldsymbol{\xi}|\boldsymbol{m}\rangle\right)\\
&= -m_\mu + \sum_{\boldsymbol{\xi}}\rho(\boldsymbol{\xi})\xi^\mu \tanh\left(\beta\langle\boldsymbol{\xi}|\boldsymbol{m}\rangle\right) \equiv -m_\mu + \mathbb{E}_{\boldsymbol{\xi}}[\xi^\mu \tanh\left(\beta\langle\boldsymbol{\xi}|\boldsymbol{m}\rangle\right)].
\end{aligned} \tag{3.30}$$

In the last step we have introduced

$$\rho(\boldsymbol{\xi}) \coloneqq \frac{1}{N}\sum_i \delta_{\boldsymbol{\xi},\boldsymbol{\xi}_i} = \rho(\boldsymbol{\xi}).$$

distribution of patterns $\boldsymbol{\xi}_i$, and used the notation $\mathbb{E}_{\boldsymbol{\xi}}[f(\boldsymbol{\xi})] \coloneqq \sum_{\boldsymbol{\xi}} f(\boldsymbol{\xi})\rho(\boldsymbol{\xi})$. The final equation can be written in vectorial form,

$$\partial_t \boldsymbol{m} = -\boldsymbol{m} + \mathbb{E}_{\boldsymbol{\xi}}[\boldsymbol{\xi}\tanh(\beta\langle\boldsymbol{\xi}|\boldsymbol{m}\rangle)]. \tag{3.31}$$

Here, the average over all sites for a specific realization of patterns has been replaced with the *average over the distribution of patterns at one particular site* (as if there were many systems). This is a simple consequence of the law of large numbers for $2^P \ll N$, however, it is a non-trivial statement if P increases (it will generally apply only to so-called *self-averaging* quantities). The steady solution will be found by imposing

$$\boldsymbol{m} = \mathbb{E}_{\boldsymbol{\xi}}[\boldsymbol{\xi}\tanh(\beta\langle\boldsymbol{\xi}|\boldsymbol{m}\rangle)]. \tag{3.32}$$

For $\beta < 1$, the only solution is $\boldsymbol{m} = \boldsymbol{0}$, as it can be shown using the fact that $x\tanh(\alpha x) \leq \alpha x^2$ for $\alpha > 0$:

$$\|\boldsymbol{m}\|^2 = \mathbb{E}_{\boldsymbol{\xi}}\left[\langle\boldsymbol{\xi}|\boldsymbol{m}\rangle\tanh(\beta\langle\boldsymbol{\xi}|\boldsymbol{m}\rangle)\right] \leq \beta\mathbb{E}_{\boldsymbol{\xi}}[\langle\boldsymbol{\xi}|\boldsymbol{m}\rangle^2] = \beta\sum_{\mu\nu} m_\mu m_\nu \underbrace{\mathbb{E}_{\boldsymbol{\xi}}[\xi^\mu\xi^\nu]}_{\delta_{\mu\nu}} = \beta\boldsymbol{m}^2. \tag{3.33}$$

As β is lowered below 1 (i.e., the temperature is increased above 1), bifurcations away from zero are expected.

Expanding Eq. (3.32) for small m_μ, and using

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\xi}}[\xi^\nu\xi^\mu] &= \delta_{\mu\nu}\\
\mathbb{E}_{\boldsymbol{\xi}}[\xi^\mu\xi^\nu\xi^\rho\xi^\lambda] &= \delta_{\mu\nu}\delta_{\rho\lambda} + \delta_{\mu\rho}\delta_{\nu\lambda} + \delta_{\mu\lambda}\delta_{\rho\nu} - 2\delta_{\mu\nu}\delta_{\nu\rho}\delta_{\rho\lambda}
\end{aligned} \tag{3.34}$$

we get

$$
\begin{aligned}
m_\mu &= \beta \sum_{\nu=1}^{P} \mathbb{E}_{\boldsymbol{\xi}}[\xi^\nu \xi^\mu] m_\nu - \frac{\beta^3}{3} \sum_{\nu,\rho,\lambda} m_\nu m_\rho m_\lambda \mathbb{E}_{\boldsymbol{\xi}}[\xi^\mu \xi^\nu \xi^\rho \xi^\lambda] \\
&= \beta m_\mu - \frac{\beta^3}{3}\left[3 m_\mu \sum_\rho m_\rho^2 - 2 m_\mu^3\right] = \beta m_\mu (1 - \beta^2 \|\boldsymbol{m}\|^2) + \frac{2\beta^3}{3} m_\mu^3
\end{aligned} \tag{3.35}
$$

i.e., putting aside for a moment the solution $m_\mu = 0$,

$$
\beta - \beta^3 \|\boldsymbol{m}\|^2 + \frac{2\beta^3}{3} m_\mu^2 = 1 \Rightarrow m_\mu^2 = \frac{3}{2}\left[\left(\frac{1}{\beta^2} - 1\right) + \|\boldsymbol{m}\|^2\right] \tag{3.36}
$$

From this equation we see that all m_μ take three possible values $m_\mu \in \{-m, 0, m\}$, so we can further simplify the equation denoting by n the number of components of $\boldsymbol{m}$ which are *not* zero:

$$
m^2 = \frac{3}{2}\left[\left(\frac{1}{\beta^2} - 1\right) + n m^2\right] \Rightarrow m \equiv m_n = \sqrt{\frac{3}{3n-2}\left(1 - \frac{1}{\beta^2}\right)} \xrightarrow{\beta \to +\infty} \sqrt{\frac{3}{3n-2}}. \tag{3.37}
$$

If we remember that m_μ is the overlap with the configuration μ, this shows that the first states to bifurcate away from zero at criticality, are symmetric mixtures of patterns, i.e., states in which $m_\mu \neq 0$ for n patterns over P.

The stability of the steady states can be studied via linearising the equation of motion, exactly as we did in the other cases, starting from Eq. (3.31) about a fixed point $\boldsymbol{m}^\star$. One sets $\boldsymbol{m} = \boldsymbol{m}^\star + \boldsymbol{\epsilon}$ and expands for small $\boldsymbol{\epsilon}$

$$
\partial_t \boldsymbol{\epsilon} = -\underbrace{\mathbb{E}_{\boldsymbol{\xi}}\Big[\boldsymbol{I} - \alpha(\boldsymbol{m}^\star)|\boldsymbol{\xi}\rangle\langle\boldsymbol{\xi}|\Big]}_{\boldsymbol{A}(\boldsymbol{m}^\star)} \boldsymbol{\epsilon}, \quad \alpha(\boldsymbol{m}^\star) \coloneqq \beta - \beta \tanh^2(\beta \langle \boldsymbol{\xi} | \boldsymbol{m}^\star \rangle). \tag{3.38}
$$

This implies that $\boldsymbol{\epsilon}(t) = \mathrm{e}^{-\boldsymbol{A}(\boldsymbol{m}^\star)t}\, \boldsymbol{\epsilon}(0)$. For $T > 1$, where $\boldsymbol{m}^\star = \boldsymbol{0}$, one has $\boldsymbol{A}(\boldsymbol{0}) = (1-\beta)\boldsymbol{I}$, hence we have exponential approach to $\boldsymbol{m}^\star = \boldsymbol{0}$. For $T < 1$, it is possible to show that a symmetric mixture of the type

$$
\boldsymbol{m}^\star = \sqrt{\frac{3(1-T^2)}{3n-2}}(1, 0, 1, 1, 0, \ldots, 0),
$$

with n nonzero components, is stable for n odd and $T < T^\star(n)$, with $T^\star(n)$ decreasing as n increases. In particular, all mixtures with odd-n are stable at $T = 0$, however, only pure states (i.e., states with $n = 1$) remain stable as $T \to 1^-$.

CHAPTER 4

The Fokker–Planck equation and the Langevin equation

ABSTRACT. In the previous chapters, we have considered simple Markov processes for which the master equation could be solved analytically. Most of the time, however, an analytical solution is not available. However, in certain limits, one can carry out an expansion and cast the master equation into a partial differential equation, known as the Fokker–Planck equation. Popular expansion schemes are the Kramers–Moyal expansion and the Van Kampen expansions, which we will discuss in this chapter.

1. The Kramers–Moyal expansion

In this chapter, we will consider the evolution of a *continuous* Markovian random process X_t taking values in $\Omega \subseteq \mathbb{R}$. For this process, we cannot work with probabilities, due to the cardinality of the set Ω, but we can introduce a *density* of probability $p(x,t)$ such that $\mathbb{P}[\mathsf{X}_t \in (x, x+\mathrm{d}\,x)] = p(x,t)\,\mathrm{d}\,x$. Interestingly we can derive a general expression an equation for $p(x,t)$, called *Fokker–Planck equation*, via a *Kramers-Moyal expansion* of the master equation for a continuous random variable X_t. Let us start precisely from the master equation, that takes the form we have anticipated in Eq. (2.8)

(4.1)
$$\partial_t p(x,t) = \int [W(x|x')p(x',t) - W(x'|x)p(x,t)]\,\mathrm{d}\,x' \qquad \text{with} \quad W(x'|x) = \lim_{\tau\to 0} \frac{p_{1|1}(x', t+\tau|x,t)}{\tau}.$$

We introduce now a function w defined as $w(\eta|x) \coloneqq W(x+\eta|x)$: in other words, $w(\eta|x)$ is the rate at which, starting from x the system is *displaced* by η. We can write the master equation as

$$\partial_t p(x,t) = \int [w(\eta|x-\eta)p(x-\eta,t) - w(\eta|x)p(x,t)]\,\mathrm{d}\,\eta, \tag{4.2}$$

Up to now, this is just a rewriting. Let us assume now that $w(\eta|x)$ is smooth and decays very fast in η, i.e., that large displacements are very unlikely. We can expand for small η, writing

$$w(\eta|x-\eta)p(x-\eta,t) = \sum_{k=0}^{\infty} \frac{(-\eta)^k}{k!} \partial_x^k [w(\eta|x)p(x,t)] \tag{4.3}$$

so that,

$$w(\eta|x-\eta)p(x-\eta,t) - w(\eta|x)p(x,t) = \sum_{k=1}^{\infty} \frac{(-\eta)^k}{k!} \partial_x^k [w(\eta|x)p(x,t)]. \tag{4.4}$$

Substituting in the master equation, we obtain the *Kramers–Moyal expansion*

$$\partial_t p(x,t) = \sum_{k=1}^{\infty} \frac{1}{k!} \partial_x^k \left[\left(\int (-\eta)^k w(\eta|x)\,\mathrm{d}\,\eta \right) p(x,t) \right] \equiv \sum_{k=1}^{\infty} \frac{1}{k!} \partial_x^k \left[a^{(k)}(x) p(x,t) \right]. \tag{4.5}$$

where we defined the jump moments

$$a^{(k)}(x) \coloneqq \int \eta^k w(\eta|x)\,\mathrm{d}\,\eta. \tag{4.6}$$

If the terms for $k > 2$ are negligible, we can truncate the expansion to the second order, obtaining the general form of the Fokker–Planck equation

$$\partial_t p(x,t) = -\partial_x \left[a^{(1)}(x)p(x,t)\right] + \frac{1}{2}\partial_x^2 \left[a^{(2)}(x)p(x,t)\right]. \tag{4.7}$$

The coefficients $a^{(1)}(x)$ and $a^{(2)}(x)$ are called the *drift* and *diffusion* coefficient, respectively. The drift and diffusion coefficients can be calculated recalling the definition of rates:

$$\begin{aligned} a^{(k)}(x) &= \lim_{\tau\to 0}\frac{1}{\tau}\int \mathrm{d}\,\eta\,\eta^k p_{1|1}(x+\eta, t+\tau|x,t) \\ &= \lim_{\tau\to 0}\frac{1}{\tau}\int \mathrm{d}\,x'\,(x'-x)^k p_{1|1}(x', t+\tau|x,t) \\ &= \lim_{\tau\to 0}\frac{\mathbb{E}[(\mathsf{X}_{t+\tau}-\mathsf{X}_t)^k|\mathsf{X}_t = x]}{\tau}. \end{aligned} \tag{4.8}$$

Hence, calculation of $a^{(1)}$ and $a^{(2)}$ only needs the knowledge of $\mathbb{E}[\mathsf{X}_{t+\tau}-\mathsf{X}_t|\mathsf{X}_t = x]$ and $\mathbb{E}[(\mathsf{X}_{t+\tau}-\mathsf{X}_t)^2|\mathsf{X}_t = x]$ to linear order in τ. As the master equation, the Fokker–Planck equation can be solved analytically for a few special cases; however, it has two alluring features when compared with the master equation:

(1) it is a partial differential equation rather than a *integro*-differential equation (so it is easier to solve numerically);
(2) it does not require the knowledge of the entire kernel $W(x|x')$, but only of two functions $a^{(1)}(x)$ and $a^{(2)}(x)$.

When the step size in the master equation cannot be made arbitrarily small (for example, in the study of chemical reactions, or in population dynamics) the Kramers–Moyal expansion may *not* give a good approximation. The truncation to $k = 2$ in the Kramers–Moyal expansion may seem rather arbitrary. In this case, a different expansion can be carried out when the system has a large parameter (volume, number of particles, etc.): this expansion, due to Van Kampen, will be presented below.

1.1. The stationary solution. Any Fokker–Planck equation has the general form of a continuity equation, i.e., $\partial_t p(x,t) = -\partial_x J(x,t)$. The Fokker–Planck equation states that the probability density is conserved:

$$\partial_t p(x,t) + \partial_x J(x,t) = 0, \tag{4.9}$$

where $J(x,t)$ is

$$J(x,t) \coloneqq a^{(1)}(x)p(x,t) - \frac{1}{2}\partial_x[a^{(2)}(x)p(x,t)]. \tag{4.10}$$

At stationarity, $p(x,t) \equiv \pi(x)$, hence $\partial_x J(x) = 0$, giving $J(x) = a^{(1)}(x)\pi(x) - \frac{1}{2}\partial_x[a^{(2)}(x)\pi(x)] =$ constant. If there is no flux of probability at the boundary, or detailed balance is satisfied, then at stationarity $J(x) = 0\ \forall\ x.$ and therefore

$$a^{(1)}(x)\pi(x) = \frac{1}{2}\partial_x[a^{(2)}(x)\pi(x)] \Rightarrow \pi(x) = \frac{1}{Za^{(2)}(x)}\exp\left[\int \frac{2a^{(1)}(y)}{a^{(2)}(y)}\,\mathrm{d}\,y\right], \tag{4.11}$$

where Z is a proper normalisation. Note that, for unbounded systems, i.e., systems evolving on $\mathbb{R}$, normalization of the probability distribution implies $\lim_{x\to\pm\infty}\pi(x) = \lim_{x\to\pm\infty}\partial_x p(x,t) = 0$,

so $\lim_{x\to\pm\infty} J(x,t) = 0$, meaning that, at stationarity, J must vanish everywhere. For bounded systems, confined in a finite interval $\Omega = [a,b]$, one can have different boundary conditions:

(1) reflecting barrier, $J(a,t) = J(b,t) = 0$ (current through barriers is zero);
(2) absorbing barrier $p(a,t) = p(b,t) = 0\ \forall\ t$;
(3) periodic boundary conditions, $p(a,t) = p(b,t)\ \forall\ t$.

For systems with reflecting boundary conditions, since the probability current vanishes at the boundaries, it will vanish everywhere at stationarity, as for unbounded systems. For these systems, the steady-state solution is thus given by (4.11).

In the special case in which $a^{(2)}(x) = 2D$ is a constant and $a^{(1)}(x) = -\partial_x V(x)$, the steady-state solution in Eq. (4.11) takes the form

$$\pi(x) = \frac{1}{Z}\,\mathrm{e}^{-\frac{V(x)}{D}}, \tag{4.12}$$

which is a typical Boltzmann form. If $V(x)$ is the external potential, then this can be identified with a Boltzmann distribution with temperature $k_{\mathrm{B}}T = D$. This result already suggests a relevant fact: the temperature T enters in the diffusion constant, i.e., is related, in the Einstein's picture of the Brownian motion, to the motility of the molecules surrounding Brown's pollen grains, as we will see below.

1.2. Equation for the moments. Similarly to the master equation, the Fokker–Planck equation can be used to get equations for averages and fluctuations. By multiplying the Fokker–Planck equation by x and integrating over x, we obtain

$$\partial_t \int x p(x,t)\,\mathrm{d}\,x = -\int x\partial_x[a^{(1)}(x)p(x,t)]\,\mathrm{d}\,x + \frac{1}{2}\int x\partial_x^2[a^{(2)}p(x,t)]\,\mathrm{d}\,x. \tag{4.13}$$

Using integration by parts and assuming p and its derivatives to be zero at the boundaries, this can be written as

$$\partial_t \mathbb{E}[\mathsf{X}_t] = \mathbb{E}[a^{(1)}(\mathsf{X}_t)]. \tag{4.14}$$

Similarly, by multiplying by x^2 and repeating the same steps, we have

$$\partial_t \mathbb{E}[\mathsf{X}_t^2] = 2\mathbb{E}[a^{(1)}(\mathsf{X}_t)\mathsf{X}_t] + \mathbb{E}[a^{(2)}(\mathsf{X}_t)].$$

The above derivation can be generalised to multi-component processes, $\mathbf{X}_t = (\mathsf{X}_t^\mu)_{\mu=1}^N$, obtaining

$$\partial_t p(\boldsymbol{x},t) = -\sum_i \partial_\mu[a_\mu^{(1)}(\boldsymbol{x})p(\boldsymbol{x},t)] + \frac{1}{2}\sum_{\mu\nu}\partial_{\mu\nu}^2[a_{ij}^{(2)}p(\boldsymbol{x},t)] \tag{4.15}$$

where $\partial_\mu \equiv \frac{\partial}{\partial x^\mu}$ and

$$a_{\mu_1,\dots,\mu_k}^{(k)}(\boldsymbol{x}) = \lim_{\tau\to 0}\frac{\mathbb{E}\Big[\prod_{j=1}^k(\mathsf{X}_{t+\tau}^{\mu_j} - \mathsf{X}_t^{\mu_j})\Big|\mathbf{X}_t = \boldsymbol{x}\Big]}{\tau} \tag{4.16}$$

1.3. The drift term. The presence of $a^{(1)}$ is a conquence of the presence of a *drift* and the term $a^{(1)}(x)$ determines a *deterministic* evolution. To justify this statement, suppose that $a^{(2)}(x) \equiv 0$, so we can focus on the effect of $a^{(1)}(x)$ only. The resulting equation is the Liouville equation

$$\partial_x p(x,t) = -\partial_x[a^{(1)}(x)p(x,t)].$$

This equation admits a solution of the type $p(x,t) = \delta(x - u(t))$, i.e., a solution in which $\mathsf{X}_t = u(t)$ *deterministically* with a given initial condition $\mathsf{X}_0 = u(0)$: the system follows a very

precise trajectory with no randomness. To obtain the expression of u, let us plug this ansatz in the equation, multiply by x and integrate:

$$\partial_t \int x\delta(x-u(t))\,\mathrm{d}\,x = -\int x\partial_x[a^{(1)}(x)\delta(x-u(t))]\,\mathrm{d}\,x \Leftrightarrow \partial_t u = a^{(1)}(u). \tag{4.17}$$

This is the equation for the trajectory u: it is the solution of the equation $\dot{u} = a^{(1)}(u)$, which is a deterministic equation for some given initial condition $u(0)$.

2. The Brownian motion

In 1827, the botanist Robert Brown observed, when suspended in water, small pollen grains are subject to an erratic and irregular motion, that, since then, took on the name of *Brownian motion*. The origin of this phenomenon was debated for many years, and the first convincing theory for it was given by Albert Einstein in 1905. Einstein's theory was refined in the very subsequent years in papers by Marian Smoluchowski (1906) and Paul Langevin (1908). By the 1950s it became clear that Brownian motion constituted a paradigm theory for many-body systems in classical mechanics and the theory developed for it could be applied to many different observables in macroscopic systems.

Einstein assumed that Brownian motion is caused by the *frequent impacts by molecules of the liquid* in which the pollen is suspended. The motion of these molecules is so complicated that its effect on the pollen grain can only be described probabilistically in terms of frequent statistically independent impacts, and the position X_t of the pollen is a stochastic process. Einstein made the following assumptions:

- each individual particle executes a motion that is independent of the motion of all other particles;
- the movements of the same particle in different time intervals $[t, t+\mathrm{d}\,t]$ and $[t+\tau, t+\tau+\mathrm{d}\,t]$ are independent processes, as long as τ is not too small.

From the practical point of view, Einstein assumed that the dynamics is the one of a *Markov chain in continuous space*, with evolution taking place at times $t = 0, \tau, 2\tau, \cdots$, with τ small, but nevertheless large enough to guarantee the mentioned independence of the dynamical evolution. He then took the $\tau \to 0$ limit, exactly as we did to obtain the equations describing continuous-time random walks. We will look, for simplicity, to the motion in one dimension. Let X_t be the position of the particle at time t, and let us suppose that the particle as a *drift velocity* v. Then

$$\mathsf{X}_{t+\tau} = \mathsf{X}_t + v\tau + \eta_\tau. \tag{4.18}$$

The important object here is η_τ. It is regarded as a random variable, incorporating the effect of the collisions of the molecules on our Brownian particle. It is a random displacement induced by the environment. We expect it to have zero average $\mathbb{E}[\eta_t] = 0$ (there is no preferred direction in the "kinks") and even distribution, $\mathbb{P}[\eta_t > 0] = \mathbb{P}[\eta_t < 0]$. Finally, its distribution does not depend on t (the effect of the environment is stationary), just on τ (for how long we observe the particle displacing). We will need only another element in our calculation. We call the *variance* of η_τ In other words,

$$\mathbb{E}[\eta_\tau^2] =: a_\tau^2 > 0. \tag{4.19}$$

We can now compute the coefficient to plug in our Fokker–Planck equation. In particular

$$a^{(1)}(x) = \lim_{\tau\to 0} \frac{\mathbb{E}[\mathsf{X}_{t+\tau} - \mathsf{X}_t | \mathsf{X}_t = x]}{\tau} = v. \tag{4.20}$$

Also we call

$$a^{(2)}(x) = \lim_{\tau\to 0} \frac{\mathbb{E}[(\mathsf{X}_{t+\tau} - \mathsf{X}_t)^2 | \mathsf{X}_t = x]}{\tau} = \lim_{\tau\to 0} \frac{\mathbb{E}[\eta_t^2]}{\tau} \equiv 2D. \tag{4.21}$$

The limit D is called *diffusion constant* and we finally have our equation

$$\partial_t p(x,t) = -v\partial_x p(x,t) + D\partial_x^2 p(x,t). \tag{4.22}$$

If $v = 0$, this equation is simply called *diffusion equation* and the process X_t is called *Wiener process*. The diffusion equation is also called *heat equation*, and it is ubiquitous in science. It had been originally derived to describe how the distribution of temperature evolves over time in a solid medium, as heat spontaneously flows from places where it is higher towards places where it is lower (since $D > 0$, $p(x,t)$ is an increasing function of time when it is convex and decreasing when it is concave).

2.0.1. *Solution by Fourier transform.* The diffusion equation can be solved using the Fourier transform, for example assuming initial condition $p(x,0) = \delta(x)$ (the initial position of the particle is in the origin, $\mathsf{X}_0 = 0$). Using the definition of Fourier transform $\hat{p}(q,t) = \int_{-\infty}^{\infty} p(x,t)\,\mathrm{e}^{iqx}\,\mathrm{d}\,x$, we write the diffusion equation in Fourier space

$$\frac{\partial \hat{p}(q,t)}{\partial t} = -Dq^2\hat{p}(q,t) \Rightarrow \hat{p}(q,t) = \hat{p}(q,0)\,\mathrm{e}^{-Dtq^2} = \mathrm{e}^{-Dtq^2}\,.$$

Taking the inverse Fourier transform leads to the result

$$p(x,t) = \int_{-\infty}^{\infty} \frac{\mathrm{d}\,q}{2\pi}\hat{p}(q,t)\,\mathrm{e}^{-iqx} = \frac{1}{\sqrt{4\pi Dt}}\,\mathrm{e}^{-\frac{x^2}{4Dt}}\,.$$

By using this Gaussian distribution, we obtain the moments of the process

$$\mathbb{E}[\mathsf{X}_t] = 0, \qquad \mathbb{E}[\mathsf{X}_t^2] = 2Dt.$$

The scaling for the average displacement $\mathsf{X}_t \sim \sqrt{t}$ is *one of the central results in statistical physics.* For a generic initial condition $p(x,0)$, one has, simply

$$p(x,t) = \int \mathrm{d}\,x \int \mathrm{d}\,q\,p(x',0)\,\mathrm{e}^{-iq(x-x')-Dq^2t} = \int p(x',0)\,\mathrm{e}^{-\frac{(x-x')^2}{4Dt}}\,\mathrm{d}\,x',$$

so that the Gaussian is the propagator of the process.

If $v \neq 0$, one can solve Eq. (4.22) using the Fourier transform, as for the $v = 0$ case. Alternatively, one can first eliminate the drift term via Galilean invariance, i.e., searching for a solution in the form $p(x,t) = g(x - vt, t)$. Plugging this form into the equation, one gets

$$\partial_t g - v\partial_x g = -v\partial_x g + D\partial_x^2 g \Leftrightarrow \partial_t g = D\partial_x^2 g \tag{4.23}$$

which brings the Fokker–Planck equation into the form we have already solved. For $p(x,0) = g(x,0) = \delta(x)$, the solution is again in Gaussian form, so that

$$p(x,t) = g(x - vt, t) = \frac{1}{\sqrt{4\pi Dt}}\,\mathrm{e}^{-\frac{(x-vt)^2}{4Dt}}\,.$$

This is now a 'travelling' Gaussian, where the average $\mathbb{E}[\mathsf{X}_t] = vt$ moves at constant speed, due to the presence of the drift v.

EXAMPLE A basic explanation of why it is harder to breathe, as one climbs mountains, relies specifically on the Fokker–Planck equation derived above. Oxygen molecules can be assimilated to Brownian particles that diffuse in the atmosphere and are subject to a constant force $F = -mg$, where g is the earth-surface gravity acceleration. Denoting X_t the position of an oxygen molecule above the earth's surface, we have

$$\partial_t p(x,t) = -\frac{mg}{\gamma}\partial_x p(x,t) + D\partial_x^2 p(x,t) \equiv -\partial_x J(x,t),$$

where we have assumed that there is a viscosity γ such that $-\gamma v = mg$, and

$$J(x,t) \coloneqq -\frac{mg}{\gamma} p(x,t) + D\partial_x p(x,t).$$

At stationarity, $J(x,t) \equiv J(x)$ and $\partial_x J(x) = 0$ (because of the Fokker–Planck equation). Since there is no probability current at the boundaries, i.e. $J(0) = J(\infty) = 0$, one has $J(x) = 0 \ \forall \ x$. The stationary solution $\pi(x)$ thus follows therefore from

$$J(x) = 0 \Rightarrow \frac{mg}{\gamma}\pi(x) = -D\partial_x \pi(x) \Rightarrow \pi(x) = \frac{1}{Z}\,\mathrm{e}^{-\frac{mgx}{D\gamma}}\,. \tag{4.24}$$

Despite the crude approximation, we have recovered the form of the basic barometric equation, which is used to model the change of air pressure P with altitude

$$P(x) = P(0)\,\mathrm{e}^{-\frac{Mgx}{RT}}$$

where M is the molar weight of dry air, T is the temperature and R the ideal gas constant. From recognising the Boltzmann form $\pi(x) = \frac{1}{Z}\,\mathrm{e}^{-\beta H(x)}$ of (4.24), we identify $H(x) = mgx$, so that $\beta \equiv \frac{1}{D\gamma}$. This equivalence tells that, in equilibrium, the fluctuation term D is related to the dissipation term γ via the equilibrium temperature

$$D = \frac{T}{\gamma},$$

which is a basic form of the fluctuation-dissipation theorem.

Q The diffusion equation obtained by Einstein can be obtained in a slightly different way. Suppose that a random walker is moving along a 1-dimensional lattice, with lattice spacing a. When the random walker is located in the site k, its position will be ak. The walker jumps at discrete times $t = n\tau$, $n = 0, 1, 2\ldots$, with the following transition probabilities

$$Q_{k\pm 1\,k}(\tau) = \mathbb{P}_{1|1}[\mathsf{X}_{t+\tau} = (k\pm 1)a | \mathsf{X}_t = ka] = \frac{1}{2} \pm \epsilon, \qquad \epsilon \in [0, 1/2].$$

We also take all other transition probabilities equal to zero. This is a discrete-time Markov process of the type we have studied in the first chapter (but on an infinite set). The idea is to take $a \to 0$ and $\tau \to 0$ so that we can obtain a process that is in the continuum and in continuous time. If, as in the first chapter, $P_k(t) \equiv \mathbb{P}[\mathsf{X}_t = ka]$, using the Markov property we can write

$$\begin{aligned} P_k(t+\tau) &= \sum_{k'} Q_{k\,k'}(\tau) P_{k'} \\ &= \left(\frac{1}{2} - \epsilon\right) P_{k-1}(t) + \left(\frac{1}{2} + \epsilon\right) P_{k+1}(t). \end{aligned} \tag{4.25}$$

Let us define now

$$P_k(t) =: a p(x,t), \qquad x = ka. \tag{4.26}$$

The quantity $p(x,t)$ is a density of probability: instead of assigning probability $P_k(t)$ to the site k, we can imagine that we assign a density $p(x,t)$ to the interval $(x - a/2, x + a/2)$ with $x = ka$. The equation then read

$$p(x, t+\tau) = \left(\frac{1}{2} - \epsilon\right) p(x-a,t) + \left(\frac{1}{2} + \epsilon\right) p(x+a,t). \tag{4.27}$$

As we did in deriving the master equation, we can subtract on both sides $p(x,t)$,

$$p(x,t+\tau) - p(x,t) = \left(\frac{1}{2} - \epsilon\right)\Big[p(x-a,t) - p(x,t)\Big] + \left(\frac{1}{2} + \epsilon\right)\Big[p(x+a,t) - p(x,t)\Big]. \tag{4.28}$$

Taylor expanding for small a and τ, the equation becomes

$$\partial_t p(x,t) = \frac{2a\epsilon}{\tau}\partial_x p(x,t) + \frac{a^2}{2\tau}\partial_x^2 p(x,t). \tag{4.29}$$

Taking the limits $\tau \to 0$ and $a \to 0$ in such a way that

$$D \coloneqq \frac{a^2}{2\tau}, \qquad v \coloneqq -\frac{2a\epsilon}{\tau},$$

we recover Eq. (4.22). In particular, if $\epsilon = 0$ (i.e., the probability of going left or right are the same), $v = 0$ and we have exactly the diffusion equation obtained by Einstein.

2.1. A related process: the Ornstein–Uhlenbeck process. An important linear process, for which we are able to calculate the time-dependent solution of the Fokker–Planck equation explicitly, is the *Ornstein–Uhlenbeck* process. The Fokker–Planck equation associated with this problem has

$$a^{(1)}(x) = -\gamma x \qquad a^{(2)}(x) = 2D, \qquad \gamma, D > 0. \tag{4.30}$$

so that it takes the form

$$\partial_t p(x,t) = \gamma \partial_x [x p(x,t)] + D\partial_x^2 p(x,t).$$

Originally used to model the velocity of a massive Brownian particle experiencing friction, this process describes a random walk with a tendency to move back towards the average value and it is widely used in financial mathematics to model interest rates, exchange rates etc. In this context, D represents the volatility caused by shocks and γ the rate at which shocks are dissipated. The stationary solution

$$\pi(x) = \sqrt{\frac{\gamma}{2\pi D}} \exp\left(-\frac{\gamma x^2}{2D}\right) \tag{4.31}$$

retrieves the form of the *Maxwell distribution* for the velocity of the molecules of a perfect gas at temperature T, $\rho(v) = \sqrt{\frac{m}{4\pi k_{\mathrm{B}} T}} \exp\left(-\frac{mv^2}{2k_{\mathrm{B}}T}\right)$ given their mass m, upon identifying ${}^{D}\!/_{\gamma} = {}^{k_{\mathrm{B}}T}\!/_{m}$. The time-dependent solution, for an initial condition $p(x,0) = \delta(x - x_0)$, is found by using the Fourier transform $\hat{p}(k,t)$ of $p(x,t)$. Multiplying the Fokker–Planck equation by e^{-ikx}, and integrating over x by parts (setting to zero the contributions from the boundaries) we have

$$\begin{aligned} \partial_t \hat{p}(k,t) &= \int \mathrm{e}^{-ikx} \left[\gamma \partial_x [x p(x,t)] + D\partial_x^2 p(x,t)\right] \mathrm{d}\, x \\ &= -\gamma k \partial_k \hat{p}(k,t) - Dk^2 \hat{p}(k,t). \end{aligned} \tag{4.32}$$

This equation can be solved by using the method of characteristics. Finally, using the inverse Fourier transform one obtains

$$\hat{p}(x,t) = \sqrt{\frac{\gamma}{2\pi D(1 - \mathrm{e}^{-2\gamma t})}} \exp\left[\frac{-\gamma (x - x_0\, \mathrm{e}^{-\gamma t})^2}{2D(1 - \mathrm{e}^{-2\gamma t})}\right].$$

For $\gamma t \gg 1$ this converges to the steady-state distribution (4.31).

3. Large size expansion

If the "step" of a stochastic process, let us call it N_t, in a master equation cannot be made arbitrarily small (for example, the process under investigation N_t is the number of molecules at some time, that, of course, can only take integer values in a population dynamics or chemical reaction kinetics), one may still be able to carry out a Kramers–Moyal expansion for the *concentration* $\mathsf{X}_t = \frac{1}{V}\mathsf{N}_t$, where V is a large parameter, e.g., the volume of the system or the population size. Starting from the master equation for a jump process

$$\dot{P}_n = r_{n-1} P_{n-1} + \ell_{n+1} P_{n+1} - (\ell_n + r_n) P_n,$$

we change variable from n to x. For $n = xV$, let us introduce

$$P_n(t) =: p(x,t)\frac{1}{V}.$$

From the definition, $p(x,t)$ is a sort of *density*, corresponding to a portion $1/V$ of the volume. Similarly, we write

$$r_n =: Vr(x), \qquad \ell_n =: V\ell(x). \tag{4.33}$$

The master equation becomes

$$\frac{\partial_t p(x,t)}{V} = r\Big(x - 1/V\Big)p\Big(x - 1/V, t\Big) + \ell\Big(x + 1/V\Big)p\Big(x + 1/V, t\Big) - (\ell(x) + r(x))p(x,t). \tag{4.34}$$

For large V we can make a Taylor expansion obtaining

$$\partial_t p(x,t) = -\partial_x\left[(r(x) - \ell(x))p(x,t)\right] + \frac{1}{2}\partial_x^2\left[\frac{r(x) + \ell(x)}{V}p(x,t)\right], \tag{4.35}$$

which is a Fokker–Planck equation with

$$a^{(1)}(x) = r(x) - \ell(x) \qquad a^{(2)}(x) = \frac{r(x) + \ell(x)}{V}.$$

As we know, the steady-state solution is given by

$$\pi(x) = \frac{1}{Z}\frac{V}{\ell(x) + r(x)}\,\mathrm{e}^{-V\phi(x)}, \qquad \phi(x) \coloneqq -2\int_0^x \frac{r(y) - \ell(y)}{r(y) + \ell(y)}\,\mathrm{d}\,y.$$

For large V, the exponent of the steady-state distribution is dominated by the minimum $x^\star$ of $\phi(x)$, and we can make a Laplace approximation

$$\phi(x) = \phi(x^\star) + \frac{1}{2}\phi''(x^\star)(x - x^\star)^2$$

leading to a Gaussian approximation

$$\pi(x) = \sqrt{\frac{V}{2\pi\sigma^2}}\,\mathrm{e}^{-V\frac{(x-x^\star)^2}{2\sigma^2}}, \qquad \sigma^2 \coloneqq \frac{1}{\phi''(x^\star)}.$$

Note that the distribution is only normalizable for $\phi''(x^\star) > 0$, i.e., when $x^\star$ is a stable fixed point of the noiseless dynamics. Hence, this approach accounts for fluctuations $\Theta(\sqrt{V})$ about metastable states, but it does not capture *large* fluctuations, i.e., $\Theta(V)$, which lead to rare events.

3.1. Van Kampen's expansion. An alternative approach to the derivation above is due to Van Kampen. This approach assumes that N_t has a time-dependent average, $\mathbb{E}[\mathsf{N}_t] = V\rho(t)$, and fluctuations about the average are $\Theta(\sqrt{V})$, so that we can decompose

$$\mathsf{N}_t = V\rho(t) + \sqrt{V}\mathsf{X}_t$$

with V volume and ρ density of the system. The goal is to obtain a set of equations for the density and to describe the stochastic fluctuations X_t. To illustrate this approach, we focus on a simple decay process

$$\bullet \to \varnothing$$

governed by the master equation with $\ell_n = 1$ and $r_n = 0$,

$$\dot{P}_n = (n+1)P_{n+1} - nP_n. \tag{4.36}$$

It is convenient to define the *raising* and *lowering* operators, such that

$$\mathcal{E}f(n) = f(n+1), \qquad \mathcal{E}^{-1}f(n) = f(n-1), \tag{4.37}$$

and write the master equation in terms of these

$$\dot{P}_n = (\mathcal{E} - 1)nP_n. \tag{4.38}$$

Since ρ is deterministic, the statistics of N_t is governed by the statistics of X_t, and we introduce

$$P_n(t) = \frac{1}{\sqrt{V}} p(x,t). \tag{4.39}$$

The function $p(x,t)$ is a density of values the quantity X_t can assume. The effect of lower and raising operators is deduced by noting that $\mathcal{E}n = n + 1$ so,

$$\mathcal{E}f(x) = \mathcal{E}f\left(\frac{n - V\rho}{\sqrt{V}}\right) = f\left(\frac{n+1-V\rho}{\sqrt{V}}\right) = f\left(x + \frac{1}{\sqrt{V}}\right) = \sum_{k=0}^{\infty} \frac{\partial_x^k f(x)}{k! V^{k/2}}, \tag{4.40}$$

hence we can say that

$$\mathcal{E} = \sum_{k=0}^{\infty} \frac{1}{k! V^{k/2}} \partial_x^k = 1 + \frac{1}{\sqrt{V}}\partial_x + \frac{1}{2V}\partial_x^2 + \dots. \tag{4.41}$$

This expression of $\mathcal{E}$ can help us to rewrite the RHS of the master equation. On the LHS, on the other hand, we have to observe that, in an interval of time $\Delta t \to 0$, $\Delta n \to 0$ and

$$\Delta n = V\Delta\rho + \sqrt{V}\Delta x \xrightarrow{\Delta t \to 0} 0 = V\,\mathrm{d}\,\rho + \sqrt{V}\,\mathrm{d}\,x \Rightarrow \frac{\mathrm{d}\,x}{\mathrm{d}\,t} = -\frac{\dot{\rho}}{\sqrt{V}}. \tag{4.42}$$

so, rewriting $\frac{\mathrm{d}}{\mathrm{d}\,t} = \partial_t + \frac{\mathrm{d}\,x}{\mathrm{d}\,t}\partial_x$ we can write

$$\sqrt{V}\dot{P}_n(t) = \partial_t\left[p\left(V^{-1/2}n - V^{1/2}\rho, t\right)\right] = \partial_t p(x,t) - \frac{\dot{\rho}}{\sqrt{V}}\partial_x p(x,t).$$

Substituting in the master equation the expressions we have obtained, we have

$$\partial_t p(x,t) - \frac{\dot{\rho}}{\sqrt{V}}\partial_x p(x,t) = \left(\frac{1}{\sqrt{V}}\partial_x + \frac{1}{2V}\partial_x^2\right)(V\rho + \sqrt{V}x)p(x,t).$$

At order $\sqrt{V}$ we have the equation for the average

$$\dot{\rho}\partial_x p = -\rho\partial_x p \Rightarrow \dot{\rho} = -\rho \Rightarrow \rho(t) = \rho(0)\,\mathrm{e}^{-t}.$$

This result makes sense with what we expect: the process is a pure death process and therefore the density decays exponentially. At order 1, we obtain an equation for the fluctuations part

$$\partial_t p = \partial_x(xp) + \frac{\rho}{2}\partial_x^2 p$$

which is a Fokker–Planck–like equation. Observe that this approach suffers from the same limitations as the Kramers–Moyal approach, although it makes the evolution of the system clearer.

4. Spectral analysis of the Fokker–Planck equation [⋆]

We can write the Fokker–Planck equation (4.7) as

$$\frac{\partial p(x,t)}{\partial t} = \mathcal{W}p(x,t) \tag{4.43}$$

with the differential *linear* operator

$$\mathcal{W} = -\partial_x\left[a^{(1)}(x)\bullet\right] + \frac{1}{2}\partial_x^2\left[a^{(2)}(x)\bullet\right].$$

The formal solution of (4.43) is

$$p(x,t) = \mathrm{e}^{t\mathcal{W}(x)}\,p(x,0), \tag{4.44}$$

where $\mathcal{Q}(x,t) \coloneqq \mathrm{e}^{t\mathcal{W}(x)}$ is the so-called *forward propagator*. This is in complete analogy with what what we discussed for the case of a master equation involving a matrix $\boldsymbol{W}$. As the propagator $\boldsymbol{Q}(t)$ satisfied the equation $\frac{\mathrm{d}\boldsymbol{Q}(t)}{\mathrm{d}t} = \boldsymbol{W}\boldsymbol{Q}(t)$ in the case of continuous-time processes in discrete space, here $\mathcal{Q}(x,t)$ satisfies the equation

$$\partial_t \mathcal{Q}(x,t) = \mathcal{W}(x)\mathcal{Q}(x,t). \tag{4.45}$$

The operator $\mathcal{W}$ acts on the space $L^2(D)$ of functions $f\colon D \to \mathbb{C}$ with $D \subseteq \mathbb{R}$, such that the integral of their modulus squared is finite,

$$L^2(D) \coloneqq \left\{ f\colon D \to \mathbb{C} \,\middle|\, \int |f(x)|^2 \,\mathrm{d}\,x < +\infty \right\}, \tag{4.46}$$

so that, if $f \in L^2$,

$$\mathcal{W}(x)f(x) \coloneqq -\partial_x \left[a^{(1)}(x)f(x)\right] + \frac{1}{2}\partial_x^2 \left[a^{(2)}(x)f(x)\right]. \tag{4.47}$$

The space L^2 can be equipped with a *inner product* that we denote using the same notation adopted for vectors

$$\langle f|g\rangle \coloneqq \int f^\star(x) r(x) \,\mathrm{d}\,x, \qquad f,g \in L^2(\mathbb{R}), \tag{4.48}$$

where $f^\star$ is the complex conjugate of f. By means of this inner product we can define the *adjoint operator* $\mathcal{W}^\dagger$ to $\mathcal{W}$ as the operator such that, for two functions $f,g \in L^2$, has

$$\langle \mathcal{W}^\dagger f|g\rangle = \langle f|\mathcal{W}g\rangle. \tag{4.49}$$

The operator $\mathcal{W}^\dagger$ can be find explicitely:

$$\begin{aligned} \langle f|\mathcal{W}g\rangle &= -\int f^\star(x)\partial_x \left[a^{(1)}(x)f(x)\right] \mathrm{d}\,x + \frac{1}{2}\int f^\star(x)\partial_x^2 \left[a^{(2)}(x)f(x)\right] \mathrm{d}\,x \\ &= \int r(x) \left[a^{(1)}(x)\partial_x + \frac{a^{(2)}(x)}{2}\partial_x^2\right] f^\star(x) \,\mathrm{d}\,x \equiv \langle \mathcal{W}^\dagger f|g\rangle, \end{aligned} \tag{4.50}$$

so that

$$\mathcal{W}^\dagger = a^{(1)}(x)\partial_x + \frac{a^{(2)}(x)}{2}\partial_x^2. \tag{4.51}$$

In a sense, $\mathcal{W}^\dagger$ expresses the action of the operator $\mathcal{W}$ if "multiplied on the right". We say that $\mathcal{W}$ is *Hermitian* if $\mathcal{W} = \mathcal{W}^\dagger$. If this is the case, then $\mathcal{W}$ has a complete set of of eigenfunctions. This means that there exists a set of functions $\phi^a(x)$ and $\psi^a(x)$ such that

$$\mathcal{W}(x)\psi^a(x) = \mu_a \psi^a(x) \qquad \mathcal{W}^\dagger(x)\phi^a(x) = \mu_a^\star \phi^a(x)$$

with a set of *real* eigenvalues μ_a, so that $\psi^a = \phi^a$,

$$\langle \psi^a|\psi^b\rangle = \delta_{ab}, \qquad \sum_a \psi^{a\star}(x)\psi^a(x) = \delta(x),$$

where $\delta(x)$ is the Dirac delta.

Suppose now that $\mathcal{W}$ has indeed a complete set of eigenfuntions. We can perform an eigenfunction decomposition of the Fokker–Planck equation, similar to what we did for the master equation, expanding $p(x,0) \equiv p_0(x)$ in the basis of eigenfunctions

$$p(x,0) = \sum_a \langle \phi^a|p_0\rangle \psi_a(x)$$

and substituting in (4.44) we have

$$p(x,t) = \sum_a \langle \phi^a|p_0\rangle \,\mathrm{e}^{\mu_a t}\, \psi^a(x). \tag{4.52}$$

However, $\mathcal{W}$ is generally non-Hermitian, so the existence of a complete set of orthonormal eigenfunctions is not ensured and eigenvalues are not need to be real. If $\mathcal{W}$ is Hermitian, then $\mu_a \in \mathbb{R}$: as in the case of master equations, the convergence to the stady-state solution depends on the spectral properties of the operator $\mathcal{W}$.

4.1. The case $a^{(2)}(x) \equiv 2D$. We will prove now the convergence to the steady state in the special case in which $a^{(2)} \equiv 2D$, i.e.,

$$\mathcal{W}(x) = \partial_x \left[V'(x)\bullet\right] + D\partial_x^2. \tag{4.53}$$

We have also written $a^{(1)}(x) = -\partial_x V(x)$, which is always possible in one dimension. We know that in this case the steady-state solution (in the case of zero probability flow on the boundaries) is the one in Eq. (4.12) In order to prove convergence to this steady-state let us first write the density $p(x,t)$ as

$$p(x,t) = q(x,t)\, \mathrm{e}^{-\frac{1}{D}V(x)}$$

Inserting the form above in $\partial_t p = \mathcal{W}p$, we get, after little algebra, an equation for q,

$$\partial_t q = \frac{1}{2} q \partial_x^2 V - \frac{(\partial_x V)^2}{4D} q + D\partial_x^2 q \equiv \hat{\mathcal{W}} q$$

with

$$\hat{\mathcal{W}} \coloneqq \frac{1}{2}\partial_x^2 V - \frac{(\partial_x V)^2}{4D} + D\partial_x^2.$$

Exactly as we did for the general Fokker–Planck equation, the equation for q has a formal solution given by

$$q(x,t) = \mathrm{e}^{t\hat{\mathcal{W}}}\, q(x,0). \tag{4.54}$$

It is easy to verify that $\hat{\mathcal{W}}$ is Hermitian, and therefore it will have a complete set of orthonormal eigenfunctions $\hat{\psi}^a$, such that $\hat{\mathcal{W}}\hat{\psi}^a = \hat{\mu}_a \hat{\psi}^a$, and $\hat{\mu}_a \in \mathbb{R}$. The crucial observation now is that the operator $\hat{\mathcal{W}}$ and the operator $\mathcal{W}$ have the same set of eigenvalues. Infact, if $\hat{\psi}_a(x)$ is an eigenfunction of $\hat{\mathcal{W}}$ with eigenvalue $\hat{\mu}_a$, then $\psi_a(x) = \hat{\psi}_a(x)\, \mathrm{e}^{-\frac{1}{D}V(x)}$ is an eigenfunction of $\mathcal{W}$ with the same eigenvalue. This can be shown easily using the fact that

$$\mathcal{W}\left[f(x)\, \mathrm{e}^{-\frac{1}{D}V(x)}\right] = \mathrm{e}^{-\frac{1}{D}V(x)}\, \hat{\mathcal{W}} f(x). \tag{4.55}$$

We can show something more: the operator $\hat{\mathcal{W}}$ is such that

$$\hat{\mathcal{W}} = -\mathcal{A}^\dagger \mathcal{A} \tag{4.56}$$

where $\mathcal{A}$ is the differential operator

$$\mathcal{A} \coloneqq \sqrt{D}\partial_x + \frac{1}{2\sqrt{D}}\partial_x V.$$

This implies that $\hat{\mu}_a \leq 0$, as, for any eigenfunction $\hat{\psi}^a$,

$$\hat{\mu}_a = \langle \hat{\psi}^a | \hat{\mathcal{W}} \hat{\psi}^a \rangle = -\langle \hat{\psi}^a | \mathcal{A}^\dagger \mathcal{A} \hat{\psi}^a \rangle = -\langle \mathcal{A}\hat{\psi}^a | \mathcal{A}\hat{\psi}^a \rangle \leq 0. \tag{4.57}$$

Since $\mathcal{W}$ and $\hat{\mathcal{W}}$ must have the same eigenvalues, this means that the eigenvalues of $\mathcal{W}$ are also non-positive, and we know that $\mathcal{W}$ has an eigenvalue $\hat{\mu}_0 = 0$ (with associated eigenfunction $\pi(x)$). By consequence, assuming that there is a *unique* eigenvalue equal to zero, by Eq. (4.52) the only surviving term is indeed the one corresponding to $\mu_0 = 0$, i.e., $\lim_{t\to 0} p(x,t) = \pi(x)$.

✎ Verify (4.56).

EXAMPLE Let us apply this very general framework to the simplest case, i.e., the heat equation

$$\partial_t p(x,t) = D\partial_x^2 p(x,t),$$

with reflecting and absorbing boundary conditions in 0 and L, so that the domain is $D = [0, L]$. The operator $\mathcal{W}$ is very simple,

$$\mathcal{W} = D\partial_x^2 \tag{4.58}$$

so that formally $p(x,t) = \mathrm{e}^{t\partial_x^2} p(x,0)$. The operator $\mathcal{W}$ is also Hermitian, $\mathcal{W}^\dagger = \mathcal{W}$, so that it will have a complete basis of eigenfuctions with real eigenvalues.

For absorbing conditions, we have the constraint $p(0,t) = p(L,t) = 0$ to be satisfied $\forall\ t$. The eigenfunctions are found by solving

$$\mathcal{W}\psi^a(x) = D\partial_x^2\psi^a(x) = \mu_a\psi^a(x),$$

but we have to take into account the constraint $\psi^a(0) = \psi^a(L) = 0$. This equation is solved by

$$\psi_a(x) = c_a \sin(\omega_a x) + c'_a \cos(\omega_a x), \qquad \mu_a = -\frac{\omega_a^2}{D}. \tag{4.59}$$

but we have indeed to impose $\psi^a(0) = \psi^a(L) = 0$, that implies $c'_a = 0$ and $\omega_a = \frac{\pi a}{L}$ for $a \in \mathbb{N}$. Also, to have $\langle\psi^a|\psi^a\rangle = 1$ we have $c_a \equiv \frac{1}{\sqrt{L}}$. In conclusion

$$\psi_a(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{\pi a x}{L}\right), \qquad \mu_a = -\frac{\pi^2 a^2}{DL}, \quad a \in \mathbb{N}_0. \tag{4.60}$$

This means that

$$p(x,t) = \sqrt{\frac{2}{L}} \sum_{a=0}^{\infty} \sin\left(\frac{\pi a x}{L}\right) \langle\psi^a|p_0\rangle \mathrm{e}^{-\frac{\pi^2 a^2 t}{DL}}, \quad \langle\psi^a|p_0\rangle = \sqrt{\frac{2}{L}} \int p(x,0) \sin\left(\frac{\pi a x}{L}\right) \mathrm{d}x. \tag{4.61}$$

Note that $p(x,t)$ vanishes in the limit $t \to \infty$, due to the leaking through the boundaries.

The derivation for reflecting conditions is very similar. In this case, the boundary conditions are $J(0) = J(L) = 0$, so that we need $\partial_x p(x,t)|_0 = \partial_x p(x,t)|_{x=L} = 0$, leading to

$$\psi^a(x) = \begin{cases} \frac{1}{\sqrt{L}}, \text{ with eigenvalue } \mu_0 = 0 & a = 0, \\ \sqrt{\frac{2}{L}} \cos\left(\frac{a\pi x}{L}\right), \text{ with eigenvalue } \mu_a = -\frac{\pi^2 a^2}{DL} & a \in \mathbb{N}. \end{cases}$$

This time

$$p(x,t) = \frac{1}{\sqrt{L}} + \sqrt{\frac{2}{L}} \sum_{a=1}^{\infty} \cos\left(\frac{\pi a x}{L}\right) \langle\psi^a|p_0\rangle \mathrm{e}^{-\frac{\pi^2 a^2 t}{DL}}, \quad \langle\psi^a|p_0\rangle = \sqrt{\frac{2}{L}} \int p(x,0) \cos\left(\frac{\pi a x}{L}\right) \mathrm{d}x. \tag{4.62}$$

For $t \to \infty$ this converges to $\lim_{t\to+\infty} p(x,t) = 1/L$ which is indeed the stationary solution found from $\partial_x^2\pi(x) = 0$ imposing the boundary condition $\partial_x\pi(x)|_{x=0} = \partial_x\pi(x)|_{x=L} = 0$ and normalization $\int_0^L \pi(x)\,\mathrm{d}\,x = 1$. This shows that on a finite interval, diffusion converges to the uniform distribution.

CHAPTER 5

The Langevin equation

1. The Brownian motion revisited: the Langevin equation

A Brownian particle immersed in a fluid is subject to two forces, characterising the action of the fluid on it: a *viscous force*, with friction coefficient $\gamma > 0$, representing the effect of the resistance of the fluid to the motion of the particle; a random force $\eta(t)$ which results from the impact of the fluid's molecules on the Brownian particle . Newton's equation of motion for the position X_t for the particle of mass m is therefore

$$m\frac{\mathrm{d}^2\mathsf{X}}{\mathrm{d}\,t^2} = -\gamma\frac{\mathrm{d}\,\mathsf{X}}{\mathrm{d}\,t} + \eta,$$

or, upon introducing $\mathsf{V} \coloneqq \frac{\mathrm{d}\,\mathsf{X}}{\mathrm{d}\,t}$,

$$m\frac{\mathrm{d}\,\mathsf{V}}{\mathrm{d}\,t} = -\gamma\mathsf{V} + \eta. \tag{5.1}$$

From now on we will set to $m = 1$ for simplicity. Historically, the above equation constitutes the first example of a *Langevin equation* or *stochastic differential equation*, i.e., a differential equation involving stochastic processes. Solving a Langevin equation thus means determining the statistical properties of the process V_t.

We will assume that the fluid (or thermal bath) is in a stationary state, so that

(1) $\mathbb{E}[\eta(t)]$ does not depend on time;
(2) $\mathbb{E}[\eta(t)\eta(t')]$ depends only on the time difference $t - t'$.

Since in the absence of external forces, $\lim_{t\to\infty}\mathbb{E}[\mathsf{V}_t] = 0$ at stationarity, one must have

$$\mathbb{E}[\eta(t)] = 0. \tag{5.2}$$

Indeed, this ensures that, assuming that $\mathsf{V}_0 = v_0$,

$$\partial_t\mathbb{E}[\mathsf{V}] = -\gamma\mathbb{E}[\mathsf{V}] \Longrightarrow \mathbb{E}[\mathsf{V}_t] = v_0\,\mathrm{e}^{-\gamma t} \tag{5.3}$$

and

$$\lim_{t\to\infty}\mathbb{E}[\mathsf{V}_t] = 0. \tag{5.4}$$

Denoting the noise autocorrelation function

$$\langle\!\langle\eta(t+\tau)\eta(t)\rangle\!\rangle = \mathbb{E}[\eta(t+\tau)\eta(t)] = C(\tau), \tag{5.5}$$

in equilibrium, due to time-reversal symmetry, $C(\tau)$ must be an even function of its argument, $C(\tau) = C(-\tau)$, and is expected to decay for $|\tau| \gg \tau_0$, for some τ_0, referred to as the correlation time, which is of the order of the mean time interval between successive collisions of the Brownian particle with the fluid molecules. For reasons that we will appear clear below, we define

$$2D \coloneqq \int_{-\infty}^{\infty} C(\tau)\,\mathrm{d}\,\tau.$$

If τ_0 is much shorter than other characteristic times (e.g., $\tau_0 \ll \gamma^{-1}$), C will have a narrow peak around $\tau = 0$, and can be approximated with

$$\mathbb{E}[\mathsf{V}_t \mathsf{V}_{t'}] = 2D\delta(t - t') \tag{5.6}$$

1.1. Solution for given initial condition. As an ODE, the Langevin equation for the Brownian motion is a first order linear differential equation, which is solved by integrating factors. For the initial condition $\mathsf{V}_0 = v_0$, we have

$$\mathsf{V}_t = v_0 \, \mathrm{e}^{-\gamma t} + \int_0^t \eta(t') \, \mathrm{e}^{-\gamma(t-t')} \, \mathrm{d}\, t'. \tag{5.7}$$

The statistical properties of V_t follow from those of the noise $\eta(t)$, which is the only stochastic piece in the solution. As we said, the average is $\mathbb{E}[\mathsf{V}_t] = v_0 \, \mathrm{e}^{-\gamma t}$. Using the above equations, the velocity correlation function is then

$$\begin{aligned} C_v(t, t') \coloneqq \langle\!\langle \mathsf{V}_t \mathsf{V}_{t'} \rangle\!\rangle &= \int_0^t \mathrm{d}\, t_1 \int_0^{t'} \mathrm{d}\, t_2 \, \mathrm{e}^{-\gamma(t - t_1 + t' - t_2)} \overbrace{\mathbb{E}[\eta(t_1)\eta(t_2)]}^{2D\delta(t_1 - t_2)} \\ &= 2D \int_0^{\min(t,t')} \mathrm{e}^{-\gamma(t + t' - 2t_1)} \, \mathrm{d}\, t_1 \\ &= \frac{D}{\gamma} \left(\mathrm{e}^{-\gamma|t - t'|} - \mathrm{e}^{-\gamma(t + t')} \right). \end{aligned}$$

The upper bound of the integral in the second line has been obtained by noting that the double integral in t_1 and t_2 is on a rectangular domain, $[0, t] \times [0, t']$ but only takes contributions along the diagonal $t_1 = t_2$, thus up to values of t_1 equal to whichever is smaller between t and t'. For large times, $t, t' \gg \gamma^{-1}$ we have

$$C_v(t, t') \to \frac{D}{\gamma} \, \mathrm{e}^{-\gamma|t - t'|} \equiv C_v(|t - t'|)$$

which is indeed a function of the time difference only, as expected at stationarity. From the above, we can calculate the average kinetic energy of the Brownian particle at equilibrium

$$\mathbb{E}[E(\mathsf{V})] \coloneqq \lim_{t \to \infty} \frac{\mathbb{E}[\mathsf{V}^2(t)]}{2} = \frac{D}{2\gamma}.$$

Assimilating the Brownian particle to a molecule of an ideal gas, Einstein used the equipartition theorem (see box below) to derive the relation

$$\frac{1}{2} k_{\mathrm{B}} T = \frac{1}{2} \mathbb{E}[\mathsf{V}^2] = \frac{D}{2\gamma} \Longrightarrow \frac{D}{\gamma} = k_{\mathrm{B}} T. \tag{5.8}$$

Very remarkably, the relation

$$D = k_{\mathrm{B}} T \gamma = \frac{RT}{N_A} \gamma$$

links a quantity which is experimentally accessible on the macroscopic scale, D, to the microscopic quantity, the Avogadro number N_{A}, by means of the ideal gas constant[1].

Q At the beginning of the XX century, it was unclear whether atoms where real entities or a useful tool to solve problems. Einstein's relation was regarded as a proof of existence of atoms.

[1] Remember that $N_{\mathrm{A}} k_{\mathrm{B}} = R$.

As we have noted studying the stationary solutions of the Fokker–Planck equation, Eq. (5.8) is a statement of the Fluctuation-Dissipation Theorem, showing that the size of fluctuations D and the damping γ (which have two opposite effects on the Brownian particle's velocity) must balance in equilibrium. The quantity D can be accessed experimentally as it builds up an observable mean square displacement

$$\begin{aligned}\mathbb{E}[(\mathsf{X}_t-\mathsf{X}_0)^2] &= \mathbb{E}\left[\left(\int_0^t \mathsf{V}_{t_1}\,\mathrm{d}\,t_1\right)^2\right] = \int_0^t \mathrm{d}\,t_1 \int_0^t \mathrm{d}\,t_2 \mathbb{E}[\mathsf{V}_{t_1}\mathsf{V}_{t_2}] \\ &= \int_0^t \mathrm{d}\,t_1 \int_0^t \mathrm{d}\,t_2 \left[v_0^2\,\mathrm{e}^{-\gamma(t_1+t_2)} + \frac{D}{\gamma}\left(\mathrm{e}^{-\gamma|t_1-t_2|} - \mathrm{e}^{-\gamma(t_1+t_2)}\right)\right] \\ &= \left(v_0^2 - \frac{D}{\gamma}\right)\frac{(1-\mathrm{e}^{-\gamma t})^2}{\gamma^2} + \frac{2D}{\gamma^2}\left(t - \frac{1-\mathrm{e}^{-\gamma t}}{\gamma}\right).\end{aligned} \tag{5.9}$$

For $t \gg 1/\gamma$,

$$\mathbb{E}[(\mathsf{X}_t-\mathsf{X}_0)^2] \simeq \frac{2D}{\gamma^2}t.$$

Note that in the opposite limit $t \ll 1/\gamma$

$$\mathbb{E}[(\mathsf{X}_t-\mathsf{X}_0)^2] \simeq v_0^2 t^2$$

so on short timescales the particle describes a ballistic motion.

THE EQUIPARTITION THEOREM Suppose that we have a free particle of mass m, travelling inside a cube of side L in 3 dimensions. The variation of momentum along the x coordinate when hitting elastically a wall is

$$\Delta p_x = 2mv_x$$

where v_x is the x-component of the particle's velocity and p_x its momentum in the same direction. The x-component of the force applied by the particle on the wall is thus

$$F_x = \frac{\Delta p_x}{\Delta t} \quad \text{with} \quad \Delta t = \frac{2L}{v_x}.$$

The contribution to the pressure along the x direction is

$$P_x = \frac{F_x}{L^2} = \frac{mv_x^2}{L^3} = \frac{mv_x^2}{V}$$

where V is the volume of the cube. Thus, the total pressure exerted by N particles in a cube, considering their action is distributed on the three coordinates is

$$P = N\frac{m\mathbb{E}[\mathsf{V}^2]}{3}$$

where we have taken the average of $\mathsf{V}^2 = \mathsf{V}_x^2 + \mathsf{V}_y^2 + \mathsf{V}_z^2$. On the other hand, from the ideal gas law one has

$$PV = Nk_{\mathrm{B}}T$$

where as usual k_{B} is the Boltzmann constant. Thus, in combination, we have

$$\frac{1}{2}m\mathbb{E}[\mathsf{V}^2] = \frac{3}{2}k_{\mathrm{B}}T$$

for three degrees of freedom. This is refereed to as the *Equipartition theorem*. It states that the average thermal energy carried by each degree of freedom in equilibrium systems at temperature T is $\frac{1}{2}k_{\mathrm{B}}T$.

2. The noise term

2.1. Noise distribution. We have so far assumed

$$\mathbb{E}[\eta(t)] = 0, \qquad \mathbb{E}[\eta(t)\eta(t')] = 2D\delta(t-t').$$

One can characterize further the distribution of η by making additional assumptions. Noting that η is the result of the effects of a great number of independent processes, it is plausible to assume that it some sort of Gaussian character. However, $\eta(t)$ is a function of time. How to define a "Gaussian distribution" for a *function*? We can start discretizing the time span t of a trajectory in little intervals of size $\tau = {}^{t}/_{n}$, $t_i = i\tau$, $i = 0, \dots, n$, and think of $\eta(t)$ as sampled at these times. Letting $\eta_i = \eta(t_i)$ we can thus approximate the path $\eta(t)$ with a collection of variables $\boldsymbol{\eta} \coloneqq (\eta_1, \eta_2, \dots, \eta_n)$. We assume the elements of the vector being such that $\mathbb{E}[\eta_i\eta_j] = \frac{2D}{\tau}\delta_{ij}$, discretised version of the condition $\mathbb{E}[\eta(t)\eta(t')] = 2D\delta(t-t')$. The multivariate Gaussian distribution for n variables with zero average is

$$\begin{aligned} p(\boldsymbol{\eta}) &= \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{C}}} \exp\left(-\frac{1}{2}\sum_{ij}\eta_i[\boldsymbol{C}^{-1}]_{ij}\eta_j\right) \\ &\equiv \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{C}}} \exp\left(-\frac{1}{2}\langle\boldsymbol{\eta}|\boldsymbol{C}^{-1}|\boldsymbol{\eta}\rangle\right), \end{aligned} \tag{5.10}$$

where $\boldsymbol{C}$ is the covariance matrix, with elements $C_{ij} = \mathbb{E}[\eta_i\eta_j]$. To fix the value of these elements, we note that, for a given function $f(t)$, denoting $f_i = f(i\tau)$, we want

$$\tau\sum_{i=0}^{n} f_i\mathbb{E}[\eta_i\eta_k] \xrightarrow{t'=k\tau\in(0,t)} \int_0^t f(s)\mathbb{E}[\eta(s)\eta(t')]\,\mathrm{d}\,s \equiv 2Df(t') \tag{5.11}$$

i.e., we want that $\tau\sum_{i=0}^{n} f_i\mathbb{E}[\eta_i\eta_k] = 2Df_k$. This is realised taking

$$\mathbb{E}[\eta_i\eta_j] = C_{ij} = \frac{2D}{\tau}\delta_{ij} \Rightarrow [\boldsymbol{C}^{-1}]_{ij} = \frac{\tau}{2D}\delta_{ij}.$$

Substituting into the general multivariate expression

$$p(\boldsymbol{\eta}) = \left(\frac{\tau}{4\pi D}\right)^{\frac{n}{2}} \exp\left(-\frac{\tau}{4D}\sum_{i=1}^{n}\eta_i^2\right) \equiv \left(\frac{\tau}{4\pi D}\right)^{\frac{n}{2}} \mathrm{e}^{-\frac{\tau\|\boldsymbol{\eta}\|^2}{4D}}. \tag{5.12}$$

Taking the continuous time limit $\tau \to 0$ and using Riemann approximation of an integral, we can write

$$p(\boldsymbol{\eta})\prod_{i=1}^{n}\mathrm{d}\,\eta_i \propto \exp\left(-\frac{1}{4D}\int_0^t \eta^2(t')\,\mathrm{d}\,t'\right)\mathcal{D}\eta(t), \qquad \mathcal{D}\eta(t) \coloneqq \lim_{\tau\to 0}\prod_{i=1}^{{}^{t}/_{\tau}}\frac{\mathrm{d}\,\eta_i}{\sqrt{4\pi^{D}/_{\tau}}}. \tag{5.13}$$

Noise averages are then defined via the functional integration or *path integral*

$$\mathbb{E}[A[\eta(t)]] \coloneqq \int A[\eta(t)]\varrho[\eta(t)]\mathcal{D}\eta(t), \quad \varrho[\eta(t)] \propto \exp\left(-\frac{1}{4D}\int_0^t \eta^2(t')\,\mathrm{d}\,t'\right), \tag{5.14}$$

where the integration is over a trajectory (or path) rather than over a variable. For all practical purposes, this can be thought of as the limit of a multivariate integral, where the number of integration variables is taken to infinity.

2.2. Moments. The Gaussian assumption on the noise distribution allows us to simplify the calculation of higher order moments, by virtue of *Wick's theorem*, which applies to zero-mean Gaussian variables $\boldsymbol{\eta}$ with covariance $\boldsymbol{C}$.

🔍 To derive the content of the theorem, let us start calculating the characteristic function of a $p(\boldsymbol{\eta})$ assuming, for a moment, a generic covariance $\boldsymbol{C}$. For a general distribution $p(\boldsymbol{\eta})$ we have

$$\begin{aligned} G(\boldsymbol{k}) &= \int p(\boldsymbol{\eta})\, \mathrm{e}^{i\langle \boldsymbol{k}|\eta\rangle}\, \mathrm{d}\,\boldsymbol{\eta} = \mathbb{E}[\mathrm{e}^{i\langle \boldsymbol{k}|\boldsymbol{\eta}\rangle}] \\ &= \sum_{m_1\dots m_n} \frac{(ik_1)^{m_1}\dots(ik_n)^{m_n}}{m_1!\dots m_n!} \mathbb{E}[\eta_1^{m_1}\dots\eta_n^{m_n}] \end{aligned} \tag{5.15}$$

so that each correlation $\mathbb{E}\left[\prod_i \eta_i^{m_i}\right]$ appears multiplied by a $\prod_i k_i^{m_i}$. On the other hand, for zero-mean Gaussian variables

$$\begin{aligned} G(\boldsymbol{k}) &= \int \frac{\mathrm{d}\,\boldsymbol{\eta}}{\sqrt{(2\pi)^n \det \boldsymbol{C}}}\, \mathrm{e}^{-\frac{1}{2}\langle \boldsymbol{\eta}|\boldsymbol{C}^{-1}|\boldsymbol{\eta}\rangle + i\langle \boldsymbol{k}|\boldsymbol{\eta}\rangle} = \mathrm{e}^{-\frac{1}{2}\langle \boldsymbol{k}|\boldsymbol{C}|\boldsymbol{k}\rangle} \\ &= \prod_{pq} \mathrm{e}^{\frac{1}{2}(ik_p)(ik_q)\mathbb{E}[\eta_p\eta_q]} = \prod_{pq}\left[1 + \frac{1}{2}(ik_p)(ik_q)\mathbb{E}[\eta_p\eta_q] + \dots\right] \end{aligned} \tag{5.16}$$

Note that only terms with an even number of ks factors show up in (5.16), and only two-terms correlations. Terms like $\mathbb{E}[\eta_a\eta_b\eta_c\dots]$ in (5.15) corresponds to the combination of products of all suitable *pairs* in (5.16). For example,

$$\mathbb{E}[\eta_a\eta_b\eta_c\eta_d] = \mathbb{E}[\eta_a\eta_b]\mathbb{E}[\eta_c\eta_d] + \mathbb{E}[\eta_a\eta_c]\mathbb{E}[\eta_b\eta_d] + \mathbb{E}[\eta_a\eta_d]\mathbb{E}[\eta_b\eta_c].$$

To summarise, for any set of k indices $\{i_j\}_{j=1}^k$,

$$\mathbb{E}[\eta_{i_1}\eta_{i_2}\dots\eta_{i_{2k-1}}] = \begin{cases} 0 & \text{if } k = 2n+1 \text{ for some } n \\ \sum_{\mathcal{P}} \prod_{j=1}^n \mathbb{E}[\eta_{i_{\mathcal{P}(2j-1)}}\eta_{i_{\mathcal{P}(2j)}}] & \text{if } k = 2n \text{ for some } n \end{cases}$$

where the sum runs over all $\frac{(2n)!}{2^n n!}$ distinct pairings of $2n$ elements[a]. For non-zero mean Gaussian variables, one can apply Wick's theorem to the shifted variable $\delta\eta = \eta - \mathbb{E}[\eta]$.

[a] There are $(2n)!$ permutations of the indices, but this has to be divided by the $n!$ ordering of the pairs, which is irrelevant; plus, the ordering of the elements inside each one of the n pairs is not relevant, and this is the reason of the 2^n in the denominator.

Wick's theorem allows us to compute higher order moments of the stochastic process V_t. Let us introduce

$$\mathsf{u}(t) \coloneqq \mathsf{V}_t - \mathbb{E}[\mathsf{V}_t] = \int_0^t \eta(\tau)\, \mathrm{e}^{-\gamma(t-\tau)}\, \mathrm{d}\,\tau.$$

We have that, for any $n \in \mathbb{N}$,

$$\mathbb{E}\left[\mathsf{u}^{2n+1}(t)\right] = \left[\prod_{i=1}^{2n+1} \int_0^t \mathrm{d}\,t_i\right] \mathbb{E}\left[\prod_{i=1}^{2n+1} \eta(t_i)\right] \mathrm{e}^{-\gamma(2n+1)t - \gamma\sum_i t_i} \equiv 0, \tag{5.17}$$

and similarly

$$\begin{aligned}\mathbb{E}\left[\mathsf{u}^{2n}(t)\right] &= \left[\prod_{i=1}^{2n}\int_0^t \mathrm{d}\,t_i\right]\mathbb{E}\left[\prod_{i=1}^{2n}\eta(t_i)\right]\mathrm{e}^{-\gamma(2n+1)t-\gamma\sum_i t_i}\\ &= \frac{(2n)!}{2^n n!}\left(\iint_0^t \mathrm{d}\,t_1\,\mathrm{d}\,t_2\;\mathrm{e}^{-\gamma(2t-t_1-t_2)}\overbrace{\mathbb{E}[\eta(t_1)\eta(t_2)]}^{2D\delta(t_1-t_2)}\right)^n\\ &= \frac{(2n)!}{2^n n!}\left(2D\int_0^t \mathrm{d}\,t_1\,\mathrm{e}^{-2\gamma(t-t_1)}\right)^n = \frac{(2n)!}{2^n n!}\left(D\frac{1-\mathrm{e}^{-2\gamma t}}{\gamma}\right)^n.\end{aligned} \tag{5.18}$$

From the moments, we can reconstruct the full distribution via the moment generating function. According to our formulas above then

$$\mathbb{E}[\mathsf{u}^k(t)] = \begin{cases}0 & \text{if } k=2n+1\\ \frac{(2n)!}{2^n n!}\left(D\frac{1-\mathrm{e}^{-2\gamma t}}{\gamma}\right)^n & \text{if } k=2n+1.\end{cases} \tag{5.19}$$

Then, the generating function of $\mathsf{u}(t)$ is

$$\begin{aligned}G_{\mathsf{u}}(k,t) &= \int p(u,t)\,\mathrm{e}^{iku}\,\mathrm{d}\,u = \sum_{n=0}^{\infty}(ik)^n\frac{\mathbb{E}[\mathsf{u}^n(t)]}{n!} = \sum_{n=0}^{\infty}(ik)^{2n}\frac{\mathbb{E}[\mathsf{u}^{2n}(t)]}{(2n)!}\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\left(-k^2D\frac{1-\mathrm{e}^{-2\gamma t}}{2\gamma}\right)^n = \exp\left(-k^2D\frac{1-\mathrm{e}^{-2\gamma t}}{2\gamma}\right);\end{aligned} \tag{5.20}$$

by taking the inverse Fourier transform

$$p(u,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G_{\mathsf{u}}(k,t)\,\mathrm{e}^{-iku}\,\mathrm{d}\,k = \sqrt{\frac{\gamma}{2\pi D(1-\mathrm{e}^{-2\gamma t})}}\exp\left(-\frac{\gamma u^2}{2D(1-\mathrm{e}^{-2\gamma t})}\right)$$

so that, replacing $u \to v - \mathbb{E}[\mathsf{V}_t] = v - v_0\,\mathrm{e}^{-\gamma t}$ we obtain

$$p(v,t) = \sqrt{\frac{\gamma}{2\pi D(1-\mathrm{e}^{-2\gamma t})}}\exp\left(-\frac{\gamma(v-v_0\,\mathrm{e}^{-\gamma t})^2}{2D(1-\mathrm{e}^{-2\gamma t})}\right) \equiv p_{1|1}(v,t|v_0,0).$$

Unsurprisingly, this coincides with the solution to the Fokker–Planck equation for the Ornstein–Uhlenbeck process, which describes the evolution of the velocity of a Brownian particle (i.e. the same process that is modelled by the Langevin equation (5.1)). For $t \gg \gamma^{-1}$ this retrieves the Maxwell distribution, upon identification of $D/\gamma = k_B T$.

2.3. Noise averages. So far, all the noise averages we have calculated followed from the assumptions $\mathbb{E}[\eta(t)] = 0$ and $\mathbb{E}[\eta(t)\eta(t')] = 2D\delta(t-t')$. The Gaussian assumption on the (functional) distribution, is useful to compute more complicated objects, for example $\mathbb{E}[\eta(t')\mathsf{V}_t]$. This can be calculated by carrying out integration by parts:

$$\begin{aligned}\mathbb{E}[\mathsf{V}_t\eta(t')] &= \int \eta(t')v(t)\,\mathrm{e}^{-\frac{1}{4D}\int_0^t\eta^2(\tau)\,\mathrm{d}\,\tau}\,\mathcal{D}\eta(t)\\ &= -2D\int v(t)\frac{\delta\,\mathrm{e}^{-\frac{1}{4D}\int_0^t\eta^2(\tau)\,\mathrm{d}\,\tau}}{\delta\eta(t')}\mathcal{D}\eta(t)\\ &= 2D\int\frac{\delta v(t)}{\delta\eta(t')}\,\mathrm{e}^{-\frac{1}{4D}\int_0^t\eta^2(\tau)\,\mathrm{d}\,\tau}\,\mathcal{D}\eta(t) = 2D\mathbb{E}\left[\frac{\delta\mathsf{V}_t}{\delta\eta(t')}\right].\end{aligned} \tag{5.21}$$

Let us consider now a particle moving in presence of an additional perturbation $h(t)$, so that the equation of motion becomes

$$\frac{\mathrm{d}\,\mathsf{V}}{\mathrm{d}\,t} = -\gamma\mathsf{V} + \eta + h.$$

In this equation, η and h play the same role, hence taking functional derivatives of V_t with respect to $\eta(t')$ or $h(t')$ is equivalent

$$\frac{\delta \mathsf{V}_t}{\delta \eta(t')} = \frac{\delta \mathsf{V}_t}{\delta h(t')}.$$

To show this explicitly, let us assume as usual the initial condition $v(0) = v_0$. Then

$$\mathsf{V}_t = v_0\,\mathrm{e}^{-\gamma t} + \int_0^t (\eta(\tau) + h(\tau))\,\mathrm{e}^{-\gamma(t-\tau)}\,\mathrm{d}\,\tau$$

and therefore we can introduce the *response function* $R(t,t')$, measuring how V_t typically changes in response to an external stimulus $h(t')$,

$$R(t,t') \coloneqq \frac{\delta\langle \mathsf{V}_t\rangle}{\delta h(t')} = \int_0^t \delta(t''-t')\,\mathrm{e}^{-\gamma(t-\tau)}\,\mathrm{d}\,\tau = \theta(t-t')\,\mathrm{e}^{-\gamma(t-t')} \equiv \mathbb{E}\left[\frac{\delta \mathsf{V}_t}{\delta\eta(t')}\right] \tag{5.22}$$

The equal-time response $R(t,t)$ is undefined, due to the discontinuity of the step function: different conventions can be used for $\theta(0)$, so prescriptions are needed to fix its value. Two common ones are the Stratonovich convention,

$$\theta(0) = \frac{1}{2} \Longrightarrow R(t,t) = \frac{1}{2} \tag{5.23}$$

and the Itô convention,

$$\theta(0) = 0 \Longrightarrow R(t,t) = 0 \tag{5.24}$$

Itô then regards $\eta(t)$ as acting *immediately after* that V_t has been updated, so the latter is independent of the concomitant noise. Conversely, Stratonovich regards concomitant V and η as "half-correlated", i.e. their correlation is half of the noise autocorrelation function.

At this point, we can express $\mathbb{E}[\mathsf{V}_t\eta(t')]$ as

$$\mathbb{E}\left[\frac{\partial \mathsf{V}_t}{\partial\eta(t')}\right] = \frac{\partial \mathbb{E}[\mathsf{V}_t]}{\partial h(t')} =: R(t,t') \Longrightarrow \mathbb{E}[\mathsf{V}_t\eta(t')] = 2DR(t,t').$$

that is a new form of the fluctuation–dissipation theorem.

2.4. From the Langevin to the Fokker-Planck equation. Let us now establish the link between the Langevin and the Fokker-Planck approaches. We recall that the Fokker-Planck equation for a stochastic process V_t is given by

$$\partial_t p(v,t) = -\partial_v\left[a^{(1)}(v,t)p(v,t)\right] + \frac{1}{2}\partial_v^2[a^{(2)}(v,t)p(v,t)] \tag{5.25}$$

with

$$\begin{aligned} a^{(1)}(v,t) &= \lim_{\tau\to 0}\frac{\mathbb{E}[\mathsf{V}_{t+\tau} - \mathsf{V}_t]_{\mathsf{V}_t=v}}{\tau} \\ a^{(2)}(v,t) &= \lim_{\tau\to 0}\frac{\mathbb{E}[(\mathsf{V}_{t+\tau} - \mathsf{V}_t)^2]_{\mathsf{V}_t=v}}{\tau} \end{aligned} \tag{5.26}$$

Integrating $\dot{\mathsf{V}} = -\gamma \mathsf{V} + \eta$ over a small time interval τ, using the forward Euler's integration method

$$\mathsf{V}_{t+\tau} - \mathsf{V}_t = -\gamma \int_t^{t+\tau} \mathsf{V}_{t'} \, \mathrm{d}\, t' + \int_t^{t+\tau} \eta(t') \, \mathrm{d}\, t' \simeq -\gamma \mathsf{V}_t \tau + \int_t^{t+\tau} \eta(t') \, \mathrm{d}\, t'.$$

Taking the average at fixed $\mathsf{V}_t = v$

$$a^{(1)}(v,t) = \lim_{\tau \to 0} \frac{\mathbb{E}[\mathsf{V}_{t+\tau} - \mathsf{V}_t]_{\mathsf{V}_t = v}}{\tau} = -\gamma v.$$

Squaring, taking the average and using $\mathbb{E}[\eta(t)\eta(t')] = 2D\delta(t-t')$ we can also compute directly

$$\mathbb{E}[(\mathsf{V}_{t+\tau} - \mathsf{V}_t)^2]_{\mathsf{V}_t = v} = \gamma^2 v^2 \tau^2 + 2D\tau \Longrightarrow a^{(2)}(v,t) = 2D$$

This results in the Fokker-Planck equation for the Ornstein-Uhlenbeck process

$$\partial_t p(v,t) = \gamma \partial_v (v\, p(v,t)) + D\partial_v^2 p(v,t).$$

As mentioned in Lectures 4, this is widely used in financial mathematics to model interest and exchange rates, as it describes a random walk with a tendency to move back towards its average value. In finance, T may represent the volatility caused by shocks, and γ the rate at which shocks are dissipated.

3. Non-linear Langevin equations

The Langevin equation for the Brownian motion given in Eq. (5.1) is *linear*, as the drift is a linear function of the random variable. Next, we consider Langevin equations with non-linear drifts.

3.1. Non-linear drift, additive noise. We start considering Langevin equations for the stochastic process X_t, with non-linear drifts, where the noise term is independent of the random variable and simply added to the equation, as in (5.1):

$$\frac{\mathrm{d}\,\mathsf{X}}{\mathrm{d}\,t} = A(\mathsf{X}) + \eta, \qquad \mathbb{E}[\eta(t)] = 0, \quad \mathbb{E}[\eta(t)\eta(t')] = 2D\delta(t-t'), \quad D > 0.$$

This type of noise is commonly referred to as 'additive'. In this case, we can calculate the drift and diffusion coefficients similarly to the way we did for the linear Langevin equation. With the usual assumption of Gaussian white noise, we obtain for the drift

$$\Delta \mathsf{X} := \mathsf{X}_{t+\tau} - \mathsf{X}_t = \int_t^{t+\tau} A(\mathsf{X}_{t'}) \, \mathrm{d}\, t' + \int_t^{t+\tau} \eta(t') \, \mathrm{d}\, t' \Longrightarrow \mathbb{E}[\Delta \mathsf{X}]_{\mathsf{X}_t = x} = A(x)\tau + O(\tau^2).$$

For the diffusion term

$$\mathbb{E}[(\Delta \mathsf{X})^2]_{\mathsf{X}_t = x} = \overbrace{\mathbb{E}\left[\left(\int_t^{t+\tau} A(\mathsf{X}_{t'}) \mathrm{d}t'\right)^2\right]_{\mathsf{X}_t = x}}^{A^2(x)\tau^2} + \overbrace{2\int_t^{t+\tau} \mathrm{d}t_1 \int_t^{t+\tau} \mathrm{d}t_2 \langle A(\mathsf{X}_{t_1}) \eta(t_2)\rangle}^{\text{use } A(\mathsf{X}_{t_i}) = A(x) + A'(x)[\mathsf{X}_{t_i} - x]} + \int_t^{t+\tau} \mathrm{d}\, t_1 \int_t^{t+\tau} \mathrm{d}\, t_2 \, \langle \eta(t_1)\eta(t_2)\rangle = 2D\tau + O(\tau^2), \tag{5.27}$$

leading to

$$\partial_t p(x,t) = -\partial_x[A(x)p(x,t)] + D\partial_x^2 p(x,t).$$

In particular, we already saw that if

$$A(x) = -\beta D \partial_x H(x) \tag{5.28}$$

for some function $H(x)$ and constant β, the steady state solution of the Fokker–Planck equation has the Boltzmann form

$$\pi(x) = \frac{1}{Z}\,\mathrm{e}^{-\beta H(x)}\,.$$

By setting the kinetic coefficient $\Gamma \coloneqq \beta D$, and by identifying $\beta^{-1} = k_{\mathrm{B}}T$, in this case we can write the Langevin equation as

$$\dot{\mathsf{X}} = -\Gamma\partial_x H(x) + \eta(t), \qquad \mathbb{E}[\eta(t)] = 0, \quad \mathbb{E}[\eta(t)\eta(t')] = 2\Gamma k_{\mathrm{B}}T\delta(t-t'), \tag{5.29}$$

which describes a gradient-descent dynamics on the energy function $H(x)$ in presence of some noise.

If the coefficient of the term η in the Langevin equation is non-constant, it is said that we have "multiplicative" noise,

$$\dot{\mathsf{X}} = A(\mathsf{X}) + B(\mathsf{X})\eta, \qquad \mathbb{E}[\eta(t)] = 0, \quad \mathbb{E}[\eta(t)\eta(t')] = 2D\delta(t-t'), \quad D > 0. \tag{5.30}$$

In the equation above, $B(x) > 0$ for all values of x. This Langevin equation equation can be mapped to one with additive noise as follows

$$\frac{\dot{\mathsf{X}}}{B(\mathsf{X})} = \frac{A(\mathsf{X})}{B(\mathsf{X})} + \eta$$

and introducing the new variable

$$\dot{\mathsf{y}} = \frac{\dot{\mathsf{X}}}{B(\mathsf{X})} \Leftrightarrow \mathsf{y} = \int^{\mathsf{X}} \frac{\mathrm{d}\,x}{B(x)} \equiv \phi(\mathsf{X})$$

Note that $\phi'(x) = \frac{1}{B(x)} > 0$. We can define

$$\frac{A(\phi^{-1}(\mathsf{y}))}{B(\phi^{-1}(\mathsf{y}))} =: A_1(\mathsf{y})$$

and the Langevin equation can be rewritten again in the form

$$\dot{\mathsf{y}} = A_1(\mathsf{y}) + \eta(t),$$

which we know corresponds to the Fokker–Planck equation for the density p_{y} of the variable y

$$\partial_t p_{\mathsf{y}}(y,t) = -\partial_y[A_1(y)p_{\mathsf{y}}(y,t)] + D\partial_x^2 p_{\mathsf{y}}(y,t)$$

To recover the density $p(x,t)$ we can start observing that

$$\begin{aligned} p_{\mathsf{y}}(y,t) &= \int p(x,t)\delta(\phi(x)-y)\,\mathrm{d}\,x = \int \mathrm{d}\,x\, p(x,t)\frac{\delta(x-\phi^{-1}(y))}{|\phi'(\phi^{-1}(y))|} \\ &= \frac{p(\phi^{-1}(y),t)}{|\phi'(\phi^{-1}(y))|} = p(\phi^{-1}(y),t)B(\phi^{-1}(y)), \end{aligned} \tag{5.31}$$

i.e., $p_{\mathsf{y}}(y,t) \xrightarrow{y=\phi(x)} p(x,t)B(x)$. Using now that

$$\frac{\partial}{\partial y} = \frac{\partial x}{\partial y}\frac{\partial}{\partial x} = B(x)\frac{\partial}{\partial x}$$

we can complete the rewriting of our equation as

$$\begin{aligned} \partial_t[B(x)p(x,t)] &= -B(x)\partial_x[A_1(\phi(x))B(x)p(x,t)] + DB(x)\partial_x[B(x)\partial_x(B(x)p(x,t)] \\ &= -B(x)\partial_x[A(x)p(x,t)] + DB(x)\partial_x[B(x)\partial_x(B(x)p(x,t)] \end{aligned} \tag{5.32}$$

which can be rewritten as

$$\partial_t p(x,t) = -\partial_x[(A(x) + DB(x)B'(x))p(x,t)] + D\partial_x^2[B^2(x)p(x,t)]. \tag{5.33}$$

In other words, $a^{(1)} = A + DBB'$ and $a^{(2)} = DB^2$. The drift term is normally referred to as the 'anomalous drift', due to the extra term DBB', which was not originally included in the original drift of the Langevin equation.

Q One can show that the above Fokker-Planck equation is compatible with the *Stratonovich* interpretation of the noise, whereas in the Itô convention the drift in the Fokker–Planck equation consists of only A. The source of the such 'disagreement' arises from the fact that the noise $\eta(t)$ can be regarded as a sequence of instantaneous 'kinks' on the system arriving at random times t_i, causing jumps in the random variable X. At every kink, it is not clear if the term $B(\mathsf{X})$ in the Langevin equation should be evaluated for the value that X takes *before* or right *after* the kink, or an average (possibly weighted) of the two should be used. Integration of the Langevin equation thus requires a prescription as we can get different results:

$$\int_t^{t+\mathrm{d}\,t} B(\mathsf{X}_\tau)\eta(\tau)\,\mathrm{d}\,\tau = \begin{cases} \text{Itô} & B(\mathsf{X}_t)\int_t^{t+\mathrm{d}\,t}\eta(\tau)\,\mathrm{d}\,\tau \\ \text{Stratonovich} & B\left(\frac{\mathsf{X}_t+\mathsf{X}_{t+\mathrm{d}\,t}}{2}\right)\int_t^{t+\mathrm{d}\,t}\eta(\tau)\,\mathrm{d}\,\tau. \end{cases} \tag{5.34}$$

The two prescriptions lead to different writings of the Fokker-Planck equation:

$$\begin{aligned} \text{Stratonovich}\quad & \partial_t p(x,t) = -\partial_x[(A(x) + DB(x)B'(x))p(x,t)] + D\partial_x^2[B^2(x)p(x,t)], \\ \text{Itô}\quad & \partial_t p(x,t) = -\partial_x[A(x)p(x,t)] + D\partial_x^2[B^2(x)p(x,t)]. \end{aligned} \tag{5.35}$$

Itô's interpretation is seemingly incompatible with standard variable transformations (in fact, Itô's calculus is required to perform it). On the other hand, if the noise has a finite correlation time, the equation of motion is non-singular, and one would expect to be able to carry out any transformation by using standard rules of calculus. Hence, the Stratonovich convention seems appropriate when physical noise with finite correlation time is considered. However, Itô's convention is easily interpreted as a Euler forward integration and it leads to simpler equations, so it is widely used with the idealised (i.e., delta-correlated) noise.

It is easy to see that the two prescriptions lead to different values of the equal-time response function, as anticipated in Lectures 4. For example, by considering the simple equation

$$\dot{\mathsf{X}} = \mathsf{X}\eta$$

one has

$$\begin{aligned} \text{Stratonovich}\quad & \partial_t p(x,t) = -D\partial_x[xp(x,t)] + D\partial_x^2[x^2 p(x,t)] \Rightarrow \partial_t\langle\mathsf{X}\rangle = D\langle\mathsf{X}\rangle, \\ \text{Itô}\quad & \partial_t p(x,t) = D\partial_x^2[x^2 p(x,t)] \Rightarrow \partial_t\langle\mathsf{X}\rangle = 0. \end{aligned} \tag{5.36}$$

Inserting in the equation of motion, one has $\langle \mathsf{X}_t\eta(t)\rangle = 0$ according to the Itô prescription, and $\langle \mathsf{X}_t\eta(t)\rangle \neq 0$ in the Stratonovich convention.

4. Many-particles systems

So far, we have considered stochastic processes described by one degree of freedom, X. We consider now stochastic systems with many interacting degrees of freedom, described therefore by a vectorial variable $\mathbf{X}$.

4.1. Linear systems. We start by considering a system of N *interacting* units, labelled by $i = 1, \dots, N$. The 'units' can be thought as charged Brownian particles, neurons in a neural networks or species in an ecosystem. We will loosely call them 'particles'. We assume that each particle i can be described by a coordinate x_i (this could be the position of a Brownian particle, the action potential of a neuron or the population density of a specie) and we will assume that

each coordinate evolves according to a Langevin dynamics with an independent zero-average white noise, so for each $i = 1, \dots, N$

$$\dot{\mathsf{X}}_i = -\gamma \mathsf{X}_i + \sum_j J_{ij} \mathsf{X}_j + \eta_i \tag{5.37}$$

with $\langle \eta_i(t) \rangle = 0$ and $\langle \eta_i(t) \eta_j(t') \rangle = 2T\delta_{ij}\delta(t-t')\ \forall\ i, j$.

Q The noiseless version of this model is one of the most popular models introduced in ecology to study the stability of diverse ecosystems with many interacting specie, studied by Mark Gardner and Ross Ashby in 1970 and by Robert May in 1972. Similiar, but noisy, models, with the replacement $\sum_k J_{ik}\mathsf{X}_k \to \sum_k J_{ik} g(\mathsf{X}_k)$ for some sigmoid function $g(x)$, have served as popular neural networks models, as proposed by Haim Sompolinski, Andrea Crisanti, and Sommers in 1988, or, more recently, to model financial markets Kartik Anand, Jonathan Khedair, and Reimer Kühn in 2017. Ecological systems and banking systems indeed share important similarities, as pointed out by Andrew Haldane and Robert May in 2011.

To study the system of coupled Langevin equations (5.37), it is convenient to write equations in vector notation

$$\partial_t |\mathbf{X}\rangle = (\boldsymbol{J} - \gamma \boldsymbol{I})|\mathbf{X}\rangle + |\boldsymbol{\eta}\rangle.$$

We will first look at symmetric interactions $J_{ij} = J_{ji}$ and will comment later on asymmetric interactions. If the J_{ij}'s are symmetric, then there exist a complete set of orthonormal eigenvectors $|\boldsymbol{v}^a\rangle$, $a = 1, \dots, N$, so that

$$\boldsymbol{J} = \sum_{a=1}^{N} \lambda_a |\boldsymbol{v}^a\rangle\langle \boldsymbol{v}^a|, \qquad \langle \boldsymbol{v}^a | \boldsymbol{v}^b \rangle = \delta_{ab}.$$

The matrix $\boldsymbol{J}$ is diagonalised by a similarity transformation

$$\boldsymbol{O}^{-1} \boldsymbol{J} \boldsymbol{O} = \boldsymbol{J}_D$$

where $\boldsymbol{O}$ is an orthogonal matrix, i.e., $\boldsymbol{O}^{-1} = \boldsymbol{O}^\top$ whose columns are constituted by the eigenvectors of $\boldsymbol{J}$, i.e., if we denote

$$|\boldsymbol{e}^k\rangle = (\delta_{ki})_{i=1}^N = (0, \dots, 0, \underbrace{1}_{\text{position } k}, 0, \dots, 0) \in \mathbb{R}^N, \tag{5.38}$$

then $\boldsymbol{O} = \sum_{a=1}^N |\boldsymbol{v}^a\rangle\langle \boldsymbol{e}^a|$. On the other hand, we have written

$$\boldsymbol{J}_D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_N \end{pmatrix} = \sum_a \lambda_a |\boldsymbol{e}^a\rangle\langle \boldsymbol{e}^a|. \tag{5.39}$$

It is then convenient to transform the coordinates as

$$|\mathsf{y}\rangle = \boldsymbol{O}^\top |\mathbf{X}\rangle, \qquad |\boldsymbol{\xi}\rangle = \boldsymbol{O}^\top |\boldsymbol{\eta}\rangle$$

so that the equation of motion, in transformed coordinates reads as

$$\partial_t |\mathsf{y}\rangle = (\boldsymbol{J}_D - \gamma \boldsymbol{I})|\mathsf{y}\rangle + |\boldsymbol{\xi}\rangle$$

and can be described by a set of independent equations

$$\dot{\mathsf{y}}_a = (\lambda_a - \gamma)\mathsf{y}_a + \xi_a. \tag{5.40}$$

Each component y^a is simply the projection of $\mathbf{X}$ over $\boldsymbol{v}$,

$$\mathsf{y}^a = \sum_i O_{ia} x_i = \sum_i v_i^a x_i = \langle \boldsymbol{v}^a | \boldsymbol{x} \rangle. \tag{5.41}$$

On the other hand

$$\mathsf{X}_i = \sum_a O_{ia}\mathsf{y}_a = \sum_a v_i^a \mathsf{y}_a. \tag{5.42}$$

Here we still have to specify the statistical properties of the transformed noise $\boldsymbol{\xi}$. Each component ξ_i is

$$\xi_a = \sum_k O_{ka}\eta_k = \sum_k v_k^a \eta_k = \langle \boldsymbol{v}^a | \boldsymbol{\eta} \rangle, \tag{5.43}$$

and therefore

$$\mathbb{E}[\xi_a(t)] = 0 \qquad \forall a,$$
$$\mathbb{E}[\xi_a(t)\xi_b(t')] = \sum_{ij} v_i^a v_j^b \mathbb{E}[\eta_i(t)\eta_j(t')] = 2T\delta(t-t')\sum_i v_i^a v_i^b = 2T\delta_{ab}\delta(t-t')$$

Hence, each Langevin equation (5.40) reduces to the equation for the single Brownian particle and can be solved independently using the same methods. In particular, each Langevin equation is equivalent to a Fokker–Planck equation for the density p_a of the variable y_a,

$$\partial_t p_a(y,t) = -\partial_y[(\lambda_a - \gamma) y p_a(y,t)] + T\partial_y^2 p_a(y,t),$$

whose stationary solution is

$$p_a(y) = \frac{1}{Z_a}\, \mathrm{e}^{-\frac{1}{2T}(\gamma-\lambda_a)y^2}$$

Recalling the relation between the original and the transformed variables $|\mathbf{y}\rangle = \boldsymbol{O}^\top |\mathbf{X}\rangle$, i.e., $\mathsf{y}_a = \sum_i v_i^a \mathsf{X}_i$, we can express the distribution of $\mathbf{X}$ in terms of the distribution of $\mathbf{y}$, which is in factorised form $p(\boldsymbol{y}) = \prod_a p_a(y_a)$, due to the independence of the transformed variables

$$\begin{aligned} p(\boldsymbol{x}) &= \prod_a \left[\int dy_a\, p_a(y_a)\delta\left(y_a - \langle \boldsymbol{v}^a | \boldsymbol{x}\rangle\right)\right] \\ &= \prod_a p_a\left(\langle \boldsymbol{v}^a | \boldsymbol{x}\rangle\right) \\ &= \frac{\exp\left[-\frac{1}{2T}\sum_a (\gamma-\lambda_a)\langle \boldsymbol{x}|\boldsymbol{v}^a\rangle\langle \boldsymbol{v}^a|\boldsymbol{x}\rangle\right]}{\prod_a Z_a} \\ &= \frac{\exp\left[-\frac{1}{2T}\langle \boldsymbol{x}|\left(\gamma\sum_a |\boldsymbol{v}^a\rangle\langle \boldsymbol{v}^a| - \sum_a \lambda_a |\boldsymbol{v}^a\rangle\langle \boldsymbol{v}^a|\right)|\boldsymbol{x}\rangle\right]}{Z} \\ &= \frac{1}{Z}\, \mathrm{e}^{-\frac{1}{2T}\langle \boldsymbol{x}|\gamma \boldsymbol{I} - \boldsymbol{J}|\boldsymbol{x}\rangle}, \end{aligned} \tag{5.44}$$

which is in Boltzmann form

$$\Pi(\boldsymbol{x}) = \frac{1}{Z}\, \mathrm{e}^{-\beta H(\boldsymbol{x})},$$

with Hamiltonian

$$H(\boldsymbol{x}) = \frac{1}{2}\langle \boldsymbol{x}|\gamma \boldsymbol{I} - \boldsymbol{J}|\boldsymbol{x}\rangle = \frac{\gamma}{2}\sum_i x_i^2 - \frac{1}{2}\sum_{ij} J_{ij}x_i x_j.$$

The convergence to the Boltzmann distribution is not surprsing as the Langevin dynamics in Eq. (5.37) can be seen as a gradient-descent on the Hamiltonian $H(\boldsymbol{x})$ in the presence of noise

$$\dot{\mathbf{X}} = -\nabla H(\mathbf{X}) + \boldsymbol{\eta}. \tag{5.45}$$

During Lectures 4, we have shown that this type of dynamics converges to the Boltzmann distribution, for a single degree of freedom. The derivation above generalises this result to many degrees of freedom.

4.2. Correlation and response functions. It is possible to solve each Langevin equation (5.40) independently from all the others, as it describes a Brownian motion of a single particle, so that

$$\mathsf{y}_a(t) = \mathsf{y}_a(0)\,\mathrm{e}^{-(\gamma-\lambda_a)t} + \int_0^t \eta_k(\tau)\,\mathrm{e}^{-(\gamma-\lambda_k)(t-\tau)}\,\mathrm{d}\,\tau,$$

and therefore

$$\mathbb{E}[\mathsf{y}_a(t)\mathsf{y}_a(t')] = \mathsf{y}_a^2(0)\,\mathrm{e}^{-(\gamma-\lambda_a)(t+t')} + 2T\int_0^{\min(t,t')} \mathrm{e}^{-(\gamma-\lambda_k)(t+t'-2\tau)}\,\mathrm{d}\,\tau$$

From these, we can calculate useful observables for the original system, such as the global correlation and the response function,

$$\begin{aligned} C(t,t') &\coloneqq \frac{1}{N}\sum_i \Big(\mathbb{E}[\mathsf{X}_i(t)\mathsf{X}_i(t')] - \mathbb{E}[\mathsf{X}_i(t)]\mathbb{E}[\mathsf{X}_i(t')]\Big) \\ R(t,t') &\coloneqq \frac{1}{N}\sum_i \frac{\delta\mathbb{E}[\mathsf{X}_i(t)]}{\delta h_i(t')}\Big|_{h=0} \end{aligned} \tag{5.46}$$

which give information on the global coherence of the system and its susceptibility to external perturbations. From

$$\frac{1}{N}\sum_i \mathbb{E}[\mathsf{X}_i(t)\mathsf{X}_i(t')] = \frac{1}{N}\sum_i\sum_{ab} v_i^a v_i^b \mathbb{E}[\mathsf{y}_a(t)\mathsf{y}_b(t')] = \frac{1}{N}\sum_i \mathbb{E}[\mathsf{y}_i(t)\mathsf{y}_i(t')], \tag{5.47}$$

where it has been used the fact that $\sum_i v_i^a v_i^b = \delta_{ab}$. Collecting all pieces we obtain then

$$\begin{aligned} C(t,t') &= \frac{1}{N}\sum_a 2T\int d\lambda\,\delta(\lambda-\lambda_a)\int_0^{\min(t,t')} \mathrm{e}^{-(\gamma-\lambda)(t+t'-2s)}\,\mathrm{d}\,a \\ &= T\sum_\lambda \rho(\lambda)\Big(\mathrm{e}^{-(\gamma-\lambda)|t-t'|} - \mathrm{e}^{-(\gamma-\lambda)(t+t')}\Big) \end{aligned} \tag{5.48}$$

where we have introduced the spectral density (i.e., the distribution of eigenvalues) of the interaction matrix

$$\rho(\lambda) = \lim_{N\to\infty}\frac{1}{N}\sum_a \delta(\lambda-\lambda_a).$$

Similarly, we obtain for the response function

$$R(t,t') = \sum_\lambda \rho(\lambda)\frac{\delta\mathbb{E}[\mathsf{X}_\lambda(t)]}{\delta h_\lambda(t')}\Big|_{h=0} = \sum_\lambda \rho(\lambda)\,\mathrm{e}^{-(\gamma-\lambda)(t-t')}\,. \tag{5.49}$$

This shows that the system's behavior is fully encoded in the spectral density of the interaction matrix. In particular, one can see that small perturbations will be damped if the largest eigenvalue $\max_a \lambda_a$ is smaller than γ (this has indeed been related to qualitative stability in ecosystems, economic models etc.). In addition we observe that for large t, t', where the second term in (5.48) vanishes, correlation and response functions are related by the fluctuation–dissipation theorem

$$R(t,t') = -\frac{1}{T}\frac{\partial C(t,t')}{\partial t}.$$

In conclusion, we have shown that for systems with two-body symmetric interactions and linear drift, the Langevin dynamics can be solved by diagonalizing the interaction matrix, the system converges to the Boltzmann distribution and FDT is satisfied. For asymmetric interactions, we

have a more complex scenario, due to the presence of complex eigenvalues (normally leading to oscillations) and left and right eigenvectors. However, one can still prove convergence to a steady state which is not Boltzmann, where FDT is generally violated.

5. Numerical integration

We conclude this chapter by illustrating how to numerically integrate stochastic differential equations like

$$\dot{\mathsf{X}} = A(\mathsf{X}, t) + B(\mathsf{X}, t)\eta.$$

The easiest way is to discretize time according to Itô's convention

$$\mathsf{X}_{t+\tau} = \mathsf{X}_t + A(\mathsf{X}, t)\tau + B(\mathsf{X}, t)\tilde{\eta}(t), \tag{5.50}$$

where we have defined the integrated noise $\tilde{\eta}(t) = \int_t^{t+\tau} \eta(t')\,\mathrm{d}\,t'$. This satisfies

$$\mathbb{E}\tilde{\eta}(t)] = 0 \qquad \mathbb{E}[\tilde{\eta}(t)\tilde{\eta}(t')] = 2T\tau\delta(t - t'), \tag{5.51}$$

and is Gaussian if η was Gaussian. In the case of Gaussian noise, we can therefore consider the following equation

$$\mathsf{X}_{t+\tau} = \mathsf{X}_t + \tau A(\mathsf{X}, t) + B(\mathsf{X}, t)\sqrt{2T\tau}\zeta,$$

where ζ is Gaussian with zero mean and unit variance. This suggests the following algorithm:

(1) Initialize $t = 0$, $x = x_0$, desired initial condition. Set τ to a small number.
(2) Calculate $A(x, t)$, $B(x, t)$. Draw $\zeta \sim \mathcal{N}(0, 1)$, i.e., Gaussian with zero mean and unit variance.
(3) Update $x + A(x, t)\tau + B(x, t)\sqrt{2T\tau}\zeta \mapsto x$.
(4) Set $t + \tau \mapsto t$ and go back to (2).

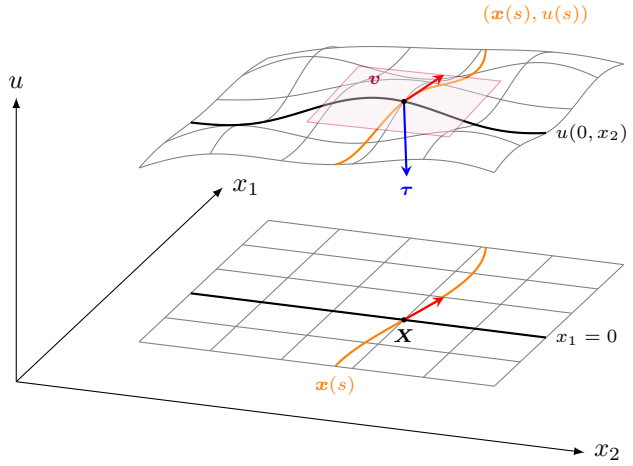
APPENDIX A

The method of characteristics

The method of characteristics is a method for solving linear, semilinear, or quasilinear PDEs of the first order, i.e., equations for the function $u(\boldsymbol{x})$, $u\colon \mathbb{R}^d \to \mathbb{R}$, in the form

$$\sum_{j=1}^{d} a_j(\boldsymbol{x}, u)\frac{\partial u}{\partial x_j} = f\left(\boldsymbol{x}, u\right), \tag{A.1}$$

where a_j and f are some known functions of $d+1$ variables. The main goal of the method is to find special curves, called *characteristics*, along which the PDE becomes a family of ODE. Once the ODEs are found and solved along the characteristics curves they can be related to the solution of the original PDE. The rationale of this idea comes from the observation that we can imagine the solution $u(\boldsymbol{x})$ of the problem as a hypersurface $\mathcal{S} \subset \mathbb{R}^{d+1}$, identified by the equation $u = u(\boldsymbol{x})$.



Given a point $(\boldsymbol{x}, u(\boldsymbol{x}))$ of this surface, the vector

$$\boldsymbol{\tau} \coloneqq \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \ldots, \frac{\partial u}{\partial x_d}, -1\right) \equiv (\nabla u, -1)$$

is orthogonal to $\mathcal{S}$ in $\boldsymbol{x}$. On the other hand, Eq. (A.1) says that the vector

$$\boldsymbol{v} = (a_1(\boldsymbol{x}, u), \ldots, a_d(\boldsymbol{x}, u), f(\boldsymbol{x}, u))$$

is orthogonal to the vector $\boldsymbol{\tau}$: indeed, the equation can be written as $\langle \boldsymbol{v} | \boldsymbol{\tau} \rangle = 0$. This means that $\boldsymbol{v}$ is tangent to $\mathcal{S}$, and the surface $\mathcal{S}$ can be thought of as the union of many curves going through the surface so that for each point of the surface $(\boldsymbol{X}, u(\boldsymbol{X}))$ there is only one curve $(\boldsymbol{x}(s), u(s))$ parametrised by $s \in \mathbb{R}$ passing through it and solving the set of equations

$$\frac{\mathrm{d}\, x_i}{\mathrm{d}\, s} = a_i(\boldsymbol{x}, u), \text{ for } i = 1, \ldots, d \qquad \frac{\mathrm{d}\, u}{\mathrm{d}\, s} = f(\boldsymbol{x}, u)$$

that can be rewritten in the form of *Lagrange-Charpit equations*

$$\frac{\mathrm{d}\, x_1}{a_1\left(\boldsymbol{x}, u\right)} = \cdots = \frac{\mathrm{d}\, x_d}{a_d\left(\boldsymbol{x}, u\right)} = \frac{\mathrm{d}\, u}{f\left(\boldsymbol{x}, u\right)}. \tag{A.2}$$

Equations (A.2) are called *characteristic equations* and describe d-parametric family (coming from d integrations) of characteristic curves. If the PDE is equipped with some initial or boundary data, then one can eliminate all the constants and find unique solution for (A.1). Otherwise, general solution can be written by assuming that all except one free parameters are expressed as some unknown functions of a last one.

EXAMPLE Consider the initial value problem for **Burgers' equation**

$$\begin{cases} \partial_t u + u \partial_x u = 0, \\ u\left(x, 0\right) = \phi(x). \end{cases}$$

The characteristic equations are given by

$$\frac{\mathrm{d}\, t}{\mathrm{d}\, s} = 1 \qquad \frac{\mathrm{d}\, x}{\mathrm{d}\, s} = u \qquad \frac{\mathrm{d}\, u}{\mathrm{d}\, s} = 0,$$

with initial conditions

$$t(0) = 0, \qquad x(0) = X, \qquad u(0) = \phi(X), \tag{A.3}$$

i.e., for $s = 0$ (at the origin of the characteristic curve) x takes the value X, $t = 0$ and $u \equiv \phi(x) = \phi(X)$. All equations are immediately solved as

$$t = s, \qquad x = us + X, \qquad u = \phi(X),$$

implying that $X = x - tu$ and therefore u satisfies the implicit equation

$$u(x, t) = \phi\left(x - u(x, t)t\right).$$