Stochastic Models of Complex Systems

CO905 - MSc in Complexity Science

Stefan Grosskinsky

Warwick, 2011

These notes and other information about the course are available on
`http://www.warwick.ac.uk/~masgav/teaching/co905.html`

Contents

References

[GS] G. Grimmett, D. Stirzaker: Probability and Random Processes (3rd edition), Oxford 2001
[Ga] C.W. Gardiner: Handbook of Stochastic Methods (3rd edition), Springer 2004
[G] G. Grimmett: Probability on Graphs, CUP 2010
`http://www.statslab.cam.ac.uk/~grg/books/pgs.html`

Introduction

In this module we will cover the basics to study models of complex systems with stochastic time evolution. There are two different origins of stochasticity:

- Classical mechanics: stochasticity due to lack of information
  In principle all components involved in the system are believed to follow a deterministic system of equations of motion. But in practice all microscopic details are not accessible and the unknown influences on the dynamics are approximated as effective random noise with a certain postulated distribution. The actual origin of the noise may be related to chaotic motion, i.e. deterministic time evolution with random initial data such as a dice or pendulum, or neglected interactions in a large system such as gases or fluids leading to a stochastic time evolution.

- Quantum mechanics: inherent stochasticity
  Even simple systems can only be described stochastically and the full microscopic details are inherently inaccessible (uncertainty principle). Mathematically, the state of such a system is therefore given by a complex probability density function (wave function), rather than a single element in the set of all possible configurations.

Examples. (gas, branching process)

In this course we only cover classical stochastic systems. After a general introduction to stochastic processes we will study some examples of particle systems with thermal interactions. The first and most classical example of this phenomenon is *Brownian motion* (see Gardiner, Section 1.2). In 1827 Robert Brown observed the irregular motion of small pollen grains suspended in water. A first satisfactory theoretical description of this phenomenon was given by Einstein in 1905. A mathematically idealized version of this is called the Wiener process and can be described by the theory of stochastic calculus which was developed in the 1950s by Itô. Due to the continuous state space of the system this theory is rather involved, and will be discussed towards the end of the module. Simpler to analyse are models with a discrete state space such as birth-death processes, which appear for example in predator-prey models in biology (see Gardiner, Section 1.3). In the first part of the course we concentrate on *Markov chains* (following [GS] Chapter 6), which are certain stochastic processes with discrete state space. We conclude the introductory section by two general definitions.

Definition 0.1 A *stochastic process* $X = (X_t : t \in \mathbb{T})$ is a family of random variables $X_t : \Omega \to S$ with *state space* S and time index set $\mathbb{T} \subseteq \mathbb{R}$.

A stochastic process $X : \mathbb{T} \times \Omega \to S$ is a function of two variables, time t and $\omega \in \Omega$. For fixed ω, the function $t \mapsto X_t(\omega)$ is called a *sample path*. The probability space Ω is arbitrary, but has to be big enough to encode all possible time evolutions. A canonical choice is the set of all possible sample paths $\Omega = \{f : \mathbb{T} \to S\}$, or often one requires some regularity of the functions f, such as continuity.

Definition 0.2 A stochastic process is a *Markov process* if for all $t_1 < t_2 < \ldots < t_n \in \mathbb{T}$, $n \in \mathbb{N}$, for all $s_1, \ldots, s_{n-1} \in S$ and all (measurable) $A \subseteq S$,

$$\mathbb{P}\big(X_{t_n} \in A_n \big| X_{t_1} = s_1, \ldots, X_{t_{n-1}} = s_{n-1}\big) = \mathbb{P}\big(X_{t_n} \in A_n \big| X_{t_{n-1}} = s_{n-1}\big) \,. \tag{0.1}$$

A Markov process is called *homogeneous* if for all (measurable) $A, B \subseteq S$ and $t > t' \in \mathbb{T}$

$$\mathbb{P}\big(X_t \in A \,\big|\, X_{t'} \in B\big) = \mathbb{P}\big(X_{t-t'} \in A \,\big|\, X_0 \in B\big) \,. \tag{0.2}$$

A homogeneous Markov process is called a *Markov chain*, if S is discrete.

In this course we will only deal with homogeneous Markov processes. We will concentrate on the choices $\mathbb{T} = \mathbb{N}$, $\mathbb{Z}$ for discrete time and $\mathbb{T} = [0, \infty)$, $\mathbb{R}$ for continuous time processes. Typical choices for state spaces are $S = \mathbb{Z}$ (e.g. random walk, birth-death processes), $\mathbb{N}$ (e.g. counting processes), $\mathbb{R}^d$ (e.g. Brownian motion).

Examples. (RW, Brownian motion, deck of cards)

1 Markov chains

1.1 General properties

Definition 1.1 For a Markov chain we define the *transition probabilities*

$$p_{ij}(t) := \mathbb{P}(X_t = j | X_0 = i) \in [0,1] \quad \text{for all } i,j \in S\,, \tag{1.1}$$

and the *transition 'matrices'* (which might be infinite)

$$P(t) := \big(p_{ij}(t) \,:\, i,j \in S\big) \in [0,1]^{|S|\times|S|}\,. \tag{1.2}$$

A homogeneous Markov chain starting at time $t = 0$ is uniquely determined by an initial distribution $\boldsymbol{\pi}(0)$ with $\pi_i(0) = \mathbb{P}(X_0 = i)$, $i \in S$ and the transition probabilities, because every joint probability can be written as

$$\begin{aligned}\mathbb{P}\big(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\big) &= \\ &= \sum_{i_0\in S, i_1\in A_1, .., i_n \in A_n} \pi_{i_0}(0) p_{i_0 i_1}(t_1) p_{i_1 i_2}(t_2 - t_1) \cdots p_{i_{n-1} i_n}(t_n - t_{n-1})\end{aligned} \tag{1.3}$$

for all $0 < t_1 < \ldots < t_n \in \mathbb{T}$ and $A_1, \ldots, A_n \subseteq S$. In particular, the distribution at time t is

$$\pi_j(t) = \mathbb{P}(X_t = j) = \sum_{i\in S} \pi_i(0)\, p_{ij}(t)\,, \quad \text{so} \quad \boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, P(t)\,. \tag{1.4}$$

Example. (RW)

Proposition 1.1 *$P(0) = Id$ and the family $\big(P(t) \,:\, t \geq 0\big)$ satisfies the **Chapman-Kolmogorov equations**,*

$$P(t+t') = P(t)\, P(t') \quad \textit{for all } t, t', t+t' \in \mathbb{T}\,. \tag{1.5}$$

Proof. $p_{ij}(0) = \delta_{ij}$ by definition, and for all $t, t', t+t' \in \mathbb{T}$

$$\begin{aligned}p_{ij}(t+t') &= \mathbb{P}(X_{t+t'} = j | X_0 = i) = \\ &= \sum_{k\in S} \mathbb{P}(X_{t+t'} = j | X_0 = i, X_t = k)\, \mathbb{P}(X_t = k | X_0 = i) = \sum_{k\in S} p_{ik}(t)\, p_{kj}(t')\,,\end{aligned} \tag{1.6}$$

using the total probability sum rule, the Markov property and homogeneity. □

For discrete time with $\mathbb{T} = \mathbb{N}$ this leads to

$$\boxed{P(n+1) = P(1)\,P(n) = P(n)\,P(1) \quad \Rightarrow \quad P(n) = P^n\,,} \tag{1.7}$$

where we denote $P = P(1)$. Therefore a discrite time Markov chain is uniquely determined by the initial distribution $\boldsymbol{\pi}(0)$ and the transition matrix P and in particular

$$\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\,P^n\,. \tag{1.8}$$

Example. (RW with various BCs)

For continuous time with $\mathbb{T} = [0, \infty)$ we require some regularity of the function $t \mapsto P(t)$ at $t = 0$. We only study processes where it is continuous and differentiable, i.e.

$$\lim_{t \searrow 0} P(t) = P(0) = Id \quad \text{and} \quad G := \lim_{t \searrow 0} \frac{P(t) - Id}{t} \quad \text{exists}\,, \tag{1.9}$$

so that $P(t) = Id + t\,G + o(t)$ for small t. Together with the Chapman-Kolmogorov equations this implies that

$$\frac{P(t+\Delta t) - P(t)}{\Delta t} = \frac{P(\Delta t) - Id}{\Delta t}\,P(t) = P(t)\frac{P(\Delta t) - Id}{\Delta t}\,, \tag{1.10}$$

and thus taking $\Delta t \searrow 0$, $P(t)$ is differentiable for all $t \geq 0$ and fulfills

$$\boxed{\frac{d}{dt}P(t) = G\,P(t) = P(t)\,G \quad \Rightarrow \quad P(t) = \exp(t\,G)\,.} \tag{1.11}$$

These are called *backward* and *forward equation*, respectively. Subject to the boundary conditions $P(0) = Id$, they often have a unique solution given by the matrix exponential

$$P(t) = \exp(t\,G) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\,G^k\,. \tag{1.12}$$

For example this is the case if $|S| < \infty$, and subject to certain technical conditions also for inifite state space S. Therefore a continuous-time Markov chain is uniquely determined by the initial

distribution and the matrix G which is called the *generator* of the process.
The distribution at time t is then given by

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, \exp(t\, G)\ . \tag{1.13}$$

What is the meaning of the entries of G and how do the sample paths of X look?

Entries of G. Assume that $X_t = i$. For small times Δt we have from (1.12)

$$p_{ij}(\Delta t) = g_{ij}\Delta t + o(\Delta t) \quad \text{for all } i \neq j \in S\ . \tag{1.14}$$

So during a time interval $(t, t+\Delta t)$ the chain jumps from state i to j with probability $g_{ij}\Delta t$, and $g_{ij} \geq 0$ can be interpreted as a jump rate. On the diagonal we have

$$p_{ii}(\Delta t) = 1 + g_{ii}\Delta t + o(\Delta t) \quad \text{for all } i \in S\ , \tag{1.15}$$

which gives the probability that nothing happens in the time interval $(t, t+\Delta t)$. By normalization we have

$$1 = \sum_{j\in S} p_{ij}(\Delta t) = 1 + \Delta t \sum_{j\in S} g_{ij} \quad \Rightarrow \quad \sum_{j\in S} g_{ij} = 0 \quad \text{for all } i \in S\ . \tag{1.16}$$

Therefore the diagonal entries of G are

$$g_{ii} = -\sum_{j\neq i} g_{ij} \leq 0 \quad \text{for all } I \in S\ , \tag{1.17}$$

and $|g_{ii}| \geq 0$ can be interpreted as the total rate to leave state i.

Sample paths. Assume that $X_0 = i$ and define the *holding time*

$$W_i := \inf\left\{t \geq 0\ :\ X_t \neq i\right\}\ , \tag{1.18}$$

i.e. the (random) time until a jump occurs. This is actually independent of t by homogeneity. If i is absorbing, $g_{ij} = 0$ for all $j \in S$ and $W_i = \infty$.

Proposition 1.2 *The random variable W_i is exponentially distributed with parameter $|g_{ii}|$ and if $|g_{ii}| > 0$, the probability that the chain jumps to $j \neq i$ after time W_i is $g_{ij}/|g_{ii}|$.*

Proof. W_i has 'loss of memory' property, i.e. for all $s, u > 0$

$$\mathbb{P}(W_i > s+u | W_i > s) = \mathbb{P}(W_i > s+u | X_s = i) = \mathbb{P}(W_i > u)\ , \tag{1.19}$$

where we have used the Markov property and homogeneity. Therefore

$$\mathbb{P}(W_i > s+u) = \mathbb{P}(W_i > u)\, \mathbb{P}(W_i > s)\ . \tag{1.20}$$

Analogous to the Chapman-Kolmogorov equations (1.5) this can be used to derive a differential equation for $\bar{F}(s) = \mathbb{P}(W_i > s)$ which has an exponential solution

$$\bar{F}(s) = \mathbb{P}(W_i > s) = e^{\lambda s} \quad \text{where} \quad \lambda = \bar{F}'(0)\ . \tag{1.21}$$

Together with (1.15) we get

$$\bar{F}'(0) = \lim_{\Delta t \searrow 0} \frac{\mathbb{P}(W_i > \Delta t) - 1}{\Delta t} = \lim_{\Delta t \searrow 0} \frac{p_{ii}(\Delta t) + o(\Delta t) - 1}{\Delta t} = g_{ii} \leq 0 \,, \tag{1.22}$$

and therefore $\mathbb{P}(W_i > s) = e^{-|g_{ii}|s}$ and $W_i \sim Exp(|g_{ii}|)$.
Now the probability that the chain jumps to j, conditioned on the event that it actually jumps somewhere in the time interval $(t, t + \Delta t]$, is given by

$$\mathbb{P}(X_{t+\Delta t} = j | X_t = i, W_i < \Delta t) = \frac{\mathbb{P}(X_{t+\Delta t} = j | X_t = i)}{\mathbb{P}(W_i < \Delta t | X_t = i)} \simeq$$

$$\simeq \frac{p_{ij}(\Delta t)}{1 - p_{ii}(\Delta t)} \to \frac{g_{ij}}{-g_{ii}} \quad \text{as } \Delta t \searrow 0 \,. \tag{1.23}$$

So conditioned on jumping at time t we get the required result. □

Picture of sample path.

The chain jumps at the *jump time* $J_n = \sum_{k=0}^{n-1} W_{Y_k}$ to state $Y_n = X_{J_n}$.
$Y = (Y_n : n \in \mathbb{N})$ is called the *jump chain*, and it is a discrete time Markov chain with transition Matrix P^Y given by

$$p_{ij}^Y = \begin{cases} 0 & , \, i = j \\ g_{ij}/|g_{ii}| & , \, i \neq j \end{cases} \quad \text{if } g_{ii} > 0 \,, \quad \text{and} \quad p_{ij}^Y = \delta_{ij} \quad \text{if } g_{ii} = 0 \,. \tag{1.24}$$

So a continuous-time Markov chain can also be characterized by its jump chain Y and a sequence of independent exponentially distributed holding times $(W_{Y_n} : n \in \mathbb{N})$.

Examples. (Poisson, CTRW)

For the *Poisson process* there exists also another characterization.

Proposition 1.3 *$X = (X_t : t \geq 0)$ is a Poisson process with rate λ if and only if it has stationary, independent increments, i.e.*

$$X_{t+t'} - X_{t'} \text{ is distributed like } X_t - X_0 \text{ and independent of } (X_s : s \leq t') \,, \tag{1.25}$$

and for each t, X_t has Poisson distribution with parameter λt, i.e. $\mathbb{P}(X_t = k) = \frac{(\lambda t)^k}{k!}\, e^{-\lambda t}$.

Proof. (differentiate $\pi_n(t)$)

Using the forward equation (1.11) we can also get an evolution equation for the distribution,

$$\frac{d}{dt}\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\,\frac{d}{dt}P(t) = \boldsymbol{\pi}(0)\,P(t)\,G = \boldsymbol{\pi}(t)\,G\,. \tag{1.26}$$

This is called the *Master equation* and using (1.17) the coordinate form is given by

$$\boxed{\frac{d}{dt}\pi_i(t) = \sum_{j\neq i}\Big(\pi_j(t)\,g_{ji} - \pi_i(t)\,g_{ij}\Big)\,.} \tag{1.27}$$

1.2 Stationary distributions and reversibility

Definition 1.2 A probability distribution $\boldsymbol{\pi}^*$ is called *stationary* if $\boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^*$ for all $t \geq 0$.

This will play an important role in the long-time behaviour of Markov chains, since 'often' $\boldsymbol{\pi}(t) \to \boldsymbol{\pi}^*$. How and when this is true will be seen later.

Proposition 1.4 *π^* is stationary if and only if*

$$\begin{aligned} \pi^* P &= \pi^* \qquad \textit{for a discrete-time chain with transition matrix } P\,, \\ \pi^* G &= \mathbf{0} \qquad \textit{for a continuous-time chain with generator } G\,. \end{aligned} \tag{1.28}$$

Proof. Assume finite state space S.For discrete time this follows directly from $P(n) = P^n$. For continuous time we have

$$\begin{aligned} \boldsymbol{\pi}^* G = (0,\ldots,0) \quad &\Rightarrow \quad \boldsymbol{\pi}^* G^k = (0,\ldots,0) \quad \text{for all } k \geq 1 \\ &\Rightarrow \quad \sum_{k=1}^{\infty} \frac{t^k}{k!} \boldsymbol{\pi}^* G^k = (0,\ldots,0) \quad \text{for all } t \geq 0 \\ &\Rightarrow \quad \boldsymbol{\pi}^* \sum_{k=0}^{\infty} \frac{t^k}{k!} G^k = \boldsymbol{\pi}^* \quad \text{for all } t \geq 0 \text{ since } G^0 = Id \\ &\Rightarrow \quad \boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^* \quad \text{for all } t \geq 0\,. \end{aligned} \tag{1.29}$$

The backward direction follows directly from differentiating $\boldsymbol{\pi}^* P(t) = \boldsymbol{\pi}^*$ at $t = 0$. □

Theorem 1.5 (Existence)
A Markov chain with finite state space S has at least one stationary distribution.

Proof. Based on linear algebra (an extended version of the Perron-Frobenius Theorem):
Since P and G have row sum 1 and 0, respectively, we have $P\mathbf{1} = \mathbf{1}$ and $G\mathbf{1} = \mathbf{0}$, where $\mathbf{1}$ is the column vector with all entries 1. So 1 and 0 are eigenvalues of P and G, respectively, and thus there exist also corresponding left eigenvectors. These can be shown to have non-negative entries and can be normalized to be a stationary distribution.

Definition 1.3 State $i \in S$ *communicates with* state $j \in S$ if $p_{ij}(t) > 0$ for some $t \in \mathbb{T}$, and we write $i \to j$. States i and j are *connected* if $i \to j$ and $j \to i$, and we write $i \leftrightarrow j$.
The Markov chain is called *irreducible* if $i \leftrightarrow j$ for all $i, j \in S$.
A state i is called *absorbing*, if $i \not\to j$ for all $j \neq i$.

Remark. The state space of a Markov chain can be decomposed into *communicating classes* S_k which are disjoint subsets of connected states such that $S = \cup_k S_k$. The chain is irreducible if there is only one such class $S_1 = S$.

Theorem 1.6 (Uniqueness)
An irreducible Markov chain has at most one stationary distribution.

Proof. Again based on linear algebra and Perron-Frobenius (see e.g. [GS], Section 6.4). The irreducible structure of the transition matrix P (discrete time) and the generator G (continuous time) can be shown to imply that the eigenvector to eigenvalue 1 and 0, respectively, are unique up to normalization.

Therefore, a finite state, irreducible Markov chain has a unique stationary distribution.

For inifinite state space S the chain can 'disappear at infinity' and there is no stationary distribution (see handout).

Examples. (RW with absorbing BC, on Z)

Definition 1.4 A Markov chain $(X_t : t \in \mathbb{T})$ is called *reverisble* w.r.t. $\boldsymbol{\pi}$ (or $\boldsymbol{\pi}$ is called *reverisble* for $(X_t : t \in \mathbb{T})$) if

$$\pi_i p_{ij}(t) = \pi_j p_{ji}(t) \quad \text{for all } t \in \mathbb{T} \quad \text{and} \quad i \neq j \in S \,. \tag{1.30}$$

Proposition 1.7 *If $\boldsymbol{\pi}$ is reversible then it is also stationary.*

Proof. For all $j \in S$ we have $\quad \big(\boldsymbol{\pi} P(t)\big)_j = \sum_{i \in S} \pi_i p_{ij}(t) = \sum_{i \in S} \pi_j p_{ji}(t) = \pi_j \,.$ □

Proposition 1.8 *$\boldsymbol{\pi}$ is reversible for the Markov chain $(X_t : t \in \mathbb{T})$ if and only if it fulfilles the **detailed balance** conditions*

$$\begin{aligned} &\textit{(discrete time)} && \pi_i p_{ij} = \pi_j p_{ji} \\ &\textit{(continuous time)} && \pi_i g_{ij} = \pi_j g j i \quad \textit{for all } i, j \in S \,. \end{aligned} \tag{1.31}$$

Proof. analogous to the proof of Prop. 1.4.

Proposition 1.9 *Let $\big(X_t : t \in \begin{smallmatrix}\mathbb{Z}\\ \mathbb{R}\end{smallmatrix}\big)$ be a finite state irreducible Markov chain with transition matrix P^X (discrete time) or generator G^X (continuous time). Assume further that the chain is stationary (i.e. $X_t \sim \boldsymbol{\pi}^*$ for all t) which makes it possible to define it also for negative times. Then the **reversed chain** $Y = \big(Y_t : t \in \begin{smallmatrix}\mathbb{Z}\\ \mathbb{R}\end{smallmatrix}\big)$ with $Y_t = X_{-t}$ is a stationary Markov chain with*

$$\begin{aligned} &\textit{transition matrix} && p^Y_{ij} = \frac{\pi^*_j}{\pi^*_i}\, p^X_{ji} && \textit{(discrete time)} \\ &\textit{generator} && g^Y_{ij} = \frac{\pi^*_j}{\pi^*_i}\, g^X_{ji} && \textit{(continuous time)} \quad \textit{for all } i,j \in S\,. \end{aligned} \tag{1.32}$$

Proof. Using stationarity and the Markov property of X we get for discrete time

$$\begin{aligned} &\mathbb{P}\big(Y_{n+1} = i_{n+1} \,\big|\, Y_n = i_n, \ldots, Y_0 = i_0\big) = \\ &\quad = \frac{\mathbb{P}(Y_k = i_k, 0 \le k \le n+1)}{\mathbb{P}(Y_k = i_k, 0 \le k \le n)} = \frac{\mathbb{P}(X_{N-k} = i_k, 0 \le k \le n+1)}{\mathbb{P}(X_{N-k} = i_k, 0 \le k \le n)} = \\ &\quad = \frac{\pi^*_{i_{n+1}} p_{i_{n+1} i_n} \cdots p_{i_1 i_0}}{\pi^*_{i_n} p_{i_n i_{n-1}} \cdots p_{i_1 i_0}} = \frac{\pi^*_{i_{n+1}} p_{i_{n+1} i_n}}{\pi^*_{i_n}} \end{aligned} \tag{1.33}$$

as required. Continuous time works analogously. □

Note that in general a time-reversed Markov chain is not necessarily a Markov chain, this only holds for stationary chains. Obviously, π^* is then also the stationary distribution for the reversed chain Y.

Prop. 1.9 together with (1.31) implies that a reversible Markov chain and its time-reversal are indistinguishable, i.e. $\big(Y_t : t \in \begin{smallmatrix}\mathbb{Z}\\ \mathbb{R}\end{smallmatrix}\big) \sim \big(X_t : t \in \begin{smallmatrix}\mathbb{Z}\\ \mathbb{R}\end{smallmatrix}\big)$ since they have the same transition probabilities

$$p^Y_{ij} = \frac{\pi^*_j}{\pi^*_i}\, p^X_{ji} = \frac{\pi^*_i}{\pi^*_i}\, p^X_{ij} = p^X_{ij}\,. \tag{1.34}$$

An analogous relation holds for rates in continuous time. The detailed balance relations (1.31) can be a useful tool to find stationary distributions. For certain Markov chains 'without loops', i.e. with a 'tree-like' structure of allowed transitions, every stationary distribution is also reversible.

Examples. (RW with PBC, semi-infinite RW)

1.3 Ergodicity

Definition 1.5 Consider a Markov chain $(X_t : t \in \mathbb{T})$. Suppose that $X_0 = i \in S$ and define the *return time* (or *recurrence time*) T_i by

$$\begin{aligned} T_i &:= \min\{n \geq 1 : X_n = i\} \qquad \text{(discrete time)} \\ T_i &:= \inf\{t \geq J_1 : X_t = i\} \qquad \text{(continuous time)} . \end{aligned} \tag{1.35}$$

The *mean recurrence time* of state $i \in S$ is then $\mu_i = \mathbb{E}(T_i)$.

For continuous-time processes the condition of being larger than the first jump time ensures that T_i is non-zero and gives the time of first return to i after an excursion. Note that $T_i = \infty$ can happen with positive probability, e.g. if the chain gets stuck in an absorbing state and does not return to i. If $T_i < \infty$ with probability 1 still $\mu_i = \infty$ is possible, i.e. the distribution of T_i does have a heavy tail with infinite expectation.

Theorem 1.10 *For an irreducible Markov chain with finite state space $\mu_i \in (0, \infty)$ for all $i \in S$, and the unique stationary distribution $\boldsymbol{\pi}^*$ is given by*

$$\begin{aligned} \pi_i^* &= \frac{1}{\mu_i} \qquad\qquad\qquad \textit{(discrete time)} \\ \pi_i^* &= \frac{\mathbb{E}(W_i)}{\mu_i} = \frac{1}{\mu_i |g_{ii}|} \qquad \textit{(continuous time)} , \end{aligned} \tag{1.36}$$

where $W_i \sim Exp(|g_{ii}|)$ is the holding time of state i.

Proof. see e.g. [GS] pp 229 - 230.

In particular, this implies that $\pi_i^* > 0$ for all $i \in S$ and is determined by the average fraction of the time the chain spends in state i. For discrete-time processes the holding time can be thought of being 1 (one time step).

Picture of sample path.

Definition 1.6 A discrete time Markov chain is called *aperiodic* if for all $i \in S$, $p_{ii}(n)$ is eventually positive, i.e.

$$\text{there exists } N_i \in \mathbb{N} \text{ such that } p_{ii}(n) > 0 \text{ for all } n \geq N_i . \tag{1.37}$$

Remark. Note that for irreducible Markov chains, aperiodicity implies that for all $i, j \in S$, $p_{ij}(n)$ is eventually positive.

Example. (RW)

Note that for continuous time there is no issue of periodicity, since

$$\text{if } i \to j \quad \text{then} \quad p_{ij}(t) > 0 \quad \text{for all } t > 0\,. \tag{1.38}$$

This is because $i \to j$ is equivalent to

$$g_{ii_1} g_{i_1 i_2} \cdots g_{i_{n-1} j} > 0 \quad \text{for some } i_1, \ldots i_{n-1} \in S, n \in \mathbb{N}\,, \tag{1.39}$$

which implies that $\qquad p_{ij}(t) \geq p_{ii_1}(t/n) \cdots p_{i_{n-1} j}(t/n) > 0\,.$

Theorem 1.11 *An irreducible (aperiodic) Markov chain with finite state space is* ***ergodic****, i.e. it has a unique stationary distribution* $\boldsymbol{\pi}^*$ *and*

$$p_{ij}(t) = \mathbb{P}(X_t = j \,|\, X_0 = i) \to \pi^*_j \quad \textit{as } t \to \infty\,, \quad \textit{for all } i, j \in S\,. \tag{1.40}$$

Proof. For discrete time this follows from the **Perron-Frobenius Theorem**:
If P is the transition matrix of a finite state, aperiodic, irreducible Markov chain then
(i) $\lambda_1 = 1$ is a single eigenvalue of P
(ii) and the remaining (complex) eigenvalues $\lambda_2, \ldots, \lambda_{|S|}$ satisfy $|\lambda_j| < 1$.
Note that (i) includes uniqueness of the stationary distribution claimed in Theorem 1.6.
Suppose further that all the eigenvalues are distinct, then P can be diagonalized, i.e.

$$B\,P\,B^{-1} = \Lambda = \begin{pmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_{|S|} \end{pmatrix} \tag{1.41}$$

where the rows of B are the left and the columns of B^{-1} are the right eigenvectors of P, normalized such that $B\,B^{-1} = Id$. Thus

$$P^n = \left(B^{-1}\Lambda B\right)^n = B^{-1} \begin{pmatrix} \lambda_1^n & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_{|S|}^n \end{pmatrix} B \to B^{-1} \begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{pmatrix} B \tag{1.42}$$

as $n \to \infty$, since $\lambda_1 = 1$ and $|\lambda_i| < 1$ for all $i > 1$. Since the first column of B^{-1} is $\mathbf{1}$ (right eigenvector to $\lambda_1 = 1$), the right-hand side is equal to the matrix

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix} B = \begin{pmatrix} \pi^*_1 \dots \pi^*_{|S|} \\ \vdots \\ \pi^*_1 \dots \pi^*_{|S|} \end{pmatrix} \quad \text{which implies the statement.} \tag{1.43}$$

The proof can be extended to more general cases and works similar for continuous time by using the corresponding jump chain. □

Remark. Theorem 1.11 implies that for every initial distribution $\boldsymbol{\pi}(0)$,

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, P(t) \to \boldsymbol{\pi}^* \quad \text{as } t \to \infty\,. \tag{1.44}$$

In general, the distribution at discrete time n is the solution to a linear first order equation. Therefore it is given by a linear combination of left eigenvectors $\mathbf{v}_i$ of the transition matrix P the form

$$\boldsymbol{\pi}(n) = A_1 \mathbf{v}_1 \lambda_1^n + \ldots + A_{|S|} \mathbf{v}_{|S|} \lambda_{|S|}^n \tag{1.45}$$

with $A_1 = 1$ and $\mathbf{v}_1 = \boldsymbol{\pi}^*$. As $n \to \infty$ all contributions with $|\lambda_i| < 1$ decay, and with $\lambda_1 = 1$ the limit is $\boldsymbol{\pi}^*$. Analogously, for continuous time

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}^* e^{t\lambda_1} + \ldots + A_{|S|} \mathbf{v}_{|S|} e^{t\lambda_{|S|}}\,, \tag{1.46}$$

where $\lambda_i \in \mathbb{C}$ are the eigenvalues of the generator G with $\lambda_1 = 0$ and $\mathrm{Re}\lambda_i < 0$ for $i > 1$.

Example. (RW on finite interval)

Theorem 1.12 (Ergodic Theorem)
Let $X = (X_t, t \in \mathbb{T})$ be an ergodic Markov chain with unique stationary distribution π^*. Then for every observable $f : S \to \mathbb{R}$

$$\frac{1}{t}\int_0^t f(X_s)\, ds \quad \textit{or} \quad \frac{1}{N}\sum_{n=0}^{N} f(X_n) \quad \longrightarrow \quad \mathbb{E}_{\pi^*}(f) \quad \textit{as } t, N \to \infty\,. \tag{1.47}$$

Proof. see e.g. [GS], chapter 9.5

So stationary expectations can be approximated by time averages over long periods. This is the basis for Markov chain Monte Carlo (MCMC) which is used to sample from the stationary

distribution of a chain. In particular, using indicator functions $f(X_t) = \delta_{X_t,i}$ the right-hand side of (1.47) is equal to π_i^*. To improve the speed of convergence in practice, the chain is run for a while before starting to sample (equilibration).

Further remarks on periodicity (non-examinable)
The *period* $d(i)$ of a state $i \in S$ is defined as

$$d(i) := gcd\big\{t \geq 1 \,:\, p_{ii}(t) > 0\big\}\,, \tag{1.48}$$

the greatest common divisor of the epochs at which return is possible.
For an irreducible Markov chain all states have the same period (for a proof see p.224 in [GS]). If the chain is aperiodic we have $d(i) = 1$ for all $i \in S$. Note, however, that the requirement $d = 1$ is weaker than Def. 1.6 and is often used as an alternative definition of aperiodicity.
Ergodicity follows again from a more general version of the **Perron-Frobenius Theorem**:
If P is the transition matrix of a finite state irreducible Markov chain with period d then
(i) the d complex roots of unity are eigenvalues of P,

$$\lambda_1 = w^0 = 1,\ \lambda_2 = w^1, \ldots, \lambda_d = w^{d-1} \quad \textit{where} \quad w = e^{2\pi i/d}\,, \tag{1.49}$$

(ii) and the remaining eigenvalues $\lambda_{d+1}, \ldots, \lambda_{|S|}$ satisfy $|\lambda_j| < 1$.

1.4 Countably infinite state spaces

For infinite state space S, the Markov chain can 'get lost at infinity', and therefore not have a stationary probability distribution.
Let $X_0 = i$ and T_i be the time of first return to state i as defined in Def. 1.5 for continuous and discrete time.

Definition 1.7 A state $i \in S$ is called

$$\begin{aligned}
\textit{transient}, \text{ if} \quad & \mathbb{P}(T_i = \infty) > 0\,, \\
\textit{null recurrent}, \text{ if} \quad & \mathbb{P}(T_i < \infty) = 1 \quad \text{and} \quad \mathbb{E}(T_i) = \infty\,, \\
\textit{positiv recurrent}, \text{ if} \quad & \mathbb{P}(T_i < \infty) = 1 \quad \text{and} \quad \mathbb{E}(T_i) < \infty\,.
\end{aligned} \tag{1.50}$$

Theorem 1.13 *Let X be an irreducible Markov chain. Then all states are either transient, null recurrent or positive recurrent. X has a unique stationary distribution if and only if it is positive recurrent. In this case X is also ergodic (subject to aperiodicity in case of discrete time).*

Proof. see Section 6.2 in [GS]

Examples. (BD process)

So the positive recurrent Markov chains behave like chains with finite state space concerning their stationary distributions and long time behaviour. Transient continuous time chains can get lost at infinity even in finite time. This phenomenon is called *explosion*. Define the *explosion time*

$$J_\infty := \lim_{n\to\infty} J_n = \sum_{i=1}^{\infty} W_i \in (0,\infty] \,. \tag{1.51}$$

This is a random variable that usually takes the value ∞, and we say that the chain is *non-explosive* if $\mathbb{P}(J_\infty = \infty) = 1$. For example this is the case if $|S| < \infty$ or $\sup_{i\in S} |g_{ii}| < \infty$.

Example. (birth chain)

Theorem 1.14 Polya's Theorem.
The simple random walk (in discrete or continuous time) on $S = \mathbb{Z}^d$ is null recurrent for $d = 1, 2$ and transient for $d \geq 3$.

Proof. The simplest proof follows an interesting connection between Markov chains and electrical network theory. Consider a continuous time SRW on $\mathbb{Z}^d$ with rate q across each bond (discrete time works analogously). The *resistence* of a bond is then given by $r = 1/q$, and *Kirchhoff's laws* apply, i.e. the total resistence r is given by

$$r = r_1 + r_2 \qquad \text{for resistances in series (sums of waiting times)} \tag{1.52}$$

$$r = \big(1/r_1 + 1/r_2\big)^{-1} \qquad \text{for resistences in parallel (sums of rates) .} \tag{1.53}$$

Denote by $R(n)$ the total resistence between the origin and the set $\big\{x \in \mathbb{Z}^d : \|x\|_1 = n\big\}$, and by $R = \lim_{n\to\infty} R(N)$ the resistence to infinity. It can be shown that the SRW is transient if and only if $R < \infty$, and a rough estimates gives

$$R(n) \approx \big(n^{d-1}\,\tfrac{1}{n}\big)^{-1} = n^{d-2} \tag{1.54}$$

counting n^{d-1} parallel paths of length n to reach distance n from the origin. This implies $R = \infty$ for $d = 1$ and $R = 0$ for $d \geq 3$, for $d = 2$ a more careful analysis reveals logarithmic corrections that lead also to $R = \infty$.
For more details and infinite expected return times see e.g. [G]. □

2 Stochastic particle systems

2.1 Basic examples

Stochastic particle systems, or often also called interacting particle systems (IPS), are Markov chains where the state space has a particular structure. Let $\Lambda = \{1, \dots, L\}$ be a finite set of L points, which we call *lattice*. The state space is given by the set of all configurations

$$\eta = \big(\eta(x) : x \in \Lambda\big) \in S = \{0,1\}^L \quad \big(\text{often also written as } \{0,1\}^\Lambda\big)\big) \,, \tag{2.1}$$

where $\eta(x) = 1$ is interpreted as the presence of a particle or an infection at site x. The dynamics of the IPS we consider is given by local continuous-time transitions of the following two kinds:

$$\begin{aligned} \eta &\to \eta^x \quad \text{with rate} \quad c(\eta, \eta^x) \,, \\ \eta &\to \eta^{xy} \quad \text{with rate} \quad c(\eta, \eta^{xy}) \end{aligned} \tag{2.2}$$

for all $x, y \in \Lambda$, where we use the shorthand notation

$$\eta^x(z) = \begin{cases} \eta(z) & , \, z \neq x \\ 1 - \eta(x) & , \, z = x \end{cases} \qquad \text{and} \qquad \eta^{xy}(z) = \begin{cases} \eta(z) & , \, z \neq x, y \\ \eta(y) & , \, z = x \\ \eta(x) & , \, z = y \end{cases} \,, \tag{2.3}$$

so that η^x corresponds to *creation/annihilation* of a particle at site x and η^{xy} to *motion* of a particle between x and y. In the following $q_{x,y} \geq 0$ will be transition rates of a continuous time random walk on Λ, and (Λ, Q) can be interpreted as a graph with adjacency matrix $Q = (q_{x,y} : x, y \in \Lambda)$. To avoid degeneracies we will always assume that

$$\textbf{the walk on } \Lambda \textbf{ with rates } q_{x,y} \textbf{ is irreducible} \,, \tag{2.4}$$

so that particles/infections can reach all parts of the lattice.

Definition 2.1 Let $q_{x,y} \in \{0, \lambda\}$ for all $x, y \in \Lambda$ with *infection rate* $\lambda > 0$, and write $x \sim y$ if $q_{x,y} = \lambda$. The *contact process* (CP) is an IPS with rates

$$c(\eta, \eta^x) = \eta(x) + \lambda\big(1 - \eta(x)\big) \sum_{y \sim x} \eta(y) \quad \text{for all } x \in \Lambda \,. \tag{2.5}$$

Sites with $\eta(x) = 1$ are interpreted as being infected and recover independently at rate 1, and healthy sites with $\eta(x) = 0$ get infected by each of their neighbours independently at rate λ. A short way of writing this is

$$1 \xrightarrow{1} 0 \quad \text{and} \quad 0 \xrightarrow{\lambda \# \mathrm{IN}} 1 \,. \tag{2.6}$$

Properties.

- The CP has one absorbing state, $\eta = \underline{0} = (0, \ldots, 0)$ (all healthy), and is not irreducible. For finite Λ $\underline{0}$ can be reached from all $\eta \in S$ and therefore $\boldsymbol{\pi}^* = \delta_{\underline{0}}$ is the unique stationary distribution and

  $$\boldsymbol{\pi}(t) \xrightarrow{t \to \infty} \delta_{\underline{0}} \quad \text{for all } \lambda \geq 0 \text{ and all initial conditions } \boldsymbol{\pi}(0) \,. \tag{2.7}$$

  Therefore, the CP on finite lattices is ergodic and the infection eventually dies out.

- For infinite Λ, it can be shown that there exists another stationary distribution $\boldsymbol{\pi}^*$ with $\rho = \mathbb{E}_{\boldsymbol{\pi}^*}\big(\eta(x)\big) > 0$ where the infection persists, provided that the infection rate λ is greater than a critical value λ_c. The loss of ergodicity in infinite system limits is also called *ergodicity breaking*. Depending on the lattice structure there can be further different cases, for details see e.g. [G].

Graphical construction.

Definition 2.2 The *(linear) voter model* (VM) is an IPS with rates

$$c(\eta, \eta^x) = \sum_{y \sim x} q_{x,y} \Big(\eta(x)\big(1-\eta(y)\big) + \big(1-\eta(x)\big)\eta(y) \Big) \quad \text{for all } x \in \Lambda \,. \tag{2.8}$$

$\eta(x) = 0, 1$ are interpreted as two opinions, and x adopts the opinion of y independently at rate $q_{x,y}$, in short

$$1_x 0_y \overset{q_{x,y}}{\longrightarrow} 0_x 0_y \quad \text{and} \quad 0_x 1_y \overset{q_{x,y}}{\longrightarrow} 1_x 1_y \,. \tag{2.9}$$

Properties.

- The linear VM is symmetric under relabelling opinions $0 \leftrightarrow 1$ and is *dual* to a system of coalescing random walkers, which can be seen from the graphical construction in reversed time (see below).
- The VM is not ergodic and has two absorbing states $\eta = \underline{0}, \underline{1}$ (all of the same opinion). It is therefore also not ergodic (even on finite lattices), and

  $$\boldsymbol{\pi}(t) \overset{t\to\infty}{\longrightarrow} \alpha\delta_{\underline{0}} + (1-\alpha)\delta_{\underline{1}} \quad \text{where} \quad \alpha \in [0,1] \quad \text{depends on the initial condition.} \tag{2.10}$$

  Eventually only one opinion will survive, and the r.h.s. denotes all stationary distributions of the finite process.
- For inifinite Λ there may be other stationary distributions $\boldsymbol{\pi}^*$ with $\rho = \mathbb{E}_{\boldsymbol{\pi}^*}\big(\eta(x)\big) \in (0,1)$ where both opinions persist.
- There are many generalizations of the linear VM, including non-linear majority rules or models with more than two opinions such as the Axelrod model.

Graphical construction.

Definition 2.3 The *exclusion process* (EP) is an IPS with rates

$$c(\eta, \eta^{xy}) = q_{x,y}\eta(x)\big(1-\eta(y)\big) \quad \text{for all } x, y \in \Lambda\ . \tag{2.11}$$

So particles ($\eta(x) = 1$) jump to empty sites ($\eta(y) = 0$) independently with rate $q_{x,y}$, in short

$$1_x 0_y \xrightarrow{q_{x,y}} 0_x 1_y\ . \tag{2.12}$$

For $\Lambda \subseteq \mathbb{Z}^d$ and $q_{x,y} > 0$ only for nearest neighbours $x \sim y$ the EP is called *simple* (SEP). If in addition $q_{x,y} = q_{y,x}$ for all $x, y \in \Lambda$ it is called *symmetric* (SSEP), otherwise *asymmetric* (ASEP). If $d = 1$ and $q_{x,y} = q\delta_{x+1,y}$ it is called *totally asymmetric* (TASEP).

Properties.

- The EP conserves the number of particles and is therefore not irreducible on the state space $S = \{0,1\}^\Lambda$. On the subspaces

  $$S_N = \Big\{\eta : \sum_{x\in\Lambda} \eta(x) = N\Big\} \quad \text{for } N = 0, \dots, L \tag{2.13}$$

  the EP is irreducible and ergodic with a unique stationary distribution $\boldsymbol{\pi}^*_{L,N}$.

- Of particular importance for systems with a conserved quantity is the corresponding stationary current

  $$j_{x,y} := \mathbb{E}_{\boldsymbol{\pi}^*_{L,N}}\big(c(\eta,\eta^{xy})\big) = q_{x,y}\,\mathbb{E}_{\boldsymbol{\pi}^*_{L,N}}\big(\eta(x)(1-\eta(y))\big)\ . \tag{2.14}$$

  For the spatially homogeneous SEP with $q_{x,y} = q\delta_{x\sim y}$ the stationary distributions factorize for large system size and this simplifies to

  $$j_{x,y} \equiv j(\rho) = q\rho(1-\rho) \quad \text{with particle density} \quad \rho = N/L \in [0,1]\ . \tag{2.15}$$

- The ASEP is one of the most studied so-called *driven diffusive systems* in non-equilibrium statistical mechanics and has various connections to many other models, such as surface growth, directed percolation or random polymer models.

Graphical construction.

General properties of the time evolution of IPS.
Let us focus on IPS with flip dynamics such as CP and VM, transport systems like the EP can be treated analogously. Due to the local dynamics, the master equation of an IPS with state space $S = \{0,1\}^\Lambda$ can be written as

$$\frac{d}{dt}\pi_\eta(t) = \sum_{x\in\Lambda} \Big(\pi_{\eta^x}(t)\, c(\eta^x,\eta) - \pi_\eta(t)\, c(\eta,\eta^x) \Big) \quad \text{for all } \eta \in S\,. \tag{2.16}$$

This is the coordinate form of the general vector equation $\frac{d}{dt}\boldsymbol{\pi}(t) = \boldsymbol{\pi}(t)G$, and the r.h.s. provides a compact notation for the generator G with intuitive gain and loss terms. Due to the fact that only one site can flip at a time, most of the entries of G are 0.

To get an equation for the time evolution of an observable $f : S \to \mathbb{R}$, we note that

$$\mathbb{E}_{\boldsymbol{\pi}(t)}(f) = \sum_{\eta\in S} \pi_\eta(t)\, f(\eta) = \boldsymbol{\pi}(t) \circ f \tag{2.17}$$

can be written as a scalar product. Using the master equation we get

$$\frac{d}{dt}\mathbb{E}_{\boldsymbol{\pi}(t)}(f) = \big(\boldsymbol{\pi}(t)G\big) \circ f = \boldsymbol{\pi}(t) \circ (G\, f) = \mathbb{E}_{\boldsymbol{\pi}(t)}(G\, f)\,. \tag{2.18}$$

When acting on the observable f (column vector) rather than the distribution $\boldsymbol{\pi}(t)$ (row vector) one often also writes $\mathcal{L}$ for the generator of the process as an operator on observables. The action on f leads to another function $\mathcal{L}f$ (column vector) and takes a particularly simple form:

$$(\mathcal{L}f)(\eta) = \sum_{x\in\Lambda} c(\eta,\eta^x)\Big(f(\eta^x) - f(\eta)\Big)\,, \tag{2.19}$$

which can be interpreted as a discrete derivative of the function f under the dynamics of the process. There is a fully developed mathematical theory of generators and corresponding semigroups for stochastic particle systems, see [G] and references therein for details.

Example. (number of infections for CP)

For a general continuous-time Markov chain on a state space S with jump rates g_{ij} as introduced in Section 1, the generator is given by

$$\mathcal{L}f(i) = \sum_{j \in S} g_{ij}\big(f(j) - f(i)\big) \,, \quad i \in S \,, \tag{2.20}$$

for all functions $f : S \to \mathbb{R}$. Like the master equation, this also has an intuitive interpretation as a 'discrete derivative' describing the change of f under all possible jumps out of state i weighted by the jump rates.

2.2 The Ising model and Markov chain Monte Carlo

The **Ising model** is a very basic model for magnetic behaviour from Statistical Mechanics. The state space is given by $S = \{-1, 1\}^\Lambda$ and configurations are denoted by $\sigma = (\sigma_x : x \in \Lambda)$, consisting of *spin variables* $\sigma_x = \pm 1$ interacting on a lattice (or graph) Λ. The interaction is defined by an *energy function* (or *Hamiltonian*) $H : S \to \mathbb{R}$, and the equilibrium distribution of the spins is given by

$$\pi(\sigma) = \frac{1}{Z}\, e^{-\beta H(\sigma)} \quad \text{where} \quad Z = \sum_{\sigma \in S} e^{-\beta H(\sigma)} \tag{2.21}$$

is the normalizing constant called *partition function*. $\beta \geq 0$ is interpreted of the inverse *temperature* in the system, and for

$$\begin{aligned} \beta \to \infty \quad & \text{very low temperature} \quad \Rightarrow \quad \pi \text{ concentrates on lowest energy configurations} \,, \\ \beta \to 0 \quad & \text{very high temperature} \quad \Rightarrow \quad \pi \text{ becomes uniform (energy irrelevant)} \,. \end{aligned}$$

The distribution (2.21) is also called the *Gibbs distribution*, and can be postulated from general principles of equipartition of energy in a physical system in equilibrium without referring to any dynamics.

The Hamiltonian for the Ising model is given by

$$H(\sigma) = -\sum_{x,y \in \Lambda} J_{xy} \sigma(x)\, \sigma(y) \tag{2.22}$$

where the *coupling constants* J_{xy} determine the interaction of the spins on the graph Λ. The classical situation is to consider a regular lattice $\Lambda \subseteq \mathbb{Z}^d$ with nearest neighbour interaction

$$J_{xy} = \begin{cases} 1 \,, & x \sim y \\ 0 \,, & \text{otherwise} \end{cases} \,. \tag{2.23}$$

The goal is now to compute expected values w.r.t. the Gibbs distribution (2.21) for the Ising model, which is not a simple task since the $\sigma(x)$ are correlated due to the interaction in the Hamiltonian, and the the size of the state space $|S| = 2^L$ is huge even for moderate lattice sizes $L = 100$. The trick is to invent a Markov chain $(\sigma_t : t \geq 0)$ (often also done in discrete time) with stationary distribution π, and use the Ergodic Theorem to sample from π. This method is called **Markov chain Monte Carlo**, and the conditions on $(\sigma_t : t \geq 0)$ usually imposed are the following:

- it should be ergodic, i.e. irreducible on S (which is large, but finite)
- it should be reversible w.r.t. π, i.e. the rates $c(\sigma, \sigma')$ fulfill the detailed balance relations

$$c(\sigma, \sigma')e^{-\beta H(\sigma)} = c(\sigma', \sigma)e^{-\beta H(\sigma')} \quad \text{for all } \sigma, \sigma' \in S\,. \tag{2.24}$$

  In fact, stationarity would be enough, but reversibility is easier to implement via detailed balance.

To ensure both conditions, one usually restricts to local (*spin flip*) dyanmics analogous to stochastic particle systems, where only single spins are flipped $\sigma \to \sigma^x$ with rates $c(\sigma, \sigma^x)$. The most basic choices are

- the **heat bath algorithm**, with

$$c(\sigma, \sigma^x) = \frac{e^{\beta H(\sigma)}}{e^{\beta H(\sigma)} + e^{\beta H(\sigma^x)}} \in (0, 1)\,, \tag{2.25}$$

  where the system is sampled at rate 1 and proposed flips are accepted with probability $c(\sigma, \sigma^x)$. This can be motivated on physical grounds (coupling to a heat bath), but has the disadvantage that the sampling rate is higher than necessary and acceptance probabilities are typically < 1.

- the **Metropolis algorithm**, with

$$c(\sigma, \sigma^x) = \begin{cases} e^{-\beta(H(\sigma^x) - H(\sigma))} & \text{, if } H(\sigma^x) > H(\sigma) \\ 1 & \text{, if } H(\sigma^x) \le H(\sigma) \end{cases} \in (0, 1]\,, \tag{2.26}$$

  i.e. the system is sampled at rate 1 and whenever a proposed flip does not increase the energy it is accepted with probability 1, otherwise with probability < 1. Therefore, implementations of this algorithm are in general faster than the heat bath algorithm.

There are more involved non-local dynamics where whole clusters of spins are flipped (e.g. *Swendsen-Wang algorithm*) which are much faster than either of the above when the system is close to a phase transition.

The phase transition observed in the Ising model is that for dimensions $d \ge 2$ there exists a finite *critical temperature* $T_c = 1/\beta_c \in (0, \infty)$, such that for $\beta > \beta_c$ (small temperature) the system exhibits a *spontaneous magnetization* $m^* \neq 0$ in the limit $L \to \infty$. Precisely, there exist two limiting Gibbs measures π^+ and π^-, such that $\mathbb{E}_{\pi^\pm}(\sigma(x)) = \pm m_* \neq 0$, and the limit of π will be given by a mixture of those (which depends in general on the boundary conditions). On the other hand, for $\beta < \beta_c$ (high temperature) there exists only one limit π^0 with $\mathbb{E}_{\pi^0}(\sigma(x)) = 0$ (uniqueness of the Gibbs measure).

Picture.

3 Processes with continuous state space

3.1 Brownian motion and the Central limit theorem

Let $Y_1, Y_2, \ldots \in \mathbb{R}$ be iidrvs with mean $\mathbb{E}(Y_i) = 0$ and variance $\mathrm{var}(Y_i) = \sigma^2 > 0$. Then define the discrete-time process

$$X_n := \sum_{i=1}^{n} Y_i \quad \text{with} \quad X_0 = 0 \,. \tag{3.1}$$

For example if $Y_i \sim U\big(\{-1, 1\}\big)$ then X is a simple symmetric random walk. Then by the **Central Limit Theorem** (CLT) as $n \to \infty$

$$\frac{X_n}{\sqrt{n}} \to \xi \sim N(0, \sigma^2) \qquad \text{(Gaussian rv with mean 0 and variance } \sigma^2) \,, \tag{3.2}$$

or, equivalently, for all $y \in \mathbb{R}$

$$\mathbb{P}\Big(\frac{X_n}{\sqrt{n}} \le y\Big) \to \int_{-\infty}^{y} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-x^2/(2\sigma^2)}}_{f_X(x)}\, dx \quad \text{with Gaussian pdf} \quad f_X(x) \,. \tag{3.3}$$

We can use the CLT to look at the process X_n in rescaled time $t = n\Delta t$ and space $b = x\Delta b$. According to CLT (3.2), with $\Delta t \sim 1/n$ we should choose $\Delta b \sim 1/\sqrt{n} = \sqrt{\Delta t}$ so that $\Delta b\, X_{[t/\Delta t]}$ converges to a t-dependent random variable as $\Delta t \to 0$. We define

$$B_t := \lim_{\Delta t \to 0} \sqrt{\Delta t}\, X_{[t/\Delta t]} = \lim_{\Delta t \to 0} \frac{\sqrt{t}}{\sqrt{t/\Delta t}} \sum_{i=1}^{[t/\Delta t]} Y_i = \sqrt{t}\, \xi_t \sim N(0, t\sigma^2) \,. \tag{3.4}$$

Here the $\xi_t \sim N(0,1)$ are different for each t, but they are certainly not independent. The relation between time and space rescaling which follows from the CLT can be summarized as

$$\boxed{t = \Delta t\, n \,, \quad b = (\Delta b)x \quad \text{with} \quad \Delta_b = \Delta_t^{\alpha} \,, \quad \alpha = 1/2} \,. \tag{3.5}$$

On all other spatial scales, the limiting process does either not exist or is degenerate,

$$B_t = 0 \quad \text{for } \alpha > 1/2 \,, \quad B_t \text{ is not well defined for } \alpha < 1/2 \,. \tag{3.6}$$

Distributional properties of the process $B = (B_t : t \geq 0)$.

$$B_0 = 0\,, \quad B_t \sim N(0, t\sigma^2) \quad \text{and analogously to (3.4)}$$
$$B_t - B_s = \lim_{\Delta t \to 0} \sqrt{\Delta t} \sum_{i=[s/\Delta t]}^{[t/\Delta t]} Y_i \sim N(0, (t-s)\sigma^2) \tag{3.7}$$

for all $t \geq s \geq 0$. So B has *stationary increments*, i.e. $B_t - B_s \sim B_{t-s} - B_0$, and by independence of the Y_i, B has *independent increments*, i.e.

$$B_t - B_s \quad \text{is independent of} \quad \{B_u : u \leq s\} \quad \text{for all } t \geq s \geq 0\,. \tag{3.8}$$

Regularity properties of a sample path.
From (3.4) we expect for Brownian motion

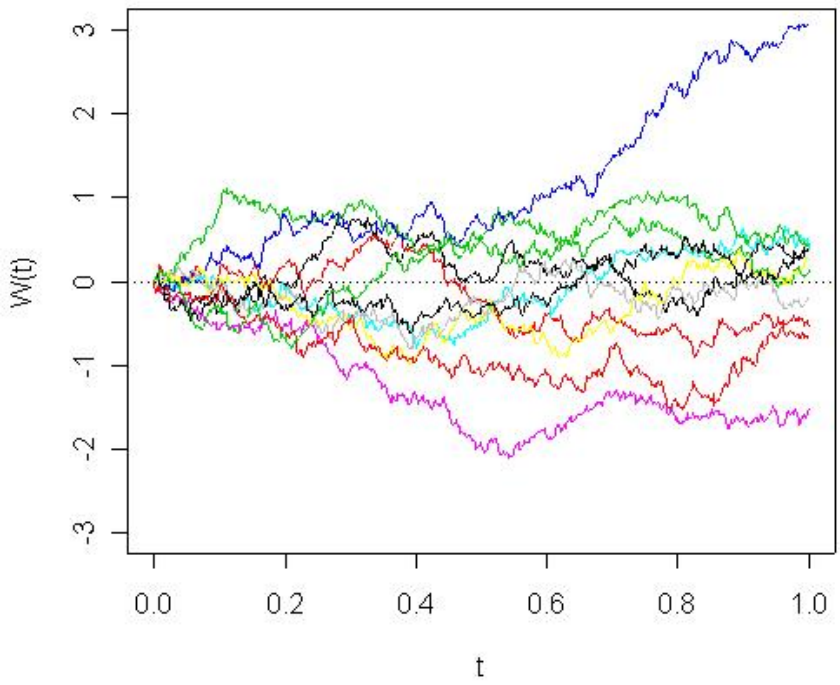
$$B_{t+h} - B_t = \sqrt{h}\,\xi \sim N(0, h\sigma^2) \to 0 \quad \text{as } h \to 0\,. \tag{3.9}$$

Therefore Brownian sample paths are continuous (and more precisely, Hölder continuous with index $1/2$). But they are nowhere differentiable, since

$$\frac{B_{t+h} - B_t}{h} = \frac{\sigma}{\sqrt{h}}\,\xi \quad \text{has no limit as } h \to 0\,. \tag{3.10}$$

These properties do not follow from (3.7) and (3.8), which can be fulfilled also by discontinuous processes. But under the restriction that $t \mapsto B_t(\omega)$ is a continous function of t for all $\omega \in \Omega$, BM can be shown to be a well defined stochastic process (see end of this section for further remarks).

Examples of sample paths.



Definition 3.1 A real-valued process $B = (B_t : t \geq 0)$ with continuous paths and stationary independent increments, such that $B_t \sim N(0,t)$ for all $t \geq 0$, is called a *standard Brownian motion* (BM) or *Wiener process*.

It suffices to look at standard BMs B with $\sigma^2 = 1$ and $B_0 = 0$, then $\sigma B + x_0$ is a BM with variance σ^2 starting in x_0. Note also that if the increments Y_i in (3.2) are not identically distributed or weakly dependent, the CLT still holds under more general conditions (see e.g. Gardiner, Section 2.8.2). So Brownian motion is the natural scaling limit for a very general class of models.

Definition 3.2 A *d-dimensional standard Brownian motion* $\mathbf{B} = (\mathbf{B}_t : t \geq 0)$ is a collection of d independent one-dimensional BMs $B^1, \ldots, B^d$ as given in Def. 3.1, i.e.

$$\mathbf{B}_t = (B_t^1, \ldots, B_t^d) \quad \text{for all } t \geq 0\,. \tag{3.11}$$

So the pdf of the increments $\mathbf{B}_t - \mathbf{B}_s$ is $\quad f_{t-s}(\mathbf{x}) = (2\pi(t-s))^{-d/2} \exp\Big(-\frac{\|\mathbf{x}\|_2^2}{2(t-s)}\Big)\,.$

3.2 General properties

This chapter is about processes with continuous state space $S = \mathbb{R}$ or $\mathbb{R}^d$ and continuous time $\mathbb{T} = [0, \infty)$. This is mathematically more complicated than Markov chains, and we will discuss some of the technical issues below. On the other hand, the sample paths are now real valued functions, our state space has an analytic structure and we will be able to use concepts from usual calculus.

For example we will often integrate over sets $A \in \mathbb{R}$ of possible values with respect to the distribution function $F(x) = \mathbb{P}(X \leq x)$ of a random variable X, e.g.

$$\mathbb{P}(X \in A) = \int_A dF(x) = \int_A f(x)\,dx \quad \text{where } f = F' \text{ is the pdf (if it exists)}\,. \tag{3.12}$$

Technical side remark: This cannot be done for all sets $A \subseteq \mathbb{R}$ but only for $A \in \mathcal{A}$, where $\mathcal{A} \subsetneq \mathcal{P}(\mathbb{R})$ is a so-called *σ-algebra*. This is a set of *measurable sets* where the measure $dF(x)$ can be consistently defined on.

Characterization of processes on $\mathbb{R}$.
As for Markov chains, the **distributional properties** of a general stochastic process are determined by fixing all *finite-dimensional distributions (fdds)*

$$F_{t_1,..,t_n}(x_1, \ldots, x_n) = \mathbb{P}\big(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n\big)\,, \tag{3.13}$$

for all $t_1, \ldots, t_n \in [0, \infty)$, $t_i \neq t_j$, $x_1, \ldots, x_n \in \mathbb{R}$ and $n \in \mathbb{N}$. We focus here on the state space $S = \mathbb{R}$ where the fdds are given by joint distribution functions F as above, in principle this can be extended to more general state spaces.

In contrast to Markov chains, for continuous state space the fdds do not determine the process uniquely. Two processes with the same fdds are called *versions* of each other, and their **sample paths** can have very different properties. This fact cannot be ignored, since it is very important when studying properties such as first-passage times (first random time to enter a given set). One is usually interested in the most regular version of the process (in order to avoid complications) and there are basically two classes of processes that are usually considered.

- *Diffusion processes* (see Section 2.4) have continuous sample paths, i.e. $t \mapsto X_t(\omega) \in C\big([0,\infty),\mathbb{R}\big)$, and the most basic example is Brownian motion.
- More generally, if one wants to allow for discontinuities in the sample paths (such as Lévy processes or as a special case also continuous time Markov chains), one restricts to the following class of paths:

Definition 3.3 A real-valued, continuous-time process X is called *càdlàg* if its sample paths are right continuous (continue à droite) and have left limits (limite à gauche), i.e.

$$\lim_{s\searrow t} X_s(\omega) = X_t(\omega) \quad \text{and} \quad \lim_{s\nearrow t} X_s(\omega) \text{ exists}\,, \quad \text{for all } \omega\in\Omega,\ t\in[0,\infty)\,. \tag{3.14}$$

Example. (Lévy processes)

Description of the dynamics.
The transition probabilities of a Markov chain can also be generalized.

Definition 3.4 Let X be a stochastic process on $\mathbb{R}$. The conditional distribution function

$$F(t,x|s,y) = \mathbb{P}(X_t \le x | X_s = y)\,, \tag{3.15}$$

is called the *transition kernel* of X. If it has a density we call this the *transition density*,

$$f(t,x|s,y) = \frac{\partial F}{\partial x}(t,x|s,y)\,. \tag{3.16}$$

Note that for a homogeneous process, the kernel is actually only a function of $t-s$.

Proposition 3.1 *The fdds of a Markov process are uniquely determined by the transition kernels and the initial distribution.*

Proof. Sample calculation for $0\le t_1\le t_2$ with densities using the Markov property,

$$\mathbb{P}\big(X_{t_1}\le x_1, X_{t_2}\le x_2\big) = \int_{-\infty}^{x_2}\int_{-\infty}^{x_1}\int_{-\infty}^{\infty} f(0,x)\, f(t_1,y|0,x)\, f(t_2,z|t_1,y)\,dx\,dy\,dz\,.$$

□

Example.

Proposition 3.2 *Let B be a standard BM. For all $t_1, \ldots, t_n$, $n \in \mathbb{N}$*

$$(B_{t_1}, \ldots, B_{t_n}) \sim N(\mathbf{0}, \Sigma) \quad \textit{with} \quad \sigma_{ij} = \min\{t_i, t_j\} \,, \tag{3.17}$$

i.e. the fdds for BM are multivariate Gaussian with zero mean and covariance matrix $\Gamma = (\sigma_{ij})_{i,j}$. In general, processes X with Gaussian fdds are called Gaussian processes, and are characterized by their mean $m(t) = E(X_t)$ and covariance function $\sigma(s,t) = \mathrm{cov}(X_s, X_t)$.
Conversely, if X is a Gaussian process with covariance function $\sigma(s,t) = \min\{s,t\}$, then X is a standard BM.

See Hand-out 3 for details on the PDF for multivariate Gaussians.

Proof. For standard BM $B_t \sim N(0,t)$ and it suffices to show that $\mathrm{cov}(B_s, B_t) = \min\{s,t\}$. Take $s < t$, then

$$\mathbb{E}(B_s B_t) = \mathbb{E}\big(B_s^2 + B_s(B_t - B_s)\big) = \mathbb{E}(B_s^2) + 0 \,, \tag{3.18}$$

since B has independent increments and $\mathbb{E}(B_s) = 0$. Thus $\mathrm{cov}(B_s, B_t) = \mathrm{var}(B_s) = s$. On the other hand, for a Gaussian process X with covariances $\sigma(s,t) = \min\{s,t\}$ we have for $t > s$ and all $u \le s$

$$\mathbb{E}\big((X_t - X_s)X_u\big) = \mathbb{E}(X_t X_u) - \mathbb{E}(X_s X_u) = u - u = 0 \,, \tag{3.19}$$

so the increments $X_t - X_s$ are uncorrelated with $(X_u : u \le s)$. So the joint distribution of $X_t - X_s$ and X_u is multivariate Gaussian with vanishing covariances, so they are in fact independent.
(Note that in general uncorrelated rvs could still be dependent, only for Gaussian rvs these concepts are equivalent, since they are fully characterized by their means and covariances.)
Finally, $X_t - X_s$ is Gaussian with mean 0 and variance (for $s < t$)

$$\mathrm{var}(X_t - X_s) = \mathrm{var}(X_t) + \mathrm{var}(X_s) - 2\mathrm{cov}(X_t, X_s) = t + s - 2s = t - s \,, \tag{3.20}$$

so the increments are also stationary. □

Further non-examinable remarks.

Theorem 3.3 *If a collection $\{F_{\mathbf{t}}\}$ of fdds fulfills the* ***Kolmogorov consistency relations***

$$\begin{aligned} F_{\mathbf{t},t_{n+1}}(\mathbf{x}, x_{n+1}) &\to F_{\mathbf{t}}(\mathbf{x}) \quad \textit{as} \quad x_{n+1} \to \infty\,, \quad \textit{and} \\ F_{\Pi\mathbf{t}}(\Pi\mathbf{x}) &= F_{\mathbf{t}}(\mathbf{x}) \quad \textit{for all permutations } \Pi \textit{ of } (1,\dots,n)\,, \end{aligned} \tag{3.21}$$

then there exists a stochastic process $X = (X_t : t \geq 0)$ (on some probability space Ω) that has fdds $\{F_{\mathbf{t}}\}$. Here we use the notation $\mathbf{t} = (t_1, \dots, t_n)$ and $\mathbf{x} = (x_1, \dots x_n)$.

Proof. Is related to the Skorohod representation theorem. Basically one takes Ω to be the path space of the process. Some hints are given in [GS] Section 8.6

Existence and uniqueness of standard BM with the properties in Definition 3.1 has been established by Wiener, via a measure on path space.

Theorem 3.4 Existence and Uniqueness of Brownian motion (Wiener)
There exists a unique probability measure $\mathcal{W}$ on the path space $C\big([0,\infty), \mathbb{R}\big)$ (called the ***Wiener measure****), such that the process with sample paths distributed according to $\mathcal{W}$ has the properties of Brownian motion as defined in 3.1.*

Proof. see e.g. Rogers and Williams, Section I.6

3.3 Brownian motion and the heat equation

We are looking for an evolution equation for the transition densities, analogous to the master equation for Markov chains. First we will derive it for Brownian motion as scaling limit from the simple random walk.

Let $(X_n : n \in \mathbb{N})$ be a simple random walk. Then the distribution at time n is given by $\boldsymbol{\pi}(n+1) = \boldsymbol{\pi}(n)\, P$, which can be written in the following incremental form

$$\boldsymbol{\pi}(n+1) - \boldsymbol{\pi}(n) = \boldsymbol{\pi}(n)(P - Id)\,, \tag{3.22}$$

where $P - Id$ is proportional to the discrete Laplacian Δ,

$$P - Id = \begin{pmatrix} \ddots & \ddots & \ddots & & \\ & \frac{1}{2} & -1 & \frac{1}{2} & \\ & & \ddots & \ddots & \ddots \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \ddots & \ddots & \ddots & & \\ & 1 & -2 & 1 & \\ & & \ddots & \ddots & \ddots \end{pmatrix} = \frac{1}{2}\Delta\,. \tag{3.23}$$

In coordinate form (3.22) this equation looks like

$$\pi_k(n+1) - \pi_k(n) = \frac{1}{2}\big(\pi_{k-1}(n) - 2\pi_k(n) + \pi_{k+1}(n)\big) \tag{3.24}$$

In the previous section we saw that under the scaling $\boxed{t = \Delta t\, n,\ x = \Delta x\, k \text{ with } \Delta x = \sqrt{\Delta t}}$, $\Delta x\, X_{[t/\Delta t]} \to B_t$ converges to Brownian motion as $\Delta t \to 0$. Therefore, we expect that $\pi_k(n)$ converges to the pdf $f(t,x)$ of B_t, i.e.

$$\lim_{\Delta t \to 0} \frac{1}{\Delta x} \pi_{x/\Delta x}(t/\Delta t) = f(t,x) = (2\pi t)^{-1/2} \exp\big(-x^2/(2t)\big)\,. \tag{3.25}$$

Here $1/\Delta x = 1/\sqrt{\Delta t}$ is the volume element to turn the probability $\pi_k(n)$ into a density.

Plugging the scaling into the discrete-time Master equation (3.24), we can derive a differential equation for f. We assume that for large n, k (i.e. small Δt, Δx), $\pi_k(n)$ is approximately given by

$$\pi_k(n) \simeq \Delta x\, f\big(n\Delta t, k\Delta x\big) = \Delta x\, f(t,x)\ . \tag{3.26}$$

Then we get by Taylor expansion

$$\begin{aligned} \frac{\pi_{k\pm1}(n)}{\Delta x} &\simeq f(t, x\pm\Delta x) = f(t,x) \pm \Delta x \frac{\partial f(t,x)}{\partial x} + \frac{(\Delta x)^2}{2}\frac{\partial^2 f(t,x)}{\partial x^2} + O\big((\Delta x)^3\big) \\ \frac{\pi_k(n+1)}{\Delta x} &\simeq f(t+\Delta t, x) = f(t,x) + \Delta t \frac{\partial f(t,x)}{\partial t} + O\big((\Delta t)^2\big)\ . \end{aligned} \tag{3.27}$$

Thus with $\Delta x = \sqrt{\Delta t}$ (otherwise the limit is again degenrate) equation (3.24) leads to

$$\begin{aligned} \frac{\partial}{\partial t} f(t,x) &= \lim_{\Delta t\to 0} \frac{\pi_k(n+1) - \pi_k(n)}{\Delta t\, \Delta x} = \lim_{\Delta t\to 0} \frac{1}{2\Delta t\, \Delta x}\big(\pi_{k-1}(n) - 2\pi_k(n) + \pi_{k+1}(n)\big) = \\ &= \lim_{\Delta t\to 0} \frac{(\Delta x)^2}{2\Delta t}\frac{\partial^2}{\partial x^2} f(t,x) + O\big(\Delta t/\Delta x\big) = \frac{1}{2}\frac{\partial^2}{\partial x^2} f(t,x)\ . \end{aligned} \tag{3.28}$$

So since standard BM starts in the origin, its pdf fulfills

$$\frac{\partial}{\partial t} f(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2} f(t,x) \quad \text{with initial condition} \quad f(0,x) = \delta_0(x)\ . \tag{3.29}$$

This PDE is the so-called *heat equation* (or *diffusion equation*) which has been well studied, and indeed (3.25) is its unique solution.

Note that $f(t,x) = f(t,x|0,0)$ is in fact a transition density with special initial condition $X_0 = 0$. Since BM and the heat equation are time-homogeneous, an analogous derivation conditioned on $B_s = y$ gives the same equation for the transition density $f(t,x|s,y)$ with the more general initial condition $f(s,x|s,y) = \delta_y(x)$ (where s is the new starting point in time).

Indeed, as we have seen before $B_t - B_s \sim N(0, t-s)$ for $t \geq s$, and therefore the transition density is given by

$$f(t,x|s,y) = \big(2\pi(t-s)\big)^{-1/2} \exp\Big(-\frac{(x-y)^2}{2(t-s)}\Big) . \tag{3.30}$$

$f(t,x|s,y)$ is also called the *heat kernel*, since for $s = 0$ it is the fundamental solution to that PDE (3.29). That means that for every intial distribution $f(0,y)$ we have

$$f(t,x) = \int_{\mathbb{R}} f(t,x|0,y)\, f(0,y)\, dy . \tag{3.31}$$

We can also derive (3.29) from the master equation (1.27) of a continuous-time Markov chain, by rescaling space as $x = \Delta x\, k$ with $\Delta x \to 0$, and speeding up the process to see a time evolution on the macroscopic scale $t' = (\Delta x)^2 t$. In these derivations the exact structure of the generator G or $P - Id$ is not important and this equation holds for a whole class of processes, including e.g. symmetric jumps of finite range, which will just influence the prefactor of the Laplacian.

3.4 Diffusion processes and Fokker-Planck equations

Definition 3.5 A Markov process X is called a *diffusion process*, if it has continuous sample paths and as $h \searrow 0$

$$\begin{aligned} \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t = x\big) &= a(t,x)\, h + o(h) , \\ \mathbb{E}\big((X_{t+h} - X_t)^2 \,\big|\, X_t = x\big) &= b(t,x)\, h + o(h) , \end{aligned} \tag{3.32}$$

for some functions $a(t,x) \in \mathbb{R}$ (*drift coefficient*) and $b(t,x) \geq 0$ (*diffusion coefficient*).

The distributional properties are uniquely characterized by the drift and the diffusion coefficient.

Theorem 3.5 *Let X be a diffusion process with drift $a(t,x)$ and diffusion coefficient $b(t,x)$. Then the transition density $f = f(t,x|s,y)$ exists and satisfies the* ***(forward) Fokker-Planck equation***

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}\big(a(t,x)\, f\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(t,x)\, f\big) \tag{3.33}$$

for all $0 \leq s \leq t$, $x, y \in \mathbb{R}$. In particular, this equation holds also for the density

$$f(t,x) = \int_{\mathbb{R}} f(t,x|0,y)\, f(0,y)\, dy \tag{3.34}$$

with general initial conditions $f(0,y)$.

Proof. by Taylor expansion for observables (see hand-out 6 if you are interested).

The Fokker-Planck equation is often written in the short form $\frac{\partial}{\partial t} f = \mathcal{L}^* f$, with the (adjoint) *generator*

$$\big(\mathcal{L}^* f\big)(t,x) = -\frac{\partial}{\partial x}\big(a(t,x)\, f\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(t,x)\, f\big) . \tag{3.35}$$

Examples. (BM with drift, OU process)

Stationary pdfs $f^*(x)$ of a **time-homogeneous diffusion process** with constant drift $a(x)$ and diffusion $b(x)$ are given by stationary solutions to (3.33), i.e.

$$0 = -\frac{\partial}{\partial x}\big(a(x)\, f^*(x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(b(x)\, f^*(x)\big)\ . \tag{3.36}$$

Integrating (3.36) and denoting the derivative by $'$ we get under mild regularity assumptions on $a(x)$ and $b(x)$ (see also later remarks)

$$\begin{aligned} 0 \ &= \ -\int_{-\infty}^{x} \big(a(y)\, f^*(y)\big)' dy + \frac{1}{2}\int_{-\infty}^{x} \big(b(y)\, f^*(y)\big)'' dy = \\ &= \ -a(x)\, f^*(x) + \frac{1}{2}\big(b(x)\, f^*(x)\big)' \quad (+const.) \end{aligned} \tag{3.37}$$

Under the assumption that $f^*(x)$ vanishes sufficiently fast as $x \to -\infty$ we can ignore the boundary terms and the above constant vanishes. If we consider processes on other state spaces, such as $S = [0, 1]$ with boundary conditions, this has to be taken into account in the derivation. The above is a first order linear differential equation and differentiating with the product rule we get

$$f^{*\prime}(x) = \frac{2a(x) - b'(x)}{b(x)}\, f^*(x)\ . \tag{3.38}$$

So the solution is

$$f^*(x) = f^*(0) \exp\Big(\int_0^x \frac{2a(y) - b'(y)}{b(y)}\, dy\Big) \tag{3.39}$$

where $f^*(0)$ is fixed by normalization $\int_{\mathbb{R}} f^*(x)\,dx = 1$.

Examples. (BM with drift, OU process)

Time evolution of observables.
Let $(X_t : t \geq 0)$ be a diffusion process on $S = \mathbb{R}$ and $g : S \to \mathbb{R}$ be an observable, such as $g(x) = x$. Then the expected value

$$\mathbb{E}\big(g(X_t)\big) = \int_{\mathbb{R}} g(x)\, f(t,x)\, dx \tag{3.40}$$

obeys the following evolution equation (using the notation (3.35)),

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\big(g(X_t)\big) &= \int_{\mathbb{R}} g(x)\, \frac{\partial f(t,x)}{\partial t}\, dx = \int_{\mathbb{R}} g(x)\, (\mathcal{L}^* f)(t,x)\, dx = \\ &= \int_{\mathbb{R}} (\mathcal{L} g)(t,x)\, f(t,x)\, dx = \mathbb{E}\big((\mathcal{L}g)(X_t)\big)\ . \end{aligned} \tag{3.41}$$

This follows from integration by parts, since

$$\int_{\mathbb{R}} g(x) \frac{\partial}{\partial x_i}\big(a_i(t,x)\, f(t,\mathbf{x})\big) dx = -\int_{\mathbb{R}} \Big(\frac{\partial}{\partial x_i} g(x)\Big) a_i(t,x)\, f(t,x) dx\ , \tag{3.42}$$

where we again assume that $f(t,x) \to 0$ as $|x| \to \infty$ sufficiently fast, so that we can ignore the boundary terms. For the diffusion part this can be done twice and leads to

$$\mathcal{L} = a(t,x)\frac{\partial}{\partial x} + \frac{1}{2} b(t,x) \frac{\partial^2}{\partial x^2}\ . \tag{3.43}$$

This operator is called the *generator* of the process $(X_t : t \geq 0)$ and describes the expected time evolution of observables. Mathematically, $\mathcal{L}^*$ acting on the PDF is the adjoint operator to $\mathcal{L}$ on the Hilbert space L^2. It is often simpler to work with the generator $\mathcal{L}$ rather than the adjoint $\mathcal{L}^*$, and therefore diffusion processes are typically characterized by defining their generator.

Examples.
The time evolution of the mean $m_t = \mathbb{E}(X_t)$ is given by choosing $g(x) = x$ and computing $\mathcal{L}x = a(t, x)$, which gives

$$\frac{d}{dt}\mathbb{E}(X_t) = \mathbb{E}\big(a(t, X_t)\big) \, . \tag{3.44}$$

This is not very surprising when looking at the definition 3.5 of diffusion processes. For the second moment with $g(x) = x^2$ we get $\mathcal{L}x^2 = 2xa(t, x) + b(t, x)$ and thus

$$\frac{d}{dt}\mathbb{E}(X_t^2) = 2\mathbb{E}\big(X_t \, a(t, X_t)\big) + \mathbb{E}\big(b(t, X_t)\big) \, , \tag{3.45}$$

so the drift also influences the time evolution of higher moments.

(OU process)

3.5 Further remarks (non-examinable)

Diffusion processes can be generalized to higher dimensions. $\mathbf{X}$ in $\mathbb{R}^d$ is called a diffusion process if in addition to the continuity property analogous to Definition 3.5

$$\begin{aligned} \mathbb{E}\big(\mathbf{X}_{t+h} - \mathbf{X}_t \,\big|\, \mathbf{X}_t = \mathbf{x}\big) &= \mathbf{a}(t,\mathbf{x})\, h + o(h)\,, \\ \mathbb{E}\big((\mathbf{X}_{t+h} - \mathbf{X}_t)(\mathbf{X}_{t+h} - \mathbf{X}_t)^T \,\big|\, \mathbf{X}_t = \mathbf{x}\big) &= b(t,\mathbf{x})\, h + o(h)\,, \end{aligned} \tag{3.46}$$

with drift vector $\mathbf{a}(t,x) \in \mathbb{R}^d$ and diffusion matrix $b \in \mathbb{R}^{d\times d}$, where

$$b_{ij} = \mathbb{E}\big((X^i_{t+h} - X^i_t)(X^j_{t+h} - X^j_t) \,\big|\, \mathbf{X}_t = \mathbf{x}\big)\,. \tag{3.47}$$

This is the covariance matrix of the increments of the process. The Fokker-Planck equation for $f = f(t,\mathbf{x})$ is now given by

$$\frac{\partial f}{\partial t} \;=\; -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}\big(a_i(t,\mathbf{x})\, f\big) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}\big(b_{ij}(t,\mathbf{x})\, f\big) = \mathcal{L}^* f\,. \tag{3.48}$$

where the right-hand side defines a linear operator $\mathcal{L}^*$ on the set of functions $f : \mathbb{R}^d \to \mathbb{R}$. $\mathcal{L}^*$ is called the *(adjoint) generator* of the process $\mathbf{X}$ and is the analogous quantity of the generator of a continuous-time Markov chain.

Definition 3.6 Let $\mathbf{X}$ be a diffusion process with time-independent drift $\mathbf{a}(\mathbf{x})$ and diffusion $b(\mathbf{x})$. $V : \mathbb{R}^d \to \mathbb{R}$ is called a *potential* for $\mathbf{X}$, if $\mathbf{a}(\mathbf{x}) = -\nabla V(\mathbf{x})$. If $b_{ij}(\mathbf{x}) = b\,\delta_{ij}$ we call $\mathbf{X}$ a (noise-perturbed) *gradient flow*.

The Fokker-Planck equation of a gradient flow is given by

$$\frac{\partial f(t,\mathbf{x})}{\partial t} = \nabla\cdot\big((\nabla V(\mathbf{x}))\, f(t,\mathbf{x})\big) + \frac{b}{2}\,\Delta f(t,\mathbf{x})\,. \tag{3.49}$$

Examples. (all 1D processes and OU)

Proposition 3.6 *Assume that $V : \mathbb{R}^d \to \mathbb{R}$ is smooth and that*

$$Z := \int_{\mathbb{R}^d} e^{-2V(\mathbf{x})/b}\, d^d x < \infty\,. \tag{3.50}$$

Then the diffusion process $\mathbf{X}$ *with (3.48) is ergodic. The unique stationary distribution is the* ***Gibbs distribution*** *with density*

$$f^*(\mathbf{x}) = \frac{1}{Z}\, e^{-2V(\mathbf{x})/b}\,, \tag{3.51}$$

and the normalization factor Z *is called* ***partition function***.

Proof. We have from (3.51) $\nabla f^* = -\frac{2}{b}(\nabla V)\, f^*$ and thus

$$\frac{b}{2}\Delta f^* = \frac{b}{2}\nabla\cdot(\nabla f^*) = -\nabla\cdot\big((\nabla V)\, f^*\big)\,. \tag{3.52}$$

Substituting this in (3.49) the right-hand side vanishes $\mathcal{L}^* f^* = 0$, and f^* is stationary.
Uniqueness and ergodicity follow from the fact gradient flows fulfill general conditions such that the Fokker-Planck equation (3.49) has a unique (time-dependent) solution. □

Examples.

Defining the *probability current* $\mathbf{J}(f)$ with i-th component

$$J_i(\mathbf{x}, f) := a_i(\mathbf{x})\, f - \frac{1}{2}\sum_{j=1}^{d}\frac{\partial}{\partial x_j}\big(b_{ij}(\mathbf{x})\, f\big)\,, \tag{3.53}$$

the Fokker-Planck equation (3.48) can be written as a *continuity equation*

$$\frac{\partial f(t,\mathbf{x})}{\partial t} + \nabla_{\mathbf{x}}\cdot\mathbf{J}\big(\mathbf{x}, f(t,\mathbf{x})\big) = 0. \tag{3.54}$$

Integrating this equation over a domain $A \subseteq \mathbb{R}^d$ and using integration by parts like above we get

$$\frac{\partial}{\partial t}\int_A f(t,\mathbf{x})\,d^dx = -\int_A \nabla_{\mathbf{x}}\cdot\mathbf{J}\big(\mathbf{x},f(t,\mathbf{x})\big)\,d^dx = -\int_{\partial A}\mathbf{J}\big(\mathbf{x},f(t,\mathbf{x})\big)\cdot d\mathbf{S}\,. \tag{3.55}$$

The second identity follows from *Stokes' theorem* (also called *Gauss' integration theorem*).
If $A=\mathbb{R}^d$ or the system is *closed* in A then $\mathbf{J}\big(\mathbf{x},f(t,\mathbf{x})\big)=\mathbf{0}$ for all $\mathbf{x}\in\partial A$. So the right-hand side of (3.55) vanishes and the total probability is conserved, i.e.

$$\mathbb{P}(\mathbf{X}_t\in A) = \int_A f(t,\mathbf{x})\,d^dx = 1\,. \tag{3.56}$$

An important class of diffusion processes with direct connections to statistical mechanics are noise-perturbed gradient flows.

For time-independent drift $\mathbf{a}(\mathbf{x})$ and diffusion $b(\mathbf{x})$ existence and uniqueness of the *initial value problem*

$$\frac{\partial f(t,\mathbf{x})}{\partial t} = (\mathcal{L}^*f)(t,\mathbf{x})\,,\quad f(0,\mathbf{x})=f_0(\mathbf{x})\,, \tag{3.57}$$

is well understood under the assumption of *uniform ellipticity*, i.e.

$$\xi\,b(\mathbf{x})\,\xi^T = \sum_{i,j=1}^d b_{ij}(\mathbf{x})\,\xi_i\,\xi_j \ge \alpha\|\xi\|_2^2 \quad \text{for some } \alpha>0 \text{ and all } \xi\in\mathbb{R}^d\,. \tag{3.58}$$

Theorem 3.7 *Under the assumption (3.58) and the growth conditions*

$$f_0(\mathbf{x}) \le Ce^{\alpha\|\mathbf{x}\|_2^2},\quad \Big|\frac{\partial a_i(\mathbf{x})}{\partial x_i}\Big| \le C_1\big(1+\|\mathbf{x}\|_2^2\big),\quad \Big|\frac{\partial^2 b_{ij}(\mathbf{x})}{\partial x_i\partial x_j}\Big| \le C_2\big(1+\|\mathbf{x}\|_2^2\big) \tag{3.59}$$

for some constants $C, C_1, C_2 > 0$, the initial value problem (3.57) has a unique ***classical*** *$C^{1,2}\big((0,\infty),\mathbb{R}^d\big)$ solution.*

4 Some stochastic calculus

4.1 Diffusion processes and SDEs

Diffusion processes can be described also by stochastic differential equations. Let X be a diffusion process in $\mathbb{R}$ with drift $a(t,x)$ and diffusion coefficient $b(t,x) = \sigma^2(t,x)$ given by

$$\begin{aligned} \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t = x\big) &= a(t,x)\,h + o(h)\,, \\ \mathbb{E}\big((X_{t+h} - X_t)^2 \,\big|\, X_t = x\big) &= \sigma^2(t,x)\,h + o(h)\,. \end{aligned} \tag{4.1}$$

In general for a random variable Y with mean μ and variance σ^2 we can write

$$Y = \mu + \sigma\xi \quad \text{where} \quad \xi = \frac{Y-\mu}{\sigma}\,. \tag{4.2}$$

Also the increments of the process X at time t are random variables with mean and variance depending on X_t and given by

$$\begin{aligned} \mathbb{E}\big(X_{t+h} - X_t \,\big|\, X_t\big) &= a(t,X_t)\,h + o(h)\,, \\ \mathrm{var}\big(X_{t+h} - X_t \,\big|\, X_t\big) &= \sigma^2(t,X_t)\,h - a(t,X_t)^2\,h^2 + o(h) = \sigma^2(t,X_t)\,h + o(h)\,. \end{aligned} \tag{4.3}$$

Therefore with $\xi_{t,t+h} = \big(X_{t+h} - X_t - a(t,X_t)\big)/\sqrt{\sigma^2(t,X_t)\,h}$ we get

$$X_{t+h} - X_t = a(t,X_t)\,h + \sigma(t,X_t)\,\sqrt{h}\,\xi_{t,t+h} + o(h)\,. \tag{4.4}$$

Then

$$\mathbb{E}(\sqrt{h}\,\xi_{t,t+h}) = 0 \quad \text{and} \quad \mathrm{var}(\sqrt{h}\,\xi_{t,t+h}) = h\,, \tag{4.5}$$

which looks an awful lot like the increment of a Brownian motion. Indeed, if the process X has independent increments also the $\xi_{t,t+h}$ are independent and

$$\xi_{t,t+h} = \sum_{k=1}^{n} \xi_{t+(k-1)/n,\,t+k/n} \tag{4.6}$$

can be written as a sum of arbitrarily many independent random variables with mean 0 and variance $1/n$. Therefore $\sqrt{h}\,\xi_{t,t+h} \sim N(0,h)$ are Gaussian and can thus be interpreted as increments of a Brownian motion. Now we can write

$$X_{t+h} - X_t = a(t,X_t)\,h + \sigma(t,X_t)(B_{t+h} - B_t) + o(h) \quad \text{for a BM } B\,. \tag{4.7}$$

Deviding by h we get in the limit $h \to 0$

$$\frac{dX_t}{dt} = a(t,X_t) + \sigma(t,X_t)\,\frac{dB_t}{dt}\,. \tag{4.8}$$

This is a differential equation for each path of X, i.e. for fixed $\omega \in \Omega$. But paths of a BM are not differentiable and therefore (4.8) is often written as

$$dX_t = a(t,X_t)\,dt + \sigma(t,X_t)\,dB_t\,. \tag{4.9}$$

Definition 4.1 (4.9) is called a *stochastic differential equation (SDE)* with *drift* $a(t,x)$ and *diffusion* $\sigma(t,x)$. Alternatively, physicists often write

$$\frac{dX_t}{dt} = a(t,X_t) + \sigma(t,X_t)\,\eta_t\,, \tag{4.10}$$

and call this a *Langevin equation*, where $\eta_t = dB_t/dt$ is called *white noise*.

The (non-existing) derivative $\eta_t = dB_t/dt$ can be understood as a normalized random force term on X uncorrelated in time. Formally it is given by a Gaussian process with mean 0 and covariance $\delta(t-s)\,Id$, which makes sense if integrated over time.

As for ordinary differential equations, it is often better to look at the integrated version of (4.9), since it requires less regularity assumptions.

Definition 4.2 A continuous process $X = (X_t : t \geq 0)$ is a *solution* of the SDE (4.9) with initial condition $X_0 = x_0$ if

$$X_t = x_0 + \int_0^t a(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \tag{4.11}$$

holds with probability 1 for all $t \geq 0$ (or $t \in [0,T]$). The solution is called *unique*, if any two solutions X^1 and X^2 satisfy

$$\mathbb{P}\big(X_t^1 = X_t^2 \text{ for all } t \geq 0\big) = 1 \qquad \big(\text{ or } t \in [0,T] \text{ resp.}\big)\,. \tag{4.12}$$

So in order to solve SDEs we have to make sense of the two *stochastic integrals* in (4.11).
Let $X = (X_t : t \geq 0)$ and $Y = (Y_t : t \geq 0)$ be two continuous processes.

We partition the time interval $[0,t]$ such that

$$0 = t_0 < t_1 < \ldots < t_n = t \quad \text{with} \quad t_k - t_{k-1} \to 0 \text{ for all } k = 1, \ldots, n, \text{ as } n \to \infty \tag{4.13}$$

Then we would like to approximate the stochastic integral $I = (I_t : t \geq 0)$ by

$$I_t^n = \sum_{k=1}^{n} Y_{t_{k-1}} (X_{t_k} - X_{t_{k-1}}) \to I_t = \int_0^t Y_s\, dX_s \quad \text{as } n \to \infty\,. \tag{4.14}$$

This is a (particular) Riemann sum approximation of the integrals in (4.11), the simple choice $X_t = t$ yields the first, and $X_t = B_t$ the second. The general question we investigate in the following is, for which processes X and Y the approximations converge and in what sense they converge. Answers turn out to be quite different for the two integrals. But it turns out that the choice of the time partition is not crucial, so we can arbitrarily choose $t_k = k/n$ to fix ideas.

4.2 Stochastic integration and Itô calculus

Proposition 4.1 *For the integrator $X_t = t$ and continuous integrand Y the limit in (4.14) exists pointwise and we can define*

$$I_t(\omega) = \int_0^t Y_s(\omega)\, ds := \lim_{n\to\infty} I_t^n(\omega) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_{t_{k-1}}(\omega)\,. \tag{4.15}$$

Proof. Usual convergence for the Riemann sum approximations holds for each fixed ω since the total variation of $X_t = t$ is finite, i.e.

$$\lim_{n\to\infty} \sum_{k=1}^{n} \left| X_{t_k}(\omega) - X_{t_{k-1}}(\omega) \right| = t < \infty \quad \text{for all } t \geq 0\,. \tag{4.16}$$

Examples.

Theorem 4.2 Itô integral
Let Y be a continuous process and $X = B$ a standard BM. If

$$\mathbb{E}\left(\int_0^t Y_s^2 ds \right) < \infty \tag{4.17}$$

for some $t \geq 0$, then

$$I_t = \int_0^t Y_s\, dB_s := \lim_{n\to\infty} I_t^n = \lim_{n\to\infty} \sum_{k=1}^{n} Y_{t_{k-1}} (B_{t_k} - B_{t_{k-1}}) \tag{4.18}$$

exists in the L^2-sense, i.e. for all $s \leq t$, $\quad \mathbb{E}\big((I_s^n - I_s)^2\big) \to 0\,.$
*If (4.17) holds for all $t \geq 0$, then $I = (I_t : t \geq 0)$ is a continuous process with $\mathbb{E}(I_t) = 0$ and is called the **(stochastic) Itô integral** of Y w.r.t. B.*

Proof. see e.g. Rogers and Williams

How do we calculate Itô integrals? Let's start with a simple example.

We see that for the Itô integral with $\alpha = 0$ we get

$$\int_{t_0}^{t} B_s \, dB_s = \frac{1}{2}\big((B_t^2 - B_{t_0}^2) - (t - t_0)\big) \, . \tag{4.19}$$

Another common choice are centred intermediate points with $\alpha = 1/2$. Here we get

$$S\int_{t_0}^{t} B_s \, dB_s = \frac{1}{2}(B_t^2 - B_{t_0}^2) \, , \tag{4.20}$$

and this integral is called the *Stratonovich integral*. The advantage of this choice is that it obeys the usual rules of calculus and arises naturally in approximations of Brownian motion by smooth processes (see later). But now dependence of Y_{τ_k} and the increment $X_{t_k} - X_{t_{k-1}}$ is more complicated, leading to several technical difficulties compared to Itô. Therefore the preferred choice is usually the Itô integral, and from this one can recover the Stratonovich version by a simple transformation. The unexpected term $(t - t_0)$ in (4.19) has to be there, since the result should have vanishing expectation. These additional terms can be easily understood by the rules of *Itô calculus*, introduced below.

It is often convenient to use the following intuitive differential notation,

$$I_t = I_{t_0} + \int_{t_0}^{t} Y_s \, dX_s \quad \Leftrightarrow \quad dI_t = Y_t \, dX_t \, . \tag{4.21}$$

For example for the integral (4.19) this gives $B_t \, dB_t = \frac{1}{2}(dB_t^2 - dt)$, leading to the rule

$$\boxed{d(B_t^2) = 2B_t \, dB_t + dt} \, . \tag{4.22}$$

From the above derivation we see that the origin of this chain rule is

$$\boxed{(dB_t)^2 = dt} \qquad \text{or more generally} \quad (\sigma dB_t)^2 = \sigma^2 dt \; . \tag{4.23}$$

For a consistent calculus all terms up to order dt have to be taken into account. For usual calculus this involves only first order expansions, but in stochastic calculus this means going up to second order.

Proposition 4.3 *The Itô integral $I_t = \int_0^t Y_s \, dB_s$ of a diffusion process $(Y_t : t \geq 0)$ w.r.t. Brownian motion is a **martingale**, i.e.*

$$\mathbb{E}\big(I_t | \{I_u : u \leq v\}\big) = I_v \quad \textit{for all} \quad 0 \leq v \leq t \; . \tag{4.24}$$

Proof.

4.3 Diffusion processes and Itô's formula

Let X be a solution of the SDE

$$dX_t = a(t, X_t) \, dt + \sigma(t, X_t) \, dB_t \; . \tag{4.25}$$

The following very useful result summarizes our findings in Section 3.2 and gives an explicit formula for time evolution of an observable $g(X_t)$.

Theorem 4.4 Itô's formula
Let X be a solution of (4.25) and $g \in C^2(\mathbb{R}, \mathbb{R})$. Then

$$dg(X_t) = g'(X_t) \, dX_t + \frac{1}{2} g''(X_t) \, \sigma^2 \, dt \; , \tag{4.26}$$

or in the (extended) integrated version

$$\begin{aligned} g(X_t) \; &= \; g(X_0) + \int_0^t g'(X_s) \, \sigma(s, X_s) dB_s \\ &\quad + \int_0^t \Big(g'(X_s) \, a(s, X_s) + \tfrac{1}{2} g''(X_s) \, \sigma^2(s, X_s) \Big) ds \; . \end{aligned} \tag{4.27}$$

Proof. Taylor expansion with terms up to order dt, using (4.23) and the Itô chain rule (4.22).

In incremental form this gives

$$\begin{aligned} dg(X_t) \; &= \; g'(X_t) \, dX_t + \frac{1}{2} \, g''(X_t) \, (dX_t)^2 = \\ &= \; g'(X_t) \big(a(t, X_t) \, dt + \sigma(t, X_t) \, dB_t \big) + \frac{1}{2} \, g''(X_t) \sigma^2(t, X_t) \, dt \; . \end{aligned} \tag{4.28}$$

Let $f(t,x)$ be the pdf of the process X that solves the SDE (4.9). Taking the expectation on both sides, we get from Itô's formula using partial integration

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\big(g(X_t)\big) &= \int_{\mathbb{R}} g(x)\,\frac{\partial}{\partial t}f(t,x)\,dx = \\ &= \int_{\mathbb{R}} \Big(g'(x)\,a(t,x) + \frac{1}{2}g''(x)\sigma^2(t,x)\Big)\,f(t,x)\,dx = \\ &= \int_{\mathbb{R}} g(x)\bigg(-\frac{\partial}{\partial x}\big(a(t,x)\,f(t,x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(\sigma^2(t,x)\,f(t,x)\big)\bigg)dx\;, \quad (4.29)\end{aligned}$$

since boundary terms and the expected value of the Itô integral vanishes (it is a martingale). This holds for arbitrary functions g, and therefore we must have

$$\frac{\partial}{\partial t}f(t,x) = -\frac{\partial}{\partial x}\big(a(t,x)\,f(t,x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(\sigma^2(t,x)\,f(t,x)\big) = \mathcal{L}^* f(t,x)\;. \quad (4.30)$$

Thus $f(t,x)$ fulfilles the Fokker-Planck equation with adjoint generator $\mathcal{L}^*$ introduced in (3.48) and X is a diffusion process with drift $a(t,x)$ and diffusion $\sigma^2(t,x)$. In Section 3.1 we have derived the opposite statement, so let us summarize both of them.

Proposition 4.5 *X is a diffusion process as defined in Def. 3.5 if and only if it is a solution of the SDE (4.25) with $\sigma^2(t,x) = b(t,x)$.*

Remember that the generator $\mathcal{L}$ of a diffusion process introduced in (3.43) is given by

$$\mathcal{L} = a(t,x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2}{\partial x^2}\;, \quad (4.31)$$

and describes the time evolution of expected values of observables (3.41). Using $\mathcal{L}$ we can rewrite Itô's formula

$$g(X_t) = g(X_0) + \int_0^t (\mathcal{L}g)(s,X_s)\,ds + \int_0^t \sigma(s,X_s)\,dB_s\;. \quad (4.32)$$

So we see that the time evolution of $g(X_t)$ is given by the generator part plus fluctuations in terms of an Itô integral, which is a martingale with vanishing expected value.

Example. (geometric BM, see [GS] p546, [Ga] p103)

Non-examinable remarks.

With considerable technical effort, the Itô integral can be generalized to non-continuous processes. But there is a particularly important class of integrators for which this can be done quite easily.

Definition 4.3 Let $N = (N_t : t \geq 0) \sim PP(\lambda)$ and $Z_1, Z_2, \ldots$ a sequence of iidrv's with $\mathbb{E}\big(|Z_k|\big) < \infty$ and distribution function F. Then

$$Q = (Q_t : t \geq 0) \quad \text{with} \quad Q_t = \sum_{k=1}^{N_t} Z_k \tag{4.33}$$

is called a *compound Poisson process* or *jump process*. For $a, \sigma \in \mathbb{R}$ and B a standard BM a process of the form

$$X_t = X_0 + a\,t + \sigma B_t + Q_t \quad \text{with stationary, independent increments} \tag{4.34}$$

is called a *Lévy process*. X is completely determined by the *Lévy triple* $(a, \sigma^2, \lambda F)$.

Lévy processes are the simplest generalization of Brownian motion to non-continuous processes with jumps. By definition $t \mapsto Q_t(\omega)$ is piecewise constant for each ω, so the integral can be computed directly. The condition of stationary, independent increments implies that F cannot be arbitrary, but has to be an *infinitely divisible law*. Examples are Gaussians or α-stable laws.

Another interesting question is, how many solutions to the SDE (4.25) there are. For time-independent drift and diffusion there is a general theorem about existence and uniqueness for SDEs.

Theorem 4.6 *Suppose that $a : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are Lipschitz-continuous, i.e.*

$$\big|a(x) - a(y)\big| \leq K|x - y| \quad \textit{for some } K > 0 \textit{ and all } x, y \in \mathbb{R}\,. \tag{4.35}$$

Then (for each probability space Ω and BM B) there exists a unique solution to the SDE

$$dX_t = a(X_t)\,dt + \sigma(X_t)\,dB_t \tag{4.36}$$

for any initial condition $X_0 = x_0 \in \mathbb{R}$.

Proof. analogous to ordinary differential equations using the contraction mapping theorem and Gronwall's Lemma.

There are more elaborate notions of 'solutions' to SDEs in a weak and strong sense, which we did not discuss. A proper understanding of these involves a substantial amount of measure theory. Itô's formula and the existence and uniqueness theorem can be extended to higher space dimensions. There is also a Stratonovich interpretation of SDEs which is directly connected to the Itô version given here. Both can be found in Gardiner, Section 4.3.