

---

# NON-EQUILIBRIUM STATISTICAL MECHANICS

---

by  
Gunnar Pruessner

Imperial College London  
Department of Mathematics  
180 Queen's Gate  
London SW7 2AZ  
United Kingdom



# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Probabilities . . . . .	7
1.2	Probability density function . . . . .	8
1.3	Moments and cumulants . . . . .	8
1.4	Generating functions . . . . .	9
1.4.1	Moment generating function of a sum . . . . .	10
1.4.2	Cumulant generating function . . . . .	11
1.5	GAUSSIANS . . . . .	11
1.6	Central Limit Theorem . . . . .	13
1.6.1	Stable distribution . . . . .	15
<b>2</b>	<b>Stochastic processes</b>	<b>17</b>
2.1	A POISSON process . . . . .	17
2.2	Events in time . . . . .	20
2.3	MARKOVIAN processes . . . . .	21
2.4	CHAPMAN-KOLMOGOROV equations . . . . .	22
<b>3</b>	<b>Random walks</b>	<b>25</b>
3.1	Pedestrian random walk in discrete time . . . . .	25
3.2	Evolution of the PDF using CHAPMAN-KOLMOGOROV . . . . .	26
3.3	Master equation approach . . . . .	27
3.4	FOKKER-PLANCK equation . . . . .	28
<b>4</b>	<b>LANGEVIN equations</b>	<b>31</b>
4.1	Random walk . . . . .	31
4.2	ORNSTEIN-UHLENBECK process . . . . .	34
<b>5</b>	<b>Critical dynamics</b>	<b>37</b>
5.1	From HAMILTONIAN to LANGEVIN equation and back . . . . .	38
5.2	The PDF of $\eta$ . . . . .	39
5.3	A FOKKER-PLANCK equation approach . . . . .	41
5.4	The HOHENBERG-HALPERIN models . . . . .	43



This is an attempt to deliver, within a couple of hours, a few key-concepts of non-equilibrium statistical mechanics. The material covered by the notes is sometimes a bit wider in scope and often more detailed than a presentation can be.

The goal is to develop some ideas of contemporary research. I start with some basic methods of statistics as they are used in statistical mechanics. Many of the ideas are illustrated or even introduced by examples. The second chapter is more specialised: First I will present the concept of LANGEVIN equations, before discussing critical dynamics (introducing the notion of relaxation to equilibrium as opposed to far-from-equilibrium) and finally (if time permits) reaction-diffusion processes.

The plan is to spend at least two hours on the first chapter and at the remaining time of up to three hours on the second chapter.

Two main sources have been used to prepare these notes:

van Kampen, N. G., 1992, *Stochastic Processes in Physics and Chemistry* (Elsevier Science B. V., Amsterdam, The Netherlands), third impression 2001, enlarged and revised edition.

Täuber, U. C., 2005, *Critical dynamics*, preprint available at <http://www.phys.vt.edu/~tauber/utaeuber.html>, (unpublished).



# Chapter 1

## Introduction

This chapter lays the theoretical foundation for the following material. The content is rather general and not restricted to non-equilibrium statistical mechanics, although some observables and techniques are characteristic for this field.

Firstly, I will introduce some basic techniques from statistics, in particular generating functions. The framework of generating functions will then be used to give a simple prove of the central limit theorem. In the last part of this chapter, some basic notions of stochastic processes will be discussed.

### 1.1 Probabilities

A few reminders regarding probability are in order:

$$P(\neg A) = 1 - P(A) \quad (1.1a)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1.1b)$$

Here  $A \cup B$  means that  $A$  or  $B$  occur (not exclusive). On the other hand  $A \cap B$  means that  $A$  and  $B$  occur simultaneously. If  $A \cap B = \emptyset$  then  $A$  and  $B$  are **mutually exclusive**. If  $P(A, B)$  (joint probability) factorises  $P(A, B) = P(A)P(B)$  then  $A$  and  $B$  are said to be **independent**. The conditional probability  $P(A|B)$  (pronounced “probability of  $A$  given  $B$ ”) is given by BAYES’s theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \quad (1.2)$$

Note that if  $P(A), P(B) \neq 0$  and  $A, B$  are independent, then they cannot be mutually exclusive, because as they are independent  $P(A \cap B) = P(A)P(B) \neq 0$ , but if  $A \cap B = \emptyset$  then  $P(A \cap B) = 0$ . That makes a lot of sense: If  $A$  excludes  $B$  from occurring, than  $A$  are obviously not independent.

## 1.2 Probability density function

We start this section by considering the **probability density function** (PDF)  $\mathcal{P}_a(x)$ , which, in loose terms, is the probability of the random variable  $x$  to have a specific value. More precisely,  $\mathcal{P}_a(x) dx$  is the probability that  $a$  is in the interval  $[x, x + dx]$  and it is very important to realise that  $\mathcal{P}_a(x)$  is a *density* (in order to keep the notation simple, I will later use the subscript  $a$  for all sorts of purposes). The integral over the density is normalised,

$$\int_{-\infty}^{\infty} dx \mathcal{P}_a(x) = 1 \quad (1.3)$$

Very often, it is more convenient to consider the **cumulative distribution function** (CDF)

$$F(z) = \int_{-\infty}^z dx \mathcal{P}_a(x) \quad (1.4)$$

which is the probability that  $x$  is in the interval  $]-\infty, z]$ . This definition is more convenient, because one does not need to introduce the whole notion of probability densities, but can stick with the concept of finite probabilities. Obviously, the lower integration limit in (1.4) should be chosen suitably and might be changed to, say, 0, if the PDF has corresponding support. A PDF is non-negative and therefore the CDF is increasing in  $z$ ; according to (1.3)  $\lim_{z \rightarrow \infty} F(z) = 1$  and from (1.4)

$$\mathcal{P}_a(x) = \frac{d}{dz} F(z) \quad (1.5)$$

In the following, we will use the PDF as the starting point for all definitions and derivations.

It is an obvious extension to introduce vector valued random variables  $\mathbf{x}$  or (equivalently) consider the **joint PDF**  $\mathcal{P}_{a,b}(x, y)$  (where  $x$  and  $y$  could also be regarded as two components of a vector). In that case one has, quite naturally,

$$\mathcal{P}_a(x) = \int_{-\infty}^{\infty} dy \mathcal{P}_{a,b}(x, y) \quad (1.6)$$

where the integration on the right is also known as **marginalisation** over the **nuisance variable**  $y$ . Two random variables are said to be **independent** if the joint PDF factorises, *i.e.* if  $\mathcal{P}_{a,b}(x, y) = \mathcal{P}_a(x) \mathcal{P}_b(y)$ .

## 1.3 Moments and cumulants

The  $n$ th moment  $\langle x^n \rangle$  is defined as

$$\langle x^n \rangle = \int_{-\infty}^{\infty} dx x^n \mathcal{P}_a(x) \quad (1.7)$$

which includes the normalisation (1.3) as the special case  $n = 0$ . By the properties of an integral we have  $\langle f(x) + g(x) \rangle = \langle f(x) \rangle + \langle g(x) \rangle$ . The extension to joint PDFs is straight forward

$$\langle x^n y^m \rangle = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy x^n y^m \mathcal{P}_{ab}(x, y) \quad (1.8)$$

If  $a$  and  $b$ <sup>1</sup> are independent, the joint probability factorises and  $\langle x^n y^m \rangle = \langle x^n \rangle \langle y^m \rangle$ . If  $\langle xy \rangle = \langle x \rangle \langle y \rangle$  then  $x$  and  $y$  are said to be **uncorrelated**. While independence entails that the variables are uncorrelated, the converse does not hold, *i.e.* if two random variables are uncorrelated, they are not necessarily independent (although many authors ignore this fact).

**Cumulants** are *defined* as derivatives of the logarithm of the moment generating function, which will be discussed below. For the time being, we introduce only two cumulants, the **first cumulant**

$$\langle x \rangle_c = \langle x \rangle \quad (1.9)$$

and the **second cumulant**

$$\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2 = \langle (x - \langle x \rangle)^2 \rangle = \sigma^2(x) \quad (1.10)$$

which equals the **variance** of the random variable  $x$ . Finally, the **third cumulant** is

$$\langle x^3 \rangle_c = \langle (x - \langle x \rangle)^3 \rangle = \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3 \quad (1.11)$$

but the fourth cumulant is (sic!)

$$\langle x^4 \rangle_c = \langle (x - \langle x \rangle)^4 \rangle - 3 \langle (x - \langle x \rangle)^2 \rangle^2 \quad (1.12)$$

It is worth noting that the notation of moments and cumulants is somewhat loose: Comparing  $\langle x^n \rangle$  and  $\langle x^n y^m \rangle$  in (1.7) and (1.8) respectively, it is not clear whether one has to integrate over both variables,  $x$  and  $y$ , or only over one, in which case  $\langle x^n y^m \rangle = y^m \langle x^n \rangle$ . Similarly,  $\langle x^2 \rangle_c$  is actually an integral over a square, as well as a square of an integral. This ambiguity is usually resolved by the context.

## 1.4 Generating functions

For many situations, the **moment generating function** (MGF) is more suitable than the “plain” PDF. The moment generating function  $\mathcal{M}_a(z)$  is defined as

$$\mathcal{M}_a(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \langle x^n \rangle \quad (1.13)$$

---

<sup>1</sup>...or equivalently  $x$  and  $y$ , which are actually the concrete values of the random variable  $a$  and  $b$  respectively — I will be quite liberal when using the subscript and the argument, but the argument is always where the function is evaluated

if the sum converges. The MGF has the very useful property that

$$\left. \frac{d^n}{dz^n} \right|_{z=0} \mathcal{M}_a(z) = \langle x^n \rangle \quad (1.14)$$

By noting that  $\langle \exp(xz) \rangle$  has exactly the same property and assuming that  $\mathcal{M}_a(z)$  is analytic around  $z = 0$ , one has

$$\mathcal{M}_a(z) = \langle e^{xz} \rangle \quad (1.15)$$

In particular,  $\mathcal{M}_a(0) = 1$  by normalisation.

Considering  $\langle \exp(xz) \rangle$  in detail, shows that it is the LAPLACE transform of the PDF, *i.e.* the MGF is the LAPLACE transform of the PDF:

$$\mathcal{M}_a(z) = \int_{-\infty}^{\infty} dx e^{xz} \mathcal{P}_a(x) \quad (1.16)$$

Differentiating  $n$  times with respect to  $z$  and evaluating at  $z = 0$  brings down  $n$  factors of  $x$ , confirming (1.14). Evaluating the MGF for imaginary arguments turns the LAPLACE transform into a Fourier transform, which is also known as the **characteristic function**.

A discretised version of the MGF is easy to write down:

$$\mathcal{M}_a(z) = \sum_x \mathcal{P}_a(x) e^{xz} \quad (1.17)$$

although in discrete problems where  $x \in \mathbb{Z}$  the factor  $\exp(xz)$  often is replaced by  $z^x$  and the moments are generated by  $\left( z \frac{d}{dz} \right)^n \Big|_{z=1}$ . Note that the coefficient in front of  $\exp(xz)$  in Eq. (1.17) are the probabilities of  $x$ .

### 1.4.1 Moment generating function of a sum

The MGF is particularly useful when calculating sums and differences of random variables. We start with the example of two independent random numbers  $x$  and  $y$  drawn from a (factorising) PDF  $\mathcal{P}_{ab}(x, y) = \mathcal{P}_a(x) \mathcal{P}_b(y)$ . The MGF of the sum of these two random numbers is

$$\begin{aligned} \mathcal{M}_{a+b}(z) &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy e^{z(x+y)} \mathcal{P}_{ab}(x, y) \\ &= \int_{-\infty}^{\infty} dx e^{zx} \mathcal{P}_a(x) \int_{-\infty}^{\infty} dy e^{zy} \mathcal{P}_b(y) = \mathcal{M}_a(z) \mathcal{M}_b(z) \end{aligned} \quad (1.18)$$

which applies, by the same derivation, to the characteristic function.

The PDF of a multiple of a random variable, say  $y = \alpha x$  can be constructed either by considering the cumulative distribution introduced above, or by using  $\mathcal{P}_x(x) dx = \mathcal{P}_y(y) dy$ , so that  $\mathcal{P}_y(y) = \mathcal{P}_x(y/\alpha) dx/dy = \mathcal{P}_x(x)/\alpha$ . The MGF of the new random variable is

$$\mathcal{M}_y(z) = \int_{-\infty}^{\infty} dy e^{zy} \mathcal{P}(y) = \int_{-\infty}^{\infty} dx e^{z\alpha x} \mathcal{P}(x) = \mathcal{M}_x(z\alpha) \quad (1.19)$$

which is not surprising, given that  $\langle y^n \rangle = \alpha^n \langle x^n \rangle$ , *i.e.* every derivative of the MGF must shed a factor  $\alpha$ . In general, the random variable  $\alpha x + \beta y$  has MGF  $\mathcal{M}_x(z\alpha)\mathcal{M}_y(z\beta)$ , which again applies similarly to the characteristic function.

The characteristic function has the immediate advantage that we are more used to FOURIER transforms and can write down its inverse immediately. Such an inverse allows us to derive the PDF from the MGF.

### 1.4.2 Cumulant generating function

Cumulants have been introduced earlier (see Eq. (1.9)-(1.12)), but generally, they are defined as the derivatives of the logarithm of the MGF. Accordingly, the **cumulant generating function** (CGL)  $\mathcal{C}_x(z)$  is defined as

$$\mathcal{C}_x(z) = \ln \mathcal{M}_x(z) . \quad (1.20)$$

so that

$$\left. \frac{d^n}{dz^n} \right|_{z=0} \mathcal{C}_x(z) = \langle x^n \rangle_c \quad (1.21)$$

similar to Eq. (1.14).

The zeroth cumulant is  $\mathcal{C}_x(0) = \ln 1 = 0$ , the first cumulant is

$$\left. \frac{d}{dz} \right|_0 \mathcal{C}_x(z) = \frac{\left. \frac{d}{dz} \right|_0 \mathcal{M}_x(z)}{\mathcal{M}_x(0)} = \langle x \rangle , \quad (1.22)$$

just the first moment. The second derivative of  $\mathcal{C}_x(z)$  in obvious notation is

$$\mathcal{C}_x(z)'' = \frac{\mathcal{M}_x(z)''}{\mathcal{M}_x(z)} - \frac{\mathcal{M}_x(z)'}{\mathcal{M}_x(z)^2} \quad (1.23)$$

and it gets more and more complicated.

Shifting the observable by  $x_0$  changes the MGF to  $\mathcal{M}_x(z) \exp(zx_0)$  and therefore the CGF becomes  $\mathcal{C}_x(z) + zx_0$ , so that only the first cumulant is affected by a shift.

**Exercise 1:** Check Eq. (1.9)–(1.12).

## 1.5 GAUSSIANS

GAUSSIANS play a central rôle in all non-equilibrium statistical mechanics and we will see below why this is so. Let's briefly recapitulate some of its properties. We define the GAUSSIAN

$$\mathcal{G}(x; x_0, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \quad (1.24)$$

which is normalised, as one can show by straight forward integration. Its mean is  $x_0$ :

$$\int_{-\infty}^{\infty} dx \mathcal{G}(x; x_0, \sigma^2) x = \int_{-\infty}^{\infty} dx \mathcal{G}(x; x_0, \sigma^2) (x - x_0) + \int_{-\infty}^{\infty} dx \mathcal{G}(x; x_0, \sigma^2) x_0 \quad (1.25)$$

but the first integral vanishes by symmetry and the second is just the normalisation times  $x_0$ . The second moment is slightly harder, see below; we note that generally

$$\int_{-\infty}^{\infty} dx \mathcal{G}(x; 0, \sigma^2) x^{2n} = (2n - 1)!! (\sigma^2)^n \quad (1.26)$$

with  $(2n - 1)!! = 1 \cdot 3 \cdot 5 \dots (2n - 1)$  is the double factorial. This can be proven, elegantly, using WICK's theorem, which states that the  $2n$ 'th moment is  $(\sigma^2)^n$  times the number of unique ways to pair  $2n$  elements. For example  $1, 2, 3, 4$  has pairs  $\{(1, 2), (3, 4)\}$ ,  $\{(1, 3), (2, 4)\}$ , and  $\{(1, 4), (2, 3)\}$ . The general statement can be shown easily: Pick any one of  $2n$  and find a partner, of which there are  $2n - 1$ . Pick any next one, and find a partner again, out of  $2n - 3$ . Repeat until only one is left, so  $(2n - 1)(2n - 3) \dots 3 \cdot 1$ . Alternatively, take all  $(2n)!$  permutations of the numbers  $1, 2, \dots, 2n$  and consider neighbouring numbers as pairs, first and second, third and fourth *etc.*. There are  $n!$  permutations for each such "configuration" which leaves the pairs unchanged, and  $2^n$  possible swaps that leave them unchanged:

$$\frac{(2n)!}{2^n n!} = \frac{1 \cdot 2 \cdot 3 \cdot \dots \cdot (2n)}{2 \cdot 4 \cdot 6 \cdot \dots \cdot (2n)} = (2n - 1)!! \quad (1.27)$$

The MGF of a GAUSSIAN is itself a (non-normalised) GAUSSIAN, which one can see by completing squares (similarly for the characteristic function),

$$\frac{(x - x_0)^2}{2\sigma^2} - xz = \frac{(x - x_0 - z\sigma^2)^2}{2\sigma^2} - \frac{(x_0 + z\sigma^2)^2 - x_0^2}{2\sigma^2} \quad (1.28)$$

so that

$$\int_{-\infty}^{\infty} dx \mathcal{G}(x; x_0, \sigma^2) e^{xz} = e^{\frac{(x_0 + z\sigma^2)^2 - x_0^2}{2\sigma^2}} \quad (1.29)$$

Taking the logarithm produces the cumulant generating function of a GAUSSIAN:

$$\mathcal{C}_g(z) = \frac{(x_0 + z\sigma^2)^2 - x_0^2}{2\sigma^2} = zx_0 + \frac{1}{2}z^2\sigma^2 \quad (1.30)$$

As required,  $\mathcal{C}_g(0) = 0$ , the first cumulant is  $x_0$  as expected and the second cumulant (the variance) is just  $\sigma^2$ . Surprisingly, and crucially, all higher cumulants vanish exactly. One might ask: Is every PDF, all cumulants of which vanish apart from the first and the second, a GAUSSIAN? Yes, that is indeed the case: Together with  $\mathcal{C}(0) = 0$ , the first and the second cumulant fix the coefficients in any second order polynomial, so such a PDF will always produce a CGF of the form Eq. (1.30). PDFs all cumulants of which vanish apart from the first can be obtained by taking the limit  $\lim_{\sigma^2 \rightarrow 0}$  for the GAUSSIAN, or

by taking the inverse LAPLACE transform of  $\exp(zx_0)$ , which gives (remember that this procedure is closely related to an inverse FOURIER transform) a DIRAC  $\delta$ -function at  $x_0$ . Finally taking away the first cumulant simply moves the peak to  $x_0 = 0$ .

Another important feature of a GAUSSIAN is that it solves the diffusion equation (remember that in the context of simple FOKKER-PLANCK equations),

$$\partial_t \phi = D \partial_x^2 \phi - v \partial_x \phi \quad (1.31)$$

on  $x \in \mathbb{R}$ , with diffusion constant  $D$ , drift velocity  $v$  and initial condition  $\lim_{t \rightarrow 0} \phi = \delta(x - x_0)$ . It is solved by  $\phi = \mathcal{G}(x - vt; x_0, 2Dt)$ . It is worth remembering that the diffusion constant  $D$  gives rise to the variance  $2Dt$ .

## 1.6 Central Limit Theorem

The CLT (central limit theorem) is of fundamental importance in all of statistics and statistical mechanics. Here, it is presented in a somewhat simplified form, but more sophisticated versions exist for correlated variables and with a more detailed analysis of precisely how the PDF of sums of random variables converge to a GAUSSIAN.

Consider the mean of  $N$  independent variables  $x_i$  with  $i = 1, 2, \dots, N$ , all drawn from the same PDF which is arbitrary, apart from having vanishing mean and finite cumulants<sup>2</sup>:

$$\mathcal{X} \equiv \frac{1}{\sqrt{N}} \sum_i^N x_i \quad (1.32)$$

Notice the somewhat unusual normalisation  $\sqrt{N}^{-1}$ , which is crucial to the argument. We will get back to that point. If the MGF of the underlying PDF<sup>3</sup> is  $\mathcal{M}_a(z)$ , then the MGF of the sum is (see Section 1.4.1)  $\mathcal{M}_X(z) = \mathcal{M}_a\left(z/\sqrt{N}\right)^N$  and the CGF is

$$\mathcal{C}_X(z) = N \mathcal{C}_a\left(z/\sqrt{N}\right). \quad (1.33)$$

The first cumulant vanishes (because the first cumulant of  $x$  vanishes), as one would expect and the second cumulant is

$$\left. \frac{d^2}{dz^2} \right|_{z=0} \mathcal{C}_X(z) = \left. \frac{d^2}{dz^2} \right|_{z=0} \mathcal{C}_a(z) \quad (1.34)$$

which means that the variance of  $\mathcal{X}$  is the same as the variance of  $x$  itself. This is true for any  $N$ . In general, the higher derivatives give

$$\langle \mathcal{X}^n \rangle_c = N^{1-n/2} \langle x^n \rangle_c \quad (1.35)$$

<sup>2</sup>If the cumulants are not finite, a CGF might still exist, but it cannot be expanded about  $z = 0$ .

<sup>3</sup>One can obtain the PDF directly through convolutions, the outcome is the same. XXX show that.

and because the cumulants of the underlying PDF are all finite, the cumulants of  $\mathcal{X}$  vanish for  $n > 2$ ,  $N \rightarrow \infty$ . Because only the second cumulant “survives”, the PDF of  $\mathcal{X}$  is a Gaussian, see Section 1.5. This is the CLT: The PDF of  $\mathcal{X}$  converges to a GAUSSian.

Going through the “proof” above, the extra condition that the average  $\langle x \rangle$  must vanish can be lifted by shifting the  $x_i$  by  $\langle x \rangle$ . One might be tempted to restate the CLT for simple averages like

$$\mathcal{X} \equiv \frac{1}{N} \sum_i^N x_i \quad (1.36)$$

or maybe

$$\mathcal{X} \equiv \sum_i^N x_i. \quad (1.37)$$

However, both these random variables do *not* produce a GAUSSian. The first one, (1.36), will produce a distribution that converges for  $N \rightarrow \infty$  to a DIRAC  $\delta$ -peak at  $\langle x \rangle$ , because with that normalisation all cumulants apart from the first one will vanish,  $\langle \mathcal{X}^n \rangle_c = N^{1-n} \langle x^n \rangle_c$ . The second cumulant vanishes like  $N^{-1}$ , but all higher cumulants vanish even faster and therefore the distribution of the average  $\mathcal{X}$  will look more and more like a Gaussian centered at  $\langle x \rangle$  with ever decreasing width. This observation is what people mean when they say that “averages converge to a Gaussian”. Strictly speaking this is wrong, because the distribution of simple averages converges to a DIRAC  $\delta$ -function, but if one accepts to consider the leading order correction in the approach to the DIRAC  $\delta$ -function, then the distribution is Gaussian (with corrections in the higher cumulants of order of at least  $N^{-2}$ ).

The second variation, (1.37) is more disastrous: All cumulants that do not vanish in the underlying PDF diverge with increasing  $N$  for the PDF of  $\mathcal{X}$ , because  $\langle \mathcal{X}^n \rangle_c = N \langle x^n \rangle_c$ . Even the centre of mass ( $\langle \mathcal{X} \rangle$ ) would run away. The observation  $\langle \mathcal{X}^n \rangle_c = N \langle x^n \rangle_c$  looks like clashing with the effect (1.37) has on moments,  $\langle \mathcal{X}^n \rangle = N^n \langle x^n \rangle + \dots$  where the dots contain lower order terms. So, if  $\langle \mathcal{X}^n \rangle \propto N^n$ , how come  $\langle \mathcal{X}^n \rangle_c \propto N$ , given that cumulants can be calculated from moments? The answer is in the dots: Calculating the  $n$ th cumulant from moments, all terms to power  $N^n, N^{n-1}, N^{n-2}, \dots, N^2$  will cancel and only the terms to order  $N$  will remain.

It is often said, that “the sum of random variables looks more and more GAUSSian the more variables are added”. That, however, is true only if one rescales the observables appropriately: If  $\sum_i^N x_i$  is not divided by a sufficiently high power of  $N$ , the resulting PDF generally broadens (indicated by increasing variance) and has a centre of mass that will generally continue to change with increasing  $N$ . Notice, that the centre of mass is moving like  $N^{1/2} \langle x \rangle$  if  $\langle x \rangle \neq 0$  in the original derivation. If, on the other hand, one divides by too high a power, the distribution collapses to a  $\delta$ -peak, the position of which might even change with  $N$ , depending on the choice of normalisation.

### 1.6.1 Stable distribution

If the PDF of the sum of two random variables is exactly identical to the PDF of the individual PDFs (after suitable rescaling), then such a PDF is known as stable. The Gaussian is the *only* stable PDF with finite variance.



## Chapter 2

# Stochastic processes

Mathematicians have a proper definition of a stochastic process. However, at the present level of rigour, it suffices to say that a stochastic process is a time-dependent observable which depends on some random events that occur in time. It is important to keep in mind that a stochastic process is characterised by an observable and a statement about the process needs to be interpreted as a statement about the observable. The observable marks a point in the **sample space**, the space of all possible outcomes.

### 2.1 A POISSON process

The POISSON process is a **point process** and can be visualised as an arrangement of  $s \geq 0$  points on the interval, say,  $[0, t]$ . The PDF for having  $s$  points at “times”  $\tau_1, \tau_2, \dots, \tau_s$  is  $Q(\tau_1, \tau_2, \dots, \tau_s)$ . We impose that any permutation of the times corresponds to exactly the same state, so when we impose normalisation, we have to discount for double counting. This can be done either by summing over an ordered set, say  $\tau_1 \leq \tau_2 \leq \dots \leq \tau_s$ , or by double counting and then dividing by the degeneracy:

$$\sum_{s=0}^{\infty} \frac{1}{s!} \int_0^t d\tau_1 \dots d\tau_s Q(\tau_1, \dots, \tau_s) = 1 \quad (2.1)$$

By imposing that the  $Q$  factorise in the form,  $Q(\tau_1, \dots, \tau_s) = e^{-\nu} q(\tau_1) \dots q(\tau_s)$  one arrives at the POISSON process (in the stationary state, *i.e.*  $q$  is constant, also known as **shot noise**). The normalisation then gives  $\exp\left(-\nu + \int_0^t d\tau q(\tau)\right) = 1$ , so

$$\nu = \int_0^t d\tau q(\tau) \quad (2.2)$$

The MGF for the moments of the number of points is easy to write down, simply multiply by a factor  $\exp(zs)$  inside the summation Eq. (2.1) and use the POISSON property

(factorisation):

$$\mathcal{M}_P(z) = e^{-\nu} \sum_{s=0}^{\infty} \frac{1}{s!} e^{zs} \nu^s = e^{(\exp(z)-1)\nu} \quad (2.3)$$

and the cumulant generating function is just  $\mathcal{C}_P(z) = (\exp(z) - 1)\nu$ , so that all cumulants greater or equal to 1 give  $\nu$ . In particular, the first moment is  $\nu$ , so that the MGF is

$$\mathcal{M}_P(z) = e^{(\exp(z)-1)\langle s \rangle} = e^{-\langle s \rangle} \sum_{n=0}^{\infty} \frac{\langle s \rangle^n}{n!} e^{nz} \quad (2.4)$$

On the other hand, the MGF is  $\langle \exp(nz) \rangle$ , see Eq. (1.15) (but notice that the averaging is now over the discrete variable  $s$ ), *i.e.*

$$\mathcal{M}_P(z) = \sum_{s=0}^{\infty} \mathcal{P}_P(s) e^{sz} \quad (2.5)$$

which by comparison of coefficients means that the probability to have  $n$  events in the interval  $[0, t]$  is given by the POISSON distribution

$$\mathcal{P}_P(s) = e^{-\langle s \rangle} \frac{\langle s \rangle^s}{s!} \quad (2.6)$$

In case of shot noise, the probability that no event takes place within a time interval  $[t, t+dt]$  is  $(1 - qdt)$ . The survival probability, *i.e.* the probability that no event takes place within a finite interval of length  $\Delta t$ , therefore is  $\lim_{dt \rightarrow 0} (1 - qdt)^{\Delta t/dt} = \exp(-q\Delta t)$ . It is important to realise that this is the probability for no event to occur for *at least*  $\Delta t$ . The probability that such a “streak” is terminated after time  $\Delta t$  is the survival probability times the probability for an event to occur in the next instance,  $\exp(-q\Delta t) qdt$ . This probability density is obviously normalised.

**Exercise 2:** ZERNIKE's "Weglängenparadoxon": Say  $q$  is a constant (shot noise) and so the probability that an event takes place is constant. Show that the average time between two events is  $1/q$  (ignore boundary effects, i.e. consider the interval  $] -\infty, \infty[$ ). Moreover, show that the average time from any randomly chosen point in time to the preceding event is  $1/q$  and to the (next) following event is  $1/q$  as well. This, however, suggests that the distance between any two events is  $2/q$ .

**Note/Solution:** The probability that time  $\alpha\Delta t$  passes with no event, then one event happens (density  $q$ ) and then time  $(1-\alpha)\Delta t$  passes is  $\exp(-q\alpha\Delta t) q \exp(-q(1-\alpha)\Delta t) = q \exp(-q\Delta t)$ . Similarly one might have any number of events  $n$  at specific times with any sequence of silences in between with probability  $q^n \exp(-q\Delta t)$ . This is the probability density to have a sequence of  $n$  events at certain times over a period of time  $\Delta t$ . The probability of having  $n$  events within  $\Delta t$  therefore is

$$\begin{aligned} \int_0^{\Delta t} dt'_1 \int_{t'_1}^{\Delta t} dt'_2 \dots \int_{t'_{n-1}}^{\Delta t} dt'_n q^n e^{-q\Delta t} \\ = \frac{1}{n!} \int_0^{\Delta t} dt'_1 \int_0^{\Delta t} dt'_2 \dots \int_0^{\Delta t} dt'_n q^n e^{-q\Delta t} \\ = \frac{1}{n!} (q\Delta t)^n e^{-q\Delta t} \quad (2.7) \end{aligned}$$

which is, after summing over all  $n$ , obviously properly normalised. Eq. (2.7) is identical to Eq. (2.6) with  $\langle s \rangle = q\Delta t$  (notice the factorial<sup>1</sup>  $1/n!$  and the additional power  $\Delta t^n$  compared to the probability of a sequence of events at certain times,  $q^n \exp(-q\Delta t)$ ). The key insight is that all that matters is the total time of silence, not the precise amount of time for each interval of silence. The state of each point in time (having or not having an event) is independent.

One might conclude that the average distance from a randomly chosen point to the next event can be derived by considering empty intervals that are terminated after time  $\Delta t$  by an event and being pierced by another one at  $\alpha\Delta t$ ,  $\alpha \in [0, 1]$ . The two events being independent, all what counts is the quiet stretch up to its termination, which has probability  $q \exp(-q\Delta t)$ , so that the average time between a randomly chosen point in time and the next event is

$$\int_0^1 d\alpha \int_0^\infty d\Delta t (\alpha\Delta t) q e^{-q\Delta t} = \frac{1}{2q}. \quad (2.8)$$

This procedure is deceptive, it seems to run over all possible configurations, namely all

<sup>1</sup>The  $1/n!$  compensates for ignoring the time-order of the integrals (generally  $1/n!$  compensates for double-counting, as events or particle are considered as indistinguishable and configurations only to be counted once): The first line in Eq. (2.7) means  $t'_1 \leq t'_2 \leq \dots \leq t'_n$ . The integrand is invariant under permutations of the  $t'_i$ , in fact, it is even independent of all  $t'_i$ . So, any arbitrary time order, say  $t'_4 \leq t'_7 \leq \dots \leq t'_2$  gives the same result and the sum over all time orders gives  $n!$  times the desired result. However, the sum over all time orders corresponds to the integral with no time order at all, the second line in Eq. (2.7).

intervals  $\Delta t$  terminated by an event and all positions  $\alpha\Delta t$  of another event within this interval. This however, turns out to be *half* of the average distance between any pair of events:  $q \exp(-q\Delta t)$  is the probability of an empty stretch occurring from some initial point, terminated by an event, but the intermediate event at  $\alpha\Delta t$  has been accounted for. The probability of *two* events occurring within  $\Delta t$ , one at  $\alpha\Delta t < \Delta t$  and one at the end is  $q^2 \exp(-q\Delta t)$ . The correct normalisation for the  $(\alpha, \Delta t)$  probability space gives an additional factor  $\Delta t$  (but note  $\int_0^\infty d\Delta t \int_0^{\Delta t} d\Delta t' q^2 \exp(-q\Delta t) = 1$ ), so correcting the arguments above, Eq. (2.8), the average distance between two events across all such events is

$$\int_0^1 d\alpha \int_0^\infty d\Delta t (\alpha\Delta t) q^2 \Delta t e^{-q\Delta t} = \frac{1}{q}. \quad (2.9)$$

What is actually missing in Eq. (2.8) is the probability density of the second, intermediate event to occur, which was missed because of  $\alpha$  being dimensionless.

More directly, the average distance from any randomly chosen point to the next event is given by the first moment of the probability of having a quiet stretch up to time  $\Delta t$ , terminated by an event

$$\int_0^\infty d\Delta t \Delta t q e^{-q\Delta t} = \frac{1}{q}. \quad (2.10)$$

This is also the distance between events, averaged over all pairs. If one wants to calculate the average length of an empty stretch chosen randomly by picking a point at random, then one has to take into account that the picking is biased towards longer stretches, which are picked with a probability  $\propto t \exp(-q\Delta t)$ . After normalising this probability, the average length of such a stretch is

$$\int_0^\infty d\Delta t \Delta t^2 q^2 e^{-q\Delta t} = 2/q. \quad (2.11)$$

## 2.2 Events in time

In the following, I will first derive some general properties of time-dependent PDFs.

Without getting to formal, we now consider the PDF of a series of “events” in the sense that an observable  $x$  (could be a high-dimensional observable, describing, for example the positions of all gas molecules in a box) has a certain value at a certain time  $t$ , *i.e.* the PDF  $\mathcal{P}_1(x, t)$ . Note that this PDF is normalised with respect to  $x$ , while  $t$  parameterises the PDF, so  $\mathcal{P}_1(x, t)$  is to be considered as a series of PDFs, each for a specific time  $t$ . If  $\mathcal{P}_1(x, t)$  is independent of  $t$ , then the process is said to be stationary. The index 1 is explained below.

One could consider the joint PDF  $\mathcal{P}_2(x_2, t_2; x_1, t_1)$ , which is the probability to observe  $x_1$  at  $t_1$  and  $x_2$  at  $t_2$ . The conditional PDF to observe  $x_2$  at  $t_2$  given that one observes  $x_1$  at  $t_1$  is

$$\mathcal{P}_{1|1}(x_2, t_2|x_1, t_1) = \frac{\mathcal{P}_2(x_2, t_2; x_1, t_1)}{\mathcal{P}_1(x_1, t_1)} = \frac{\mathcal{P}_{1|1}(x_1, t_1|x_2, t_2) \mathcal{P}_1(x_2, t_2)}{\mathcal{P}_1(x_1, t_1)}. \quad (2.12)$$

This procedure can be generalised to any number of observations and the index indicates how many observations the PDF covers and how many are conditions. To ease notation, we will occasionally drop the index and write just  $n$  for  $x_n, t_n$ .

Trivially marginalising over the nuisance variable gives

$$\mathcal{P}_{1|1}(3|1) = \int d2 \mathcal{P}_{2|1}(2, 3|1) \quad (2.13)$$

and because

$$\mathcal{P}_{2|1}(2, 3|1) = \frac{\mathcal{P}_3(1, 2, 3)}{\mathcal{P}_1(1)} = \frac{\mathcal{P}_3(1, 2, 3)}{\mathcal{P}_2(1, 2)} \frac{\mathcal{P}_2(1, 2)}{\mathcal{P}_1(1)} = \mathcal{P}_{1|2}(3|1, 2) \mathcal{P}_{1|1}(2|1) \quad (2.14)$$

we have

$$\mathcal{P}_{1|1}(3|1) = \int d2 \mathcal{P}_{1|2}(3|1, 2) \mathcal{P}_{1|1}(2|1) \quad (2.15)$$

## 2.3 MARKOVian processes

The term MARKOVian is heavily overused in the physics literature. It means that the PDF of a sequence of events (a particular observable) with  $t_1 < t_2 < t_3 < \dots < t_{n+1}$  (for  $n \geq 1$ ) has the property

$$\mathcal{P}_{1|n}(n+1|1, 2, 3, \dots, n) = \mathcal{P}_{1|1}(n+1|n) \quad (2.16)$$

*i.e.* for the probability of an observable to have value  $x_1$  at  $t_1$ , given that it had values  $x_2, x_3, \dots, x_n$  earlier, only the most recent time matters (if it matters at all). The MARKOV property can be used to compile a hierarchy. By BAYES

$$\mathcal{P}_2(1, 2) = \mathcal{P}_1(1) \mathcal{P}_{1|1}(2|1) \quad (2.17a)$$

$$\mathcal{P}_2(1, 2, 3) = \mathcal{P}_2(1, 2) \mathcal{P}_{1|2}(3|1, 2) \quad (2.17b)$$

$$\mathcal{P}_2(1, 2, 3, 4) = \mathcal{P}_3(1, 2, 3) \mathcal{P}_{1|3}(4|1, 2, 3) \quad (2.17c)$$

and therefore

$$\begin{aligned} \mathcal{P}_2(1, 2, 3, 4) &= \mathcal{P}_2(1, 2) \mathcal{P}_{1|2}(3|1, 2) \mathcal{P}_{1|3}(4|1, 2, 3) \\ &= \mathcal{P}_1(1) \mathcal{P}_{1|1}(2|1) \mathcal{P}_{1|2}(3|1, 2) \mathcal{P}_{1|3}(4|1, 2, 3) . \end{aligned} \quad (2.18)$$

The conditional PDFs on the RHS can be simplified using the MARKOV property Eq. (2.16), and so:

$$\mathcal{P}_2(1, 2, 3, 4) = \mathcal{P}_1(1) \mathcal{P}_{1|1}(2|1) \mathcal{P}_{1|1}(3|2) \mathcal{P}_{1|1}(4|3) . \quad (2.19)$$

All PDFs are therefore fully determined by the initial distribution  $\mathcal{P}_1(1)$  and the transition probability  $\mathcal{P}_{1|1}()$ . The idea is that “all that matters for later times is the state at the system

at any earlier point”, or rather “history does not matter” (the process does not have any memory). By including enough details (namely *all* it depends on) into the observable, every process can be made MARKOVian, which is why the statement that a process is MARKOVian is an empty one without specifying the observable.<sup>2</sup>

conditional an additio

additional observables might very well change the conditional PDF

Note that the MARKOV property relies on some sort of temporal order. It is always the most recent, the preceding time, that matters,  $\mathcal{P}(n+1|1,2,3,\dots,n) = \mathcal{P}(n+1|n)$ . However, given that property, one can show that  $\mathcal{P}(1|2,\dots,n+1) = \mathcal{P}(1|2)$ , using (2.19):

$$\begin{aligned} \mathcal{P}_{1|n}(1|2,\dots,n+1) &= \frac{\mathcal{P}_{n+1}(1,2,\dots,n+1)}{\mathcal{P}_n(2,\dots,n+1)} \\ &= \frac{\mathcal{P}_1(1)\mathcal{P}_{1|1}(2|1)\mathcal{P}_{1|1}(3|2)\dots\mathcal{P}_{1|1}(n+1|n)}{\mathcal{P}_1(2)\mathcal{P}_{1|1}(3|2)\dots\mathcal{P}_{1|1}(n+1|n)} = \frac{\mathcal{P}_1(1)\mathcal{P}_{1|1}(2|1)}{\mathcal{P}_1(2)} = \frac{\mathcal{P}_2(1,2)}{\mathcal{P}_1(2)} = \mathcal{P}_{1|1}(1|2), \end{aligned} \quad (2.20)$$

so the probability to see  $x_1$  at  $t_1$  depends only on what happens at  $t_2$  not on what happens later; there is no preferred direction of time in the MARKOV-property.<sup>3</sup>

## 2.4 CHAPMAN-KOLMOGOROV equations

The MARKOV has an important integral form. Going back to Eq. (2.15), this expression becomes

$$\mathcal{P}_{1|1}(3|1) = \int d2 \mathcal{P}_{1|1}(3|2) \mathcal{P}_{1|1}(2|1) \quad (2.21)$$

<sup>2</sup>One needs to be a bit careful with the emphasis here. MARKOVian means adding knowledge about the history does not change the conditional PDFs, such as Eq. (2.19). Starting with a non-MARKOVian process, one can render it MARKOVian by adding additional variables, which carry the information encoded otherwise in the history. In that sense one can translate “no history needed” into “no additional observables needed”. However, the conditional (conditional to history) PDF of a MARKOV process might easily be a function of an additional variable, *i.e.* adding that observables renders the process non-MARKOVian (the trivial example is that of a MARKOV process that combined with a non-MARKOVian process is non-MARKOVian). So adding variables might not always improve the situation towards MARKOVian. In any case, MARKOVian or non-MARKOVian is a feature of a process and its observable.

<sup>3</sup>One might wonder: If the time-direction in BOLTZMANN’s H-theorem does not have its origin in the MARKOV-property (2.16), where does it come from? If there was no inverse MARKOV-property (2.20), then one would argue that a time direction in the “increase of entropy” of the H-theorem comes from Eq. (2.16) and that the inverse of the transition matrix  $\mathcal{P}_{1|1}(2|1)$ , namely  $\mathcal{P}_{1|1}(1|2)$ , does not obey a corresponding inverse relation. Now we find out that it does, see Eq. (2.20), which suggests that the H-theorem works backwards for the inverse matrix. The riddle can be resolved by showing that  $\mathcal{P}_{1|1}(1|2)$  is not a MARKOVian matrix (unless  $\mathcal{P}_{1|1}(1|2)$  is effectively deterministic, in which case the entropy does not increase in the original model; just try  $m = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ), by having negative elements, for example. In that case,  $\mathcal{P}_{1|1}(1|2)$  could not be applied to every general “initial” distribution  $\mathcal{P}_1(2)$ , *i.e.*  $\mathcal{P}_1(1) = \int d2 \mathcal{P}_1(2) \mathcal{P}_{1|1}(1|2)$  by marginalising over 2.

where  $\mathcal{P}_{1|2}(3|1,2) = \mathcal{P}_{1|1}(3|2)$  has been used (compare to Eq. (2.15)) because of the MARKOV property. Multiplying by  $\mathcal{P}_1(1)$  and integrating over 1 gives

$$\mathcal{P}_1(3) = \int d2 \mathcal{P}_{1|1}(3|2) \mathcal{P}_1(2) . \quad (2.22)$$

The corresponding integral of Eq. (2.15) gives the more complicated expression

$$\mathcal{P}_1(3) = \int d1 \int d2 \mathcal{P}_{1|2}(3|1,2) \mathcal{P}_2(1,2) = \int d1 \int d2 \mathcal{P}_3(1,2,3) , \quad (2.23)$$

with the right hand side being a trivial marginalisation.

Eq. (2.21) is known as the CHAPMAN-KOLMOGOROV equation; it says that the PDF at  $t_3$  given the observation at  $t_1$  is the integral over all intermediate values  $x_2$  at time  $t_2$  over the product of the transitions from  $x_1$  to  $x_2$  and from  $x_2$  to  $x_3$ . It is only that last part that is different from Eq. (2.15): There we had to consider the transition to  $x_3$  given  $x_2$  *and*  $x_1$ . While Eq. (2.21) looks like the trivial statement that “one has to take an intermediate step 2 to go from 1 to 3”, it actually contains the profound statement that for the step from 2 to 3 its previous position at 1 is of no relevance.

The CHAPMAN-KOLMOGOROV equation is a consequence of the MARKOV property.<sup>4</sup>

---

<sup>4</sup>I don't find it obvious to show that the converse holds.



## Chapter 3

# Random walks

The random walk, described by the position  $\mathbf{x}(t)$  in  $d$  dimensions as a function of time  $t$ , is the prime example for a MARKOVIAN stochastic process. The position  $\mathbf{x}$  evolves under a certain transition probability distribution.

In the following, first a discrete version of the random walk is analysed, without making use of the evolution equation of the PDF as given by the CHAPMAN-KOLMOGOROV equation, Eq. (2.21). Next, we will consider those equations, allowing the position of the random walker and the time to be continuous. Finally, the concepts of a master and FOKKER-PLANCK equation are introduced.

### 3.1 Pedestrian random walk in discrete time

Instead of considering the evolution of the PDF  $\mathcal{P}_{\text{rw}}(\mathbf{x})$ , we investigate the MGF of the position. For simplicity, time and space are discretised and space is restricted to one dimension, so  $\mathbf{x}(t) = n(t) \in \mathbb{Z}$  and  $t \in \mathbb{N}$ .

Say the random walker starts at  $n(t=0) = n_0$  and in every time step, its position increases by 1 with probability  $p$  and decreases by 1 (with probability  $q \equiv 1 - p$ ) otherwise. The MGF at  $t = 0$  is just  $\mathcal{M}_{\text{rw}}(z; t = 0) = \exp(zn_0)$ . In the following time step its position is  $n_0 + 1$  with probability  $p$  and  $n_0 - 1$  with probability  $q$ , so

$$\mathcal{M}_{\text{rw}}(z; t = 1) = pe^{z(n_0+1)} + qe^{z(n_0-1)} = \mathcal{M}_{\text{rw}}(z; t = 0) (pe^z + qe^{-z}) \quad (3.1)$$

Note that the factor  $(p \exp(z) + q \exp(-z))$  can be written as

$$pe^z + qe^{-z} = p(\cosh(z) + \sinh(z)) + q(\cosh(z) - \sinh(z)) = \cosh(z) + (p - q) \sinh(z) \quad (3.2)$$

In fact, there is a general updating rule of the MGF: The term  $\exp(zn)$  indicates position  $n$  and its coefficient is the probability of the random walker to be at  $n$ , see Eq. (1.17). So, by multiplication by  $p \exp(z)$  each  $\exp(zn)$  is shifted by one to  $\exp(z(n+1))$ , which is the new position, and the probability is  $p$  times the probability to be there, which is given by the coefficient. The same can be done for the “down moves” and the collective

(i.e. summed) coefficient in front of  $\exp(zn')$  indicates the new probability to be at any  $n'$ . In other words

$$\mathcal{M}_{\text{rw}}(z; t+1) = \mathcal{M}_{\text{rw}}(z; t) p e^z + \mathcal{M}_{\text{rw}}(z; t) q e^{-z} \quad (3.3)$$

so that

$$\mathcal{M}_{\text{rw}}(z; t) = \mathcal{M}_{\text{rw}}(z; t=0) (p e^z + q e^{-z})^t = \mathcal{M}_{\text{rw}}(z; t=0) (\cosh(z) + (p-q) \sinh(z))^t \quad (3.4)$$

(which might remind one of the one-dimensional ISING model). It is trivial to write down the above result explicitly,

$$\mathcal{M}_{\text{rw}}(z; t) = \sum_{i=0}^t p^i q^{t-i} \binom{t}{i} e^{z(n_0+i-(t-i))} \quad (3.5)$$

where of course  $n_0 + i - (t - i) = n_0 + 2i - t$ , which means that parity is conserved for even  $t$ . Without loss of generality  $n_0 = 0$  and for even  $t$  the random walker can only be at an even position  $n$  (the coefficients of  $\exp(zn)$  then vanish for odd  $n$ ). With  $n = 2i$  the PDF is then just  $p^i q^{t-i} \binom{t}{i}$ .

### 3.2 Evolution of the PDF using CHAPMAN-KOLMOGOROV

Next, the condition that the position is discrete is relaxed. The random walker still evolves in discrete time and, say,  $\mathcal{P}_{\text{rw}}(x; t=0) = \delta(x - x_0)$ . According to Eq. (2.21) we only need to specify the evolution “matrix”<sup>1</sup>,  $\mathcal{P}_{1|1}(2|1)$ . If that expression depends only on the time difference  $t_2 - t_1$ , the MARKOV process is said to be homogeneous. Sometimes the term “stationary” is used as well, although it should rather be reserved for time independent PDFs.

Say the transition matrix is given by

$$\mathcal{P}_{1|1}(x_2, t_2 | x_1, t_1) = \frac{1}{\sqrt{4\pi D(t_2 - t_1)}} e^{-\frac{(x_2 - x_1)^2}{4D(t_2 - t_1)}}, \quad (3.6)$$

which is known as the all-important WIENER process. Because the initial condition is specified as a  $\delta$ -function, not much work is needed to find

$$\mathcal{P}_{\text{rw}}(x, t) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{(x - x_0)^2}{4D t}}. \quad (3.7)$$

So that is trivial, but only because  $\mathcal{P}_{1|1}(x_2, t_2 | x_1, t_1)$  was given for *all* times  $t_1 < t_2$ . In the previous example, only a single step  $t_2 - t_1 = 1$  was specified. Assuming some

---

<sup>1</sup>I will call this object a transition matrix in the following, because considering  $\int d2$  as a sum over all elements renders Eq. (2.21) a matrix equation.

sort of “conservation of effort”, one finds the real work hidden in the statement that  $\mathcal{P}_{1|1}(x_2, t_2|x_1, t_1)$  given in Eq. (3.6) obeys the CHAPMAN-KOLMOGOROV equation.

**Exercise 3:** Show that Eq. (3.6) obeys the CHAPMAN-KOLMOGOROV equation Eq. (2.21).

### 3.3 Master equation approach

The CHAPMAN-KOLMOGOROV equation is an integral equation that defines how the PDF of the process evolves in time. If the transition matrix  $\mathcal{P}_{1|1}(x_2, t_2|x_1, t_1)$  is given for all times, then deriving the distribution for any time is usually straight forward. In Section 3.1 we dealt with the situation that the evolution was only known for a single time step.

In continuum time this corresponds to knowing the transition *rates* at any moment in time. We rewrite Eq. (2.21) for homogeneous processes with  $T(x_2|x_1; t_2 - t_1) = \mathcal{P}_{1|1}(x_2, t_2|x_1, t_1)$ :

$$T(x_3|x_1; \tau + \tau') = \int dx_2 T(x_3|x_2; \tau') T(x_2|x_1; \tau) \quad (3.8)$$

where  $\tau = t_2 - t_1$  and  $\tau' = t_3 - t_2$ . Differentiating with respect to  $\tau'$  and taking the limit  $\tau' \rightarrow 0$  gives

$$\begin{aligned} \partial_{\tau'} T(x_3|x_1; \tau) &= \int dx_2 (-\alpha_0(x_2)\delta(x_3 - x_2) + W(x_3|x_2)) T(x_2|x_1; \tau) \\ &= \int dx_2 W(x_3|x_2) T(x_2|x_1; \tau) - \alpha_0(x_3) T(x_3|x_1; \tau) \end{aligned} \quad (3.9)$$

where  $\lim_{\tau \rightarrow 0} \partial_{\tau'} T(x_3|x_2; \tau) = -\alpha_0(x_2)\delta(x_3 - x_2) + W(x_3|x_2)$  has been used, where  $W(x_3|x_2)$  is thought to not explicitly depend on time (we are dealing with a homogeneous process after all). The choice of that particular form is motivated by the following expansion for small times  $\tau$ :

$$T(x_3|x_2; \tau) = (1 - \alpha_0(x_2)\tau)\delta(x_3 - x_2) + \tau W(x_3|x_2) + \mathcal{O}(\tau^2) \quad (3.10)$$

The  $\delta$ -function is there to account for the rate with which the system remains unchanged. Because it must evolve into something or stay the same, the integral over  $x_3$  of  $T(x_3|x_2; \tau)$  must be 1 for all  $\tau$ . Imposing this property order by order in  $\tau$  and given that the leading order is  $\int dx_3 \delta(x_3 - x_2) = 1$ , means that

$$\alpha_0(x_2) = \int dx_3 W(x_3|x_2) \quad (3.11)$$

and therefore produces the so-called **master equation**

$$\partial_{\tau} T(x_3|x_1; \tau) = \int dx_2 (W(x_3|x_2) T(x_2|x_1; \tau) - W(x_2|x_3) T(x_3|x_1; \tau)) , \quad (3.12)$$

where we used  $\alpha_0(x_3) = \int dx_2 W(x_2|x_3)$  (note that  $W$  is time-independent and it does not matter whether  $x_2$  is associated with an earlier time than  $x_3$ ; both are just dummy variables). If the PDF is known at the fixed time  $t_1$ , one can multiply by  $\mathcal{P}_1(x_1, t_1)$  and integrate over 1:

$$\partial_\tau \mathcal{P}_1(x_3, \tau) = \int dx_2 (W(x_3|x_2)\mathcal{P}_1(x_2, \tau) - W(x_2|x_3)\mathcal{P}_1(x_3, \tau)) , \quad (3.13)$$

which is slightly misleading, suggesting that the master equation is only about determining the PDF at later times given the PDF at an initial time  $t_1$ . Instead the master equation is an integro-differential equation for the transition probabilities given *any* initial state.

For discrete states  $n$ , the last expression has a particularly appealing form. Write  $\mathcal{P}_n(t)$  for the probability to be in  $n$  at time  $t$  and one has

$$\partial_t \mathcal{P}_n(t) = \sum_{n'} W(n|n')\mathcal{P}_{n'}(t) - W(n'|n)\mathcal{P}_n(t) \quad (3.14)$$

which are in fact just gain and loss terms for the probability to be in state  $n$ . Notice that  $W(n|n)$  cancels. Defining the matrix  $W$  as

$$W_{nn'} = W(n|n') - \delta_{nn'} \sum_{n''} W(n''|n) \quad (3.15)$$

one can rewrite Eq. (3.14) as

$$\partial_t \mathbf{p}(t) = W_{nn'} \mathbf{p}(t) \quad (3.16)$$

where the vector  $\mathbf{p}(t)$  has entries according to  $\mathcal{P}_n(t)$ . The formal solution of this expression is just  $\mathbf{p}(t) = \exp(tW_{nn'}) \mathbf{p}(0)$ .

### 3.4 FOKKER-PLANCK equation

The FOKKER-PLANCK equation is a particular type of master equation. As long as one is not concerned with discrete states, the above general master equation Eq. (3.12) and the expansion it is based on Eq. (3.10) might look slightly discomfoting. For small time differences, the transition matrix should and indeed does collapse to a  $\delta$ -function. But for finite time  $W$  mixes in and one might wonder what that object actually describes. It is transition rate, so that  $T$  is a probability again, but if the transitions take place at that rate in *every* instant, how come the transition ever collapses to a  $\delta$ -function.

The answer to this is of course the time span  $\tau$  which must be multiplied on the transition rate. What is, however, really unsettling is the fact that Eq. (3.10) states that for short times, the system might remain at *exactly* the same state with finite probability. How can that be possible in a stochastic process, supposedly operating at every instant in time? It cannot and in fact  $W(x_3|x_2)$  will cancel the term  $\alpha_0(x_2)\delta(x_3 - x_2)$  (giving rise to the negative term in Eq. (3.12)) as we will see in the following, Eq. (3.19).

In the following we write the transition matrix  $W(x'|x)$  as  $w(x, -r)$  with  $r = x - x'$ , which gives for Eq. (3.13)

$$\begin{aligned} \partial_\tau \mathcal{P}_1(x_3, \tau) &= \int dx_2 (w(x_2, x_3 - x_2) \mathcal{P}_1(x_2, \tau) - w(x_3, x_2 - x_3) \mathcal{P}_1(x_3, \tau)) \\ &= \int dr (w(x_3 - r, r) \mathcal{P}_1(x_3 - r, \tau) - w(x_3, -r) \mathcal{P}_1(x_3, \tau)) \end{aligned} \quad (3.17)$$

where  $r = x_3 - x_2$ . Next we assume that  $w(x, r)$  is sharply peaked around  $r = 0$  and varies only slowly in  $x$  and that  $\mathcal{P}_1(x, \tau)$  is also varying slowly in  $x$ . Expanding  $w(x_3 - r, r) \mathcal{P}_1(x_3 - r, \tau)$  about  $r = 0$  produces

$$\begin{aligned} w(x_3 - r, r) \mathcal{P}_1(x_3 - r, \tau) &= w(x_3, r) \mathcal{P}_1(x_3, \tau) - r \partial_x (w(x_3, r) \mathcal{P}_1(x_3, \tau)) \\ &\quad + \frac{1}{2} r^2 \partial_x^2 (w(x_3, r) \mathcal{P}_1(x_3, \tau)) + \mathcal{O}(r^3) \end{aligned} \quad (3.18)$$

and keeping only terms up to second order in  $r$  then gives

$$\begin{aligned} \partial_\tau \mathcal{P}_1(x_3, \tau) &= \int dr (w(x_3, r) \mathcal{P}_1(x_3, \tau) - r \partial_x (w(x_3, r) \mathcal{P}_1(x_3, \tau)) \\ &\quad + \frac{1}{2} r^2 \partial_x^2 (w(x_3, r) \mathcal{P}_1(x_3, \tau)) - w(x_3, -r) \mathcal{P}_1(x_3, \tau)) \end{aligned} \quad (3.19)$$

The first and the last term cancel and the integral runs over  $r$ , so that the factors  $\mathcal{P}_1(x_3, \tau)$  can be taken outside the integrals. Defining

$$A(x) = \int dr r w(x, r) \quad (3.20a)$$

$$B(x) = \int dr r^2 w(x, r) \quad (3.20b)$$

which are, effectively, the first and second moments of the displacement  $r$  we arrive at the FOKKER-PLANCK equation:

$$\partial_\tau \mathcal{P}_1(x, \tau) = -\partial_x (A(x) \mathcal{P}_1(x, \tau)) + \frac{1}{2} \partial_x^2 (B(x) \mathcal{P}_1(x, \tau)) , \quad (3.21)$$

which is an equation of motion for the entire PDF of the observable, rather than for the observable itself.

The two unknown functions can be related to the time-evolution of the first and the second cumulant. In particular, we note that  $\int dx \partial_\tau \mathcal{P}_1(x, \tau) = 0$  as surface terms vanish and

$$\partial_t \langle x \rangle = \partial_\tau \int dx x \mathcal{P}_1(x, \tau) = - \int dx x \partial_x (A(x) \mathcal{P}_1(x, \tau)) + \frac{1}{2} \int dx x \partial_x^2 (B(x) \mathcal{P}_1(x, \tau)) \quad (3.22)$$

where the first integral gives  $\int dx A(x) \mathcal{P}_1(x, \tau) = \langle A(x) \rangle$  after an integration by parts and

the second  $\int dx \partial_x (B(x)\mathcal{P}_1(x, \tau)) = 0$ . Therefore

$$\partial_t \langle x \rangle = \langle A(x) \rangle \tag{3.23}$$

## Chapter 4

# LANGEVIN equations

It is generally accepted that equilibrium thermodynamics is “solved”, while non-equilibrium thermodynamics lacks a central formalism à la MAXWELL-BOLTZMANN. To state it so strongly is an exaggeration: Equilibrium thermodynamics still contains many puzzles and the MAXWELL-BOLTZMANN formalism may not always be ideal to address certain issues. While there might be reason to be a bit more pessimistic about the state of affairs in equilibrium, there are certainly many reasons to be a bit more optimistic about non-equilibrium. Many different approaches exist in parallel and it is not always easy to translate between them or formulate the problem in a way that suits any of the methods. But in principle, methods are available ranging from LANGEVIN equations over master equations and FOKKER-PLANCK equations to the JANSSEN-DE DOMINICIS response functional. The following chapters discuss some of these methods.

LANGEVIN equations are stochastic partial differential equations which are in wide use within the physics community. They suffer from a number of mathematical shortcomings, so that there are less popular among mathematicians, who generally prefer integral formulations of the problems. However, in many physical situations, the LANGEVIN equation is the most direct way to describe the phenomenon.

### 4.1 Random walk

A LANGEVIN equation is an equation of motion, usually first order in time. It describes the time evolution of an observable, rather than the time evolution of a PDF of that observable. The observable can be a function (for example an interface that evolves in time), but for simplicity we consider the scalar situation of a particle under the influence of random displacements, also known as BROWNIAN motion. If the position at time  $t$  is  $x(t)$

then its temporal evolution is given by the LANGEVIN equation<sup>1</sup>

$$\dot{x}(t) = \eta(t) \quad (4.1)$$

where  $\eta(t)$  is a noise and therefore a rather peculiar mathematical object. To define it,  $\langle \cdot \rangle$  is interpreted as averaging over all realisations of that noise, and so the **noise-correlator** is given by

$$\langle \eta(t)\eta(t') \rangle = 2\Gamma^2\delta(t - t'). \quad (4.2)$$

Since the FOURIER transform of that correlator is a constant, *i.e.* every mode  $k$  is present with equal strength, it is known as **white noise**. In many applications, it is completely sufficient to specify only the two point correlator of the noise, rather than its full distribution.

Notice that  $2\Gamma^2$  is the amplitude of the  $\delta$ -function; the noise has *infinite* variance. For many purposes it is enough to specify only the correlator, however, the first moment is virtually always set to 0, while higher correlation functions are usually fixed by declaring that the noise is GAUSSIAN so that higher correlations are given by WICK's theorem.<sup>2</sup> While any finite integral over  $\eta$  is like the sum over infinitely many random variables and therefore subject to the central limit theorem (*i.e.* the sum is essentially GAUSSIAN), the noise itself does not need to be GAUSSIAN,  $\mathcal{P}([\eta(t)]) \propto \exp(-\frac{1}{4\Gamma^2} \int dt \eta(t)^2)$  (see Eq. (5.15)). As a matter of convenience the probability of  $\eta$  at any point should be an exponential, so that the product measure (different point in time are uncorrelated) is an exponential of the sum or rather an integral over different points. For example,  $\mathcal{P}([\eta(t)]) \propto \exp(-\frac{1}{4\Gamma^2} \int dt \eta(t)^4)$  works as well, but correlators are now (without WICK) *very* difficult to calculate. It is worth noting that the probability density of a particular realisation of GAUSSIAN noise depends on the total square displacement, not on its sum or integral. The latter is a sum of independent random variables and therefore always GAUSSIAN (even when based on  $\eta(t)^4$ ). In most physical situations, coarse graining amounts to averaging over the noise, which renders any noise GAUSSIAN.

The position  $x(t)$  in the equation of motion Eq. (4.1) is to be considered a functional of  $\eta(t)$ . In fact, we can simply integrate<sup>3</sup> and find

$$x(t) = x_0 + \int_{t_0}^t dt' \eta(t'). \quad (4.3)$$

<sup>1</sup>Notice that we traded the deterministic equation of motion for the PDF against a stochastic equation for the observable.

<sup>2</sup>Is it possible to show that a process described by the LANGEVIN equation is non-MARKOVian if the noise is not  $\delta$ -correlated? Adding observables is not allowed (such as using an O-U process to generate correlated noise and use that to drive another O-U process, and then include both process in a joint observable, which of course would still be MARKOVian, even with the second being subject to correlated noise). Yes! See Nico van Kampen, *Braz. J. Phys*, **28**, 90 (1998).

<sup>3</sup>Very often it is not that simple and the actual rules of evaluation of the right hand side, in particular if there are terms in  $x(t)$  whose instantaneous value depends on the noise at the same instant in time, are crucial.

implying  $x_0 = x(t_0)$ . Although the entire history of  $x(t)$  is fixed given  $\eta(t)$  and its value at *any* point in time, one usually sets  $t_0 = 0$  and considers only  $t \geq 0$ . The mean position is

$$\langle x(t) \rangle = \langle x_0 \rangle + \left\langle \int_{t_0}^t dt' \eta(t') \right\rangle = x_0 \quad (4.4)$$

where we have used  $\langle \eta(t') \rangle = 0$ . The correlation function  $\langle x(t_1)x(t_2) \rangle$  now is

$$\begin{aligned} \langle x(t_1)x(t_2) \rangle &= x_0^2 + \left\langle \int_{t_0}^{t_1} dt'_1 \int_{t_0}^{t_2} dt'_2 \eta(t'_1)\eta(t'_2) \right\rangle = x_0^2 + \int_{t_0}^{t_1} dt'_1 \int_{t_0}^{t_2} dt'_2 \langle \eta(t'_1)\eta(t'_2) \rangle \\ &= x_0^2 + \int_{t_0}^{t_1} dt'_1 \int_{t_0}^{t_2} dt'_2 2\Gamma^2 \delta(t'_1 - t'_2) \quad (4.5) \end{aligned}$$

To evaluate this integral we specify  $t_2 \geq t_1$ , so that the resulting  $\delta$ -function can be integrated over  $t'_2$  and is guaranteed to contribute for all  $t'_1$ . Performing the integration the other way around is more complicated because for all  $t'_2 > t_1$  the integral over  $t'_1$  would not contribute. We therefore have

$$\langle x(t_1)x(t_2) \rangle = x_0^2 + 2\Gamma^2 \min(t_1, t_2) \quad (4.6)$$

Choosing  $t_1 = t_2$  we note that the variance increases linearly in time,

$$\langle x(t)^2 \rangle_c = 2\Gamma^2 t. \quad (4.7)$$

The combination

$$\langle x(t_1)x(t_2) \rangle - \langle x(t_1) \rangle \langle x(t_2) \rangle = \langle (x(t_1) - \langle x(t_1) \rangle) (x(t_2) - \langle x(t_2) \rangle) \rangle \quad (4.8)$$

is known as the **two-time correlation function**, sometimes written just as  $\langle x(t_1)x(t_2) \rangle_c$  (notice the subscript c). In the present case

$$\langle x(t_1)x(t_2) \rangle - \langle x(t_1) \rangle \langle x(t_2) \rangle = 2\Gamma^2 \min(t_1, t_2) \quad (4.9)$$

so that the diffusion constant of the random walk is given by  $D = \Gamma^2$  (see the remark at the end of Section 1.5).

**Exercise 4:** Verify Eq. (4.8).

All higher cumulants of the noise vanish (because the noise is GAUSSIAN) and consequently all higher cumulants of  $x(t)$  vanish as well. The GAUSSIAN noise therefore generates a GAUSSIAN position  $x(t)$ , but its variance is linearly increasing in time. This is the same feature as in the central limit theorem: The observable  $y(t) = x(t)/\sqrt{t}$  has a constant variance equal to that of the noise, *i.e.* rescaling space by  $\sqrt{t}$  keeps a random walker confined.

## 4.2 ORNSTEIN-UHLENBECK process

The second process to be discussed in terms of a LANGEVIN equation is of fundamental importance: It is the *only* MARKOVian, stationary, GAUSSian process. Or to put it more clearly: Every process, which is MARKOVian, stationary and GAUSSian is (by DOBB's theorem) the ORNSTEIN-UHLENBECK process.

Its stationarity is somehow "enforced" by specifying the right initial distribution. The process converges to the same PDF for any initial condition. By picking this asymptotic PDF as the initial PDF of the process implies that it will not change over time.

The LANGEVIN equation for the ORNSTEIN-UHLENBECK process is (with any initial condition)

$$\dot{x}(t) = \eta(t) - \gamma x(t) \quad (4.10)$$

The new term compared Eq. (4.1) is the "pull-back" term  $-\gamma x(t)$ . Whenever  $x(t)$  strays away too far from the origin, its average velocity points back to the origin if  $\gamma > 0$ , because  $\langle \dot{x} \rangle(t) = -\gamma \langle x \rangle(t)$ , so that

$$\langle x \rangle(t; x_0) = x_0 e^{-\gamma t} \quad (4.11)$$

with  $x_0 = x(t=0)$  fixed.<sup>4</sup> The complete ORNSTEIN-UHLENBECK process includes a PDF of those,

$$\mathcal{P}_{OU}(x_0) = \sqrt{\frac{\gamma}{2\pi\Gamma^2}} e^{-\frac{x_0^2 \gamma}{2\Gamma^2}} \quad (4.12)$$

and by including that into the averaging one has immediately (by symmetry)  $\langle x \rangle(t) = 0$ .

To calculate higher moments, it is worthwhile writing down the formal solution of Eq. (4.10):

$$x(t; x_0) = x_0 e^{-\gamma t} + \int_0^t dt' \eta(t') e^{-\gamma(t-t')} \quad (4.13)$$

Taking the ensemble average reproduces Eq. (4.11), while the correlation function becomes

$$\langle x(t_1)x(t_2) \rangle(x_0) = x_0^2 e^{-\gamma(t_1+t_2)} + 2\Gamma^2 \int_0^{t_1} dt'_1 \int_0^{t_2} dt'_2 \delta(t'_1 - t'_2) e^{-\gamma((t_1+t_2)-(t'_1+t'_2))} \quad (4.14)$$

We recognise the first term on the right as  $x_0^2 \exp(-\gamma(t_1 + t_2)) = \langle x \rangle(t_1; x_0) \langle x \rangle(t_2; x_0)$ .

Again, we choose  $t_2 \geq t_1$ , so that the integration over  $t'_2$  can be done immediately, leaving essentially only the integral  $\int_0^{t_1} dt'_1 \exp(2\gamma t'_1) = (2\gamma)^{-1}(\exp(2\gamma t_1) - 1)$  and there-

---

<sup>4</sup>Notice the notation: The parameters  $(t; x_0)$  are outside the average, indicating that the average itself depends on it. The time  $t$  could obviously also stand inside the average, because averaging over the position at a certain time, is just the same as the position as a function of time. But we will sometimes average over  $x_0$  as well, so we want to distinguish between the situation where  $x_0$  is a parameter (standing outside) and being averaged over as well (inside). Of course, the meaning of  $\langle \cdot \rangle$  averaging over  $x_0$  or not actually changes in the two situations.

fore

$$\langle x(t_1)x(t_2) \rangle (x_0) = x_0^2 e^{-\gamma(t_1+t_2)} + \frac{\Gamma^2}{\gamma} \left( e^{-\gamma(t_2-t_1)} - e^{-\gamma(t_2+t_1)} \right) \quad (4.15)$$

so that the two point correlation function for fixed  $x_0$  becomes

$$\langle x(t_1)x(t_2) \rangle (x_0) - \langle x \rangle (t_1; x_0) \langle x \rangle (t_2; x_0) = \frac{\Gamma^2}{\gamma} \left( e^{-\gamma(t_2-t_1)} - e^{-\gamma(t_2+t_1)} \right) \quad (4.16)$$

Choosing again  $t_1 = t_2 = t$  gives a variance of the position that increases exponentially slowly in time (compare to Eq. (4.7)), converging to  $\Gamma^2/\gamma$ :

$$\langle x(t; x_0)x(t; x_0) \rangle - \langle x \rangle (t; x_0) \langle x \rangle (t; x_0) = \frac{\Gamma^2}{\gamma} (1 - e^{-2\gamma t}) \quad (4.17)$$

Note: If we take the limit  $\gamma \rightarrow 0$  of this expression or more generally Eq. (4.16), then we should obtain the two-point correlation function of the random walker studied in Section 4.1. In fact

$$\lim_{\gamma \rightarrow 0} \frac{\Gamma^2}{\gamma} \left( e^{-\gamma(t_2-t_1)} - e^{-\gamma(t_2+t_1)} \right) = 2\Gamma^2 t_1 \quad (4.18)$$

exactly the same as Eq. (4.9) with  $t_1 \leq t_2$ .

Averaging over the initial condition is a bit dangerous:<sup>5</sup> Wherever there is a  $\langle \cdot \rangle$ , we could just integrate over  $x_0$  weighted by  $\mathcal{P}_{OU}(x_0)$ . That way,  $\langle \cdot \rangle$  includes the averaging over  $x_0$ . However, when we replaced  $x_0^2 \exp(-\gamma(t_1 + t_2))$  we used  $\langle x \rangle (t; x_0) = x_0 \exp(-\gamma t)$ . When averaging over  $x_0$  we have  $\langle x \rangle (t; x_0) = 0$  and we cannot make that substitution. So, for the “real ORNSTEIN-UHLENBECK process” (including the averaging over  $x_0$ ), we have  $\langle x \rangle (t) = 0$  and what remains is the average over Eq. (4.15):

$$\langle x(t_1)x(t_2) \rangle - \langle x \rangle (t_1) \langle x \rangle (t_1) = \int dx_0 \mathcal{P}_{OU}(x_0) \left\{ x_0^2 e^{-\gamma(t_1+t_2)} + \frac{\Gamma^2}{\gamma} \left( e^{-\gamma(t_2-t_1)} - e^{-\gamma(t_2+t_1)} \right) \right\} \quad (4.19)$$

The only term that is actually affected by that averaging is the first term on the right, proportional to  $x_0^2$ , but from Eq. (4.12) we know it contributes  $\Gamma^2\gamma^{-1}$ , so the terms in  $\exp(-\gamma(t_1 + t_2))$  cancel and all what is left of the correlation function is

$$\langle x(t_1)x(t_2) \rangle - \langle x \rangle (t_1) \langle x \rangle (t_1) = \frac{\Gamma^2}{\gamma} e^{-\gamma(t_2-t_1)} \quad (4.20)$$

which is in fact stationary, because it depends only on the relative time, not the absolute. The same applies to higher moments and similar to Section 4.1, all higher cumulants van-

---

<sup>5</sup>The initial distribution is chosen to be the asymptotic distribution, so the correlation function should depend only on time differences after taking the average over  $x_0$ . Yet, there is no way to get rid of the  $\exp(-\gamma(t_2 + t_1))$  in Eq. (4.16) by averaging over  $x_0$ . The problem is that we have subtracted the average position  $x_0 \exp(-\gamma t)$  before we have taken the average over  $x_0$ . If we average naively over Eq. (4.16) using  $\mathcal{P}_{OU}(x_0)$  we would effectively confuse the average of the square of  $x_0$  with the square of its average.

ish, meaning that the PDF of  $x(t)$  is GAUSSIAN. However, this time the variance remains constant,  $\Gamma^2/\gamma$ , and the average vanishes – in fact, the PDF of  $x(t)$  equals that of  $x_0$  for all times.

An interesting question to follow up is how LANGEVIN equations with GAUSSIAN, additive noise can ever produce something non-GAUSSIAN. The only way to break free from the central limit theorem is through algebraic correlations, which have to be generated in the presence of the noise by the propagator (in the PDE sense, the solution of the LANGEVIN equation without noise, rather than in the field theoretic sense, the solution of the linear part of the LANGEVIN equation).

It will turn out that purely relaxational LANGEVIN equations (with the RHS derived from a HAMILTONIAN) have a stationary distribution given by the HAMILTONIAN. These models are “out of equilibrium”, not “far from equilibrium”. As long as  $\phi$  is only time dependent (mean field theory of model A, for instance), the resulting PDF can be read off from the HAMILTONIAN immediately. By choosing non-linear terms in the LANGEVIN equation, the PDF can be made arbitrarily non-GAUSSIAN. To stress that, such non-GAUSSIAN behaviour derived from a GAUSSIAN noise can only be produced by algebraic correlations.<sup>6</sup>

If  $\phi$  is space dependent and there is interaction in space (as in model A), then the coupling  $\gamma$  needs to be tuned to a very special value (corresponding to the critical temperature), not necessarily to 0 in order to produce non-GAUSSIAN behaviour in the large time and space limit. Again, the stationary PDF is given by the HAMILTONIAN, but it is not immediately obvious what long range and time behaviour this HAMILTONIAN produces. In fact, that latter question is subject of static critical phenomena.

---

<sup>6</sup>It would be wonderful to find an example for that, an exact solution or so.

## Chapter 5

# Critical dynamics

Dynamical, *i.e.* time-dependent behaviour can be obtained in equilibrium (*e.g.* the two point correlation function in the ideal gas), in non-equilibrium but near equilibrium (usually meaning relaxation to the equilibrium, *e.g.* model A, maybe best termed “out of equilibrium” or “near equilibrium”) or far from equilibrium where a constant flux prevents detailed balance. In this section, some general features of stochastic dynamics and dynamic scaling are presented. Critical dynamics generally are some time-dependent phenomena that display critical behaviour, *i.e.* algebraic decay of correlations at a critical point. The most common and probably the most convenient form to describe these systems is in terms of LANGEVIN equations, as introduced earlier.

Instead of considering the stationary state, one can define an “evolution rule” and ask how the system’s features change in time, for example, recover after a change in the external field. The time is a new parameter and enters the above scaling assumption just like any other parameter

$$f(\tau, h, t) = \lambda^{-d} f(\tau \lambda^{y_t}, h \lambda^{y_h}, t \lambda^{-z}) \quad (5.1)$$

Differentiating with respect to  $h$  and evaluating at  $h = 0$  and  $\tau = 0$  gives

$$m(0, 0, t) = \lambda^{y_h - d} m(0, 0, t \lambda^{-z}) \quad (5.2)$$

where we have used  $\partial_h f = -m$ . Using  $y_h - d = -\beta/\nu$  and  $t \lambda^{-z} = 1$  we have

$$m(0, 0, t) = t^{-\frac{\beta}{\nu z}} m(0, 0, 1) \quad (5.3)$$

This is a typical example for dynamical scaling behaviour: At the critical point, after having taken the thermodynamic limit, the system relaxes algebraically. Eq. (5.3) can be used to determine critical exponents. The system is prepared in a configuration away from equilibrium and over time it relaxes back to it. The only subtlety is that the single configuration the system is prepared into says very little about whether or not the system is in equilibrium. One has to repeat the procedure over and over with initial configurations

the ensemble of which is not the equilibrium ensemble.

The critical exponent  $z$  depends on the choice of the dynamics. In ferromagnetic models, GLAUBER dynamics is a typical choice. The exponent  $z$  is the “anomalous dimension” of the time and so the correlation time when simulating a ferromagnetic model diverges like  $L^z$  at the critical point. In order to determine good estimates of static quantities, correlations should be kept to a minimum and therefore evolution strategies are employed which have a dynamical exponent  $z$  close to 0. Cluster algorithms (SWENDSEN-WANG or WOLFF) perform particularly well in that respect.

In Section 5.4 various variations of the dynamics for  $\phi^4$  theories, such as the ISING model will be discussed, each characterised by its particular dynamical exponent.

## 5.1 From HAMILTONian to LANGEVIN equation and back

*A priori* a model studied in equilibrium critical phenomena does not have a dynamical updating rule associated with it. Say we are concerned with  $\phi^4$  theory, *i.e.* the HAMILTONian is

$$\mathcal{H}[\phi] = \int d^d x \frac{1}{2} (\nabla \phi)^2 + \frac{1}{2} r \phi^2 + \frac{u}{4!} \phi^4 + h(x) \phi(x) \quad (5.4)$$

which is a functional of the order parameter field  $\phi(x)$ . The external, local field is represented by  $h(x)$ . A naïve relaxation of that field minimises that Hamiltonian,

$$\dot{\phi} = -D \frac{\delta \mathcal{H}}{\delta \phi} \quad (5.5)$$

where  $\frac{\delta \mathcal{H}}{\delta \phi}$  is the functional derivative of the HAMILTONian with respect to the field  $\phi$  and  $D$  is a coefficient effectively setting the time scale, the **relaxation rate**. If  $\phi$  minimises the HAMILTONian, then a positive perturbation will lead to a positive “slope”, so that the response is negative and vice versa. For the  $\phi^4$  HAMILTONian the equation of motion therefore is

$$\dot{\phi} = D(\nabla^2 \phi - r\phi + \frac{u}{6} \phi^3 + h) \quad (5.6)$$

where we have used

$$\frac{\delta}{\delta \phi(\mathbf{x}')} (\nabla \phi(\mathbf{x}))^2 = -2 \nabla^2 \phi(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}') . \quad (5.7)$$

Eq. (5.5) describes a deterministic relaxation. To account for the noise present in a thermodynamic system due to the heat-bath (one might wonder where the noise is coming from in the bath), we add a noise term  $\eta(\mathbf{x}, t)$  and arrive at

$$\dot{\phi}(\mathbf{x}, t) = D \left( \nabla^2 \phi(\mathbf{x}, t) - r\phi(\mathbf{x}, t) + \frac{u}{6} \phi(\mathbf{x}, t)^3 + h(\mathbf{x}, t) \right) + \eta(\mathbf{x}, t) \quad (5.8)$$

This time, the noise has correlator

$$\langle \eta(\mathbf{x}, t) \eta(\mathbf{x}', t') \rangle = 2\Gamma^2 \delta(t - t') \delta(\mathbf{x} - \mathbf{x}'), \quad (5.9)$$

compare to Eq. (4.2).

Was it not for the non-linear term  $(u/6)\phi(\mathbf{x}, t)^3$ , the formal solution of this LANGEVIN could be written down using a GREEN function approach.

Eq. (5.8) is also known as model A or GLAUBER dynamics and will be discussed in a wider context in Section 5.4. For the remainder of the present section, we discuss the features of LANGEVIN equations constructed as a relaxation process towards equilibrium

$$\dot{\phi}(\mathbf{x}, t) = -D \left. \frac{\delta \mathcal{H}([\psi])}{\delta \psi(\mathbf{x})} \right|_{\phi(\mathbf{x})=\phi(\mathbf{x}, t)} + \eta(\mathbf{x}, t) \quad (5.10)$$

Note that the HAMILTONIAN is a functional of a field  $\psi(\mathbf{x})$  that depends on space only, but evaluated for one that is time-dependent as well. That is important, because, for example, a harmonic potential would be represented by  $\int d^d \mathbf{x}' \psi(\mathbf{x}')^2$ , not containing any time. Replacing  $\psi(\mathbf{x})$  by  $\phi(\mathbf{x}, t)$  and differentiating functionally with respect to the latter gives

$$\frac{\delta}{\delta \phi(\mathbf{x}, t)} \int d^d \mathbf{x}' \phi(\mathbf{x}', t')^2 = 2\phi(\mathbf{x}, t) \delta(t - t') \quad (5.11)$$

with an undesired extra factor  $\delta(t - t')$  on the right.

**Exercise 5:** Find the formal solution of Eq. (5.8) with  $u = 0$  and  $\mathfrak{h} \equiv 0$ , and determine the correlation function  $\langle \phi(\mathbf{x}, t) \phi(\mathbf{x}', t') \rangle - \langle \phi(\mathbf{x}, t) \rangle \langle \phi(\mathbf{x}', t') \rangle$  in the limit of large times  $t, t'$  as a function of the difference  $t - t' \geq 0$ . Consider, in particular,  $r = 0$ .

## 5.2 The PDF of $\eta$

In order to make further progress, we need to specify the PDF of  $\eta$ . We know that it is GAUSSIAN, has vanishing mean and a  $\delta$ -correlator. To get started, we discretise space, so that the DIRAC  $\delta$  function turn into a KRONECKER divided by the time or volume element. To ease notation, we go through the procedure only for the noise in time, Eq. (4.2), which now reads

$$\langle \eta_i \eta_j \rangle = 2\Gamma^2 \delta_{ij} \Delta t^{-1} \quad (5.12)$$

where  $\eta_i$  means  $\eta(t_i)$  at the discrete time  $t_i$ , which ticks in units of  $\Delta t$ . With the  $\eta_i$  mutually independent and having vanishing mean, a product measure seems suitable, and individual  $\eta_i$  are distributed according to

$$\mathcal{P}_i(\eta) = \sqrt{\frac{\Delta t}{4\pi\Gamma^2}} e^{-\frac{\eta^2 \Delta t}{4\Gamma^2}} \quad (5.13)$$

and so

$$\mathcal{P}(\eta_1, \dots, \eta_n) = \left( \frac{\Delta t}{4\pi\Gamma^2} \right)^{n/2} e^{-\frac{\Delta t \sum_i^n \eta_i^2}{4\Gamma^2}} \quad (5.14)$$

In order to take the continuum limit, we have to drop the normalisation,

$$\mathcal{P}([\eta(t)]) \propto e^{-\frac{1}{4\Gamma^2} \int dt \eta(t)^2}. \quad (5.15)$$

Any average over the noise can now be written as

$$\langle \cdot \rangle = \int \mathcal{D}\eta \mathcal{P}([\eta(t)]). \quad (5.16)$$

where  $\mathcal{D}\eta$  stands for  $\prod d\eta_i$  if time is discretised again. The moment generating function of the noise is  $\langle \exp(\int dt \eta h(t)) \rangle$  with  $h(t)$  a function of time. Completing the square gives

$$-\frac{1}{4\Gamma^2} \eta(t)^2 + \eta(t)h(t) = -\frac{1}{4\Gamma^2} (\eta(t) - 2\Gamma^2 h(t))^2 + \Gamma^2 h(t)^2 \quad (5.17)$$

so that  $\langle \exp(\int dt \eta h(t)) \rangle = \exp(\int dt \Gamma^2 h(t)^2)$ . Taking the logarithm and differentiating twice (functionally) with respect to  $h(t)$  gives the correlator

$$\frac{\delta^2}{\delta h(t) \delta h(t')} \ln \langle e^{\int dt \eta h(t)} \rangle = \frac{\delta^2}{\delta h(t) \delta h(t')} \int dt \Gamma^2 h(t)^2 = 2\Gamma^2 \delta(t - t') \quad (5.18)$$

reproducing the correlator Eq. (4.2).

The procedure can be generalised for  $\eta$  that is time and space dependent with

$$\mathcal{P}([\eta(\mathbf{x}, t)]) \propto e^{-\frac{1}{4\Gamma^2} \int dt d^d x \eta(\mathbf{x}, t)^2} \quad (5.19)$$

Given a LANGEVIN equation of the form  $\partial_t \phi(\mathbf{x}, t) = -\mathcal{F}[\phi] + \eta(\mathbf{x}, t)$  one can now ask for average of functionals of  $\phi(\mathbf{x}, t)$  which are solutions of the LANGEVIN equation, *i.e.* themselves functionals of the noise. Because the determinant of the JACOBIAN  $\mathcal{D}\phi/\mathcal{D}\eta$  turns out to be unity<sup>1</sup> one then has

$$\langle \cdot \rangle = \int \mathcal{D}\phi \exp \left( -\frac{1}{4\Gamma^2} \int dt d^d x [\partial_t \phi(\mathbf{x}, t) - \mathcal{F}[\phi]]^2 \right). \quad (5.20)$$

where we have used  $\eta(\mathbf{x}, t) = \partial_t \phi(\mathbf{x}, t) - \mathcal{F}[\phi]$ . Analysing Eq. (5.10) along these lines, one chooses

$$-\mathcal{F}[\phi(\mathbf{x}, t)] = D \frac{\delta \mathcal{H}([\psi])}{\delta \psi(\mathbf{x})} \Big|_{\phi(\mathbf{x}) = \psi(\mathbf{x}, t)} = D \mathcal{H}'([\phi(\mathbf{x}, t)]) \quad (5.21)$$

<sup>1</sup>Rather, it can be chosen to be unity. The choice comes from the fact that LANGEVIN equations with a noise that has infinite variance actually make little sense, and to render them mathematically well-defined one has to make a choice about their interpretation.

and arrives at the ONSAGER-MACHLUP functional

$$\langle \cdot \rangle = \int \mathcal{D}\phi \exp \left( -\frac{1}{4\Gamma^2} \int dt' d^d x' [\partial_t \phi(\mathbf{x}', t') + D\mathcal{H}'([\phi(\mathbf{x}', t')])]^2 \right). \quad (5.22)$$

### 5.3 A FOKKER-PLANCK equation approach

From Eq. (5.22) one can derive a FOKKER-PLANCK equation (partly from “Quantum Field Theory and Critical Phenomena” by ZINN-JUSTIN). To simplify the derivation, we will first derive the FOKKER-PLANCK equation for a space independent field  $\phi(t)$ . The LANGEVIN equation to be considered is

$$\dot{\phi}(t) = -D \partial_\psi \big|_{\phi(t)} \mathcal{H}(\psi) + \eta(t), \quad (5.23)$$

see Eq. (5.10). To ease notation, we write  $\partial_\psi \big|_{\phi(t)} \mathcal{H}(\psi) = \mathcal{H}'(\phi)$ .

The probability that  $\phi$  has the value  $\phi_0$  at time  $t$  is

$$\mathcal{P}_\phi(\phi_0; t) = \langle \delta(\phi(t) - \phi_0) \rangle \quad (5.24)$$

To find the time evolution of  $\mathcal{P}_\phi(\phi_0; t)$ , we differentiate with respect to  $t$ , which gives

$$\partial_t \mathcal{P}_\phi(\phi_0; t) = \partial_t \langle \delta(\phi(t) - \phi_0) \rangle \quad (5.25a)$$

$$= \left\langle \dot{\phi}(t) \frac{\partial}{\partial \phi} \delta(\phi(t) - \phi_0) \right\rangle \quad (5.25b)$$

If  $\langle \cdot \rangle$  runs over  $\eta$ , the variable  $\phi(t)$  is, via the LANGEVIN equation, after integration a functional of  $\eta$ . In that case, the LANGEVIN equation is implied whenever  $\phi(t)$  is considered. If, on the other hand, the path integral runs over  $\phi$ , one can at any time introduce a new dummy variable  $\eta(t) = \partial_t \phi(x) + D\mathcal{H}'([\phi(x)])$ , which produces, as we know, a JACOBIAN with determinant 1. However we look at it, we can always replace  $\dot{\phi}(t)$  in Eq. (5.25b) by the LANGEVIN equation.

The term  $\partial_\phi \delta(\phi - \phi_0)$  can be replaced by  $-\partial_{\phi_0} \delta(\phi - \phi_0)$ , which can be taken out of the average, so that

$$\partial_t \mathcal{P}_\phi(\phi_0; t) = -\partial_{\phi_0} \langle (-D\mathcal{H}'(\phi(t)) + \eta(t)) \delta(\phi(t) - \phi_0) \rangle \quad (5.26a)$$

Because of the  $\delta$ -function  $\mathcal{H}'$  can be equally evaluated at  $\phi_0$  and taken out of the average as well,

$$\langle -D\mathcal{H}'(\phi(t)) \delta(\phi(t) - \phi_0) \rangle = -D\mathcal{H}'(\phi_0) \langle \delta(\phi(t) - \phi_0) \rangle = -D\mathcal{H}'(\phi_0) \mathcal{P}_\phi(\phi_0; t). \quad (5.27)$$

The other term of Eq. (5.26) to be considered is  $\langle \eta(t) \delta(\phi(t) - \phi_0) \rangle$ . Writing the average

again as a path integral over  $\eta$ , we note that

$$\int \mathcal{D}\eta \frac{\delta}{\delta\eta(t)} \exp\left(-\frac{1}{4\Gamma^2} \int dt' \eta(t')^2\right) = \int \mathcal{D}\eta \left(-\frac{1}{2\Gamma^2} \eta(t)\right) \exp\left(-\frac{1}{4\Gamma^2} \int dt' \eta(t')^2\right) \quad (5.28)$$

Using a functional integration by parts, one therefore obtains

$$\begin{aligned} \langle \eta(t) \delta(\phi(t) - \phi_0) \rangle &= -2\Gamma^2 \int \mathcal{D}\eta \delta(\phi(t) - \phi_0) \frac{\delta}{\delta\eta(t)} \exp\left(-\frac{1}{4\Gamma^2} \int dt' \eta(t')^2\right) \\ &= 2\Gamma^2 \int \mathcal{D}\eta \exp\left(-\frac{1}{4\Gamma^2} \int dt' \eta(t')^2\right) \frac{\delta}{\delta\eta(t)} \delta(\phi(t) - \phi_0) \\ &= 2\Gamma^2 \left\langle \frac{\delta}{\delta\eta(t)} \delta(\phi(t) - \phi_0) \right\rangle \end{aligned} \quad (5.29)$$

As mentioned above,  $\phi(t)$  is a functional of  $\eta$  (more precisely, an integral) and one can choose

$$\frac{\delta}{\delta\eta(t)} \phi(t) = \frac{1}{2} \quad (5.30)$$

so that

$$\left\langle \frac{\delta}{\delta\eta(t)} \delta(\phi(t) - \phi_0) \right\rangle = -\frac{1}{2} \partial_{\phi_0} \langle \delta(\phi(t) - \phi_0) \rangle = -\frac{1}{2} \partial_{\phi_0} \mathcal{P}_\phi(\phi_0; t) \quad (5.31)$$

again turning the derivative with respect to  $\phi(t)$  of the DIRAC  $\delta$  function into a negative one with respect to  $\phi_0$ . Collecting all terms, one has the FOKKER-PLANCK equation

$$\partial_t \mathcal{P}_\phi(\phi_0; t) = \partial_{\phi_0} (D\mathcal{H}'(\phi_0) \mathcal{P}_\phi(\phi_0; t)) + \Gamma^2 \partial_{\phi_0}^2 \mathcal{P}_\phi(\phi_0; t). \quad (5.32)$$

In the stationary state  $\partial_t \mathcal{P}_\phi(\phi_0; t) = 0$  and therefore

$$\partial_{\phi_0} (D\mathcal{H}'(\phi_0) \mathcal{P}_\phi(\phi_0; t) + \Gamma^2 \partial_{\phi_0} \mathcal{P}_\phi(\phi_0; t)) = 0 \quad (5.33)$$

one solution<sup>2</sup> of which is

$$\mathcal{P}_{\phi; \text{stat}}(\phi) \propto e^{-\frac{D}{\Gamma^2} \mathcal{H}(\phi)} \quad (5.34)$$

This result can be extended to space dependent order parameters, so that the stationary PDF becomes the BOLTZMANN-factor

$$\mathcal{P}_{\phi; \text{stat}}([\phi(\mathbf{x})]) \propto e^{-\frac{D}{\Gamma^2} \mathcal{H}([\phi])}, \quad (5.35)$$

where  $D/\Gamma^2$  can be identified with the inverse temperature  $1/(k_B T)$ .

---

<sup>2</sup>Apparently  $D\mathcal{H}'(\phi_0) \mathcal{P}_\phi(\phi_0; t) + \Gamma^2 \partial_{\phi_0} \mathcal{P}_\phi(\phi_0; t) = C_1$  and the general solution of  $\mathcal{P}_\phi(\phi_0; t)' = C_1 - \frac{D}{\Gamma^2} \mathcal{H}'(\phi_0) \mathcal{P}_\phi(\phi_0; t)$  is  $\mathcal{P}_\phi(\phi_0; t) = \exp\left(-\frac{D}{\Gamma^2} \mathcal{H}(\phi_0)\right) \left(C_2 + C_1 \int_0^{\phi_0} d\phi \exp\left(\frac{D}{\Gamma^2} \mathcal{H}(\phi)\right)\right)$

**Exercise 6:** Find  $\langle \phi(\mathbf{x}, t)\phi(\mathbf{x}', t') \rangle - \langle \phi(\mathbf{x}, t) \rangle \langle \phi(\mathbf{x}', t') \rangle$  in the stationary state, using Eq. (5.35) with the Hamiltonian Eq. (5.4) with  $u = 0$ ,  $h \equiv 0$  and  $t = t'$  (as we consider the stationary state,  $t$  cannot be different from  $t'$  — there is no time left). Compare the result to the previous exercise.

## 5.4 The HOHENBERG-HALPERIN models

The way we constructed the relaxational dynamics (5.5) does not prescribe the way the noise enters. That choice was made later, Eq. (5.8). HOHENBERG and HALPERIN introduced (RMP 1977) a now well-established classification of models, which differ in the way the noise is implemented as well as in the general physical features.

One model has been mentioned already, namely **model A**, also known as **GLAUBER dynamics**. It is defined by the Langevin equation

$$\dot{\phi}(\mathbf{x}, t) = D \left( \nabla^2 \phi(\mathbf{x}, t) - r\phi(\mathbf{x}, t) + \frac{u}{6} \phi(\mathbf{x}, t)^3 + h(\mathbf{x}, t) \right) + \eta(\mathbf{x}, t), \quad (5.36)$$

(see Eq. (5.8)) with noise correlator

$$\langle \eta(\mathbf{x}, t)\eta(\mathbf{x}', t') \rangle = 2\Gamma^2 \delta(\mathbf{x} - \mathbf{x}') \delta(t - t'). \quad (5.37)$$

It is probably the best studied, non-trivial model in non-equilibrium statistical mechanics.

The second most important model is **model B** (also known as **KAWASAKI dynamics**) has **conserved order parameter**,

$$\dot{\phi}(\mathbf{x}, t) = -\nabla^2 D \left( \nabla^2 \phi(\mathbf{x}, t) - r\phi(\mathbf{x}, t) + \frac{u}{6} \phi(\mathbf{x}, t)^3 + h(\mathbf{x}, t) \right) + \zeta(\mathbf{x}, t), \quad (5.38)$$

with noise  $\zeta = \nabla \eta$ , so that the entire right hand side can be written as a gradient. The integral  $\int d^d x \dot{\phi}(\mathbf{x}, t) = \partial_t \int d^d x \phi(\mathbf{x}, t)$  therefore vanishes (ignoring surface terms).

**Model C and D** introduce a conserved energy density  $\rho$  and combine that with non-conserved order parameter (model C) and conserved order parameter (model D) respectively. If the order parameter is not a scalar but higher dimensional, then its dynamics is described by **model J** and by adding an anisotropy, that turns into **model E**. Finally, model E becomes **model G** by an anti-ferromagnetic coupling constant.