Introduction to the theory of stochastic processes and Brownian motion problems

Lecture notes for a graduate course,

by J. L. García-Palacios (Universidad de Zaragoza)

May 2004

These notes are an introduction to the *theory of stochastic processes* based on several sources. The presentation mainly follows the books of van Kampen [5] and Wio [6], except for the introduction, which is taken from the book of Gardiner [2] and the parts devoted to the Langevin equation and the methods for solving Langevin and Fokker–Planck equations, which are based on the book of Risken [4].

Contents

1 Historical introduction

Theoretical science up to the end of the nineteenth century can be roughly viewed as the study of solutions of differential equations and the modelling of natural phenomena by deterministic solutions of these differential equations. It was at that time commonly thought that if all initial (and contour) data could only be collected, one would be able to predict the future with certainty.

We now know that this is not so, in at least two ways. First, the advent of *quantum mechanics* gave rise to a new physics, which had as an essential ingredient a purely statistical element (the measurement process). Secondly, the concept of *chaos* has arisen, in which even quite simple differential equations have the rather alarming property of giving rise to essentially unpredictable behaviours.

Chaos and quantum mechanics are not the subject of these notes, but we shall deal with systems were limited predictability arises in the form of fluctuations due to the finite number of their discrete constituents, or interaction with its environment (the "thermal bath"), etc. Following Gardiner [2] we shall give a semi-historical outline of how a phenomenological *theory of fluctuating phenomena* arose and what its essential points are.

The experience of careful measurements in science normally gives us data like that of Fig. 1, representing the time evolution of a certain variable X. Here a quite well defined deterministic trend is evident, which is re-

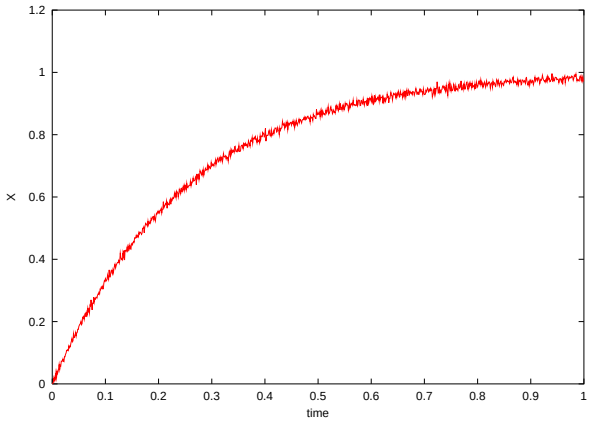

Figure 1: Schematic time evolution of a variable X with a well defined deterministic motion plus fluctuations around it.

producible, unlike the fluctuations around this motion, which are not. This evolution could represent, for instance, the growth of the (normalised) number of molecules of a substance X formed by a chemical reaction of the form $A \rightleftharpoons X$, or the process of charge of a capacitor in a electrical circuit, etc.

1.1 Brownian motion

The observation that, when suspended in water, small pollen grains are found to be in a very animated and irregular state of motion, was first systematically investigated by Robert Brown in 1827, and the observed phenomenon took the name of *Brownian motion*. This motion is illustrated in Fig. 2. Being a botanist, he of course tested whether this motion was in some way a manifestation of life. By showing that the motion was present in any suspension of fine particles —glass, mineral, etc.— he ruled out any specifically organic origin of this motion.

1.1.1 Einstein's explanation (1905)

A satisfactory explanation of Brownian motion did not come until 1905, when Einstein published an article entitled *Concerning the motion, as required by the molecular-kinetic theory of heat, of particles suspended in liquids at rest*. The same explanation was independently developed by Smoluchowski

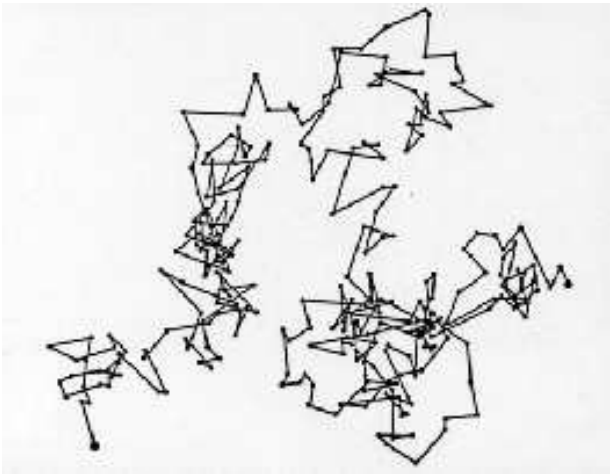

Figure 2: Motion of a particle undergoing Brownian motion.

in 1906, who was responsible for much of the later systematic development of the theory. To simplify the presentation, we restrict the derivation to a one-dimensional system.

There were two major points in Einstein's solution of the problem of Brownian motion:

- The motion is caused by the exceedingly frequent impacts on the pollen grain of the incessantly moving molecules of liquid in which it is suspended.

- The motion of these molecules is so complicated that its effect on the pollen grain can only be described probabilistically in term of exceedingly frequent statistically independent impacts.

Einstein development of these ideas contains all the basic concepts which make up the subject matter of these notes. His reasoning proceeds as follows: "It must clearly be assumed that each individual particle executes a motion which is *independent of the motions of all other particles*: it will also be considered that *the movements of one and the same particle in different time intervals are independent processes*, as long as these time intervals are not chosen too small."

"We introduce a time interval τ into consideration, which is very small compared to the observable time intervals, but nevertheless so large that in two successive time intervals τ, the motions executed by the particle can be thought of as events which are independent of each other."

"Now let there be a total of n particles suspended in a liquid. In a time interval τ, the X-coordinates of the individual particles will increase by an amount Δ, where for each particle Δ has a different (positive or negative) value. There will be a certain *frequency law* for Δ; the number $\mathrm{d}n$ of the particles which experience a shift between Δ and $\Delta + \mathrm{d}\Delta$ will be expressible by an equation of the form: $\mathrm{d}n = n\,\phi(\Delta)\mathrm{d}\Delta$, where $\int_{-\infty}^{\infty} \phi(\Delta)\mathrm{d}\Delta = 1$, and ϕ is only different from zero for very small values of Δ, and satisfies the condition $\phi(-\Delta) = \phi(\Delta)$."

"We now investigate how the diffusion coefficient depends on ϕ. We shall restrict ourselves to the case where the number of particles per unit volume depends only on x and t."

"Let $f(x,t)$ be the number of particles per unit volume. We compute the distribution of particles at the time $t + \tau$ from the distribution at time t. From the definition of the function $\phi(\Delta)$, it is easy to find the number of

particles which at time $t+\tau$ are found between two planes perpendicular to the x-axis and passing through points x and $x+\mathrm{d}x$. One obtains:

$$\boxed{f(x,t+\tau)\mathrm{d}x = \mathrm{d}x \int_{-\infty}^{\infty} f(x+\Delta,t)\phi(\Delta)\mathrm{d}\Delta \; .} \tag{1.1}$$

But since τ is very small, we can set

$$f(x,t+\tau) = f(x,t) + \tau\frac{\partial f}{\partial t} \; .$$

Furthermore, we expand $f(x+\Delta,t)$ in powers of Δ:

$$f(x+\Delta,t) = f(x,t) + \Delta\frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2!}\frac{\partial^2 f(x,t)}{\partial x^2} + \cdots \; .$$

We can use this series under the integral, because only small values of Δ contribute to this equation. We obtain

$$\boxed{f + \tau\frac{\partial f}{\partial t} = f\int_{-\infty}^{\infty}\phi(\Delta)\mathrm{d}\Delta + \frac{\partial f}{\partial x}\int_{-\infty}^{\infty}\Delta\,\phi(\Delta)\mathrm{d}\Delta + \frac{\partial^2 f}{\partial x^2}\int_{-\infty}^{\infty}\frac{\Delta^2}{2}\phi(\Delta)\mathrm{d}\Delta \; .} \tag{1.2}$$

Because $\phi(-\Delta) = \phi(\Delta)$, the second, fourth, etc. terms on the right-hand side vanish, while out of the 1st, 3rd, 5th, etc., terms, each one is very small compared with the previous. We obtain from this equation, by taking into consideration

$$\int_{-\infty}^{\infty}\phi(\Delta)\mathrm{d}\Delta = 1 \; .$$

and setting

$$\frac{1}{\tau}\int_{-\infty}^{\infty}\frac{\Delta^2}{2}\phi(\Delta)\mathrm{d}\Delta = D \; , \tag{1.3}$$

and keeping only the 1st and 3rd terms of the right hand side,

$$\boxed{\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2} \; .} \tag{1.4}$$

This is already known as the differential equation of diffusion and it can be seen that D is the diffusion coefficient."

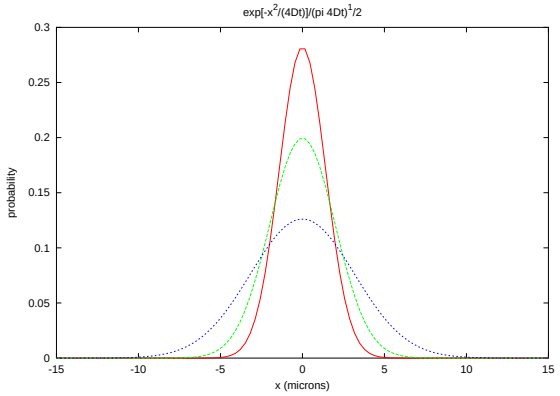

Figure 3: Time evolution of the non-equilibrium probability distribution (1.5).

"The problem, which correspond to the problem of diffusion from a single point (neglecting the interaction between the diffusing particles), is now completely determined mathematically: its solution is

$$f(x,t) = \frac{1}{\sqrt{\pi\, 4Dt}}\, e^{-x^2/4Dt} \ . \tag{1.5}$$

This is the solution, with the initial condition of all the Brownian particles initially at $x = 0$; this distribution is shown in Fig. 3 [1]

[1] We can get the solution (1.5) by using the method of the integral transform to solve partial differential equations. Introducing the space Fourier transform of $f(x,t)$ and its inverse,

$$F(k,t) = \int \mathrm{d}x\, e^{-\mathrm{i}kx} f(x,t) \ , \qquad f(x,t) = \frac{1}{2\pi} \int \mathrm{d}k\, e^{\mathrm{i}kx} F(k,t) \ ,$$

the diffusion equation (1.4) transforms into the simple form

$$\frac{\partial F}{\partial t} = -D\, k^2 F \quad \Longrightarrow \quad F(k,t) = F(k,0) e^{-D\, k^2 t} \ .$$

For the initial condition $f(x, t = 0) = \delta(x)$, the above Fourier transform gives $F(k, t = 0) = 1$. Then, taking the inverse transform of the solution in k-space, we finally have

$$f(x,t) = \frac{1}{2\pi} \int \mathrm{d}k\, e^{\mathrm{i}kx} e^{-D\, k^2 t} = \frac{e^{-x^2/4Dt}}{2\pi} \underbrace{\int \mathrm{d}k\, e^{-Dt\,(k - \mathrm{i}x/2Dt)^2}}_{\sqrt{\pi/Dt}} = \frac{e^{-x^2/4Dt}}{\sqrt{\pi\, 4Dt}} \ ,$$

where in the second step we have completed the square in the argument of the exponential

Einstein ends with: "We now calculate, with the help of this equation, the displacement λ_x in the direction of the X-axis that a particle experiences on the average or, more exactly, the square root of the arithmetic mean of the square of the displacements in the direction of the X-axis; it is

$$\lambda_x = \sqrt{\langle x^2 \rangle - \langle x_0^2 \rangle} = \sqrt{2\,D\,t}\,. \tag{1.6}$$

Einstein derivation contains very many of the major concepts which since then have been developed more and more generally and rigorously over the years, and which will be the subject matter of these notes. For example:

(i) The *Chapman–Kolgomorov equation* occurs as Eq. (1.1). It states that the probability of the particle being at point x at time $t+\tau$ is given by the sum of the probabilities of all possible "pushes" Δ from positions $x+\Delta$, multiplied by the probability of being at $x+\Delta$ at time t. This assumption is based on the independence of the push Δ of any previous history of the motion; it is only necessary to know the initial position of the particle at time t—not at any previous time. This is the *Markov postulate* and the Chapman–Kolmogorov equation, of which Eq. (1.1) is a special form, is the central dynamical equation to all Markov processes. These will be studied in Sec. 3.

(ii) *The Kramers–Moyal expansion.* This is the expansion used [Eq. (1.2)] to go from Eq. (1.1) (the Chapman–Kolmogorov equation) to the diffusion equation (1.4).

(iii) *The Fokker–Planck equation.* The mentioned diffusion equation (1.4), is a special case of a Fokker–Planck equation. This equation governs an important class of Markov processes, in which the system has a continuous sample path. We shall consider points (ii) and (iii) in detail in Sec. 4.

1.1.2 Langevin's approach (1908)

Some time after Einstein's work, Langevin presented a new method which was quite different from the former and, according to him, "infiniment plus simple". His reasoning was as follows.

$-Dk^2t + \mathrm{i}kx = -Dt\,(k - \mathrm{i}x/2Dt)^2 - x^2/4Dt$, and in the final step we have used the Gaussian integral $\int \mathrm{d}k\, e^{-\alpha(k-b)^2} = \sqrt{\pi/\alpha}$, which also holds for complex b.

From statistical mechanics, it was known that the mean kinetic energy of the Brownian particles should, in equilibrium, reach the value

$$\left\langle \tfrac{1}{2} m v^2 \right\rangle = \tfrac{1}{2} k_{\rm B} T \ . \tag{1.7}$$

Acting on the particle, of mass m, there should be two forces:

(i) a viscous force: assuming that this is given by the same formula as in macroscopic hydrodynamics, this is $-m\gamma \mathrm{d}x/\mathrm{d}t$, with $m\gamma = 6\pi\mu a$, being μ the viscosity and a the diameter of the particle.

(ii) a fluctuating force $\xi(t)$, which represents the incessant impacts of the molecules of the liquid on the Brownian particle. All what we know about it is that is indifferently positive and negative and that its magnitude is such that maintains the agitation of the particle, which the viscous resistance would stop without it.

Thus, the equation of motion for the position of the particle is given by Newton's law as

$$\boxed{m \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = -m\gamma \, \frac{\mathrm{d}x}{\mathrm{d}t} + \xi(t) \ .} \tag{1.8}$$

Multiplying by x, this can be written

$$\frac{m}{2} \frac{\mathrm{d}^2 (x^2)}{\mathrm{d}t^2} - m v^2 = -\frac{m\gamma}{2} \frac{\mathrm{d}(x^2)}{\mathrm{d}t} + \xi \, x \ .$$

If we consider a large number of identical particles, average this equation written for each one of them, and use the equipartition result (1.7) for $\langle m v^2 \rangle$, we get and equation for $\langle x^2 \rangle$

$$\frac{m}{2} \frac{\mathrm{d}^2 \langle x^2 \rangle}{\mathrm{d}t^2} + \frac{m\gamma}{2} \frac{\mathrm{d} \langle x^2 \rangle}{\mathrm{d}t} = k_{\rm B} T \ .$$

The term $\langle \xi \, x \rangle$ has been set to zero because (to quote Langevin) "of the irregularity of the quantity $\xi(t)$". One then finds the solution (C is an integration constant)

$$\frac{\mathrm{d} \langle x^2 \rangle}{\mathrm{d}t} = 2 k_{\rm B} T / m\gamma + C e^{-\gamma t} \ .$$

Langevin estimated that the decaying exponential approaches zero with a time constant of the order of $10^{-8}\,\mathrm{s}$, so that $\mathrm{d}\left\langle x^2\right\rangle/\mathrm{d}t$ enters rapidly a constant regime $\mathrm{d}\left\langle x^2\right\rangle/\mathrm{d}t = 2k_\mathrm{B}T/m\gamma$ Therefore, one further integration (in this asymptotic regime) leads to

$$\left\langle x^2\right\rangle - \left\langle x_0^2\right\rangle = 2(k_\mathrm{B}T/m\gamma)t\ ,$$

which corresponds to Einstein result (1.6), provided we identify the diffusion coefficient as

$$\boxed{D = k_\mathrm{B}T/m\gamma\ .} \tag{1.9}$$

It can be seen that Einstein's condition of the independence of the displacements Δ at different times, is equivalent to Langevin's assumption about the vanishing of $\langle \xi\, x\rangle$. Langevin's derivation is more general, since it also yields the short time dynamics (by a trivial integration of the neglected $Ce^{-\gamma t}$), while it is not clear where in Einstein's approach this term is lost.

Langevin's equation was the first example of a *stochastic differential equation*— a differential equation with a random term $\xi(t)$ and hence whose solution is, in some sense, a random function.[2] Each solution of the Langevin equation represents a different random trajectory and, using only rather simple properties of the fluctuating force $\xi(t)$, measurable results can be derived. Figure 4 shows the trajectory of a Brownian particle in two dimensions obtained by numerical integration of the Langevin equation (we shall also study numerical integration of stochastic differential equations). It is seen the growth with t of the area covered by the particle, which corresponds to the increase of $\langle x^2\rangle - \langle x_0^2\rangle$ in the one-dimensional case discussed above.

The theory and experiments on Brownian motion during the first two decades of the XX century, constituted the most important indirect evidence of the existence of atoms and molecules (which were unobservable at that time). This was a strong support for the atomic and molecular theories of matter, which until the beginning of the century still had strong opposition by the so-called energeticits. The experimental verification of the theory of Brownian motion awarded the 1926 Nobel price to Svedberg and Perrin. [3]

[2] The rigorous mathematical foundation of the theory of stochastic differential equations was not available until the work of Ito some 40 years after Langevin's paper.

[3] Astonishingly enough, the physical basis of the phenomenon was already described in the 1st century B.C.E. by Lucretius in *De Rerum Natura* (II, 112–141), a didactical poem which constitutes the most complete account of ancient atomism and Epicureanism.

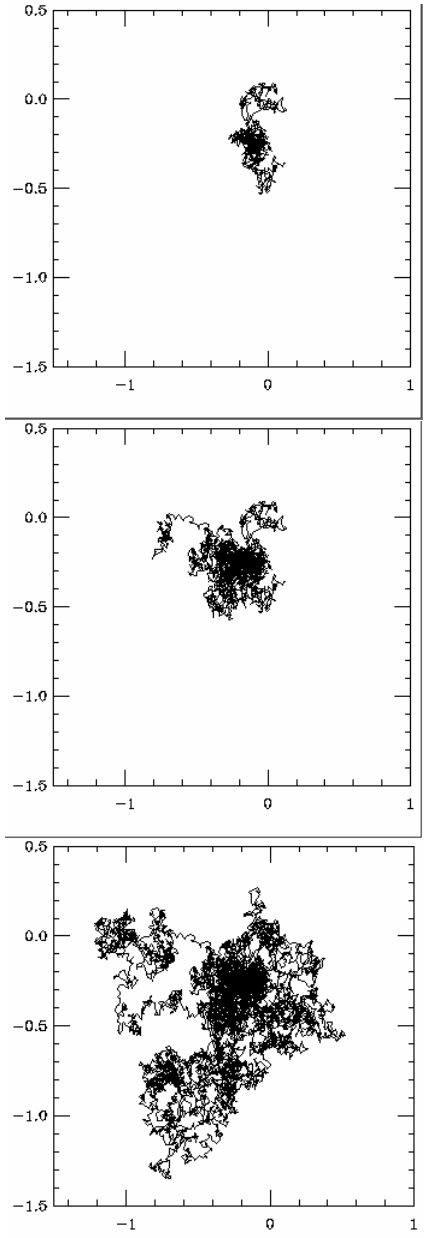

Figure 4: Trajectory of a simulated Brownian particle projected into the x-y plane, with $D = 0.16\ \mu\,\mathrm{m}^2/\mathrm{s}$. The x and y axes are marked in microns. It starts from the origin $(x, y) = (0, 0)$ at $t = 0$, and the pictures show the trajectory after 1 sec, 3 sec and 10 sec.

The picture of a Brownian particle immersed in a fluid is typical of a variety of problems, even when there are no real particles. For instance, it is the case if there is only a certain (slow or heavy) degree of freedom that interacts, in a more or less irregular or random way, with many other (fast or light) degrees of freedom, which play the role of the bath. Thus, the general concept of fluctuations describable by Fokker–Planck and Langevin equations has developed very extensively in a very wide range of situations. A great advantage is the necessity of only a few parameters; in the example of the Brownian particle, essentially the coefficients of the derivatives in the Kramers–Moyal expansion (allowing in general the coefficients a x and t dependence)

$$\int_{-\infty}^{\infty} \Delta\, \phi(\Delta) \mathrm{d}\Delta\ , \qquad \int_{-\infty}^{\infty} \frac{\Delta^2}{2} \phi(\Delta) \mathrm{d}\Delta\ , \ldots\ . \tag{1.10}$$

It is rare to find a problem (mechanical oscillators, fluctuations in electrical circuits, chemical reactions, dynamics of dipoles and spins, escape over metastable barriers, etc.) with cannot be specified, in at least some degree of approximation, by the corresponding Fokker–Planck equation, or equivalently, by augmenting a deterministic differential equation with some fluctuating force or field, like in Langevin's approach. In the following sections we shall describe the methods developed for a systematic and more rigorous study of these equations.

When observing dust particles dancing in a sunbeam, Lucretius conjectured that the particles are in such irregular motion since they are being continuously battered by the invisible blows of restless atoms. Although we now know that such dust particles' motion is caused by air currents, he illustrated the right physics but only with a wrong example. Lucretius also extracted the right consequences from the "observed" phenomenon, as one that shows macroscopically the effects of the "invisible atoms" and hence an indication of their existence.

2 Stochastic variables

2.1 Single variable case

A stochastic or random variable is a quantity X, defined by a set of possible values $\{x\}$ (the "range", "sample space", or "phase space"), and a probability distribution on this set, $P_X(x)$.[4] The range can be discrete or continuous, and the probability distribution is a non-negative function, $P_X(x) \geq 0$, with $P_X(x)\mathrm{d}x$ the probability that $X \in (x, x + \mathrm{d}x)$. The probability distribution is normalised in the sense

$$\int \mathrm{d}x\, P_X(x) = 1 \,,$$

where the integral extends over the whole range of X.

In a discrete range, $\{x_n\}$, the probability distribution consists of a number of delta-type contributions, $P_X(x) = \sum_n p_n \delta(x - x_n)$ and the above normalisation condition reduces to $\sum_n p_n = 1$. For instance, consider the usual example of casting a die: the range is $\{x_n\} = \{1, 2, 3, 4, 5, 6\}$ and $p_n = 1/6$ for each x_n (in a honest die). Thus, by allowing δ-function singularities in the probability distribution, one may formally treat the discrete case by the same expressions as those for the continuous case.

2.1.1 Averages and moments

The average of a function $f(X)$ defined on the range of the stochastic variable X, with respect to the probability distribution of this variable, is defined as

$$\langle f(X) \rangle = \int \mathrm{d}x\, f(x) P_X(x) \,.$$

The moments of the stochastic variable, μ_m, correspond to the special cases $f(X) = X^m$, i.e.,[5]

$$\mu_m = \langle X^m \rangle = \int \mathrm{d}x\, x^m P_X(x) \,, \qquad m = 1, 2, \ldots \,. \tag{2.1}$$

[4] It is advisable to use different notations for the stochastic variable, X, and for the corresponding variable in the probability distribution function, x. However, one relaxes this convention when no confusion is possible. Similarly, the subscript X is here and there dropped from the probability distribution.

[5] This definition can formally be extended to $m = 0$, with $\mu_0 = 1$, which expresses the normalisation of $P_X(x)$.

2.1.2 Characteristic function

This useful quantity is defined by the average of $\exp(\mathrm{i}kX)$, namely

$$G_X(k) = \langle \exp(\mathrm{i}kX) \rangle = \int \mathrm{d}x \, \exp(\mathrm{i}kx) P_X(x) \, . \tag{2.2}$$

This is merely the Fourier transform of $P_X(x)$, and can naturally be solved for the probability distribution

$$P_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}k \, \exp(-\mathrm{i}kx) G_X(k) \, .$$

The function $G_X(k)$ provides an alternative complete characterisation of the probability distribution.

By expanding the exponential in the integrand of Eq. (2.2) and interchanging the order of the resulting series and the integral, one gets

$$G_X(k) = \sum_{m=0}^{\infty} \frac{(\mathrm{i}k)^m}{m!} \int \mathrm{d}x \, x^m P_X(x) = \sum_{m=0}^{\infty} \frac{(\mathrm{i}k)^m}{m!} \mu_m \, . \tag{2.3}$$

Therefore, one finds that $G_X(k)$ is the *moment generating function*, in the sense that

$$\mu_m = (-i)^m \, \frac{\partial^m}{\partial k^m} G_X(k) \bigg|_{k=0} \, . \tag{2.4}$$

2.1.3 Cumulants

The cumulants, κ_m, are defined as the coefficients of the expansion of the *cumulant function* $\ln G_X(k)$ in powers of $\mathrm{i}k$, that is,

$$\ln G_X(k) = \sum_{m=1}^{\infty} \frac{(\mathrm{i}k)^m}{m!} \kappa_m \, .$$

Note that, owing to $P_X(x)$ is normalised, the $m = 0$ term vanishes and the above series begins at $m = 1$. The explicit relations between the first four cumulants and the corresponding moments are

$$\begin{aligned} \kappa_1 &= \mu_1 \\ \kappa_2 &= \mu_2 - \mu_1^2 \\ \kappa_3 &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \\ \kappa_4 &= \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4 \, . \end{aligned} \tag{2.5}$$

Thus, the first cumulant is coincident with the first moment (mean) of the stochastic variable: $\kappa_1 = \langle X \rangle$; the second cumulant κ_2, also called the *variance* and written σ^2, is related to the first and second moments via $\sigma^2 \equiv \kappa_2 = \langle X^2 \rangle - \langle X \rangle^2$.[6] We finally mention that there exists a general expression for κ_m in terms of the determinant of a $m \times m$ matrix constructed with the moments $\{\mu_i \mid i = 1, \ldots, m\}$ (see, e.g., [4, p. 18]):

$$\kappa_m = (-1)^{m-1} \begin{vmatrix} \mu_1 & 1 & 0 & 0 & 0 & \ldots \\ \mu_2 & \mu_1 & 1 & 0 & 0 & \ldots \\ \mu_3 & \mu_2 & \binom{2}{1}\mu_1 & 1 & 0 & \ldots \\ \mu_4 & \mu_3 & \binom{3}{1}\mu_2 & \binom{3}{2}\mu_1 & 1 & \ldots \\ \mu_5 & \mu_4 & \binom{4}{1}\mu_3 & \binom{4}{2}\mu_2 & \binom{4}{3}\mu_1 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{vmatrix}_m . \tag{2.6}$$

where the $\binom{i}{k}$ are binomial coefficients.

2.2 Multivariate probability distributions

All the above definitions, corresponding to one variable, are readily extended to higher-dimensional cases. Consider the n-dimensional vector of stochastic variables $\mathbf{X} = (X_1, \ldots, X_n)$, with a probability distribution $P_n(x_1, \ldots, x_n)$. This distribution is also referred to as the *joint probability distribution* and

$$P_n(x_1, \ldots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n \ ,$$

is the probability that $X_1, \ldots, X_n$ have certain values between $(x_1, x_1 + \mathrm{d}x_1)$,..., $(x_n, x_n + \mathrm{d}x_n)$.

Partial distributions. One can also consider the probability distribution for some of the variables. This can be done in various ways:

1. Take a subset of $s < n$ variables $X_1, \ldots, X_s$. The probability that they have certain values in $(x_1, x_1 + \mathrm{d}x_1), \ldots, (x_s, x_s + \mathrm{d}x_s)$, regardless of the values of the remaining variables $X_{s+1}, \ldots, X_n$, is

$$P_s(x_1, \ldots, x_s) = \int \mathrm{d}x_{s+1} \cdots \mathrm{d}x_n \, P_n(x_1, \ldots, x_s, x_{s+1}, \ldots, x_n) \ ,$$

[6] Quantities related to the third- and fourth-order cumulants have also their own names: *skewness*, $\kappa_3/\kappa_2^{3/2}$, and *kurtosis*, κ_4/κ_2^2.

which is called the *marginal distribution* for the subset $X_1, \ldots, X_s$. Note that from the normalisation of the joint probability it immediately follows the normalisation of the marginal probability.

2. Alternatively, one may attribute fixed values to $X_{s+1}, \ldots, X_n$, and consider the joint probability of the remaining variables $X_1, \ldots, X_s$. This is called the *conditional probability*, and it is denoted by

$$P_{s|n-s}(x_1, \ldots, x_s | x_{s+1}, \ldots, x_n) \ .$$

This distribution is constructed in such a way that the total joint probability $P_n(x_1, \ldots, x_n)$ is equal to the marginal probability for $X_{s+1}, \ldots, X_n$ to have the values $x_{s+1}, \ldots, x_n$, times the conditional probability that, this being so, the remaining variables $X_1, \ldots, X_s$ have the values $(x_1, \ldots, x_s)$. This is Bayes' rule, and can be considered as the *definition* of the conditional probability:

$$P_n(x_1, \ldots, x_n) = P_{n-s}(x_{s+1}, \ldots, x_n) P_{s|n-s}(x_1, \ldots, x_s | x_{s+1}, \ldots, x_n) \ .$$

Note that from the normalisation of the joint and marginal probabilities it follows the normalisation of the conditional probability.

Characteristic function: moments and cumulants. For multivariate probability distributions, the moments are defined by

$$\langle X_1^{m_1} \cdots X_n^{m_n} \rangle = \int \mathrm{d}x_1 \cdots \mathrm{d}x_n \, x_1^{m_1} \cdots x_n^{m_n} P(x_1, \ldots, x_n) \ ,$$

while the characteristic (moment generating) function depends on n auxiliary variables $\mathbf{k} = (k_1, \ldots, k_n)$:

$$\begin{aligned} G(\mathbf{k}) &= \langle \exp[\mathrm{i}(k_1 X_1 + \cdots + k_n X_n)] \rangle \\ &= \sum_0^\infty \frac{(\mathrm{i}k_1)^{m_1} \cdots (\mathrm{i}k_n)^{m_n}}{m_1! \cdots m_n!} \langle X_1^{m_1} \cdots X_n^{m_n} \rangle \ . \end{aligned} \tag{2.7}$$

Similarly, the cumulants of the multivariate distribution, indicated by double brackets, are defined in terms of the coefficients of the expansion of $\ln G$ as

$$\ln G(\mathbf{k}) = {\sum_0^\infty}' \frac{(\mathrm{i}k_1)^{m_1} \cdots (\mathrm{i}k_n)^{m_n}}{m_1! \cdots m_n!} \langle\langle X_1^{m_1} \cdots X_n^{m_n} \rangle\rangle \ ,$$

where the prime indicates the absence of the term with all the m_i simultaneously vanishing (by the normalisation of P_n).

Covariance matrix. The second-order moments may be combined into a n by n matrix $\langle X_i X_j \rangle$. More relevant is, however, the *covariance matrix*, defined by the second-order cumulants

$$\langle\langle X_i X_j \rangle\rangle = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle = \langle [X_i - \langle X_i \rangle][X_j - \langle X_j \rangle] \rangle \ .$$

Each diagonal element is called the *variance* of the corresponding variable, while the off-diagonal elements are referred to as the *covariance* of the corresponding pair of variables.[7]

Statistical independence. A relevant concept for multivariate distributions is that of *statistical independence*. One says that, e.g., two stochastic variables X_1 and X_2 are statistically independent of each other if their joint probability distribution factorises:

$$P_{X_1 X_2}(x_1, x_2) = P_{X_1}(x_1) P_{X_2}(x_2) \ .$$

The statistical independence of X_1 and X_2 is also expressed by any one of the following three equivalent criteria:

1. All moments factorise: $\langle X_1^{m_1} X_2^{m_2} \rangle = \langle X_1^{m_1} \rangle \langle X_2^{m_2} \rangle$.
2. The characteristic function factorises: $G_{X_1 X_2}(k_1, k_2) = G_{X_1}(k_1) G_{X_2}(k_2)$.
3. The cumulants $\langle\langle X_1^{m_1} X_2^{m_2} \rangle\rangle$ vanish when both m_1 and m_2 are $\neq 0$.

Finally, two variables are called *uncorrelated* when its *covariance*, $\langle\langle X_1 X_2 \rangle\rangle$, is zero, which is a condition weaker than that of statistical independence.

2.3 The Gaussian distribution

This important distribution is defined as

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu_1)^2}{2\sigma^2} \right] \ . \tag{2.8}$$

It is easily seen that μ_1 is indeed the average and σ^2 the variance, which justifies the notation. The corresponding characteristic function is

$$G(k) = \exp(\mathrm{i}\mu_1 k - \tfrac{1}{2}\sigma^2 k^2) \ , \tag{2.9}$$

[7] Note that $\langle\langle X_i X_j \rangle\rangle$ is, by construction, a symmetrical matrix.

as can be seen from the definition (2.2), by completing the square in the argument of the total exponential $\mathrm{i}kx-(x-\mu_1)^2/2\sigma^2$ and using the Gaussian integral $\int \mathrm{d}k\, e^{-\alpha(k-b)^2} = \sqrt{\pi/\alpha}$ for complex b as in the footnote in p. 7. Note that the logarithm of this characteristic function comprises terms up to quadratic in k only. Therefore, all the cumulants after the second one vanish identically, which is a property that indeed *characterises* the Gaussian distribution.

For completeness, we finally write the Gaussian distribution for n variables $\mathbf{X} = (X_1, \ldots, X_n)$ and the corresponding characteristic function

$$\begin{aligned} P(\mathbf{x}) &= \sqrt{\frac{\det \hat{A}}{(2\pi)^n}} \exp\left[-\tfrac{1}{2}(\mathbf{x}-\mathbf{x}_0)\cdot\hat{A}\cdot(\mathbf{x}-\mathbf{x}_0)\right] \\ G(\mathbf{k}) &= \exp\left(i\mathbf{x}_0\cdot\mathbf{k} - \tfrac{1}{2}\mathbf{k}\cdot\hat{A}^{-1}\cdot\mathbf{k}\right) . \end{aligned}$$

The averages and covariances are given by $\langle\mathbf{X}\rangle = \mathbf{x}_0$ and $\langle\langle X_i X_j\rangle\rangle = (\hat{A}^{-1})_{ij}$.

2.4 Transformation of variables

For a given stochastic variable X, every related quantity $Y = f(X)$ is again a stochastic variable. The probability that Y has a value between y and $y + \Delta y$ is

$$P_Y(y)\Delta y = \int\limits_{y<f(x)<y+\Delta y} \mathrm{d}x\, P_X(x) ,$$

where the integral extends over all intervals of the range of X where the inequality is obeyed. Note that one can equivalently *define* $P_Y(y)$ as[8]

$$\boxed{P_Y(y) = \int \mathrm{d}x\, \delta[y - f(x)] P_X(x) .} \tag{2.11}$$

[8] Note also that from Eq. (2.11), one can formally write the probability distribution for Y as the following average [with respect to $P_X(x)$ and taking y as a parameter]

$$P_Y(y) = \langle\delta[y - f(X)]\rangle . \tag{2.10}$$

From this expression one can calculate the characteristic function of Y:

$$\begin{aligned} G_Y(k) &\overset{\text{Eq. (2.2)}}{=} \int \mathrm{d}y\, \exp(\mathrm{i}ky) P_Y(y) \\ &= \int \mathrm{d}x\, P_X(x) \int \mathrm{d}y\, \exp(\mathrm{i}ky)\delta[y - f(x)] \\ &= \int \mathrm{d}x\, P_X(x) \exp[\mathrm{i}kf(x)] \,, \end{aligned}$$

which can finally be written as

$$\boxed{G_Y(k) = \langle \exp[\mathrm{i}kf(X)] \rangle \;.} \tag{2.12}$$

As the simplest example consider the linear transformation $Y = \alpha X$. The above equation then yields $G_Y(k) = \langle \exp(\mathrm{i}k\alpha X) \rangle$, whence

$$G_Y(k) = G_X(\alpha k) \,, \qquad (Y = \alpha X) \,. \tag{2.13}$$

2.5 Addition of stochastic variables

The above equations for the transformation of variables remain valid when X stands for a stochastic variable with n components and Y for one with s components, where s may or may not be equal to n. For example, let us consider the case of the addition of two stochastic variables $Y = f(X_1, X_2) = X_1 + X_2$, where $s = 1$ and $n = 2$. Then, from Eq. (2.11) one first gets

$$\begin{aligned} P_Y(y) &= \int \mathrm{d}x_1 \int \mathrm{d}x_2\, \delta[y - (x_1 + x_2)] P_{X_1 X_2}(x_1, x_2) \\ &= \int \mathrm{d}x_1\, P_{X_1 X_2}(x_1, y - x_1) \,. \end{aligned} \tag{2.14}$$

Properties of the sum of stochastic variables. One easily deduces the following three rules concerning the addition of stochastic variables:

1. Regardless of whether X_1 and X_2 are independent or not, one has[9]

$$\langle Y \rangle = \langle X_1 \rangle + \langle X_2 \rangle \ . \tag{2.15}$$

2. If X_1 and X_2 are *uncorrelated*, $\langle\langle X_1 X_2 \rangle\rangle = 0$, a similar relation holds for the variances[10]

$$\left\langle\left\langle Y^2 \right\rangle\right\rangle = \left\langle\left\langle X_1^2 \right\rangle\right\rangle + \left\langle\left\langle X_2^2 \right\rangle\right\rangle \ . \tag{2.16}$$

3. The characteristic function of $Y = X_1 + X_2$ is[11]

$$G_Y(k) = G_{X_1 X_2}(k, k) \ . \tag{2.17}$$

On the other hand, if X_1 and X_2 are *independent*, Eq. (2.14) and the factorization of $P_{X_1 X_2}$ and $G_{X_1 X_2}$ yields

$$P_Y(y) = \int \mathrm{d}x_1 \, P_{X_1}(x_1) P_{X_2}(y - x_1) \ , \qquad G_Y(k) = G_{X_1}(k) G_{X_2}(k) \ . \tag{2.18}$$

---

[9] Proof of Eq. (2.15):

$$\begin{aligned} \langle Y \rangle &\equiv \int \mathrm{d}y \, y P_Y(y) = \int \mathrm{d}x_1 \int \mathrm{d}x_2 \, P_{X_1 X_2}(x_1, x_2) \int \mathrm{d}y \, y \, \delta[y - (x_1 + x_2)] \\ &= \int \mathrm{d}x_1 \int \mathrm{d}x_2 \, P_{X_1 X_2}(x_1, x_2)(x_1 + x_2) = \langle X_1 \rangle + \langle X_2 \rangle \ . \quad \text{Q.E.D.} \end{aligned}$$

[10] Proof of Eq. (2.16):

$$\langle Y^2 \rangle \equiv \int \mathrm{d}y \, y^2 P_Y(y) = \int \mathrm{d}x_1 \int \mathrm{d}x_2 \, P_{X_1 X_2}(x_1, x_2)(x_1 + x_2)^2 = \langle X_1^2 \rangle + \langle X_2^2 \rangle + 2 \langle X_1 X_2 \rangle \ .$$

Therefore

$$\left\langle\left\langle Y^2 \right\rangle\right\rangle = \langle Y^2 \rangle - \langle Y \rangle^2 = \underbrace{\langle X_1^2 \rangle - \langle X_1 \rangle^2}_{\langle\langle X_1^2 \rangle\rangle} + \underbrace{\langle X_2^2 \rangle - \langle X_2 \rangle^2}_{\langle\langle X_2^2 \rangle\rangle} + 2 \underbrace{(\langle X_1 X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle)}_{\langle\langle X_1 X_2 \rangle\rangle}$$

from which the statement follows for uncorrelated variables. Q.E.D.

[11] Proof of Eq. (2.17):

$$G_Y(k) \overset{\text{Eq. (2.12)}}{=} \langle \exp[\mathrm{i} k f(X_1, X_2)] \rangle = \langle \exp[\mathrm{i} k (X_1 + X_2)] \rangle \overset{\text{Eq. (2.7)}}{=} G_{X_1 X_2}(k, k) \ . \quad \text{Q.E.D.}$$

Thus, the probability distribution of the sum of two *independent* variables is the *convolution* of their individual probability distributions. Correspondingly, the characteristic function of the sum [which is the Fourier transform of the probability distribution; see Eq. (2.2)] is the *product* of the individual characteristic functions.

2.6 Central limit theorem

As a particular case of transformation of variables, one can also consider the sum of an arbitrary number of stochastic variables. Let $X_1, \ldots, X_n$ be a set of n *independent* stochastic variables, each having the same probability distribution $P_X(x)$ with zero average and (finite) variance σ^2. Then, from Eqs. (2.15) and (2.16) it follows that their sum $Y = X_1 + \cdots + X_n$ has zero average and variance $n\sigma^2$, which grows linearly with n. On the other hand, the distribution of the arithmetic mean of the variables, $(X_1 + \cdots + X_n)/n$, becomes narrower with increasing n (variance σ^2/n). It is therefore more convenient to define a suitable scaled sum

$$Z = \frac{X_1 + \cdots + X_n}{\sqrt{n}} \ ,$$

which has variance σ^2 for all n.

The *central limit theorem* states that, even when $P_X(x)$ is not Gaussian, the probability distribution of the Z so-defined tends, as $n \to \infty$, to a Gaussian distribution with zero mean and variance σ^2. This remarkable result is responsible for the important rôle of the Gaussian distribution in all fields in which statistics are used and, in particular, in the equilibrium and non-equilibrium statistical physics.

Proof of the central limit theorem. We begin by expanding the characteristic function of an arbitrary $P_X(x)$ with zero mean as [cf. Eq. (2.3)]

$$G_X(k) = \int \mathrm{d}x \, \exp(\mathrm{i}kx) P_X(x) = 1 - \tfrac{1}{2}\sigma^2 k^2 + \cdots \ . \tag{2.19}$$

The factorization of the characteristic function of the sum $Y = X_1 + \cdots + X_n$ of statistically independent variables [Eq. (2.18)], yields

$$G_Y(k) = \prod_{i=1}^{n} G_{X_i}(k) = [G_X(k)]^n \ ,$$

where the last equality follows from the equivalent statistical properties of the different variables X_i. Next, on accounting for $Z = Y/\sqrt{n}$, and using the result (2.13) with $\alpha = 1/\sqrt{n}$, one has

$$G_Z(k) = G_Y\left(\frac{k}{\sqrt{n}}\right) = \left[G_X\left(\frac{k}{\sqrt{n}}\right)\right]^n \simeq \left(1 - \frac{\sigma^2 k^2}{2n}\right)^n \stackrel{n\to\infty}{\longrightarrow} \exp\left(-\tfrac{1}{2}\sigma^2 k^2\right) \ , \quad (2.20)$$

where we have used the definition of the exponential $e^x = \lim_{n\to\infty}(1 + x/n)^n$. Finally, on comparing the above result with Eqs. (2.8), one gets

$$P_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{z^2}{2\sigma^2}\right) \ . \quad \text{Q.E.D.}$$

Remarks on the validity of the central limit theorem. The derivation of the central limit theorem can be done under more general conditions. For instance, it is not necessary that all the cumulants (moments) exist. However, it is necessary that the moments up to at least second order exist [or, equivalently, $G_X(k)$ being twice differentiable at $k = 0$; see Eq. (2.4)]. The necessity of this condition is illustrated by the counter-example provided by the Lorentz–Cauchy distribution:

$$P(x) = \frac{1}{\pi}\frac{\gamma}{x^2 + \gamma^2} \ , \qquad (-\infty < x < \infty) \ .$$

It can be shown that, if a set of n independent variables X_i have Lorentz–Cauchy distributions, their sum also has a Lorentz–Cauchy distribution (see footnote below). However, for this distribution the conditions for the central limit theorem to hold are not met, since the integral (2.1) defining the moments μ_m, does not converge even for $m = 1$.[12]

Finally, although the condition of independence of the variables is important, it can be relaxed to incorporate a sufficiently weak statistical dependence.

[12]This can also be demonstrated by calculating the corresponding characteristic function. To do so, one can use $\int_{-\infty}^{-\infty} dx\, e^{iax}/(1 + x^2) = \pi e^{-|a|}$, which is obtained by computing the residues of the integrand in the upper (lower) half of the complex plane for $a > 0$ ($a < 0$). Thus, one gets

$$G(k) = \exp(-\gamma|k|) \ ,$$

which, owing to the presence of the modulus of k, is not differentiable at $k = 0$. Q.E.D.

We remark in passing that, from $G_{X_i}(k) = \exp(-\gamma_i|k|)$ and the second Eq. (2.18), it follows that the distribution of the sum of *independent* Lorentz–Cauchy variables has a Lorentz–Cauchy distribution (with $G_Y(k) = \exp[-(\sum_i \gamma_i)|k|]$).

2.7 Exercise: marginal and conditional probabilities and moments of a bivariate Gaussian distribution

To illustrate the definitions given for multivariate distributions, let us compute them for a simple two-variable Gaussian distribution

$$P_2(x_1, x_2) = \sqrt{\frac{1-\lambda^2}{(2\pi\sigma^2)^2}} \exp\left[-\frac{1}{2\sigma^2}\left(x_1^2 - 2\lambda\, x_1 x_2 + x_2^2\right)\right] , \qquad (2.23)$$

where λ is a parameter $-1 < \lambda < 1$, to ensure that the quadratic form in the exponent is definite positive (the equivalent condition to assume σ^2 to be positive in the one-variable Gaussian distribution (2.8). The normalisation factor can be seen to take this value by direct integration, or by comparing our distribution with the multidimensional Gaussian distribution (Sec. 2.3); here $\hat{A} = \frac{1}{\sigma^2}\begin{pmatrix} 1 & -\lambda \\ -\lambda & 1 \end{pmatrix}$ so that $\det \hat{A} = (1-\lambda^2)/\sigma^4$. Finally, if one wishes to fix ideas one can interpret $P_2(x_1, x_2)$ as the Boltzmann distribution of two harmonic oscillators coupled by a potential term $\propto \lambda\, x_1 x_2$.

Let us first rewrite the distribution in a form that will facilitate to do the integrals by completing once more the square $-2\lambda\, x_1 x_2 + x_2^2 = (x_2 - \lambda x_1)^2 - \lambda^2 x_1^2$

$$P_2(x_1, x_2) = C \exp\left[-\frac{1-\lambda^2}{2\sigma^2} x_1^2\right] \exp\left[-\frac{1}{2\sigma^2}(x_2 - \lambda x_1)^2\right] . \qquad (2.24)$$

and $C = \sqrt{(1-\lambda^2)/(2\pi\sigma^2)^2}$ is the normalisation constant. We can now compute the marginal probability of the individual variables (for one of them since they are equivalent), defined by $P_1(x_1) = \int \mathrm{d}x_2\, P_2(x_1, x_2)$

$$P_1(x_1) = C\, e^{-\frac{1-\lambda^2}{2\sigma^2} x_1^2} \underbrace{\int \mathrm{d}x_2\, e^{-(x_2 - \lambda x_1)^2/2\sigma^2}}_{\sqrt{\pi/\alpha} \quad \alpha = 1/2\sigma^2} .$$

Therefore, recalling the form of C, we merely have

$$P_1(x_1) = \frac{1}{\sqrt{2\pi\sigma_\lambda^2}} \exp\left(-\frac{x_1^2}{2\sigma_\lambda^2}\right) , \qquad \text{with} \qquad \sigma_\lambda^2 = \sigma^2/(1-\lambda^2) . \qquad (2.25)$$

We see that the marginal distribution depends on λ, which results in a modified variance. To see that σ_λ^2 is indeed the variance $\langle\langle x_1^2\rangle\rangle = \langle x_1^2\rangle - \langle x_1\rangle^2$,

note that $\langle x_1^m \rangle$ can be obtained from the marginal distribution only (this is a *general* result)

$$\langle x_1^m \rangle = \int \mathrm{d}x_1 \int \mathrm{d}x_2 \, x_1^m P_2(x_1, x_2) = \int \mathrm{d}x_1 \, x_1^m \underbrace{\int \mathrm{d}x_2 \, P_2(x_1, x_2)}_{P_1(x_1)} = \int \mathrm{d}x_1 \, x_1^m P_1(x_1)$$

Then inspecting the marginal distribution obtained [Eq. (2.25)] we get that the first moments vanish and the variances are indeed equal to σ_λ^2:

$$\boxed{\begin{array}{rclrcl} \langle x_1 \rangle & = & 0 & \langle x_2 \rangle & = & 0 \\ \langle x_1^2 \rangle & = & \sigma_\lambda^2 & \langle x_2^2 \rangle & = & \sigma_\lambda^2 \end{array}} \tag{2.26}$$

To complete the calculation of the moments up to second order we need the covariance of x_1 and x_2: $\langle\langle x_1 x_2 \rangle\rangle = \langle x_1 x_2 \rangle - \langle x_1 \rangle \langle x_2 \rangle$ which reduces to calculate $\langle x_1 x_2 \rangle$. This can be obtained using the form (2.24) for the distribution

$$\begin{aligned} \langle x_1 x_2 \rangle &= \int \mathrm{d}x_1 \int \mathrm{d}x_2 \, x_1 x_2 P_2(x_1, x_2) \\ &= C \int \mathrm{d}x_1 x_1 \exp\left[-\frac{1-\lambda^2}{2\sigma^2} x_1^2 \right] \overbrace{\int \mathrm{d}x_2 \, x_2 \exp\left[-\frac{1}{2\sigma^2} (x_2 - \lambda x_1)^2 \right]}^{\lambda x_1 \sqrt{2\pi\sigma^2}} \\ &= \lambda \sqrt{2\pi\sigma^2} C \underbrace{\int \mathrm{d}x_1 x_1^2 \exp\left[-\frac{1-\lambda^2}{2\sigma^2} x_1^2 \right]}_{\sqrt{2\pi\sigma^2/(1-\lambda^2)}\,\sigma^2/(1-\lambda^2)} \quad \Rightarrow \quad \boxed{\langle x_1 x_2 \rangle = \frac{\lambda}{1-\lambda^2} \sigma^2} \end{aligned}$$

since $C = \sqrt{(1-\lambda^2)/(2\pi\sigma^2)^2}$. Its is convenient to compute the normalised covariance $\langle x_1 x_2 \rangle / \sqrt{\langle x_1^2 \rangle \langle x_2^2 \rangle}$, which is merely given by λ. Therefore the parameter λ in the distribution (2.23) is a measure of how much correlated the variables x_1 and x_2 are. Actually in the limit $\lambda \to 0$ the variables are not correlated at all and the distribution factorises. In the opposite limit $\lambda \to 1$ the variables are maximally correlated, $\langle x_1 x_2 \rangle / \sqrt{\langle x_1^2 \rangle \langle x_2^2 \rangle} = 1$. The distribution is actually a function of $(x_1 - x_2)$, so it is favoured that x_1 and x_2 take similar values (see Fig. 5)

$$P_2|_{\lambda=0} = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-x_1^2/2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-x_2^2/2\sigma^2}, \quad P_2|_{\lambda=1} \to e^{-(x_1-x_2)^2/2\sigma^2} \ . \tag{2.27}$$

We can now interpret the increase of the variances with λ: the correlation between the variables allows them to take arbitrarily large values, with the only restriction of their difference being small (Fig. 5).

To conclude we can compute the conditional probability by using Bayes rule $P_{1|1}(x_1|x_2) = P_2(x_1, x_2)/P_1(x_2)$ and Eqs. (2.23) and (2.25)

$$\begin{aligned} P_{1|1}(x_1|x_2) &= \frac{\sqrt{\frac{1-\lambda^2}{(2\pi\sigma^2)^2}}\exp\left[-\frac{1}{2\sigma^2}\left(x_1^2 - 2\lambda\, x_1 x_2 + x_2^2\right)\right]}{\sqrt{\frac{1-\lambda^2}{2\pi\sigma^2}}\,\exp\left(-\frac{1-\lambda^2}{2\sigma^2}x_2^2\right)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}\left(x_1^2 - 2\lambda\, x_1 x_2 + [\,\not{1} - (\,\not{1} - \lambda^2)]x_2^2\right)\right] \,, \end{aligned}$$

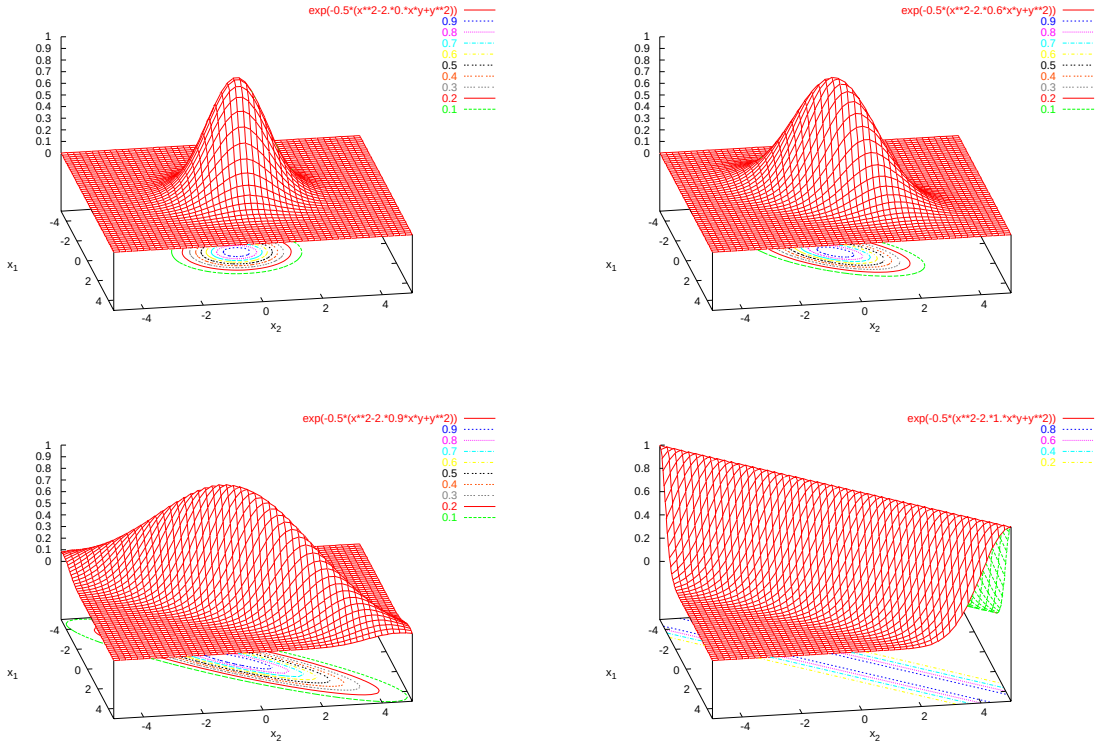

Figure 5: Gaussian distribution (2.23) for $\lambda = 0,\ 0.6,\ 0.9$ and 1 (non-normalised).

and hence (recall that here x_2 is a parameter; the known output of X_2)

$$P_{1|1}(x_1|x_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(x_1 - \lambda\, x_2\right)^2\right] \ . \tag{2.28}$$

Then, at $\lambda = 0$ (no correlation) the values taken by x_1 are independent of the output of x_2 while for $\lambda \to 1$ they are centered around those taken by x_2, and hence strongly conditioned by them.

3 Stochastic processes and Markov processes

Once a stochastic variable X has been defined, an infinity of other stochastic variables derive from it, namely, all the quantities Y defined as functions of X by some mapping. These quantities Y may be any kind of mathematical object; in particular, also functions of an auxiliary variable t:

$$Y_X(t) = f(X,t) \ ,$$

where t could be the time or some other parameter. Such a quantity $Y_X(t)$ is called a *stochastic process*. On inserting for X one of its possible values x, one gets an ordinary function of t, $Y_x(t) = f(x,t)$, called a *sample function* or *realisation* of the process. In physical language, one regards the stochastic process as the "ensemble" of these sample functions.[13]

It is easy to form averages on the basis of the underlying probability distribution $P_X(x)$. For instance, one can take n values $t_1, \ldots, t_n$, and form the nth moment

$$\langle Y(t_1) \cdots Y(t_n) \rangle = \int \mathrm{d}x \, Y_x(t_1) \cdots Y_x(t_n) P_X(x) \ . \tag{3.1}$$

Of special interest is the *auto-correlation function*

$$\begin{aligned} \kappa(t_1,t_2) \equiv \langle\langle Y(t_1) Y(t_2) \rangle\rangle &= \langle Y(t_1) Y(t_2) \rangle - \langle Y(t_1) \rangle \langle Y(t_2) \rangle \\ &= \big\langle [Y(t_1) - \langle Y(t_1) \rangle][Y(t_2) - \langle Y(t_2) \rangle] \big\rangle \ , \end{aligned}$$

which, for $t_1 = t_2 = t$, reduces to the time-dependent variance $\langle\langle Y(t)^2 \rangle\rangle = \sigma(t)^2$.

A stochastic process is called *stationary* when the moments are not affected by a shift in time, that is, when

$$\langle Y(t_1 + \tau) \cdots Y(t_n + \tau) \rangle = \langle Y(t_1) \cdots Y(t_n) \rangle \ . \tag{3.2}$$

In particular, $\langle Y(t) \rangle$ is then independent of the time, and the auto-correlation function $\kappa(t_1,t_2)$ only depends on the time difference $|t_1 - t_2|$. Often there exist a constant τ_c such that $\kappa(t_1,t_2) \simeq 0$ for $|t_1 - t_2| > \tau_c$; one then calls τ_c the *auto-correlation time* of the stationary stochastic process.

[13] As regards the terminology, one also refers to a stochastic time-dependent quantity as a *noise term*. This name originates from the early days of radio, where the great number of highly irregular electrical signals occurring either in the atmosphere, the receiver, or the radio transmitter, certainly sounded like noise on a radio.

If the stochastic quantity consist of several components $Y_i(t)$, the auto-correlation function is replaced by the correlation matrix

$$K_{ij}(t_1, t_2) = \langle\langle Y_i(t_1) Y_j(t_2)\rangle\rangle \ .$$

The diagonal elements are the *auto-correlations* and the off-diagonal elements are the *cross-correlations*. Finally, in case of a zero-average stationary stochastic process, this equation reduces to

$$K_{ij}(\tau) = \langle Y_i(t) Y_j(t+\tau)\rangle = \langle Y_i(0) Y_j(\tau)\rangle \ .$$

3.1 The hierarchy of distribution functions

A stochastic process $Y_X(t)$, defined from a stochastic variable X in the way described above, leads to a hierarchy of probability distributions. For instance, the probability distribution for $Y_X(t)$ to take the value y at time t is [cf. Eq. (2.11)]

$$P_1(y,t) = \int \mathrm{d}x \, \delta[y - \underbrace{Y_x(t)}_{f(x,t)}] P_X(x) \ .$$

Similarly, the joint probability distribution that Y has the value y_1 at t_1, and also the value y_2 at t_2, and so on up to y_n at t_n, is

$$P_n(y_1, t_1; \ldots; y_n, t_n) = \int \mathrm{d}x \, \delta[y_1 - Y_x(t_1)] \cdots \delta[y_n - Y_x(t_n)] P_X(x) \ . \qquad (3.3)$$

In this way an infinite hierarchy of probability distributions P_n, $n = 1, 2, \ldots$, is defined. They allow one the computation of all the averages already introduced, e.g.,[14]

$$\langle Y(t_1) \cdots Y(t_n)\rangle = \int \mathrm{d}y_1 \cdots \mathrm{d}y_n \, y_1 \cdots y_n P_n(y_1, t_1; \ldots; y_n, t_n) \ . \qquad (3.4)$$

[14] This result is demonstrated by introducing the definition (3.3) in the right-hand side of Eq. (3.4):

$$\begin{aligned}
&\int \mathrm{d}y_1 \cdots \mathrm{d}y_n \, y_1 \cdots y_n P_n(y_1, t_1; \ldots; y_n, t_n) \\
&\qquad = \int \mathrm{d}y_1 \cdots \mathrm{d}y_n \, y_1 \cdots y_n \int \mathrm{d}x \, \delta[y_1 - Y_x(t_1)] \cdots \delta[y_n - Y_x(t_n)] P_X(x) \\
&\qquad = \int \mathrm{d}x \, Y_x(t_1) \cdots Y_x(t_n) P_X(x) \\
&\qquad \overset{\text{Eq. (3.1)}}{=} \langle Y(t_1) \cdots Y(t_n)\rangle \ . \quad \text{Q.E.D.}
\end{aligned}$$

We note in passing that, by means of this result, the definition (3.2) of stationary processes, can be restated in terms of the dependence of the P_n on the time differences alone, namely

$$P_n(y_1, t_1 + \tau; \ldots; y_n, t_n + \tau) = P_n(y_1, t_1; \ldots; y_n, t_n) \; .$$

Consequently, a necessary (but not sufficient) condition for the stochastic process being stationary is that $P_1(y_1)$ does not depend on the time.

Although the right-hand side of Eq. (3.3) also has a meaning when some of the times are equal, one regards the P_n to be defined only when all times are different. The hierarchy of probability distributions P_n then obeys the following consistency conditions:

1. $P_n \geq 0$;
2. P_n is invariant under permutations of two pairs (y_i, t_i) and (y_j, t_j) ;
3. $\displaystyle\int \mathrm{d}y_n \, P_n(y_1, t_1; \ldots; y_n, t_n) = P_{n-1}(y_1, t_1; \ldots; y_{n-1}, t_{n-1})$;
4. $\displaystyle\int \mathrm{d}y_1 \, P_1(y_1, t_1) = 1$.

Inasmuch as the distributions P_n enable one to compute all the averages of the stochastic process [Eq. (3.4)], they constitute a complete specification of it. Conversely, according to a theorem due to Kolmogorov, it is possible to prove that the inverse is also true, i.e., that any set of functions obeying the above four consistency conditions determines a stochastic process $Y(t)$.

3.2 Gaussian processes

A stochastic process is called a *Gaussian* process, if all its P_n are multivariate Gaussian distributions (Sec. 2.3). In that case, all cumulants beyond $m = 2$ vanish and, recalling that $\langle\langle Y(t_1)Y(t_2)\rangle\rangle = \langle Y(t_1)Y(t_2)\rangle - \langle Y(t_1)\rangle \langle Y(t_2)\rangle$, one sees that a Gaussian process is fully specified by its average $\langle Y(t)\rangle$ and its second moment $\langle Y(t_1)Y(t_2)\rangle$. Gaussian stochastic processes are often used as an approximate description for physical processes, which amounts to assuming that the higher-order cumulants are negligible.

3.3 Conditional probabilities

The notion of conditional probability for multivariate distributions can be applied to stochastic processes, via the hierarchy of probability distributions introduced above. For instance, the conditional probability $P_{1|1}(y_2, t_2|y_1, t_1)$ represents the probability that Y takes the value y_2 at t_2, given that its value at t_1 "was" y_1. It can be constructed as follows: from all sample functions $Y_x(t)$ of the ensemble representing the stochastic process, select those passing through the point y_1 at the time t_1; the fraction of this *sub-ensemble* that goes through the gate $(y_2, y_2 + \mathrm{d}y_2)$ at the time t_2 is precisely $P_{1|1}(y_2, t_2|y_1, t_1)\mathrm{d}y_2$. More generally, one may fix the values of Y at n different times $t_1, \ldots, t_n$, and ask for the joint probability at m other times $t_{n+1}, \ldots, t_{n+m}$. This leads to the general definition of the conditional probability $P_{m|n}$ by Bayes' rule:

$$P_{m|n}(y_{n+1}, t_{n+1}; \ldots; y_{n+m}, t_{n+m}|y_1, t_1; \ldots; y_n, t_n) = \frac{P_{n+m}(y_1, t_1; \ldots; y_{n+m}, t_{n+m})}{P_n(y_1, t_1; \ldots; y_n, t_n)} \ . \tag{3.5}$$

Note that the right-hand side of this equation is well defined in terms of the probability distributions of the hierarchy P_n previously introduced. Besides, from their consistency conditions it follows the normalisation of the $P_{m|n}$.

3.4 Markov processes

Among the many possible classes of stochastic processes, there is one that merits a special treatment—the so-called Markov processes.

Recall that, for a stochastic process $Y(t)$, the conditional probability $P_{1|1}(y_2, t_2|y_1, t_1)$, is the probability that $Y(t_2)$ takes the value y_2, provided $Y(t_1)$ has taken the value y_1. In terms of this quantity one can express P_2 as

$$P_2(y_1, t_1; y_2, t_2) = P_1(y_1, t_1)P_{1|1}(y_2, t_2|y_1, t_1) \ . \tag{3.6}$$

The notation P_{m | n}(... | ...) means the conditional joint probability of m variables conditioned on n variables.

However, to construct the higher-order P_n one needs transition probabilities $P_{n|m}$ of higher order, e.g., $P_3(y_1, t_1; y_2, t_2; y_3, t_3) = P_2(y_1, t_1; y_2, t_2)P_{1|2}(y_3, t_3|y_1, t_1; y_2, t_2)$. A stochastic process is called a *Markov process*, if for any set of n successive times $t_1 < t_2 < \cdots < t_n$, one has

$$P_{1|n-1}(y_n, t_n|y_1, t_1; \ldots; y_{n-1}, t_{n-1}) = P_{1|1}(y_n, t_n|y_{n-1}, t_{n-1}) \ . \tag{3.7}$$

In words: the conditional probability distribution of y_n at t_n, given the value y_{n-1} at t_{n-1}, is uniquely determined, and is not affected by any knowledge of the values at earlier times.

A Markov process is therefore fully determined by the two distributions $P_1(y,t)$ and $P_{1|1}(y',t'|y,t)$, from which the entire hierarchy $P_n(y_1,t_1;\ldots;y_n,t_n)$ can be constructed. For instance, consider $t_1 < t_2 < t_3$; P_3 can be written as

$$\begin{aligned} P_3(y_1,t_1;y_2,t_2;y_3,t_3) &\overset{\text{Eq. (3.5)}}{=} P_2(y_1,t_1;y_2,t_2)P_{1|2}(y_3,t_3|y_1,t_1;y_2,t_2) \\ &\overset{\text{Eq. (3.7)}}{=} P_2(y_1,t_1;y_2,t_2)P_{1|1}(y_3,t_3|y_2,t_2) \\ &\overset{\text{Eq. (3.6)}}{=} P_1(y_1,t_1)P_{1|1}(y_2,t_2|y_1,t_1)P_{1|1}(y_3,t_3|y_2,t_2) \ . \end{aligned} \tag{3.8}$$

From now on, we shall only deal with Markov processes. Then, the only independent conditional probability is $P_{1|1}(y',t'|y,t)$, so we shall omit the subscript 1|1 henceforth and call $P_{1|1}(y',t'|y,t)$ the *transition probability*.

3.5 Chapman–Kolmogorov equation

Let us now derive an important identity that must be obeyed by the transition probability of any Markov process. On integrating Eq. (3.8) over y_2, one obtains $(t_1 < t_2 < t_3)$

$$P_2(y_1,t_1;y_3,t_3) = P_1(y_1,t_1)\int \mathrm{d}y_2\, P(y_2,t_2|y_1,t_1)P(y_3,t_3|y_2,t_2) \ ,$$

where the consistency condition 3 of the hierarchy of distribution functions P_n has been used to write the left-hand side. Now, on dividing both sides by $P_1(y_1,t_1)$ and using the special case (3.6) of Bayes' rule, one gets

$$\boxed{P(y_3,t_3|y_1,t_1) = \int \mathrm{d}y_2\, P(y_3,t_3|y_2,t_2)P(y_2,t_2|y_1,t_1) \ ,} \tag{3.9}$$

which is called the Chapman–Kolmogorov equation. The time ordering is essential: t_2 must lie between t_1 and t_3 for Eq. (3.9) to hold. This is required in order to the starting Eq. (3.8) being valid, specifically, in order to the second equality there being derivable from the first one by dint of the definition (3.7) of a Markov process.

Note finally that, on using Eq. (3.6) one can rewrite the particular case $P_1(y_2,t_2) = \int \mathrm{d}y_1\, P_2(y_2,t_2;y_1,t_1)$ of the relation among the distributions of the hierarchy as

$$\boxed{P_1(y_2,t_2) = \int \mathrm{d}y_1\, P(y_2,t_2|y_1,t_1)P_1(y_1,t_1) \ .} \tag{3.10}$$

This is an additional relation involving the two probability distributions characterising a Markov process. Reciprocally, any non-negative functions obeying Eqs. (3.9) and (3.10), define a Markov process uniquely.

3.6 Examples of Markov processes

Wiener–Lévy process. This stochastic process was originally introduced in order to describe the behaviour of the *position* of a free Brownian particle in one dimension. On the other hand, it plays a central rôle in the rigorous foundation of the stochastic differential equations. The Wiener–Lévy process is defined in the range $-\infty < y < \infty$ and $t > 0$ through [cf. Eq. (1.5)]

$$P_1(y_1, t_1) = \frac{1}{\sqrt{2\pi t_1}} \exp\left(-\frac{y_1^2}{2t_1}\right) , \tag{3.11a}$$

$$P(y_2, t_2|y_1, t_1) = \frac{1}{\sqrt{2\pi(t_2 - t_1)}} \exp\left[-\frac{(y_2 - y_1)^2}{2(t_2 - t_1)}\right] , \quad (t_1 < t_2) . \tag{3.11b}$$

This is a non-stationary (P_1 depends on t), Gaussian process. The second-order moment is

$$\langle Y(t_1)Y(t_2)\rangle = \min(t_1, t_2) , \tag{3.12}$$

Proof: Let us assume $t_1 < t_2$. Then, from Eq. (3.4) we have

$$\begin{aligned} \langle Y(t_1)Y(t_2)\rangle &= \int \mathrm{d}y_1 \int \mathrm{d}y_2\, y_1 y_2 P_2(y_1, t_1; y_2, t_2) \\ &= \int \mathrm{d}y_1\, y_1 P_1(y_1, t_1) \underbrace{\int \mathrm{d}y_2\, y_2 P(y_2, t_2|y_1, t_1)}_{y_1 \text{ by Eq. (3.11b)}} = \underbrace{\int \mathrm{d}y_1\, y_1^2 P_1(y_1, t_1)}_{t_1 \text{ by Eq. (3.11a)}} , \end{aligned}$$

where we have used that t_1 is the time-dependent variance of P_1. Q.E.D.

Ornstein–Uhlenbeck process. This stochastic process was constructed to describe the behaviour of the *velocity* of a free Brownian particle in one dimension (see Sec. 4.5). It also describes the position of an overdamped particle in an harmonic potential. It is defined by ($\Delta t = t_2 - t_1 > 0$)

$$P_1(y_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}y_1^2\right) , \tag{3.13a}$$

$$P(y_2, t_2|y_1, t_1) = \frac{1}{\sqrt{2\pi(1 - e^{-2\Delta t})}} \exp\left[-\frac{(y_2 - y_1 e^{-\Delta t})^2}{2(1 - e^{-2\Delta t})}\right] . \tag{3.13b}$$

The Ornstein–Uhlenbeck process is stationary, Gaussian, and Markovian. According to a theorem due to Doob, it is essentially the only process with these three properties. Concerning the Gaussian property, it is clear for $P_1(y_1)$. For $P_2(y_2,t_2;y_1,t_1) = P_1(y_1)P(y_2,t_2|y_1,t_1)$ [Eq. (3.6)], we have

$$P_2(y_2,t_2;y_1,t_1) = \frac{1}{\sqrt{(2\pi)^2(1-e^{-2\Delta t})}} \exp\left[-\frac{y_1^2 - 2y_1y_2e^{-\Delta t} + y_2^2}{2(1-e^{-2\Delta t})}\right] . \quad (3.14)$$

This expression can be identified with the bivariate Gaussian distribution (2.23) and the following parameters

$$\lambda = e^{-\Delta t} , \qquad \sigma^2 = 1 - e^{-2\Delta t} ,$$

with the particularity that $\sigma^2 = 1-\lambda^2$ in this case. Therefore, we immediately see that the Ornstein–Uhlenbeck process has an exponential auto-correlation function $\langle Y(t_1)Y(t_2)\rangle = e^{-\Delta t}$, since $\sigma^2\lambda/(1-\lambda^2) = \lambda$ in this case.[15]

The evolution with time of the distribution $P_2(y_2,t_2;y_1,t_1)$, seen as the velocity of a Brownian particle, has a clear meaning. At short times the velocity is strongly correlated with itself: then $\lambda \sim 1$ and the distribution would be like in the lower right panel of Fig. 5 [with a shrinked variance $\sigma^2 = (1-\lambda^2) \to 0$]. As time elapses λ decreases and we pass form one panel to the previous and, at long times, $\lambda \sim 0$ and the velocity has lost all memory of its value at the initial time due to the collisions and hence $P_2(y_2,t_2;y_1,t_1)$ is completely uncorrelated.

Exercise: check by direct integration that the transition probability (3.13b) obeys the Chapman–Kolgomorov equation (3.9).

[15] This result can also be obtained by using Eqs. (3.13) directly:

$$\begin{aligned}
\kappa(t_1,t_2) &= \langle Y(t_1)Y(t_2)\rangle - \overbrace{\langle Y(t_1)\rangle\langle Y(t_2)\rangle}^{0} \\
&= \int \mathrm{d}y_1\mathrm{d}y_2\, y_1y_2 \underbrace{P_2(y_1,t_1;y_2,t_2)}_{P_1(y_1)P(y_2,t_2|y_1,t_1)} \\
&= \int_{-\infty}^{\infty} \mathrm{d}y_1\, y_1 \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}y_1^2\right) \underbrace{\int_{-\infty}^{\infty} \mathrm{d}y_2\, y_2 \frac{1}{\sqrt{2\pi(1-e^{-2\Delta t})}} \exp\left[-\frac{(y_2 - \overbrace{y_1e^{-\Delta t}}^{\mu_1})^2}{2\underbrace{(1-e^{-2\Delta t})}_{\sigma^2}}\right]}_{y_1e^{-\Delta t}} \\
&= e^{-\Delta t} \underbrace{\int_{-\infty}^{\infty} \mathrm{d}y_1\, y_1^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}y_1^2\right)}_{1} . \quad \text{Q.E.D.}
\end{aligned}$$

4 The master equation: Kramers–Moyal expansion and Fokker–Planck equation

The Chapman–Kolmogorov equation (3.9) for Markov processes is not of much assistance when one searches for solutions of a given problem, because it is essentially a property of the solution. However, it can be cast into a more useful form—the *master equation*.

4.1 The master equation

The master equation is a differential equation for the transition probability. Accordingly, in order to derive it, one needs first to ascertain how the transition probability behaves for short time differences.

Firstly, on inspecting the Chapman–Kolmogorov equation (3.9) for equal time arguments one finds the natural result

$$P(y_3,t_3|y_1,t) = \int \mathrm{d}y_2\, P(y_3,t_3|y_2,t)P(y_2,t|y_1,t) \quad \Rightarrow \quad P(y_2,t|y_1,t) = \delta(y_2-y_1)\ ,$$

which is the zeroth-order term in the short-time behaviour of $P(y',t'|y,t)$. Keeping this in mind one *adopts* the following expression for the short-time transition probability:

N.b., P(y2, t | y1, t) = delta(y2 - y1). Hence from 4.1, partial_{t'} P(y2, t' | y1, t) evaluated at t' = t is -a^0(y1)delta(y2-y1) + W_t(y2 | y1).

$$\boxed{P(y_2,t+\Delta t|y_1,t) = \delta(y_2-y_1)[1-a^{(0)}(y_1,t)\Delta t] + W_t(y_2|y_1)\Delta t + O[(\Delta t)^2]\ ,} \tag{4.1}$$

where $W_t(y_2|y_1)$ is interpreted as the *transition probability per unit time* from y_1 to y_2 at time t. Then, the coefficient $1-a^{(0)}(y_1,t)\Delta t$ is to be interpreted as the probability that no "transition" takes place during Δt. Indeed, from the normalisation of $P(y_2,t_2|y_1,t_1)$ one has:

W is nowhere negative.

$$1 = \int \mathrm{d}y_2\, P(y_2,t+\Delta t|y_1,t) \simeq 1 - a^{(0)}(y_1,t)\Delta t + \int \mathrm{d}y_2\, W_t(y_2|y_1)\Delta t\ .$$

Therefore, to first order in Δt, one gets[16]

This is used below to allow us to write the master equation solely in terms of W's, without any a^0.

$$a^{(0)}(y_1,t) = \int \mathrm{d}y_2\, W_t(y_2|y_1)\ , \tag{4.2}$$

[16] The reason for the notation $a^{(0)}$ will become clear below.

which substantiates the interpretation mentioned: $a^{(0)}(y_1,t)\Delta t$ is the total probability of escape from y_1 in the time interval $(t, t+\Delta t)$ and, thus, $1 - a^{(0)}(y_1,t)\Delta t$ is the probability that no transition takes place during this time.

Now we can derive the differential equation for the transition probability from the Chapman–Kolmogorov equation (3.9). Insertion of the above short-time expression for the transition probability ~~in~~ into it yields

$$\begin{aligned} P(y_3,t_2+\Delta t|y_1,t_1) &= \int \mathrm{d}y_2 \overbrace{P(y_3,t_2+\Delta t|y_2,t_2)}^{\delta(y_3-y_2)[1-a^{(0)}(y_2,t_2)\Delta t]+W_{t_2}(y_3|y_2)\Delta t} P(y_2,t_2|y_1,t_1) \\ &\simeq [1-a^{(0)}(y_3,t_2)\Delta t]P(y_3,t_2|y_1,t_1) \\ &\quad + \Delta t \int \mathrm{d}y_2\, W_{t_2}(y_3|y_2)P(y_2,t_2|y_1,t_1)\ . \end{aligned}$$

Next, on using Eq. (4.2) to write $a^{(0)}(y_3,t_2)$ in terms of $W_{t_2}(y_2|y_3)$, one has

$$\begin{aligned} &\frac{1}{\Delta t}\left[P(y_3,t_2+\Delta t|y_1,t_1) - P(y_3,t_2|y_1,t_1)\right] \\ &\qquad \simeq \int \mathrm{d}y_2\ \left[W_{t_2}(y_3|y_2)P(y_2,t_2|y_1,t_1) - W_{t_2}(y_2|y_3)P(y_3,t_2|y_1,t_1)\right]\ , \end{aligned}$$

which in the limit $\Delta t \to 0$ yields, after some changes in notation ($y_1,t_1 \to y_0,t_0$, $y_2,t_2 \to y',t$, and $y_3 \to y$), the *master equation*

$$\boxed{\frac{\partial}{\partial t}P(y,t|y_0,t_0) = \int \mathrm{d}y'\ \left[W_t(y|y')P(y',t|y_0,t_0) - W_t(y'|y)P(y,t|y_0,t_0)\right]\ ,} \tag{4.3}$$

which is an integro-differential equation.

The master equation is a differential form of the Chapman–Kolmogorov equation (and sometimes it is referred to as such). Therefore, it is an equation for the transition probability $P(y,t|y_0,t_0)$, but not for $P_1(y,t)$. However, an equation for $P_1(y,t)$ can be obtained by using the concept of "extraction of a sub-ensemble". Suppose that $Y(t)$ is a stationary Markov process characterised by $P_1(y)$ and $P(y,t|y_0,t_0)$. Let us define a new, non-stationary Markov process $Y^*(t)$ for $t \geq t_0$ by setting

$$P_1^*(y_1,t_1) = P(y_1,t_1|y_0,t_0)\ , \tag{4.4a}$$

$$P^*(y_2,t_2|y_1,t_1) = P(y_2,t_2|y_1,t_1)\ . \tag{4.4b}$$

You define a stochastic rate matrix K from W by

K(y | y') = W(y | y') - delta(y - y') \int dy'' W(y'' | y')

So Eq. 4.6 is equivalent to setting

partial P(y, t) / partial t = sum_y' K(y } y') P(y', t).

Intuitively, K is identical to W, just with the diagonal term set to a

negative value to ensure normalization whereas W(y } Y) = 0.

This is a sub-ensemble of $Y(t)$ characterised by taking the sharp value y_0 at t_0, since $P_1^*(y_1, t_0) = \delta(y_1 - y_0)$. More generally, one may extract a sub-ensemble in which at a given time t_0 the values of $Y^*(t_0)$ are distributed according to a given probability distribution $p(y_0)$:

$$P_1^*(y_1, t_1) = \int \mathrm{d}y_0 \, P(y_1, t_1 | y_0, t_0) p(y_0) \ , \tag{4.5}$$

and $P^*(y_2, t_2 | y_1, t_1)$ as in Eq. (4.4b). Physically, the extraction of a sub-ensemble means that one "prepares" the system in a certain non-equilibrium state at t_0.

By construction, the above $P_1^*(y_1, t_1)$ obey the same differential equation as the transition probability (with respect its first pair of arguments), that is, $P_1^*(y_1, t_1)$ obeys the master equation. Consequently, we may write, suppressing unessential indices,

Normalization of probability means the integral over y of the RHS must equal 0 - and it does so automatically, no matter what W is. So in particular, there is no need for W to integrate to 0, 1 or anything else

$$\frac{\partial P(y,t)}{\partial t} = \int \mathrm{d}y' \ [W(y|y')P(y',t) - W(y'|y)P(y,t)] \ . \tag{4.6}$$

If the range of Y is a discrete set of states labelled with n, the equation reduces to

$$\frac{\mathrm{d}p_n(t)}{\mathrm{d}t} = \sum_{n'} [W_{nn'} p_{n'}(t) - W_{n'n} p_n(t)] \ . \tag{4.7}$$

In this form the meaning becomes clear: *the master equation is a balance (gain–loss) equation for the probability of each state.* The first term is the "gain" due to "transitions" from other "states" n' to n, and the second term is the "loss" due to "transitions" into other configurations. Remember that $W_{n'n} \geq 0$ and that the term with $n = n'$ does not contribute.

Owing to $W(y|y')\Delta t$ is the transition probability in a short time interval Δt, it can be computed, for the system under study, by means of any available method valid for short times, e.g., by Dirac's time-dependent perturbation theory leading to the "golden rule". Then, the master equation serves to determine the time evolution of the system over long time periods, at the expense of *assuming* the Markov property.

The master equation can readily be extended to the case of a multi-component Markov process $Y_i(t)$, $i = 1, 2, \ldots, N$, on noting that the Chapman–Kolmogorov equation (3.9) is valid as it stands by merely replacing y by $\mathbf{y} = (y_1, \cdots, y_N)$. Then, manipulations similar as those leading to Eq. (4.6)

yield the multivariate counterpart of the master equation

$$\boxed{\frac{\partial P(\mathbf{y},t)}{\partial t} = \int \mathrm{d}\mathbf{y}' \; [W(\mathbf{y}|\mathbf{y}')P(\mathbf{y}',t) - W(\mathbf{y}'|\mathbf{y})P(\mathbf{y},t)] \; .} \tag{4.8}$$

Example: the decay process. Let us consider an typical example of master equation describing a decay process, in which $p_n(t)$ determines the probability of having at time t, n surviving "emitters" (radioactive nuclei, excited atoms emitting photons, etc.). The transition probability in a short interval is

$$W_{n,n'}\Delta t = \begin{cases} 0 & \text{for} \quad n > n' \\ \gamma n' \Delta t & \text{for} \quad n = n' - 1 \\ \mathcal{O}(\Delta t)^2 & \text{for} \quad n < n' - 1 \end{cases}$$

That is, there are not transitions to a state with more emitters (they can only decay; reabsortion is negligible), and the decay probability of more that one decay in Δt is of higher order in Δt. The decay parameter γ can be computed with quantum mechanical techniques. The corresponding master equation is Eq. (4.7) with $W_{n,n'} = \gamma n' \delta_{n,n'-1}$

$$\frac{\mathrm{d}p_n(t)}{\mathrm{d}t} = W_{n,n+1}\, p_{n+1}(t) - W_{n-1,n}\, p_n(t) \; .$$

and hence

$$\frac{\mathrm{d}p_n(t)}{\mathrm{d}t} = \gamma(n+1)\, p_{n+1}(t) - \gamma n\, p_n(t) \; . \tag{4.9}$$

Without finding the complete solution for $p_n(t)$, we can derive the equation for the average number of surviving emitters $\langle N \rangle (t) = \sum_{n=0}^{\infty} n\, p_n(t)$

$$\begin{aligned} \sum_{n=0}^{\infty} n(\mathrm{d}p_n/\mathrm{d}t) &= \gamma \overbrace{\sum_{n=0}^{\infty} n(n+1)p_{n+1}}^{k=n+1} - \gamma \overbrace{\sum_{n=0}^{\infty} n^2 p_n}^{n=0\to 1} \\ &= \gamma \sum_{k=1}^{\infty} [(\cancel{k} - 1)k - \cancel{k^2}]p_k = -\gamma \overbrace{\sum_{k=1}^{\infty} k p_k}^{\langle N \rangle} \; . \end{aligned}$$

Therefore the differential equation for $\langle N \rangle$ and its solution are:

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle N \rangle = -\gamma \langle N \rangle \; , \quad \Rightarrow \quad \boxed{\langle N \rangle (t) = n_0 e^{-\gamma t} \; .} \tag{4.10}$$

4.2 The Kramers–Moyal expansion and the Fokker–Planck equation

The Kramers–Moyal expansion of the master equation casts this integro-differential equation into the form of a differential equation of infinite order. It is therefore not easier to handle but, under certain conditions, one may break off after a suitable number of terms. When this is done after the second-order terms one gets a partial differential equation of second order for $P(y,t)$ called the Fokker–Planck equation.

Let us first express the transition probability W as a function of the size r of the jump from one configuration y' to another one y, and of the starting point y':

$$W(y|y') = W(y';r)\ , \qquad r = y - y'\ . \tag{4.11}$$

The master equation (4.6) then reads,

$$\frac{\partial P(y,t)}{\partial t} = \int \mathrm{d}r\, W(y-r;r)P(y-r,t) - P(y,t)\int \mathrm{d}r\, W(y;-r)\ , \tag{4.12}$$

where the sign change associated with the change of variables $y' \to r = y - y'$, is absorbed in the boundaries (integration limits), by considering a symmetrical integration interval extending from $-\infty$ to ∞:

$$\int_{-\infty}^{\infty} \mathrm{d}y'\, f(y') = -\int_{y+\infty}^{y-\infty} \mathrm{d}r\, f(y-r) = -\int_{\infty}^{-\infty} \mathrm{d}r\, f(y-r) = \int_{-\infty}^{\infty} \mathrm{d}r\, f(y-r)\ .$$

Moreover, since finite integration limits would incorporate an additional dependence on y, we shall restrict our attention to problems to which the boundary is irrelevant.

Let us now assume that the changes on y occur via small jumps, i.e., that $W(y';r)$ is a sharply peaked function of r but varies slowly enough with y'. A second assumption is that $P(y,t)$ itself also varies slowly with y. It is then possible to deal with the shift from y to $y-r$ in the first integral in Eq. (4.12) by means of a Taylor expansion:

$$\begin{aligned}
\frac{\partial P(y,t)}{\partial t} &= \int \mathrm{d}r\, W(y;r)P(y,t) + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!}\int \mathrm{d}r\, r^m \frac{\partial^m}{\partial y^m}[W(y;r)P(y,t)] \\
&\quad - P(y,t)\int \mathrm{d}r\, W(y;-r) \\
&= \sum_{m=1}^{\infty} \frac{(-1)^m}{m!}\frac{\partial^m}{\partial y^m}\left\{\left[\int \mathrm{d}r\, r^m W(y;r)\right] P(y,t)\right\}\ ,
\end{aligned}$$

The FP equation can also be formulated as a conventional "continuity equation" (as that term is defined in E&M), where the "flux" is *probability* flux rather than (for example) flux of charge, given by a^1 P + (1/2) partial_y [a^2 P]. See comment on p. 1 of esposito.van.den.broeck.fokker.planck.pdf

So in particular, for zero diffusion coefficient one gets deterministic flow, and a^1(y) is the speed of the flow of the probability at y. N.b. ,in this case all jump moments above the first are zero; for all y, W(y;r) has nonzero first moment but all higher moments are zero. See previous page comment on deterministic dynamics.

where we have used that the first and third terms on the first right-hand side cancel each other.[17] Note that the dependence of $W(y;r)$ on its second argument r is fully kept; an expansion with respect to it, is not useful as W varies rapidly with r. Finally, on introducing the *jump moments*

$$a^{(m)}(y,t) = \int \mathrm{d}r\, r^m W(y;r)\,, \tag{4.13}$$

one gets the *Kramers–Moyal expansion of the master equation*:

$$\boxed{\frac{\partial P(y,t)}{\partial t} = \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \frac{\partial^m}{\partial y^m} \left[a^{(m)}(y,t) P(y,t)\right]\,.} \tag{4.14}$$

Formally, Eq. (4.14) is identical with the master equation and is therefore not easier to deal with, but it suggest that one may break off after a suitable number of terms. For instance, there could be situations where, for $m > 2$, $a^{(m)}(y,t)$ is identically zero or negligible. In this case one is left with

$$\boxed{\frac{\partial P(y,t)}{\partial t} = -\frac{\partial}{\partial y}\left[a^{(1)}(y,t)P(y,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left[a^{(2)}(y,t)P(y,t)\right]\,,} \tag{4.15}$$

Expanding, the RHS is

P [A'+ B" / 2]
+
P' [A + B']
+
P" B / 2

where A is shorthand for a^(1) while B is shorthand for a^(2)

which is the celebrated *Fokker–Planck equation*. The first term is called the *drift* or *transport* term and the second one the *diffusion* term, while $a^{(1)}(y,t)$ and $a^{(2)}(y,t)$ are the drift and diffusion "coefficients".

It is worth recalling that, being derived from the master equation, the Kramers–Moyal expansion, and the Fokker–Planck equation as a special case of it, involve the transition probability $P(y,t|y_0,t_0)$ of the Markov stochastic process, not its one-time probability distribution $P_1(y,t)$. However, they also apply to the $P_1^*(y,t)$ of every subprocess that can be extracted from a Markov stochastic process by imposing an initial condition [see Eqs. (4.4) and (4.5)].

4.3 The jump moments

The transition probability per unit time $W(y'|y)$ enters in the definition (4.13) of the jump moments. Therefore, in order to calculate $a^{(m)}(y,t)$, we

[17] This can be shown upon interchanging $-r$ by r and absorbing the sign change in the integration limits, as discussed above.

must use the relation (4.1) between $W(y'|y)$ and the transition probability for short time differences.

Firstly, from Eq. (4.11) one sees that $W(y';r) = W(y|y')$ with $y = y' + r$. Accordingly, one can write

$$W(y;r) = W(y'|y) \ , \qquad y' = y + r \ .$$

On inserting this expression in Eq. (4.13) one can write the jump moments as[18]

$$\boxed{a^{(m)}(y,t) = \int \mathrm{d}y' \, (y'-y)^m W(y'|y) \ .} \tag{4.16}$$

In order to calculate the jumps moments we introduce the quantity

$$\mathcal{A}^{(m)}(y;\tau,t) = \int \mathrm{d}y' \, (y'-y)^m P(y', t+\tau|y,t) \ , \qquad (m \geq 1) \ ,$$

which is the average of $[Y(t+\tau) - Y(t)]^m$ with sharp initial value $Y(t) = y$ (conditional average). Then, by using the short-time transition probability (4.1), one can write

$$\begin{aligned} \mathcal{A}^{(m)}(y;\tau,t) &= \int \mathrm{d}y' \, (y'-y)^m \left\{ \delta(y'-y)[1 - a^{(0)}(y,t)\tau] + W(y'|y)\tau + \mathcal{O}(\tau^2) \right\} \\ &= \tau \int \mathrm{d}y' \, (y'-y)^m W(y'|y) + \mathcal{O}(\tau^2) \\ &= a^{(m)}(y,t)\tau + \mathcal{O}(\tau^2) \ , \qquad (m \geq 1) \ , \end{aligned}$$

where the integral involving the first term in the short-time transition probability vanishes due to the presence of the Dirac delta. Therefore, one can calculate the jump moments from the derivatives of the conditional averages as follows

$$a^{(m)}(y,t) = \frac{\partial}{\partial \tau} \mathcal{A}^{(m)}(y;\tau,t) \bigg|_{\tau=0} \ .$$

Finally, on writing

$$\mathcal{A}^{(m)}(y;\Delta t,t) = \int \mathrm{d}y' \, (y'-y)^m P(y', t+\Delta t|y,t) = \Big\langle [Y(t+\Delta t) - Y(t)]^m \Big\rangle \Big|_{Y(t)=y} \ ,$$

[18] This equation makes clear the notation employed. The quantity $a^{(0)}(y,t)$ [Eq. (4.2)], which was introduced in Eq. (4.1), is indeed the $m = 0$ jump moment.

one can alternatively express the jump moments as

$$\boxed{a^{(m)}(y,t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left\langle [Y(t+\Delta t) - Y(t)]^m \right\rangle \Big|_{Y(t)=y} \ .} \tag{4.17}$$

In Sec. 5 below, which is devoted to the Langevin equation, we shall calculate the corresponding jump moments in terms of the short-time conditional averages by means of this formula.

4.4 Expressions for the multivariate case

The above formulae can be extended to the case of a multi-component Markov process $Y_i(t)$, $i = 1, 2, \ldots, N$. Concerning the *Kramers–Moyal expansion* one only needs to use the multivariate Taylor expansion to get

$$\frac{\partial P}{\partial t} = \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \sum_{j_1 \ldots j_m} \frac{\partial^m}{\partial y_{j_1} \cdots \partial y_{j_m}} \left[a^{(m)}_{j_1,\ldots,j_m}(\mathbf{y},t) P \right] \ , \tag{4.18}$$

while the *Fokker–Planck equation* is then given by

RHS is a sum over a gradient plus a sum over a Hessian.

$$\boxed{\frac{\partial P}{\partial t} = -\sum_i \frac{\partial}{\partial y_i} \left[a^{(1)}_i(\mathbf{y},t) P \right] + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j} \left[a^{(2)}_{ij}(\mathbf{y},t) P \right] \ .} \tag{4.19}$$

In these equations, the jump moments are given by the natural generalisation of Eq. (4.16), namely

$$a^{(m)}_{j_1,\ldots,j_m}(\mathbf{y},t) = \int \mathrm{d}\mathbf{y}' \, (y'_{j_1} - y_{j_1}) \cdots (y'_{j_m} - y_{j_m}) W(\mathbf{y}'|\mathbf{y}) \ , \tag{4.20}$$

and can be calculated by means of the corresponding generalisation of Eq. (4.17):

$$\boxed{a^{(m)}_{j_1,\ldots,j_m}(\mathbf{y},t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left\langle \prod_{\mu=1}^{m} [Y_{j_\mu}(t+\Delta t) - Y_{j_\mu}(t)] \right\rangle \Big|_{Y_k(t)=y_k} \ ,} \tag{4.21}$$

that is, by means of the derivative of the corresponding conditional average.

4.5 Examples of Fokker–Planck equations

Diffusion equation for the position. In Einstein's explanation of Brownian motion he arrived at an equation of the form [see Eq. (1.4)]

$$\boxed{\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2}\,.} \tag{4.22}$$

Comparing with the Fokker–Planck equation (4.15), we see that in this case $a^{(1)}(x,t) \equiv 0$, since no forces act on the particle and hence the net drift is zero. Similarly $a^{(2)}(x,t) = 2D$, which is independent of space and time. This is because the properties of the surrounding medium are homogeneous [otherwise $D = D(x)$]. The solution of this equation for $P(x, t = 0) = \delta(x)$ was Eq. (1.5), which corresponds to the Wiener–Lévy process (3.11).

More precisely, that solution is the single-moment marginal of the full Wiener *process*.

This equation is a special case of the *Smoluchowski equation* for a particle with large damping coefficient γ (overdamped particle), the special case corresponding to no forces acting on the particle.

I.e., corresponding to the dynamics quickly settling to a velocity for which the acceleration equals zero - terminal velocity.

Diffusion equation in phase space (x, v)**.** The true diffusion equation of a free Brownian particle is

$$\boxed{\frac{\partial P}{\partial t} = -v\frac{\partial P}{\partial x} + \gamma\left(\frac{\partial}{\partial v}v + \frac{k_{\rm B}T}{m}\frac{\partial^2}{\partial v^2}\right)P\,.} \tag{4.23}$$

This equation is the no potential limit of the *Klein–Kramers equation* for a particle with an arbitrary damping coefficient γ. From this equation one can obtain the diffusion equation (4.22) using *singular perturbation theory*, as the leading term in a expansion in powers of $1/\gamma$. Alternatively, we shall give a proof of this in the context of the Langevin equations corresponding to these Fokker–Planck equations.[19]

So this approach doesn't work in the limit of small gamma.

We have stated without proof that the Ornstein–Uhlenbeck process describes the time evolution of the transition probability of the velocity of a free Brownian particle. We shall demonstrate this, by solving the equation for the marginal distribution for v obtained from (4.23). The marginal

[19] We shall see that the Langevin equation $m\ddot{x} = -m\gamma\,\dot{x} + \xi(t)$ [Eq. (1.8)] leads to Eq. (4.23), while the overdamped approximation $m\gamma\,\dot{x} \simeq \xi(t)$ corresponds to Eq. (4.22).

The reason that large gamma leads to the approximation that the LHS = 0 is that under the full differential equation, the system will quickly evolve to a velocity where the LHS ~ zero.

probability is $P_V(v,t) = \int \mathrm{d}x\, P(x,v,t)$. Integrating Eq. (4.23) over x, using $\int \mathrm{d}x\, \partial_x P(x,v,t) = 0$, since $P(x = \pm\infty, v, t) = 0$, we find

$$\frac{\partial P_V}{\partial t} = \gamma \left(\frac{\partial}{\partial v} v + \frac{k_\mathrm{B} T}{m} \frac{\partial^2}{\partial v^2} \right) P_V \ . \tag{4.24}$$

We will see that this equation also describes the position of an overdamped particle in an harmonic potential. Thus, let us find the solution of the generic equation

$$\boxed{\tau \partial_t P = \partial_y (yP) + D \partial_y^2 P \ .} \tag{4.25}$$

Introducing the characteristic function (2.2) (so we are solving by the Fourier transform method)

$$G(k,t) = \int \mathrm{d}y\, e^{\mathrm{i}ky} P(y,t) \ , \qquad P(y,t) = \frac{1}{2\pi} \int \mathrm{d}k\, e^{-\mathrm{i}ky} G(k,t) \ ,$$

the second order partial differential equation (4.25) transforms into a first order one

$$\tau \partial_t G + k \partial_k G = -Dk^2 G \ , \tag{4.26}$$

which can be solved by means of the method of characteristics.[20]

In this case the subsidiary system is

$$\frac{\mathrm{d}t}{\tau} = \frac{\mathrm{d}k}{k} = -\frac{\mathrm{d}G}{Dk^2 G} \ .$$

Two integrals are easily obtained considering the systems t, k and k, G:

$$\begin{aligned} \frac{\mathrm{d}t}{\tau} &= \frac{\mathrm{d}k}{k} \quad \rightarrow \quad k = a\, e^{t/\tau} \quad \rightarrow \quad u = k e^{-t/\tau} = a \\ -Dk\mathrm{d}k &= \mathrm{d}G/G \ \rightarrow \ -\tfrac{1}{2} Dk^2 = \ln G + \mathrm{c} \ \rightarrow \ v = e^{-\frac{1}{2}Dk^2} G = b \end{aligned}$$

[20] In brief, if we have a differential equation of the form

$$P \frac{\partial f}{\partial x} + Q \frac{\partial f}{\partial y} = R \ ,$$

and $u(x,y,f) = a$ and $v(x,y,f) = b$ are two solutions of the subsidiary system

$$\frac{\mathrm{d}x}{P} = \frac{\mathrm{d}y}{Q} = \frac{\mathrm{d}f}{R} \ ,$$

the general solution of the original equation is an arbitrary function of u and v, $h(u,v) = 0$.

Then, the solution $h(u,v)=0$ can be solved for v as $v=\phi(u)$ with still an arbitrary function ϕ, leading the desired general solution of Eq. (4.26)

$$G = e^{-\frac{1}{2}Dk^2}\phi(ke^{-t/\tau})\ , \tag{4.27}$$

by means of the methods of characteristics.

The solution for sharp initial values $P(y,t=0)=\delta(y-y_0)$ leads to $G(k,t=0)=\exp(\mathrm{i}ky_0)$, from which we get the functional form of ϕ: $\phi(k)=\exp(\mathrm{i}ky_0+\frac{1}{2}Dk^2)$. Therefore, one finally obtains for $G(k,t)$

$$G(k,t)=\exp\left[\mathrm{i}y_0e^{-t/\tau}\,k-\tfrac{1}{2}D(1-e^{-2t/\tau})\,k^2\right]\ , \tag{4.28}$$

which is the characteristic function of a Gaussian distribution [see Eq. (2.9)], with $\mu_1=y_0e^{-t/\tau}$ and $\sigma^2=D(1-e^{-2t/\tau})$. Therefore, the probability distribution solving Eq. (4.25) is

$$\boxed{P(y,t|y_0,0)=\frac{1}{\sqrt{2\pi D(1-e^{-2t/\tau})}}\exp\left[-\frac{(y-y_0e^{-t/\tau})^2}{2D(1-e^{-2t/\tau})}\right]}\ . \tag{4.29}$$

which, as stated, is the transition probability of the Ornstein–Uhlenbeck process [Eq. (3.13b)]. Q.E.D.

Note finally that the parameters for the original equation for P_V [Eq. (4.24)], are simply $\mu_1=v_0e^{-t/\tau}$ and $\sigma^2=(k_{\mathrm{B}}T/m)(1-e^{-2t/\tau})$. Thus, at long times we have $P_V\propto\exp(-\frac{1}{2}mv^2/k_{\mathrm{B}}T)$ which is simply the statistical mechanical equilibrium Boltzmann distribution for free particles.

Diffusion equation for a dipole. The diffusion equation for a dipole moment p in an electric field E is (neglecting inertial effects)

$$\boxed{\zeta\frac{\partial P}{\partial t}=\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\left[\sin\vartheta\left(k_{\mathrm{B}}T\frac{\partial P}{\partial\vartheta}+pE\sin\vartheta P\right)\right]}\ . \tag{4.30}$$

This equation was introduced by Debye in the 1920's, and constitutes the first example of rotational Brownian motion. ζ is the viscosity coefficient (the equivalent to γ in translational problems). It is easily seen that $P_0\propto\exp(pE\cos\vartheta/k_{\mathrm{B}}T)$ is the stationary solution of Eq. (4.30), which leads to the famous result for the average dipole moment of an assembly of dipoles, $\langle\cos\vartheta\rangle=\coth\alpha-1/\alpha$ with $\alpha=pE/k_{\mathrm{B}}T$ and to *Curie's paramagnetic law* at low fields. However, Eq. (4.30) also governs non-equilibrium situations, and in particular the time evolution between different equilibrium states.

See pruessner.master.fokker.planck.slides

5 The Langevin equation

5.1 Langevin equation for one variable

The Langevin equation for one variable is a "differential equation" of the form [cf. Eq. (1.8)]

$$\boxed{\frac{\mathrm{d}y}{\mathrm{d}t} = A(y,t) + B(y,t)\xi(t)}\,, \tag{5.1}$$

where $\xi(t)$ is a given stochastic process. The choice for $\xi(t)$ that renders $y(t)$[21] a Markov process is that of the Langevin "process" (white noise), which is Gaussian and its statistical properties are

$$\langle \xi(t) \rangle = 0\,, \tag{5.2a}$$
$$\langle \xi(t_1)\xi(t_2) \rangle = 2D\delta(t_1 - t_2)\,. \tag{5.2b}$$

Since Eq. (5.1) is a *first-order* differential equation, for each sample function (realisation) of $\xi(t)$, it determines $y(t)$ uniquely when $y(t_0)$ is given. In addition, the values of the fluctuating term at different times are statistically independent, due to the delta-correlated nature of $\xi(t)$. Therefore, the values of $\xi(t)$ at previous times, say $t' < t_0$, cannot influence the conditional probabilities at times $t > t_0$. From these arguments it follows the Markovian character of the solution of the Langevin equation (5.1).

The terms $A(y,t)$ and $B(y,t)\xi(t)$ are often referred to as the *drift* (*transport*) and *diffusion* terms, respectively. Due to the presence of $\xi(t)$, Eq. (5.1) is a *stochastic differential equation*, that is, a differential equation comprising random terms with given stochastic properties. To solve a Langevin equation then means to determine the statistical properties of the process $y(t)$.

Finally, the higher-order moments of $\xi(t)$ are obtained from the second order ones (5.2), by assuming relations like those of the multivariate Gaussian case, i.e., all odd moments of $\xi(t)$ vanish and, e.g.,

$$\begin{aligned}\langle \xi(t_1)\xi(t_2)\xi(t_3)\xi(t_4) \rangle = (2D)^2 \Big[& \langle \xi(t_1)\xi(t_2) \rangle \langle \xi(t_3)\xi(t_4) \rangle \\ & + \langle \xi(t_1)\xi(t_3) \rangle \langle \xi(t_2)\xi(t_4) \rangle \\ & + \langle \xi(t_1)\xi(t_4) \rangle \langle \xi(t_2)\xi(t_3) \rangle \Big]\,.\end{aligned} \tag{5.3}$$

All those expectations on the RHS are just delta functions, saying the two associated times must be equal.

[21] Hereafter, we use the same symbol for the stochastic process $Y(t)$ and its realisations $y(t)$.

To check this result, we shall demonstrate a slightly more general result known as Novikov theorem.

5.1.1 Novikov theorem and Wick formula

Novikov theorem states that for a multivariate Gaussian distribution (Sec. 2.3) with zero mean

$$P(\mathbf{x}) = \sqrt{\frac{\det \hat{A}}{(2\pi)^n}} \exp\left(-\tfrac{1}{2}\mathbf{x}\cdot\hat{A}\cdot\mathbf{x}\right) , \tag{5.4}$$

the averages of the type $\langle x_i f(\mathbf{x})\rangle$, can be obtained as

$$\boxed{\langle x_i f(\mathbf{x})\rangle = \sum_m \langle x_i x_m\rangle \left\langle \frac{\partial f}{\partial x_m}\right\rangle .} \tag{5.5}$$

Applying this result to $f(\mathbf{x}) = x_j x_k x_\ell$ and using $\partial x_i/\partial x_m = \delta_{im}$, we have

$$\langle x_i x_j x_k x_\ell\rangle = \sum_m \langle x_i x_m\rangle \Big\langle \underbrace{\delta_{jm}x_k x_\ell + x_j\delta_{km}x_\ell + x_j x_k \delta_{\ell m}}_{\partial f/\partial x_m} \Big\rangle .$$

Therefore, using the Kronecker's delta to do the sum, we get Wick's formula

$$\langle x_i x_j x_k x_\ell\rangle = \langle x_i x_j\rangle\langle x_k x_\ell\rangle + \langle x_i x_k\rangle\langle x_j x_\ell\rangle + \langle x_i x_\ell\rangle\langle x_j x_k\rangle . \tag{5.6}$$

Equation (5.3) then follows because $\xi(t)$ is assumed to be a Gaussian process, which *by definition* means that the n times probability distribution $P_n(\xi_1, t_1; \ldots; \xi_n, t_n)$ is a multivariate Gaussian distribution.

Proof of Novikov theorem. We shall demonstrate this theorem in three simple steps:
(1) If we denote by $E(\mathbf{x}) = \frac{1}{2}\sum_{ij} x_i A_{ij} x_j$ minus the exponent in the Gaussian distribution (5.4), we have

$$\begin{aligned}
\frac{\partial E}{\partial x_m} &= \frac{1}{2}\frac{\partial}{\partial x_m}\sum_{ij} x_i A_{ij} x_j = \tfrac{1}{2}\sum_{ij}(\delta_{im}A_{ij}x_j + x_i A_{ij}\delta_{jm}) \\
&= \tfrac{1}{2}\sum_j A_{mj}x_j + \tfrac{1}{2}\sum_i x_i A_{im} \\
[A_{ij}\text{ sym.}] \quad &= \sum_j A_{mj}x_j \quad \Rightarrow \quad \boxed{x_i = \sum_m (A^{-1})_{im}\frac{\partial E}{\partial x_m}} .
\end{aligned}$$

(2) Using the definition of average for $\langle x_i f(\mathbf{x})\rangle$, inserting the above expression for x_i and integrating by parts, we have

$$\begin{aligned}
\langle x_i f(\mathbf{x})\rangle &= \mathrm{C}\int \mathrm{d}\mathbf{x}\, x_i f(\mathbf{x}) e^{-E} \\
&= \mathrm{C}\sum_m (A^{-1})_{im}\int \mathrm{d}\mathbf{x}\, f(\mathbf{x})\overbrace{\frac{\partial E}{\partial x_m}e^{-E}}^{-\partial_{x_m}e^{-E}} \\
&= \sum_m (A^{-1})_{im}\mathrm{C}\int \mathrm{d}\mathbf{x}\,\frac{\partial f}{\partial x_m}e^{-E} \quad \Rightarrow \quad \boxed{\langle x_i f(\mathbf{x})\rangle = \sum_m (A^{-1})_{im}\left\langle \frac{\partial f}{\partial x_m}\right\rangle}
\end{aligned}$$

(3) Finally, we demonstrate that $\langle X_i X_j\rangle = (\hat{A}^{-1})_{ij}$ (a particular case of the result $\langle\langle X_i X_j\rangle\rangle = (\hat{A}^{-1})_{ij}$ given without proof in Sec. 2.3). Indeed, using the above result for $f = x_j$ and $\partial x_j/\partial x_m = \delta_{jm}$, we have $\langle x_i x_j\rangle = \sum_m (A^{-1})_{im}\delta_{jm} = (A^{-1})_{ij}$. Insertion of this in the above result completes the proof of Novikov's theorem.

5.2 The Kramers–Moyal coefficients for the Langevin equation

Since the solution of the Langevin equation is a Markov process, it obeys a master equation, which may be written in the Kramers–Moyal form (4.14). Let us calculate the successive coefficients (4.17) occurring in that expansion. We first cast the differential equation (5.1) into the form of an integral equation

N.b., a determinstic integral plus a stochastic integral

$$y(t+\Delta t) - y = \int_t^{t+\Delta t} \mathrm{d}t_1\, A[y(t_1), t_1] + \int_t^{t+\Delta t} \mathrm{d}t_1\, B[y(t_1), t_1]\xi(t_1)\ , \qquad (5.7)$$

where y stands for the initial value $y(t)$. On expanding according to

$$\begin{aligned}
A[y(t_1), t_1] &= A(y, t_1) + A'(y, t_1)[y(t_1) - y] + \cdots\ , \\
B[y(t_1), t_1] &= B(y, t_1) + B'(y, t_1)[y(t_1) - y] + \cdots\ ,
\end{aligned}$$

where the prime denotes partial derivative with respect to y evaluated at the initial point:

$$A'(y,t) \equiv \left.\frac{\partial A}{\partial y}\right|_y \qquad B'(y,t) \equiv \left.\frac{\partial B}{\partial y}\right|_y\ ,$$

one gets

$$\begin{aligned} y(t+\Delta t)-y &= \int_t^{t+\Delta t} \mathrm{d}t_1\, A(y,t_1) \\ &\quad + \int_t^{t+\Delta t} \mathrm{d}t_1\, A'(y,t_1)[y(t_1)-y] + \cdots \\ &\quad + \int_t^{t+\Delta t} \mathrm{d}t_1\, B(y,t_1)\xi(t_1) \\ &\quad + \int_t^{t+\Delta t} \mathrm{d}t_1\, B'(y,t_1)[y(t_1)-y]\xi(t_1) + \cdots . \end{aligned} \tag{5.8}$$

For $y(t_1)-y$ in the above integrands we iterate Eq. (5.8) to get

$$\begin{aligned} y(t+\Delta t)-y &= \int_t^{t+\Delta t} \mathrm{d}t_1\, A(y,t_1) \\ &\quad + \int_t^{t+\Delta t} \mathrm{d}t_1\, A'(y,t_1) \int_t^{t_1} \mathrm{d}t_2\, A(y,t_2) \\ &\quad + \int_t^{t+\Delta t} \mathrm{d}t_1\, A'(y,t_1) \int_t^{t_1} \mathrm{d}t_2\, B(y,t_2)\xi(t_2) + \cdots \\ &\quad + \int_t^{t+\Delta t} \mathrm{d}t_1\, B(y,t_1)\xi(t_1) \\ &\quad + \int_t^{t+\Delta t} \mathrm{d}t_1\, B'(y,t_1)\xi(t_1) \int_t^{t_1} \mathrm{d}t_2\, A(y,t_2) \\ &\quad + \int_t^{t+\Delta t} \mathrm{d}t_1\, B'(y,t_1)\xi(t_1) \int_t^{t_1} \mathrm{d}t_2\, B(y,t_2)\xi(t_2) + \cdots \end{aligned} \tag{5.9}$$

If we take the average of this equation for fixed $y = y(t)$, by using the statistical properties (5.2), we obtain the conditional average required to get $a^{(1)}(y,t)$

$$\begin{aligned} \langle y(t+\Delta t)-y \rangle &= \int_t^{t+\Delta t} \mathrm{d}t_1\, A(y,t_1) + \int_t^{t+\Delta t} \mathrm{d}t_1\, A'(y,t_1) \int_t^{t_1} \mathrm{d}t_2\, A(y,t_2) \\ &\quad + 2D \int_t^{t+\Delta t} \mathrm{d}t_1\, B'(y,t_1) \int_t^{t_1} \mathrm{d}t_2\, B(y,t_2)\delta(t_2-t_1) + \cdots . \end{aligned}$$

Next, on using for the Dirac delta the result $\int_{t_0}^{t_1} \mathrm{d}t\, \delta(t-t_0) f(t) = \frac{1}{2} f(t_0)$, we obtain

$$\int_t^{t_1} \mathrm{d}t_2\, B(y,t_2)\delta(t_2-t_1) = \tfrac{1}{2} B(y,t_1)\ . \tag{5.10}$$

Finally, on considering that $a^{(1)} = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \langle y(t+\Delta t) - y\rangle \,|_{y(t)=y}$, for the calculation of which only terms through order Δt need to be retained, one finally gets

$$a^{(1)}(y,t) = A(y,t) + D\, B(y,t) \frac{\partial B(y,t)}{\partial y}\ .$$

Other integrals not written down in the above formulae do not contribute in the limit $\Delta t \to 0$. This can be seen as follows: each Langevin fluctuating term on the right-hand side of Eq. (5.9), is accompanied by an integral. The lowest-order terms are written in that expression, whereas higher-order terms can be of two types: (i) Integrals of the form of, e.g.,

$$\left\langle \int_t^{t+\Delta t} \mathrm{d}t_1 \cdots \xi(t_1) \int_t^{t_1} \mathrm{d}t_2 \cdots \xi(t_2) \int_t^{t_2} \mathrm{d}t_3 \cdots \xi(t_3) \int_t^{t_3} \mathrm{d}t_4 \cdots \xi(t_4) \right\rangle\ ,$$

which can only give a contribution proportional to $(\Delta t)^2$, as it is seen by using the splitting of $\langle \xi(t_1)\xi(t_2)\xi(t_3)\xi(t_4)\rangle$ in sum of products of the form $\langle \xi(t_i)\xi(t_j)\rangle \langle \xi(t_k)\xi(t_\ell)\rangle$ [Eq. (5.3)]. (ii) Integrals containing no Langevin terms, which are proportional to $(\Delta t)^n$, where n is the number of simple integrals. Both types of terms clearly vanish when dividing by Δt and taking the limit $\Delta t \to 0$.

On using the same type of arguments to identify some vanishing integrals one can compute the second coefficient in the Kramers–Moyal expansion, $a^{(2)} = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \langle [y(t+\Delta t) - y]^2\rangle \,|_{y(t)=y}$, obtaining

$$a^{(2)}(y,t) = \lim_{\Delta t\to 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} \mathrm{d}t_1\, B(y,t_1) \int_t^{t+\Delta t} \mathrm{d}t_2\, B(y,t_2) 2D\delta(t_1-t_2) = 2D\, B^2(y,t)\ ,$$

whereas all the coefficients $a^{(m)}$ vanish for $m \geq 3$. Thus, on collecting all

N.b.; it's not that those terms are being approximated away, but that they equal zero identically. I think ....

these results one can finally write

$$\boxed{\begin{aligned} a^{(1)}(y,t) &= A(y,t) + D\,B(y,t)\frac{\partial B(y,t)}{\partial y}\,, \\ a^{(2)}(y,t) &= 2D\,B^2(y,t)\,, \\ a^{(m)}(y,t) &= 0\,, \qquad \text{for} \quad m \geq 3\,. \end{aligned}} \tag{5.11}$$

5.3 Fokker–Planck equation for the Langevin equation

From Eq. (5.11) it follows that, for the Markov stochastic process determined by the Langevin equation (5.1) with Gaussian δ-correlated $\xi(t)$, the Kramers–Moyal expansion includes up to second-order terms. Therefore, the distribution of probability obeys a Fokker–Planck equation [Eq. (4.15)], which in terms of the above jump moments is explicitly given by

$$\boxed{\frac{\partial P}{\partial t} = -\frac{\partial}{\partial y}\left\{\left[A(y,t) + D\,B(y,t)\frac{\partial B(y,t)}{\partial y}\right]P\right\} + D\frac{\partial^2}{\partial y^2}\left[B^2(y,t)P\right]\,.} \tag{5.12}$$

Note that, along with the deterministic drift $A(y,t)$, $a^{(1)}(y,t)$ contains a term, $DB(y,t)B'(y,t)$, which is called the *noise-induced drift*. This equation is very important, since it allows one to construct the Fokker–Planck equation directly in terms of the coefficients appearing in the equation of motion. In some cases, it can even be done by simply inspection of that equation.

Intuitively, if noise is bigger at y + deltay than at y, then the probability mass at y + delta y will "spill over", causing a net negative drift at y. That's why the derivative of B matters.

5.3.1 Multivariate case

The stochastic differential (Langevin) equation for a multi-component process $\mathbf{y} = (y_1, \cdots, y_N)$ has the form

$$\boxed{\frac{\mathrm{d}y_i}{\mathrm{d}t} = A_i(\mathbf{y},t) + \sum_k B_{ik}(\mathbf{y},t)\xi_k(t)\,,} \tag{5.13}$$

where the $\xi_k(t)$ are $N_{\rm L}$ white-noise terms.[22] The statistical properties of the $\xi_k(t)$ are

$$\langle \xi_k(t) \rangle = 0 \,, \tag{5.14a}$$
$$\langle \xi_k(t_1) \xi_\ell(t_2) \rangle = 2D \delta_{k\ell} \delta(t_1 - t_2) \,. \tag{5.14b}$$

Again, the higher-order moments are obtained from these ones, on assuming relations like those of the (multivariate) Gaussian case.

The successive coefficients (4.21) occurring in the Kramers–Moyal expansion (4.18) can be calculated by using arguments entirely analogous to those employed above to identify some vanishing integrals. On doing so, one gets the following generalisation of Eqs. (5.11) in the multivariate case:

$$\begin{aligned} a_i^{(1)}(\mathbf{y},t) &= A_i(\mathbf{y},t) + D \sum_{jk} B_{jk}(\mathbf{y},t) \frac{\partial B_{ik}(\mathbf{y},t)}{\partial y_j} \,, \\ a_{ij}^{(2)}(\mathbf{y},t) &= 2D \sum_k B_{ik}(\mathbf{y},t) B_{jk}(\mathbf{y},t) \,, \\ a_{j_1,\ldots,j_m}^{(m)}(\mathbf{y},t) &= 0 \,, \qquad \text{for} \quad m \geq 3 \,. \end{aligned} \tag{5.15}$$

Again, for the Markov stochastic process defined by the set (5.13) of Langevin equations, the Kramers–Moyal expansion of the master equation includes up to second-order terms, so that the probability distribution obeys a Fokker–Planck equation

$$\begin{aligned} \frac{\partial P}{\partial t} &= -\sum_i \frac{\partial}{\partial y_i} \left\{ \left[A_i(\mathbf{y},t) + D \sum_{jk} B_{jk}(\mathbf{y},t) \frac{\partial B_{ik}(\mathbf{y},t)}{\partial y_j} \right] P \right\} \\ &\quad + D \sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j} \left\{ \left[\sum_k B_{ik}(\mathbf{y},t) B_{jk}(\mathbf{y},t) \right] P \right\} \,, \end{aligned} \tag{5.16}$$

RHS is sums over a gradient and over a Hessian.

which is entirely determined by the coefficients of the Langevin equation.

[22] The number of Langevin sources, $N_{\rm L}$, does not need to be equal to the number of equations. For example, the sum in k in Eq. (5.13) can even have one term, $N_{\rm L} = 1$ —the case of "scalar noise".

5.4 Examples of Langevin equations and derivation of their Fokker–Planck equations

5.4.1 Diffusion in phase-space: Klein–Kramers equation

Let us consider the following generalisation of the original Langevin equation (1.8), in order to account for the presence of an external potential $U(x,t)$ (e.g., gravity in the Brownian motion problem)

$$\boxed{m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -m\gamma\,\frac{\mathrm{d}x}{\mathrm{d}t} - \frac{\partial U}{\partial x} + m\xi(t)\;.} \tag{5.17}$$

This is simply Newton equation augmented by the fluctuating force [for convenience we have extracted m from $\xi(t)$].

with a velocity-dependent force as well as a conservative one.

Let us divide by the mass, introduce $V = U/m$ and the notation $V' = \partial V/\partial x$, and write (5.17) as a pair of first-order differential equations

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v \tag{5.18}$$

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -(\gamma\,v + V') + \xi(t)\;. \tag{5.19}$$

Then, comparing with the multivariate Langevin equation (5.13), we identify $\xi_x(t) \equiv 0$ and $\xi_v(t) = \xi(t)$, as well as

$$\begin{array}{lll} A_x = v & B_{xx} \equiv 0 & B_{xv} \equiv 0 \\ A_v = -(\gamma\,v + V') & B_{vx} \equiv 0 & B_{vv} = 1 \end{array}$$

Inserting these results in the general Fokker–Planck equation (5.16), one gets (note $\partial_j B_{ik} \equiv 0$)

VERY sloppy here and in Eq. 5.20; the P on the RHS is part of the arguments of the partial derivatives on the RHS.

$$\frac{\partial P}{\partial t} = \left[-\frac{\partial}{\partial x}v - \frac{\partial}{\partial v}[-(\gamma\,v + V')] + D\frac{\partial^2}{\partial v^2}\right] P\;.$$

That's from physics; T is temperature.

Gathering the Hamiltonian terms and identifying $D/\gamma = k_{\mathrm{B}}T/m$, we finally find the famous *Klein–Kramers equation*

$$\boxed{\frac{\partial P}{\partial t} = -v\frac{\partial P}{\partial x} + V'\frac{\partial P}{\partial v} + \gamma\left(\frac{\partial}{\partial v}v + \frac{k_{\mathrm{B}}T}{m}\frac{\partial^2}{\partial v^2}\right) P\;.} \tag{5.20}$$

...into Eq. 5.20 ...

The result $D/\gamma = k_{\rm B}T/m$ comes from inserting the Boltzmann distribution $P_0 \propto \exp[-(\frac{1}{2}mv^2 + mV)/k_{\rm B}T]$ and finding the conditions for it to be a stationary solution. This is equivalent to P. Langevin recourse to the equipartition theorem (Sec. 1.1.2) to find $\langle mv^2 \rangle = k_{\rm B}T$. Note finally that in the absence of potential, the Klein–Kramers equation leads to the equation for free diffusion (4.23), with solution (for the marginal distribution $P_V(v,t) = \int {\rm d}x\, P(x,v,t)$) given by the Ornstein–Uhlenbeck process (4.29).

5.4.2 Overdamped particle: Smoluchowski equation

Let us consider the overdamped limit of the Newton–Langevin equation (5.17)

the "underdamped limit" is just 5.17, with all terms included.

I.e., take m -> 0 while gamma -> infinity, in such a way that their product is finite, and then divide both sides of 5.17 by m gamma.

$$\boxed{\frac{{\rm d}x}{{\rm d}t} = -V'/\gamma + \xi(t)/\gamma \,.} \tag{5.21}$$

N.b., since the coefficient of the noise term is uniform across x, by Eq. 5.12, the force term - V' - is the FP *drift* term (up to the prop. constant, 1/gamma). So force equals average velocity, in the overdamped regime.

Comparing with the univariate Langevin equation (5.1), we identify

$$A = -V'/\gamma \,, \qquad B \equiv 1/\gamma \,.$$

Inserting these results in the Fokker–Planck equation (5.12), one gets (putting again $D/\gamma = k_{\rm B}T/m$) the *Smoluchowski equation*

$$\boxed{\frac{\partial P}{\partial t} = \left(\frac{1}{\gamma}\frac{\partial}{\partial x}V' + \frac{k_{\rm B}T}{m\gamma}\frac{\partial^2}{\partial x^2} \right) P \,.} \tag{5.22}$$

The result $D/\gamma = k_{\rm B}T/m$ can also be obtained on inserting the marginal Boltzmann distribution $P_0 \propto \exp[-mV(x)/k_{\rm B}T]$ and finding the conditions for it to be a stationary solution.

In the absence of potential, the Smoluchowski equation leads to the equation for free diffusion (4.22) or Einstein's Eq. (1.4), with solution given by the Wiener–Lévy process [Eqs. (1.5) or (3.11)]. Note also that in an harmonic potential $V(x) = \frac{1}{2}\omega_0^2 x^2$, the equation is equivalent to Eq. (4.25), with parameters $\tau = \gamma/\omega_0^2$, $D = k_{\rm B}T/m\omega_0^2$, whose solution is the Ornstein–Uhlenbeck process (4.29). Therefore we can at once write for the overdamped harmonic oscillator

$$P(x,t) = \sqrt{\frac{m\omega_0^2}{2\pi k_{\rm B}T(1-e^{-2t/\tau})}} \exp\left[-\frac{m\omega_0^2 (x - x_0 e^{-t/\tau})^2}{2k_{\rm B}T(1-e^{-2t/\tau})} \right] \,, \qquad \tau = \gamma/\omega_0^2 \,. \tag{5.23}$$

Thus, at long times we have $P \propto \exp(-\frac{1}{2}m\omega_0^2 x^2/k_{\rm B}T)$ which is simply the statistical mechanical equilibrium Boltzmann distribution for the harmonic oscillator. In addition Eq. (5.23) tells us how the relaxation to the equilibrium state proceeds.

5.4.3 Equations for a classical spin (dipole)

For classical spins, the Langevin equation is the *stochastic Landau–Lifshitz equation*, which for instance describes the dynamics of the magnetic moment of a magnetic nanoparticle. Written in simplified units (fields in frequency units), it reads

$$\boxed{\frac{{\rm d}\vec{s}}{{\rm d}t} = \vec{s} \wedge \left[\vec{B} + \vec{\xi}(t)\right] - \lambda \vec{s} \wedge \left(\vec{s} \wedge \vec{B}\right) \ ,} \tag{5.24}$$

Here, $\vec{B} = -\partial\mathcal{H}/\partial\vec{s}$ is the effective field associated with the Hamiltonian of the spin $\mathcal{H}(\vec{s})$ (the equivalent to $F = -\partial U/\partial x$ in mechanical problems), and the double vector product is the damping term, which rotates $\vec{s}$ towards the potential minima (preserving its length). The stochastic properties of the components of $\vec{\xi}(t)$ are the usual ones [Eq. (5.14)], but $\vec{\xi}(t)$ is now interpreted as a fluctuating field. Finally, the damping coefficient λ measures the relative importance of the relaxation and precession terms.

The stochastic Landau–Lifshitz equation (5.24), can be cast into the form of the general system of Langevin equations (5.13), by identifying

$$A_i = \sum_{jk} \epsilon_{ijk}\, s_j B_k + \lambda \sum_k (s^2\delta_{ik} - s_i s_k) B_k \ , \tag{5.25}$$

$$B_{ik} = \sum_j \epsilon_{ijk} s_j \ . \tag{5.26}$$

where ϵ_{ijk} is the antisymmetrical unit tensor of rank three (Levi-Civita symbol)[23] and we have expanded the triple vector products $-\vec{s}\wedge(\vec{s}\wedge\vec{B})$ by using the rule $\vec{a} \wedge \left(\vec{b} \wedge \vec{c}\right) = \vec{b}\left(\vec{a}\cdot\vec{c}\right) - \vec{c}\left(\vec{a}\cdot\vec{b}\right)$ ("BAC-CAB" rule).

To calculate the *noise-induced* drift coefficient of the Fokker–Planck equation [the term accompanying A_i in Eq. (5.16)] we need the derivative of the

[23] This tensor is defined as the tensor antisymmetrical in all three indices with $\epsilon_{xyz} = 1$. Therefore, one can write the vector product of $\vec{A}$ and $\vec{B}$ as $\left(\vec{A}\wedge\vec{B}\right)_i = \sum_{jk} \epsilon_{ijk} A_j B_k$. In

diffusion "coefficient":

$$\frac{\partial B_{ik}}{\partial s_j} = \epsilon_{ijk} \ . \tag{5.29}$$

On multiplying by Eq. (5.26) for B_{ik}, summing, and using the second contraction property (5.28) for the ϵ_{ijk}, one finds

$$\sum_{jk} B_{jk} \frac{\partial B_{ik}}{\partial s_j} = \sum_{\ell} \overbrace{\Big(\sum_{jk} \epsilon_{j\ell k} \epsilon_{ijk} \Big)}^{-2\delta_{i\ell}} s_\ell \ , \quad \Rightarrow \quad \boxed{D \sum_{jk} B_{jk} \frac{\partial B_{ik}}{\partial s_j} = -2Ds_i}$$

Let us compute now the coefficient in the diffusion term $\sum_k B_{ik} B_{jk}$:

$$\begin{aligned} \sum_k B_{ik} B_{jk} &= \sum_k \Big(\sum_r \epsilon_{irk} s_r \Big) \Big(\sum_s \epsilon_{jsk} s_s \Big) = \sum_{r,s} s_r s_s \overbrace{\sum_k \epsilon_{irk} \epsilon_{jsk}}^{\delta_{ij}\delta_{rs} - \delta_{is}\delta_{rj}} \\ &= s^2 \delta_{ij} - s_i s_j \ , \quad \Rightarrow \quad \sum_k B_{ik} B_{jk} = s^2 \delta_{ij} - s_i s_j \ , \end{aligned}$$

where we have employed the contraction rule (5.27).

Therefore the Langevin equation associated to the stochastic Landau–Lifshitz equation (5.24), reads

$$\begin{aligned} \frac{\partial P}{\partial t} &= -\sum_i \frac{\partial}{\partial s_i} \Bigg\{ \Bigg[\sum_{jk} \epsilon_{ijk} \, s_j B_k + \lambda \sum_k (s^2 \delta_{ik} - s_i s_k) B_k - 2D s_i \Bigg] P \Bigg\} \\ &\quad + D \sum_{ij} \frac{\partial^2}{\partial s_i \partial s_j} \Bigg[\Big(s^2 \delta_{ij} - s_i s_j \Big) P \Bigg] \ . \end{aligned}$$

Taking the s_j-derivative in the last term by using $\sum_j \partial_j (s^2 \delta_{ij} - s_i s_j) = \sum_j (2 s_j \delta_{ij} - \delta_{ij} s_j - s_i \delta_{jj}) = 2 s_i - s_i - 3 s_i$, we obtain $D \sum_i \partial_i [-2 s_i P + \sum_j (s^2 \delta_{ij} -$

addition, one has the useful contraction property

$$\sum_k \epsilon_{ijk} \epsilon_{i'j'k} = \delta_{ii'} \delta_{jj'} - \delta_{ij'} \delta_{ji'} \tag{5.27}$$

$$\sum_{jk} \epsilon_{ijk} \epsilon_{i'jk} = 2 \delta_{ii'} \ , \qquad \sum_{ijk} \epsilon_{ijk} \epsilon_{ijk} = 6 \ . \tag{5.28}$$

where the last two are obtained by repeated contraction of the first one.

$s_i s_j)\partial_j P]$. The first term cancels $-2Ds_i$ in the drift and the 2nd can be combined with $\sum_k (s^2\delta_{ik} - s_i s_k)B_k$. Finally, returning to a vector notation

$$\boxed{\frac{\partial P}{\partial t} = -\frac{\partial}{\partial \vec{s}} \cdot \left\{ \vec{s} \wedge \vec{B} - \lambda\, \vec{s} \wedge \left[\vec{s} \wedge \left(\vec{B} - k_{\rm B}T \frac{\partial}{\partial \vec{s}} \right) \right] \right\} P \ ,} \tag{5.30}$$

where $(\partial/\partial\vec{s}) \cdot \vec{J} = \sum_i (\partial J_i/\partial s_i)$ (divergence) and, by analogy with the mechanical problems, we have set $D = \lambda k_{\rm B}T$. This equation can be seen as the rotational counterpart of the Klein–Kramers equation.

The electric case corresponds to the precession term dominated by the damping term (a sort of Smoluchowski equation)

$$\frac{\partial P}{\partial t} = \lambda \frac{\partial}{\partial \vec{p}} \cdot \left\{ \vec{p} \wedge \left[\vec{p} \wedge \left(\vec{E} - k_{\rm B}T \frac{\partial}{\partial \vec{p}} \right) \right] \right\} P \ . \tag{5.31}$$

Then, introducing spherical coordinates (ϑ, φ) and assuming the Hamiltonian to be axially symmetric, the equation above reduces to

$$\frac{1}{\lambda}\frac{\partial P}{\partial t} = \frac{1}{\sin\vartheta}\frac{\partial}{\partial \vartheta}\left[\sin\vartheta \left(\frac{\partial \mathcal{H}}{\partial \vartheta} P + k_{\rm B}T \frac{\partial P}{\partial \vartheta} \right) \right] \ . \tag{5.32}$$

This equation corresponds to Eq. (4.30) by identifying $1/\lambda \to \zeta$ and $\mathcal{H} = -pE\cos\vartheta$, which is the Hamiltonian of the dipole in an external field.

Let us solve Eq. (4.30) for a time dependent field $E(t)$ with $E(t<0) = \Delta E$, while $E(t>0) = 0$. That is, in the distant past the system was equilibrated in the presence of a small field ΔE which at $t = 0$ is removed, and we seek for the time evolution of $P(\vartheta, t)$. For $t < 0$ it is easily seen that $P_0 = N\exp(p\Delta E\cos\vartheta/k_{\rm B}T)$ is the stationary solution of Eq. (4.30), since $\partial P_0/\partial\vartheta = -(p\Delta E\sin\vartheta/k_{\rm B}T)P_0$. Then, introducing the notation $\alpha = p\Delta E/k_{\rm B}T$, and using that $\alpha \ll 1$, we have

$$P_0(\vartheta) \simeq N(1 + \alpha\cos\vartheta) \ , \quad (t<0) \ . \tag{5.33}$$

For $t > 0$ the field is removed and for the solution we use the ansatz

$$P(\vartheta, t) = N\left[1 + \alpha g(t)\cos\vartheta\right] \ , \quad (t>0) \ , \tag{5.34}$$

with $g(t)$ a function to be determined by inserting this P in the Fokker–Planck equation (4.30) with $E(t>0) = 0$:

$$\zeta N\alpha \frac{{\rm d}g}{{\rm d}t}\cos\vartheta = -k_{\rm B}TN\alpha\, g \frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin^2\vartheta\right) \ .$$

Therefore, defining the *Debye relaxation time*

$$\tau_{\rm D} = \zeta/2k_{\rm B}T \ , \tag{5.35}$$

we have

$$\mathrm{d}g/\mathrm{d}t = -g/\tau_{\rm D} \qquad \Longrightarrow \qquad g(t) = \overbrace{g(0)}^{1} e^{-t/\tau_{\rm D}} \ . \tag{5.36}$$

The initial condition $g(0) = 1$, comes from the matching of the distributions (5.33) and (5.34) at $t = 0$. Since the normalisation constant follows from $1 = \int_0^\pi \mathrm{d}\vartheta \sin\vartheta P_0 = 2N$, we finally have

$$\boxed{P(\vartheta, t) = \frac{1}{2}\left[1 + \frac{p\Delta E}{k_{\rm B}T} e^{-t/\tau_{\rm D}} \cos\vartheta\right] \ .} \tag{5.37}$$

It is easy to compute now the average dipole moment along the field direction $\langle p\cos\vartheta\rangle = \int_0^\pi \mathrm{d}\vartheta \sin\vartheta P(\vartheta, t) p\cos\vartheta$, by the change of variables $z = \cos\vartheta$

$$\langle p\cos\vartheta\rangle = \frac{p}{2}\int_{-1}^{1} \mathrm{d}z \left[z + \frac{p\Delta E}{k_{\rm B}T} e^{-t/\tau_{\rm D}} z^2\right] \quad \Rightarrow \quad \boxed{\langle p\cos\vartheta\rangle = \frac{p^2\Delta E}{3k_{\rm B}T} e^{-t/\tau_{\rm D}} \ .} \tag{5.38}$$

This result goes from the *Curie law* for the linear response of a paramagnet $p^2\Delta E/3k_{\rm B}T$ in the initial equilibrium regime, to zero at long times, corresponding to the final equilibrium state in the absence of field. The solution of the Fokker–Planck equation (5.37) provides also a complete description the intermediate non-equilibrium regime.

6 Linear response theory, dynamical susceptibilities, and relaxation times (Kramers' theory)

In this section we shall consider some general results that can be obtained for the response of a very general class of dynamical systems (and in particular those having a "Fokker–Planck dynamics"), results which hold when the external perturbation is weak enough. In the context of this *linear response theory*, due to Kubo, the definition of dynamical *response functions* appears naturally (in the time and frequency domain), and sometimes, associated with them, quantities characterising the lifetime of certain states—the relaxation times.

6.1 Linear response theory

Let us consider a system governed by an evolution equation of the type

$$\boxed{\partial_t P = \mathcal{L} P \ ,} \tag{6.1}$$

where $\mathcal{L}$ is a linear operator (linear in its action on P), and P characterises the system. The natural example in this context is the Fokker–Planck operator

$$\mathcal{L} = -\partial_y \left(a^{(1)} \cdot \right) + \tfrac{1}{2} \partial_y^2 \left(a^{(2)} \cdot \right) \tag{6.2}$$

Well, let us apply an external perturbation, and separate the part of the Fokker–Planck operator accounting for the coupling with the perturbation, and the unperturbed part, which we assume to have a stationary solution P_0:

$$\mathcal{L} = \mathcal{L}_{,0} + \mathcal{L}_{\,\mathrm{ext}}(t) \ , \qquad \mathcal{L}_{,0} P_0 = 0 \ .$$

These conditions are quite general. For instance $\mathcal{L}$ could be the Liouville operator of a mechanical system, P_0 the equilibrium Boltzmann distribution, and $\mathcal{L}_{\,\mathrm{ext}}$ the part corresponding to the external potential. In the Fokker–Planck case $\mathcal{L}_{\,\mathrm{ext}}$ could be $V'_{\mathrm{ext}} \partial_v$ in the Klein–Kramers equation or $\gamma^{-1} \partial_x (V'_{\mathrm{ext}} \cdot)$ in the Smoluchowski equation. However, the external perturbation does not need to be a force or a field. For instance, external modulations of the system parameters, like the bath temperature $\Delta T(t)$ are also possible; then in the Smoluchowski equation we will have $\mathcal{L}_{\,\mathrm{ext}} \propto \Delta T \partial_x^2$.

If the perturbation is weak enough, we can write the deviation from the stationary state as $P = P_0 + p$, and the evolution equation would lead to first order to

$$\underbrace{\partial_t P_0}_{0} + \partial_t p = \left[\mathcal{L}_{,0} + \mathcal{L}_{\rm ext}(t)\right](P_0 + p) \simeq \underbrace{\mathcal{L}_{,0} P_0}_{0} + \mathcal{L}_{,0} p + \mathcal{L}_{\rm ext}(t) P_0 \ ,$$

(we have disregarded $\mathcal{L}_{\rm ext}\, p$). The resulting equation can be solved formally[24]

$$\partial_t p = \mathcal{L}_{,0} p + \mathcal{L}_{\rm ext}(t) P_0 \quad \rightarrow \quad \boxed{p = \int_{-\infty}^{t} {\rm d}s\, e^{(t-s)\mathcal{L}_{,0}} \mathcal{L}_{\rm ext}(s) P_0 \ .} \tag{6.3}$$

This equation gives the formal solution to the problem of time evolution in the presence of a weak time-dependent perturbation. **ToDo, warn on the order of the operators**

6.2 Response functions

6.2.1 Time domain

Let us consider any function $c(y)$ of the variables of the system y, and compute the variation of its average with respect to the unperturbed state

$$\Delta C(t) \equiv \langle c \rangle (t) - \langle c \rangle_0 \ . \tag{6.4}$$

To this end we extract the time dependent part of $\mathcal{L}_{\rm ext}(y,t) = \mathcal{L}_{\rm ext}(y) F(t)$ (factorisation is the common case) and use the solution (6.3)

$$\begin{aligned} \Delta C(t) &= \int {\rm d}y\, c(y) P(y,t) - \int {\rm d}y\, c(y) P_0(y) \\ &= \int {\rm d}y\, c(y)\, p(y,t) \\ &= \int_{-\infty}^{t} {\rm d}s \left[\int {\rm d}y\, c\, e^{(t-s)\mathcal{L}_{,0}} \mathcal{L}_{\rm ext} P_0 \right] F(s) \ . \end{aligned}$$

[24] Indeed, taking the t derivative of the presumed solution, we have

$$\partial_t p = \underbrace{\mathcal{L}_{\rm ext}(t) P_0}_{\text{integrand at } s=t} + \underbrace{\mathcal{L}_{,0} \int_{-\infty}^{t} {\rm d}s\, e^{(t-s)\mathcal{L}_{,0}} \mathcal{L}_{\rm ext}(s) P_0}_{\mathcal{L}_{,0} p} = \mathcal{L}_{\rm ext}(t) P_0 + \mathcal{L}_{,0} p \quad \text{Q.E.D.}$$

Then, introducing the response function for the quantity c

$$R_c(t) = \begin{cases} \int \mathrm{d}y\, c(y)\, e^{t\mathcal{L},^0} \mathcal{L}_{\mathrm{ext}} P_0(y) & t > 0 \\ 0 & t < 0 \end{cases} , \tag{6.5}$$

we can write the response $\Delta C(t)$ simply as

$$\boxed{\Delta C(t) = \int_{-\infty}^{\infty} \mathrm{d}s\, R_c(t-s) F(s) \; .} \tag{6.6}$$

The following linear response functions are used [$\Theta(t)$ is the step function $\Theta(t<0) = 0$ and $\Theta(t>0) = 1$]:

$$F(t) = \begin{cases} \delta(t) & \text{pulse response function } \Delta C_{\mathrm{p}}(t) \\ \Theta(t) & \text{excitation function } \Delta C_{\mathrm{e}}(t) \\ \Theta(-t) & \text{relaxation function } \Delta C_{\mathrm{r}}(t) \end{cases} \tag{6.7}$$

Let us consider the last one, also called after-effect function, which corresponds to switch a constant excitation off at $t = 0$

$$\Delta C_{\mathrm{r}}(t) = \int_{-\infty}^{\infty} \mathrm{d}s\, R_c(t-s)\Theta(-s) = \int_{-\infty}^{0} \mathrm{d}s\, R_c(\underbrace{t-s}_{s'}) = \int_{t}^{\infty} \mathrm{d}s'\, R_c(s') \; .$$

Thus, we see that $R_c(t)$ can be obtained as the derivative of $\Delta C_{\mathrm{r}}(t)$

$$\Delta C_{\mathrm{r}}(t) = \int_{t}^{\infty} \mathrm{d}s\, R_c(s) \qquad \Longrightarrow \qquad \boxed{R_c(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \Delta C_{\mathrm{r}} \; .} \tag{6.8}$$

6.2.2 Frequency domain

Introducing now the Fourier transforms of $\Delta C(t)$, $F(t)$, and $R_c(t)$

$$\Delta\tilde{C}(\omega) = \int \mathrm{d}t\, e^{-\mathrm{i}\omega t} \Delta C(t), \quad \tilde{F}(\omega) = \int \mathrm{d}t\, e^{-\mathrm{i}\omega t} F(t), \quad \chi_c(\omega) = \int \mathrm{d}t\, e^{-\mathrm{i}\omega t} R_c(t) \; ,$$

the convolution in Eq. (6.6) relating those quantities, reduces to a simple product

$$\boxed{\Delta\tilde{C}(\omega) = \chi_c(\omega)\tilde{F}(\omega) \; .} \tag{6.9}$$

The complex quantity $\chi_c(\omega)$ is known as the susceptibility. It can be seen that corresponds to the usual definition: if we excite with a perturbation $F(t) = e^{\mathrm{i}\omega t}$, the corresponding response function $\Delta C(t)$ oscillates in the stationary state with $e^{\mathrm{i}\omega t}$ with proportionality coefficient $\chi_c(\omega)$

$$\Delta C(t) = \int_{-\infty}^{\infty} \mathrm{d}s\, R_c(t-s) e^{\mathrm{i}\omega s} = \underbrace{\left[\int_{-\infty}^{\infty} \mathrm{d}s\, R_c(t-s) e^{-\mathrm{i}\omega(t-s)} \right]}_{\chi_c(\omega)} e^{\mathrm{i}\omega t} \, .$$

To conclude, we shall demonstrate a famous linear response relation between the dynamic susceptibility and the relaxation (after-effect) function. From the definition of $\chi_c(\omega)$, using that $R_c(t<0) = 0$ and its relation with the relaxation function, we have

$$\begin{aligned} \chi_c(\omega) &\overset{\text{def.}}{=} \int_{-\infty}^{\infty} \mathrm{d}t\, e^{-\mathrm{i}\omega t} R_c(t) = \int_0^{\infty} \mathrm{d}t\, e^{-\mathrm{i}\omega t} \overbrace{R_c(t)}^{-\mathrm{d}(\Delta C_\mathrm{r})/\mathrm{d}t} \\ &= -\Delta C_\mathrm{r}(t) e^{-\mathrm{i}\omega t}\Big|_0^{\infty} + \int_0^{\infty} \mathrm{d}t\, \Delta C_\mathrm{r}(t) \frac{\mathrm{d}}{\mathrm{d}t} e^{-\mathrm{i}\omega t} \\ &= \Delta C_\mathrm{r}(0) - \mathrm{i}\omega \int_0^{\infty} \mathrm{d}t\, \Delta C_\mathrm{r}(t) e^{-\mathrm{i}\omega t} \, . \end{aligned}$$

Then, extracting the static (thermal-equilibrium) susceptibility $\chi_c^{\mathrm{eq}} = \chi_c(0) = \Delta C_\mathrm{r}(0)$, we have

$$\boxed{\chi_c(\omega) = \chi_c^{\mathrm{eq}} \left[1 - \mathrm{i}\omega \int_0^{\infty} \mathrm{d}t\, \frac{\Delta C_\mathrm{r}(t)}{\Delta C_\mathrm{r}(0)} e^{-\mathrm{i}\omega t} \right] \, ,} \tag{6.10}$$

which gives the dynamical susceptibility in terms of the normalised relaxation function $\Delta C_\mathrm{r}(t)/\Delta C_\mathrm{r}(0)$.

Example: Debye relaxation. We can immediately apply these linear response results to the example of the dynamics of the electrical dipole [Eq. (4.30)]. There, we calculated the time evolution of the average of the field projection of the dipole, Eq. (5.38). Thus, in this case

$$c = p\cos\vartheta \, , \qquad \Delta P_\mathrm{r}(t) = \frac{p^2}{3k_\mathrm{B}T} e^{-t/\tau_\mathrm{D}} \, ,$$

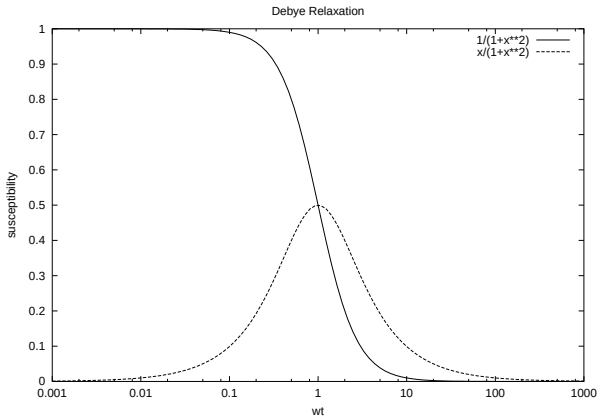

Figure 6: Real part (full line) and imaginary part (dashed line) of the Debye formula (6.11) for the dynamic susceptibility (normalised).

and hence $\chi_p^{\rm eq} = (p^2/3k_{\rm B}T)$ (Curie law). Therefore, we have $\Delta P_{\rm r}(t)/\Delta P_{\rm r}(0) = e^{-t/\tau_{\rm D}}$, so that

$$\chi_p(\omega) = \frac{p^2}{3k_{\rm B}T}\left[1 - {\rm i}\omega \overbrace{\int_0^\infty {\rm d}t\, e^{-({\rm i}\omega + 1/\tau_{\rm D})t}}^{1/({\rm i}\omega+1/\tau_{\rm D})}\right] \quad \Rightarrow \quad \boxed{\chi_p(\omega) = \frac{p^2}{3k_{\rm B}T}\frac{1}{1+{\rm i}\omega\tau_{\rm D}}\,.} \tag{6.11}$$

which is the famous formula displaying "Debye" relaxation (see Fig. 6).

6.3 Relaxation times

When we considered various examples of Fokker–Planck equations, we obtained the solution for the Smoluchowski equation of an harmonic oscillator

$$P(x,t) = \sqrt{\frac{m\omega_0^2}{2\pi k_{\rm B}T(1-e^{-2t/\tau})}}\exp\left[-\frac{m\omega_0^2(x-x_0e^{-t/\tau})^2}{2k_{\rm B}T(1-e^{-2t/\tau})}\right]\,, \qquad \tau = \gamma/\omega_0^2\,.$$

In this equation we see that the time scale for the relaxation to the equilibrium state $P_0 \propto \exp(-\frac{1}{2}m\omega_0^2x^2/k_{\rm B}T)$, is given by $\tau = \gamma/\omega_0^2$. This quantity is the *relaxation time*. In this problem it depends on the system parameters γ and ω_0, but it is independent of the temperature.

Now, in the example of the dielectric dipole, a natural relaxation time has also appeared $\tau_{\rm D} = \zeta/2k_{\rm B}T$ [Eq. (5.35)]. In this problem the relaxation

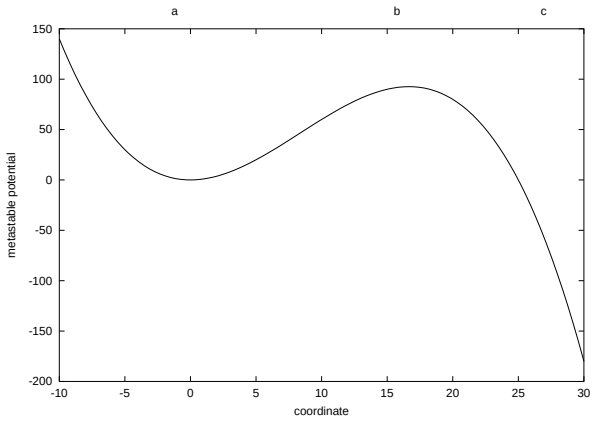

Figure 7: Sketch of a metastable potential for the Kramers calculation of the relaxation time. The point a is at the potential minimum, b at the maximum, and c is a point at the right-side of the barrier.

time depends on T, which is the common case, however, the dependence is not very strong.

It is very common to find expressions for different relaxation times that depend exponentially on T (Arrhenius law), which is the generic behaviour when to establish the equilibrium potential barriers need to be overcome. This was not the case of the previous examples, and such dependence was absent. We shall solve now a simple problem with potential barriers to see how the exponential dependence arises (the theoretical study of this problems was initiated by Kramers in 1940 to study the relaxation rate of chemical reactions). **ToDo, warn on low temperature assumption**

To simplify the calculation let us consider an overdamped particle, described by Smoluchowski equation

$$\frac{\partial P}{\partial t} = \frac{1}{\gamma}\left(\frac{\partial}{\partial x}V' + \frac{k_{\rm B}T}{m}\frac{\partial^2}{\partial x^2}\right)P = -\frac{\partial J}{\partial x}\ . \tag{6.12}$$

where the last equality defines the current of probability J, and the metastable potential is depicted in Fig. 7. At very low temperatures the probability of escape from the metastable minimum is very low (zero at $T = 0$; deterministic system). Therefore the flux of particles over the barrier is very slow, and we can solve the problem as if it were stationary. Then the expression for J,

$$-\gamma J = V'P + \frac{k_{\rm B}T}{m}\frac{\partial P}{\partial x}\ , \tag{6.13}$$

is assumed to be independent of x and the differential equation for P can be integrated ($U = mV$, $D = k_{\rm B}T/m\gamma$)

$$P = e^{-U/k_{\rm B}T}\left[C_1 - \frac{J}{D}\int_c^x {\rm d}x'\, e^{U(x')/k_{\rm B}T}\right], \tag{6.14}$$

were c is an arbitrary point. The integration constant is $C_1 = P(c)e^{U(c)/k_{\rm B}T}$. If we choose c well outside the barrier region $c \to \infty$, we have $P(c) \simeq 0$, and we find for the current J

$$J = -D\frac{P(x)e^{U(x)/k_{\rm B}T}}{\displaystyle\int_c^x {\rm d}x'\, e^{U(x')/k_{\rm B}T}}\ . \tag{6.15}$$

Since $\partial J/\partial x \simeq 0$ we can choose x at will; we set $x = a$ (the metastable minimum), so the integral in the denominator covers the entire maximum. The main contribution to that integral comes from a small region about the maximum $x = b$, so we expand there $U(x) = U_b - \frac{1}{2}m\omega_b^2(x-b)^2$, being $m\omega_b^2 = U''(b)$. The integration limits can be shifted to $\pm\infty$, so the resulting Gaussian integral leads

$$J = D\frac{P(a)e^{U_a/k_{\rm B}T}}{e^{U_b/k_{\rm B}T}\sqrt{2\pi k_{\rm B}T/m\omega_b^2}} = DP(a)\sqrt{\frac{m\omega_b^2}{2\pi k_{\rm B}T}}e^{-(U_b-U_a)/k_{\rm B}T}\ . \tag{6.16}$$

To compute $P(a)$ we use the following argument. The fraction of particles close to the potential minimum N_a can be obtained integrating $P(x)$ in an interval around a, with the distribution approximated as $P = C_2 e^{-U(x)/k_{\rm B}T}$ with $U(x) = U_a + \frac{1}{2}m\omega_a^2(x-a)^2$, where $m\omega_a^2 = U''(a)$. Then for the particles in the well we have $N_a = C_2 e^{-U_a/k_{\rm B}T}\sqrt{2\pi k_{\rm B}T/m\omega_a^2}$, so that $P(a)/N_a = \sqrt{m\omega_a^2/2\pi k_{\rm B}T}$.

The relaxation rate is defined as J/N_a (so the number of particles in the well N_a, times the escape rate, gives the flux J leaving the well). Then, introducing the above expression for $P(a)/N_a$ into Eq. (6.16) divided by N_a, and using $D/k_{\rm B}T = 1/m\gamma$ we finally have

$$\boxed{\frac{1}{\tau} = \frac{\omega_a\omega_b}{2\pi\gamma}e^{-(U_b-U_a)/k_{\rm B}T}\ .} \tag{6.17}$$

This formula has the typical exponential dependence on the barrier height over the temperature. Although the calculation in other cases (intermediate to weak damping) is much more elaborated, this exponential dependence always appears.

7 Methods for solving Langevin and Fokker–Planck equations (mostly numerical)

In general the Fokker–Planck or Langevin equations cannot be solved analytically. In some cases one can use approximate methods, in others numerical methods are preferable. Here we shall discuss some of these methods.

7.1 Solving Langevin equations by numerical integration

We shall start with methods to integrate the Langevin equations numerically. These are the counterpart of the known methods for the deterministic differential equations.

7.1.1 The Euler scheme

In order to integrate the system of Langevin equations

$$\frac{\mathrm{d}y_i}{\mathrm{d}t} = A_i(\mathbf{y},t)+\sum_k B_{ik}(\mathbf{y},t)\xi_k(t)\ , \qquad \langle\xi_k(t)\rangle = 0\ , \quad \langle\xi_k(t)\xi_\ell(s)\rangle = 2D\delta_{k\ell}\delta(t-s)$$

starting at $t = t_0$ with the values $\mathbf{y}_0$, to the time $t_0 + T$, one first divides the time interval $[t_0, t_0 + T]$ into N_s time steps of length Δt, i.e., $t_n = t_0 + n\Delta t$. The stochastic variables at a later time $\mathbf{y}(t_{n+1})$, are calculated in terms of $\mathbf{y}(t_n)$ according to

$$\boxed{y_i(t_{n+1}) = y_i(t_n) + a_i^{(1)}[\mathbf{y}(t_n), t_n]\Delta t + \sum_k B_{ik}[\mathbf{y}(t_n), t_n]\Delta W_{kn}\ ,} \qquad (7.1)$$

where $a_i^{(1)} = A_i + D\sum_{jk} B_{jk}\partial_j B_{ik}$ is the first jump moment, ΔW_{kn}, $k = 1,\ldots,N_\mathrm{L}$ (the number of Langevin sources), $n = 1,\ldots,N_\mathrm{s}$, are independent Gaussian numbers with zero mean and variance $2D\Delta t$, i.e.,

$$\langle\Delta W_{kn}\rangle = 0\ , \qquad \langle\Delta W_{kn}\Delta W_{k'n'}\rangle = (2D\Delta t)\delta_{kk'}\delta_{nn'}\ . \qquad (7.2)$$

The recursive algorithm (7.1) is called the Euler scheme, in analogy with the Euler method to integrate deterministic differential equations. By construc-

tion, for $\Delta t \to 0$, the above recursive scheme, leads to the correct Kramers–Moyal coefficients.[25]

7.1.2 Stochastic Heun scheme

This is a higher-order scheme for the numerical integration of the Langevin equations given by (a sort of Runge–Kutta scheme)

$$\boxed{\begin{aligned} y_i(t+\Delta t) = y_i(t) \;&+\; \tfrac{1}{2}\left\{A_i[\tilde{\mathbf{y}}, t+\Delta t] + A_i[\mathbf{y}(t), t]\right\}\Delta t \\ &+\; \tfrac{1}{2}\sum_k \left\{B_{ik}[\tilde{\mathbf{y}}, t+\Delta t] + B_{ik}[\mathbf{y}(t), t]\right\}\Delta W_k \;, \end{aligned}} \tag{7.4}$$

with Euler-type supporting values,

$$\tilde{y}_i = y_i(t) + A_i[\mathbf{y}(t), t]\Delta t + \sum_k B_{ik}[\mathbf{y}(t), t]\Delta W_k \;. \tag{7.5}$$

Note that if one uses this support value as the numerical integration algorithm [by identifying $y_i(t+\Delta t) = \tilde{y}_i$], the result does not agree with the ordinary Euler scheme if $\partial B_{ik}/\partial y_j \neq 0$ (or equivalently if $a_i^{(1)} \neq A_i$).

The Euler scheme only requires the evaluation of A_i and B_{ik} at one point per time step, while the Heun scheme requires two, increasing the computational effort. Nevertheless, the Heun scheme substitutes the derivatives of B_{ik} by the evaluation of B_{ik} at different points. Besides, it treats the deterministic part of the differential equations with a second-order accuracy in Δt,

---

[25] Let us prove this in the simple one-variable case. Then

$$y(t+\Delta t) = y(t) + a^{(1)}(y,t)\Delta t + B(y,t)\Delta W \;. \tag{7.3}$$

To obtain the Kramers–Moyal coefficients, we average this equation for fixed initial values $y(t)$ (conditional average). To do so, one can use $\langle \Delta W \rangle = 0$ and $\langle \Delta W^2 \rangle = 2D\Delta t$, to get

$$\langle B\Delta W \rangle = 0 \;, \qquad \langle a^{(1)}\Delta t B \Delta W \rangle = 0 \;, \qquad \langle B\Delta W B \Delta W \rangle = 2DB^2\Delta t \;.$$

Therefore, one obtains

$$\begin{aligned} \langle y(t+\Delta t) - y(t) \rangle &= a^{(1)}\Delta t \\ \left\langle \left[y(t+\Delta t) - y(t)\right]^2 \right\rangle &= (a^{(1)})^2\Delta t^2 + 2a^{(1)}\Delta t B \langle \Delta W \rangle + B^2 \langle \Delta W^2 \rangle = 2DB^2\Delta t + \mathcal{O}[(\Delta t)^2] \;, \end{aligned}$$

which lead to the Kramers–Moyal coefficients (5.11) via Eq. (4.17). Q.E.D.

being numerically more stable. Thus, the computational advantage of the Euler scheme, may disappear if it needs to be implemented with a smaller integration step (Δt) to avoid numerical instabilities.

7.1.3 Gaussian random numbers

The *Gaussian* random numbers required to simulate the variables ΔW, can be constructed from *uniformly* distributed ones by means of the Box–Muller algorithm (see, e.g., Ref. [3, p. 280]). This method is based on the following property: if r_1 and r_2 are random numbers uniformly distributed in the interval $(0, 1)$ (as those pseudo-random ones provided by a computer), the transformation

$$w_1 = \sqrt{-2\ln(r_1)}\cos(2\pi r_2)\ , \quad w_2 = \sqrt{-2\ln(r_1)}\sin(2\pi r_2)\ , \tag{7.6}$$

outputs w_1 and w_2, which are Gaussian-distributed independent random numbers of zero mean and variance unity. Then, if one needs Gaussian numbers with variance σ^2, these are immediately obtained by multiplying the above w_i by σ (e.g., $\sigma = \sqrt{2D\Delta t}$ in the Langevin equations).

7.1.4 Example I: Brownian particle

The Langevin equations for a particle subjected to fluctuations and dissipation evolving in a potential $V(x)$ are

$$\left\{ \begin{array}{rcl} \mathrm{d}x/\mathrm{d}t & = & v \\ \mathrm{d}v/\mathrm{d}t & = & -V' - \gamma\, v + \xi(t)\ , \qquad \langle \xi(t)\xi(s)\rangle = 2D\delta(t-s) \end{array} \right. \tag{7.7}$$

with $D = \gamma(k_{\mathrm{B}}T/m)$. For the potential we consider that of a constant force field F plus a periodic substrate potential of the form

$$V(x) = -d\,[\sin x + (\rho/2)\sin 2x] - Fx\ . \tag{7.8}$$

For the familiar case $\rho = 0$, we have the cosine potential and the Langevin equations describe a variety of systems:

(i) Non-linear pendulum:

$$\ddot{\phi} + \gamma\,\dot{\phi} + (g/\ell)\sin\phi = N + \xi(t)\ . \tag{7.9}$$

In this case we have $x = \phi$ (the angle of the pendulum with respect to the vertical direction), $d = g/\ell$ (gravity over length of the pendulum), $F = N$ (external torque), and $D = \gamma(k_{\mathrm{B}}T/m\ell)$.

(ii) Josephson junction (RCSJ model):

$$C\ddot{\varphi} + \frac{1}{R}\dot{\varphi} + \frac{2e}{\hbar} I_{\rm c} \sin\varphi = \frac{2e}{\hbar}\left[I + \xi(t)\right] \ . \tag{7.10}$$

Here $x = \varphi$ (the phase across the junction), $\gamma = 1/RC$, $d = 2eI_{\rm c}/\hbar C$ (essentially the critical current), $F = 2eI/\hbar C$ (external current), and $D = k_{\rm B}T/R$.

(iii) Others: superionic conductors, phase-locked loops, etc.

When $\rho \neq 0$ in Eq. (7.8), V is called a ratchet potential, where it is more difficult to surmount the barrier to the left than to the right (like a saw tooth; see Fig. 8). Ratchet potentials have been used to model directional motion in diverse systems, one of them the molecular motors in the cell.

If Fig. 9 we show the average velocity vs. force for a system of independent particles in a ratchet potential, obtained by numerical integration of the Langevin equation (7.7) with a fourth order algorithm. It is seen that the depinning (transition to a state with non-zero velocity), occurs at lower forces to the right that to the left, as can be expected from the form of the potential. It is also seen that for lower damping, the transition to the running state is quite sharp, and the curve quickly goes to the limit velocity curve $\gamma \langle v \rangle = F$. The reason is that for high damping, if the particle has crossed the barrier, it will not necessarily pass to a running state, but can be trapped in the next well, while the weakly damped particle has more chances to travel, at least a number of wells.

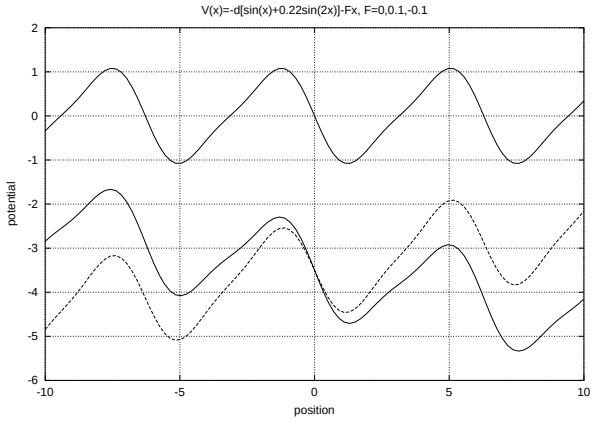

Figure 8: Ratchet potential (7.8), with asymmetry parameter $\rho = 0.44$, at zero force and $F/d = \pm 0.15$ (displaced for clarity).

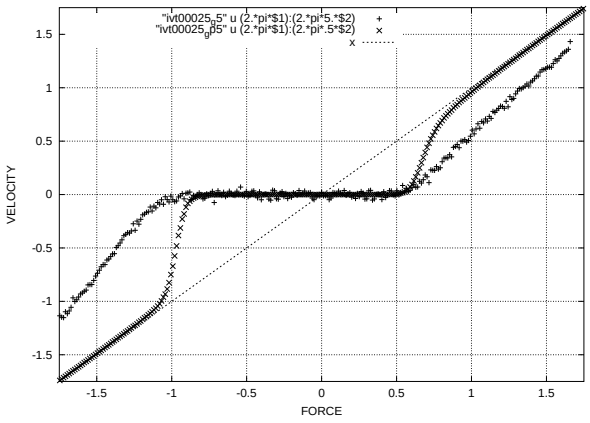

Figure 9: $\gamma \langle v \rangle$ vs. F for a particle in a ratchet potential at $T/d = 0.1$ with $\gamma = 0.5$ and 5. The dotted line shows the limit curve $\gamma \langle v \rangle = F$. Results obtained by numerical integration of the Langevin equation with a Runge–Kutta-like 4th order algorithm for a system of 1000 independent particles.

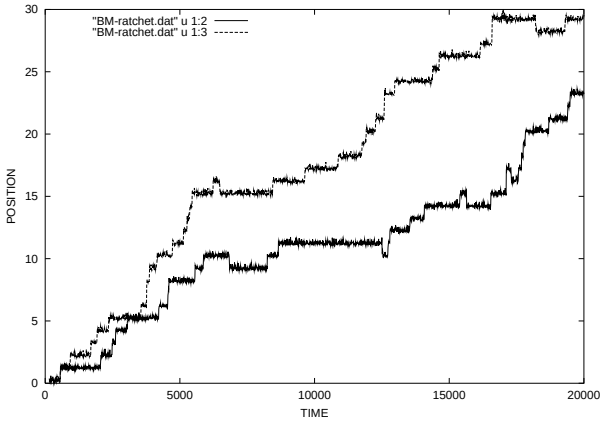

Figure 10: Time evolution of the position of two independent Brownian particles in a ratchet potential corresponding to Fig. 9. The particles are in a weak force $F/d \sim 0.03$ so their random-walk is biased in the force direction. The other parameters are $T/d \sim 0.1$, $\gamma = 0.5$.

The smooth graphs in Fig. 9 are obtained averaging the results for 1000 particles. The individual trajectories of the particles, however, are quite irregular. In Fig. 10, the trajectories of two of them are shown. It is seen that to the overall trend of advancing in the direction of the force, there are

superimposed Brownian fluctuations (biased random walk), and indeed we see that the particles can be trapped in the wells for some time and even to return to the previous well.

7.1.5 Example II: Brownian spins and dipoles.

The Langevin equation for a spin subjected to fluctuations and dissipation is the Landau–Lifshitz equation

$$\frac{\mathrm{d}\vec{s}}{\mathrm{d}t} = \vec{s}\wedge\left[\vec{B}+\vec{\xi}(t)\right] - \lambda\vec{s}\wedge\left(\vec{s}\wedge\vec{B}\right) \ , \quad \langle\xi_k(t)\xi_\ell(s)\rangle = 2D\delta_{k\ell}\delta(t-s) \ , \ (7.11)$$

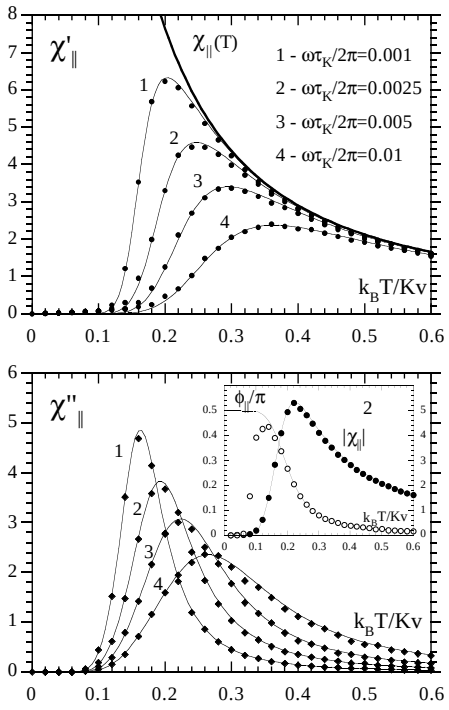

Figure 11: Symbols: $\chi(\omega, T)$ vs. T obtained by numerical integration of the stochastic Landau–Lifshitz equation with the Heun scheme (7.4). Thin lines: simple Debye approximation. Thick line: thermal-equilibrium susceptibility.

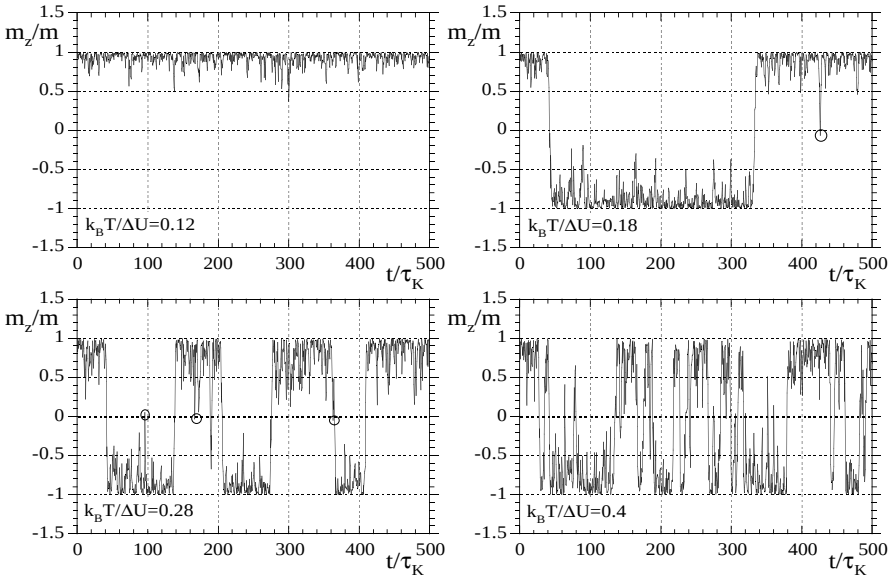

Figure 12: Projection of $\vec{s}(t)$ onto the anisotropy axis for $\mathcal{H} = -\Delta U s_z^2$ in zero field, for various temperatures.

with $D = \lambda k_{\rm B}T/s$???? [cf. $D = \gamma(k_{\rm B}T/m)$ for the particle] and $\vec{B} = -\partial\mathcal{H}/\partial\vec{s}$ (cf. $-\partial V/\partial x$). For the Hamiltonian we choose

$$\mathcal{H}(\vec{s}) = -\Delta U s_z^2 - \vec{s}\cdot\vec{B}_0 \ , \qquad \vec{B} = 2\Delta U s_z \hat{z} + \vec{B}_0 \ , \tag{7.12}$$

with the term of coupling with the external field $\vec{B}_0$ and a *magnetic anisotropy* term (magnetocrystalline, magnetostatic, etc.), which favours orientation of the spin along $\pm\hat{z}$. The anisotropy and the field play the role of the substrate potential and the force for the particle problem.

The stochastic Landau–Lifshitz equation describes a variety of systems

(i) Magnetic systems: magnetic nanoparticles and approximately, magnetic molecular clusters.

(ii) Electric systems (taking formally the limit $\lambda \gg 1$): permanent dipole molecules, nematic liquid crystals, and relaxor ferroelectrics.

Figure 11 displays the results for the dynamical susceptibility vs. the temperature for an ensemble of 1000 spins with parallel anisotropy axes for

$\lambda = 0.1$. At low temperatures, the relaxation time obeys $\tau \gg 2\pi/\omega$, so that the probability of over-barrier rotations is almost zero. The spins rotate close to the bottom of the potential wells (see the panel $k_{\mathrm{B}}T/\Delta U = 0.12$ of Fig. 12) and the response is small. As T is increased the spins can depart from the minima (panel $k_{\mathrm{B}}T/\Delta U = 0.18$) and rotate over the barrier. However, since the jump mechanism is not efficient enough, the response lags behind the probing field, and an appreciable imaginary component of the response arises. If T is further increased, the very thermal agitation, which up to this point was responsible for the growth of the response, reaches a level that simultaneously disturbs the alignment of the magnetic moments in the field direction. The response has then a maximum and starts to decrease. Finally, at still higher temperatures the jumps are so frequent (panel $k_{\mathrm{B}}T/\Delta U = 0.4$) that the spins quickly redistribute according to the conditions set by the instantaneous field and the response tends to the thermal equilibrium one.

7.2 Methods for the Fokker–Planck equation

There are several methods to solve Fokker–Planck equations not amenable for an analytical treatment. Among them we can mention the method of the eigenvalues, which consist in finding the eigenvalues of the Fokker–Planck operator occurring in the dynamical equation $\partial_t P = \mathcal{L}_{\mathrm{FP}} P$ However, the operator $\mathcal{L}_{\mathrm{FP}}$ is not in general Hermitian. If the stationary distribution is known P_0, and $\mathcal{L}_{\mathrm{FP}}$ fulfills the *detailed balance* condition, it can be seen that the transformed operator $\overline{\mathcal{L}}_{\mathrm{FP}} = P_0^{-1/2} \mathcal{L}_{\mathrm{FP}} P_0^{1/2}$ is Hermitian, so the problem can be reduced to an ordinary eigenvalue problem.

$$\overline{\mathcal{L}}_{\mathrm{FP}} p_n = -\lambda_n p_n \ , \tag{7.13}$$

from which we will have $P(y,t) = \sum_n a_n p_n(y) e^{-\lambda_n t}$. However, along with relying on the knowledge of P_0 and the detailed balance condition, the method has the drawback that the eigenvalue problem may be difficult to solve (as happens for the Schrödinger equation). Concerning approximate methods, we can mention the *small noise expansion* (also for Langevin equations) and the method of *adiabatic elimination of fast variables*, which reduces the dimensionality of the problem.

In what follows we shall discuss an alternative method, the *continued fraction method*, which is a special case of the expansion into complete sets approach. The method, although limited in principle to problems with a few

variables, it is *exact* and in general *very efficient*, and, in addition, illustrates more general techniques used in non-equilibrium statistical mechanics. We shall discuss the method with two important examples: the Klein–Kramers equation and the Fokker–Planck equation for the Debye dipole.

7.2.1 Solution of the Klein–Kramers equation by continued–fractions

The Klein–Kramers equation (5.20) can be written in the compact form as

$$\partial_t P = \Big(\underbrace{\mathcal{L}_{\rm rev} + \mathcal{L}_{\rm irr}}_{\mathcal{L}_{\rm FP}} \Big) P \ , \qquad \left\{ \begin{array}{rcl} \mathcal{L}_{\rm rev} & = & -v\,\partial_x + V'\partial_v = \big\{ \mathcal{H}\,, \cdot \big\}_{\rm PB} \\ \mathcal{L}_{\rm irr} & = & \gamma\,\partial_v \left[v + (k_{\rm B}T/m)\partial_v \right] \end{array} \right. \tag{7.14}$$

Scaled quantities. To simplify the notation, we introduce a thermal rescaling of the velocity, time, damping, and potential

$$\bar{v} = v/\sqrt{k_{\rm B}T/m}\ , \quad \bar{t} = t \times \sqrt{k_{\rm B}T/m}\ , \quad \bar{\gamma} = \gamma/\sqrt{k_{\rm B}T/m}\ , \quad U = mV/k_{\rm B}T\ . \tag{7.15}$$

Well, for the v, t, and γ so defined we shall not keep the bars and we shall simply write the Klein–Kramers equation as

$$\boxed{\partial_t P = \left(\mathcal{L}_{\rm rev} + \mathcal{L}_{\rm irr} \right) P \ , \qquad \left\{ \begin{array}{rcl} \mathcal{L}_{\rm rev} & = & -v\,\partial_x + U'\partial_v \\ \mathcal{L}_{\rm irr} & = & \gamma\,\partial_v \left(v + \partial_v \right) \end{array} \right. .} \tag{7.16}$$

Expansion in an orthonormal basis of v. We can expand the probability distribution $P(x,v,t)$ in an orthonormal basis $\psi_n(v)$ (to be specified below) as follows

$$P = W \sum_n c_n(x,t)\psi_n(v)\ , \tag{7.17}$$

where $W = W(x,v)$ is some function we extract for later convenience. Due to the orthonormality of the $\psi_n(v)$, we can write $c_n = \int \mathrm{d}v\, \psi_n P/W$, and hence

OBS: partial derivatives$$\dot{c}_n = \int \mathrm{d}v\, \psi_n W^{-1} \underbrace{\partial_t P}_{\mathcal{L}_{\rm FP}P} = \int \mathrm{d}v\, \psi_n \left(W^{-1} \mathcal{L}_{\rm FP} W \right) \underbrace{P/W}_{\sum_m c_m \psi_m} \ .$$

Then, we can compactly write

$$\boxed{\dot{c}_n = \sum_m \hat{Q}_{n,m} c_m \ , \qquad \hat{Q}_{n,m} = \int \mathrm{d}v \, \psi_n \overline{\mathcal{L}}_{\mathrm{FP}} \psi_m \quad \overline{\mathcal{L}}_{\mathrm{FP}} = W^{-1} \mathcal{L}_{\mathrm{FP}} W \ .} \tag{7.18}$$

Then, the equations for the c_n are like a system of linear equations with the matrix elements of $\overline{\mathcal{L}}_{\mathrm{FP}}$. However, the $\hat{Q}_{n,m}$ still contain operators over x. Besides, it seems that all the c_m, contribute to the equation for $\dot{c}_n$.

As for the function W, by analogy with the transformation $P_0^{-1/2} \mathcal{L}_{\mathrm{FP}} P_0^{1/2}$ to simplify the problem, we try it at least for the v dependent part. Then, since $P_0 \propto \exp(-v^2/2)$ in our units[26] we put $W \propto \exp(-v^2/4)$. For the x dependent part, one sets $W \propto \exp(-\varepsilon U)$, which for $\varepsilon = 1/2$ corresponds to the x-dependent part of $P_0^{1/2}$ (if given by the Boltzmann distribution). Thus, from these considerations we set $W \propto \exp[-(v^2/4 + \varepsilon U)]$.

Calculation of $\overline{\mathcal{L}}_{\mathrm{FP}} = W^{-1} \mathcal{L}_{\mathrm{FP}} W$. Since $\mathcal{L}_{\mathrm{FP}} = \mathcal{L}_{\mathrm{rev}} + \mathcal{L}_{\mathrm{irr}}$, we have to calculate $\overline{\mathcal{L}}_{\mathrm{rev}} = W^{-1} \mathcal{L}_{\mathrm{rev}} W$ and $\overline{\mathcal{L}}_{\mathrm{irr}} = W^{-1} \mathcal{L}_{\mathrm{irr}} W$.

For $\overline{\mathcal{L}}_{\mathrm{irr}}$ we have

$$\gamma^{-1} \overline{\mathcal{L}}_{\mathrm{irr}} f = e^{v^2/4} \partial_v \left(v + \partial_v \right) e^{-v^2/4} f = - \left(-\partial_v^2 + \tfrac{1}{4} v^2 - \tfrac{1}{2} \right) f \ ,$$

which has the structure of (minus) the Hamilton operator of the harmonic oscillator in quantum mechanics. Hence, we introduce creation and annihilation operators b^+ and b, defined as

$$\left\{ \begin{array}{rcl} b & = & \partial_v + \frac{1}{2} v \\ b^+ & = & -\partial_v + \frac{1}{2} v \end{array} \right. \qquad \Longrightarrow \qquad \left[b \, , b^+ \right] = 1 \ . \tag{7.19}$$

Since the $1/2$ in $\overline{\mathcal{L}}_{\mathrm{irr}}$ cancels the "zero-point fluctuations", we finally have

$$\boxed{\overline{\mathcal{L}}_{\mathrm{irr}} = -\gamma b^+ b \ .} \tag{7.20}$$

Because of this, it seems natural to choose the $\psi_n(v)$ as the Hermite functions

$$\psi_n(v) = \frac{1}{\sqrt{n! (2\pi)^{1/2}}} \, e^{-v^2/4} H_n(v) \ . \tag{7.21}$$

[26] Undoing the thermal rescaling (7.15), we have $\exp(-v^2/2) \to \exp(-mv^2/2k_{\mathrm{B}}T)$.

For $\overline{\mathcal{L}}_{\rm rev}$, since $\mathcal{L}_{\rm rev} = -v\,\partial_x + U'\partial_v$, we need $\overline{\partial_v}$ and $\overline{\partial_x}$

$$\begin{aligned}
\overline{\partial_v} f &= e^{v^2/4}\partial_v\left(e^{-v^2/4} f\right)\ , \qquad \Rightarrow \qquad \overline{\partial_v} = -b^+ \\
\overline{\partial_x} f &= e^{\varepsilon U}\partial_x\left(e^{-\varepsilon U} f\right)\ , \qquad \Rightarrow \qquad \overline{\partial_x} = \partial_x - \varepsilon U'\ .
\end{aligned}$$

Then, for $\overline{\mathcal{L}}_{\rm rev}$ we have

$$\overline{\mathcal{L}}_{\rm rev} = -\underbrace{\left(b+b^+\right)}_{v}\underbrace{\left(\partial_x - \varepsilon U'\right)}_{\overline{\partial_x}}\underbrace{-b^+}_{\overline{\partial_v}} U' = -b\left(\partial_x - \varepsilon U'\right) - b^+\left[\partial_x + (1-\varepsilon)U'\right]\ .$$

Calling D_+ to the "coefficient" of b and D_- to that of b^+, we finally have

$$\left\{\begin{array}{rcl} D_+ &=& \partial_x - \varepsilon U' \\ D_- &=& \partial_x + (1-\varepsilon)U' \end{array}\right. \quad \Rightarrow \quad \boxed{\overline{\mathcal{L}}_{\rm rev} = -bD_+ - b^+ D_-\ .} \tag{7.22}$$

Calculation of $\hat{Q}_{n,m}$: the Brinkman hierarchy. Recall that we need $\overline{\mathcal{L}}_{\rm FP} = \overline{\mathcal{L}}_{\rm rev} + \overline{\mathcal{L}}_{\rm irr}$, to obtain its matrix elements $\hat{Q}_{n,m} = \int {\rm d}v\,\psi_n \overline{\mathcal{L}}_{\rm FP}\psi_m$, which enter in the equation of motion for the coefficients $\dot{c}_n = \sum_m \hat{Q}_{n,m}c_m$. This is an easy task now, since we have written $\overline{\mathcal{L}}_{\rm irr}$ and $\overline{\mathcal{L}}_{\rm rev}$ in terms of b and b^+. Then, using the "ladder" properties of b and b^+, that is, $b^+\psi_n = \sqrt{n+1}\psi_{n+1}$ and $b\psi_n = \sqrt{n}\psi_{n-1}$, plus the orthogonality of the ψ_n, we obtain

$$\hat{Q}_{n,m} = \int {\rm d}v\,\psi_n\left(\overline{\mathcal{L}}_{\rm rev} + \overline{\mathcal{L}}_{\rm irr}\right)\psi_m = -\sqrt{n}D_-\delta_{n-1,m} - n\gamma\delta_{n,m} - \sqrt{n+1}D_+\delta_{n+1,m}\ .$$

Therefore, the sum in $\dot{c}_n = \sum_m \hat{Q}_{n,m}c_m$, is trivially done, and one finally gets the so-called Brinkman hierarchy (1956)

$$\boxed{\dot{c}_n = -\left(\sqrt{n}D_-\,c_{n-1} + \gamma n c_n + \sqrt{n+1}D_+\,c_{n+1}\right)\ .} \tag{7.23}$$

We see that, due to (i) the choice of the basis functions to expand P *and* (ii) the extraction of the factor $\exp(-v^2/4)$, only the nearest neighbours of c_n contribute in $\dot{c}_n = \sum_m \hat{Q}_{n,m}c_m$. Writing explicitly the equations, we see

that this fact results in a tri-diagonal structure of the system

$$\boxed{\begin{array}{rcccccccccc}
-\dot{c}_0 & = & 0\gamma c_0 & + & \sqrt{1}\,D_+c_1 & + & 0 & + & \cdots & & \\
-\dot{c}_1 & = & \sqrt{1}\,D_-c_0 & + & 1\gamma c_1 & + & \sqrt{2}\,D_+c_2 & + & 0 & + & \cdots \\
-\dot{c}_2 & = & 0 & + & \sqrt{2}\,D_-c_1 & + & 2\gamma c_2 & + & \sqrt{3}\,D_+c_3 & + & 0 \\
-\dot{c}_3 & = & \cdots & + & 0 & + & \sqrt{3}\,D_-c_2 & + & 3\gamma c_3 & + & \sqrt{4}\,D_+c_4 \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots
\end{array}}$$

This equation is completely equivalent to the Klein–Kramers equation (7.16) and valid for any potential.

Continued fractions. Why are we so happy for having transformed the Klein–Kramers equation into an infinity hierarchy for the c_n? First, by taking the Laplace transform $\tilde{f}(s) \equiv \int_0^\infty \mathrm{d}t\, e^{-st} f(t)$ a differential-recurrence relation of the general form

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} + Q_i^- c_{i-1} + Q_i c_i + Q_i^+ c_{i+1} = f_i \;, \tag{7.24}$$

can be reduced to (we omit the tildes on the Laplace transforms)

$$\boxed{Q_i^- c_{i-1} + \hat{Q}_i c_i + Q_i^+ c_{i+1} = \hat{f}_i \;,} \tag{7.25}$$

where $\hat{Q}_i = Q_i + s$ and $\hat{f}_i = f_i + c_i(0)$. Then, this relation can be solved by introducing the ansatz $c_i = S_i c_{i-1} + a_i$, obtaining

$$\boxed{c_i = S_i c_{i-1} + a_i \quad \text{with} \quad S_i = -\frac{Q_i^-}{\hat{Q}_i + Q_i^+ S_{i+1}} \quad a_i = -\frac{Q_i^+ a_{i+1} - \hat{f}_i}{\hat{Q}_i + Q_i^+ S_{i+1}}} \tag{7.26}$$

It is to be remarked that the quantities involved in the relation (7.25) do not need to be scalars, but they can be vectors (c_i and f_i) and the coefficients Q_i matrices. The only change in the solution (7.26) is that the fraction bar then means matrix inversion.

The reason for the name "continued fraction", is that, if we consider for instance S_i, it is given in terms of S_{i+1} in the denominator. But this can be expressed in terms of S_{i+2}, and so on, leading to an expression of the form

$$K = \cfrac{p_1}{q_1 + \cfrac{p_2}{q_2 + \cfrac{p_3}{q_3 + \cdots}}} , \tag{7.27}$$

which is called a continued fraction. Well, if the c_i are judiciously chosen to decrease with increasing i, we can truncate at some large N, setting $c_N = 0$. This leads all the quantities to vanish at $i = N$, and we can iterate downward the continued fractions in Eq. (7.26) down to $i = 0$, storing the successive S_i and a_i. Then, starting from c_0 (usually known by some means, e.g., normalisation of the distribution), we iterate upwards with $c_i = S_i c_{i-1} + a_i$, obtaining the solution to the recurrence-relation (7.25). To ensure convergence, one can repeat the calculation with a truncation index $2N$, $4N$, etc. and check that the results do not change. Note that the steps described above are very easy to code in a computer program (even in a programmable pocket calculator). This is the answer to the question of why it was important to transform the Klein–Kramers equation into the Brinkman hierarchy.

Solving the Brinkman hierarchy. As written, the hierarchy (7.23), involves coefficients that still contain the operators $D_\pm$ over the x variables, instead of true coefficients. But we only need to find a representation of those operators in a basis of functions of x, say $u_p(x)$. Then we just calculate the matrix elements of any operator $\hat{A}$ in that basis $A_{p\,q} = \int \mathrm{d}x\, u_p^* \hat{A} u_q$, while the expansion coefficients of $c_n(x,t)$ are expressed as a column vector

$$c_n(x,t) = \sum_p c_n^p(t) u_p(x) , \qquad \Rightarrow \qquad \mathbf{C}_n = \begin{pmatrix} c_n^{-P} \\ \vdots \\ c_n^P \end{pmatrix} .$$

Then, the Brinkman hierarchy (7.23) is reduced to the following differential recurrence relation

$$\boxed{\dot{\mathbf{C}}_n = \mathbf{Q}_n^- \mathbf{C}_{n-1} + \mathbf{Q}_n \mathbf{C}_n + \mathbf{Q}_n^+ \mathbf{C}_{n+1} ,} \tag{7.28}$$

whose coefficients are not any more operators but ordinary matrices, with elements

$$\begin{aligned}(\mathbf{Q}_n^-)_{p\,q} &= -\sqrt{n}\left[(\partial_x)_{p\,q} + (1-\varepsilon)U'_{p\,q}\right] \\ (\mathbf{Q}_n)_{p\,q} &= -n\,\gamma\,\delta_{p\,q} \\ (\mathbf{Q}_n^+)_{p\,q} &= -\sqrt{n+1}\left[(\partial_x)_{p\,q} - \varepsilon U'_{p\,q}\right]\end{aligned} \tag{7.29}$$

The choice of the basis functions $u_p(x)$ is naturally dictated by the symmetry of the potential.

In this form, we can directly use the method of continued fractions to solve the Brinkman hierarchy. In principle, with the c_n^p obtained we can construct the probability distribution and compute any observable. Nevertheless, this is not even required, since common observables are directly obtained from the expansion coefficients. For instance, if $\varepsilon = 0$ in $W \propto \exp(-v^2/4)$, this quantity is after normalisation $W = \psi_0$ (see the definition of the Hermite functions (7.21)] so that for the averaged velocity $\langle v\rangle = \int \mathrm{d}x \int \mathrm{d}v\, vP(x,v,t)$ we get

$$\langle v\rangle = \sum_{n\,p} c_n^p \int \mathrm{d}x\, u_p(x) \underbrace{\int \mathrm{d}v \overbrace{\underbrace{v}_{b+b^+}\psi_0(v)}^{\psi_1}\psi_n(v)}_{\delta_{n,1}} = \sum_p c_1^p \int \mathrm{d}x\, u_p(x) = c_1^0\,,$$

where he have just assumed that the 0th element of the basis $u_p(x)$ is constant (say $u_0 = 1$), so that the result follows from orthogonality ($u_0 \perp u_n$).

Example: particles in a periodic potential. Let us consider the case of a general periodic potential $U' = \sum_q U'_q e^{\mathrm{i}qx}$, where the U'_q are the Fourier components of the potential derivative. In this case the natural choice of the basis functions $u_p(x)$ is that of plane waves $u_p(x) = e^{\mathrm{i}qx}/\sqrt{2\pi}$. Then the matrix elements needed to construct the matrices $\mathbf{Q}_n$ in Eq. (7.29) are simply

$$(\partial_x)_{p\,q} = \mathrm{i}p\delta_{p\,q}\,, \qquad (U')_{p\,q} = U'_{p-q}\,. \tag{7.30}$$

Then, the number of diagonals occupied in the matrices $\mathbf{Q}_n$ below and above the main diagonal is equal to the number of harmonics in the potential. In the example of the cosine potential, this number is only one, while in the ratchet potential (7.8), this number is two.

Well, in the example of the ratchet potential we computed the average velocity vs. force by Langevin simulation (Fig. 9). If we compute now the

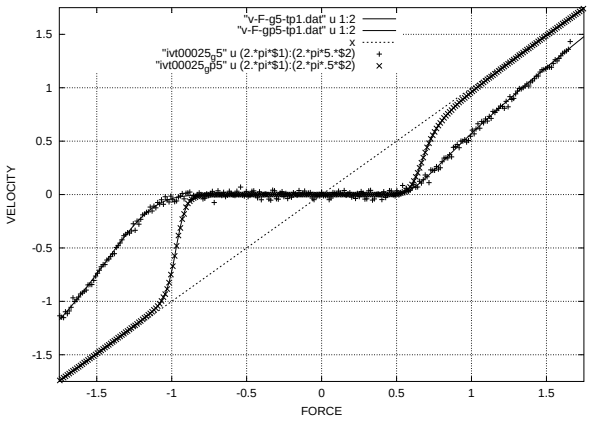

Figure 13: Depinning curves in a ratchet potential of Fig. 9, obtained with Langevin simulation (symbols), together with the results obtained with the continued-fraction method (thick lines).

same curves with the continued-fraction method, we get a complete agreement shown in Fig. 13 (without any free parameter).

As an additional example, in Fig. 14, we show the real part of the dynamical susceptibility as a function of the frequency in an ordinary (non-ratchet) periodic potential (thus, the curves could correspond to a Josephson junction, or to the ac conductivity in a transport problem). We see a peak about the frequency of oscillations in the bottom of the potential wells (which defines the frequency scale in the graph). The peak becomes more and more sharp and high the lower the damping is, showing the typical features of a resonant system.

When the continued-fraction method can be used its advantages are: (i) it has no statistical error bars, (ii) it is in general extremely efficient, (iii) it outputs the complete distribution if required, and (iv) it may be extended to problems of quantum Brownian motion. The drawbacks are that it is quite specific of the problem to be solved (we need to compute the matrix elements for each potential, and the relation between the coefficients and the observables), and in this respect the Langevin simulations do not have this problem. Besides, the simulations output trajectories, which are lost in the continued fraction approach. Finally, it can only be implemented for systems with a few variables (e.g., independent particles), while the Langevin simulations do not suffer this limitation.

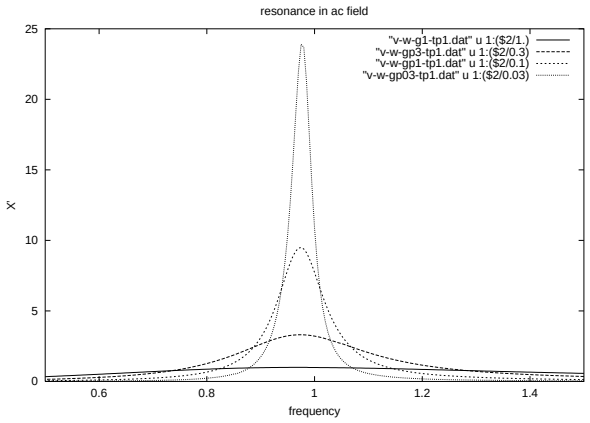

Figure 14: Real part of the dynamical susceptibility (ac conductivity) vs. frequency. The curves have been obtained with the continued-fraction method for $T/d = 0.1$ and various values of the damping $\gamma = 1$, 0.3, 0.1, 0.03. The resonance at $\omega \sim 1$ is clearly seen.

7.2.2 Solution of the Debye equation by continued–fractions

This continued-fraction method can also be applied to rotational problems. We shall illustrate this with the Fokker–Planck equation for a dipole (4.30). First, using the definition of the Debye relaxation time $\tau_{\rm D} = \zeta/2k_{\rm B}T$ [Eq. (5.35)], the abbreviation $\alpha = pE/k_{\rm B}T$ and changing to Cartesian coordinates $z = \cos\vartheta$, the Debye equation can be written as

$$\boxed{2\tau_{\rm D}\frac{\partial P}{\partial t} = \frac{\partial}{\partial z}\left[(1-z^2)\left(\frac{\partial P}{\partial z} - \alpha\, P\right)\right]\ .} \tag{7.31}$$

In this problem the natural choice for basis function is the Legendre polynomials $p_n(\cos\vartheta)$ [cf. Eq. (7.17)]

$$P = \sum_n c_n(t) p_n(z)\ . \tag{7.32}$$

Due to the orthogonality of the $p_n(z)$, $\int_{-1}^{1} {\rm d}z\, p_n(z)p_m(z) = 2\delta_{n,m}/(2n+1)$, we have $c_n = [(2n+1)/2]\int {\rm d}z\, p_n(z)\, P(z)$. Then the equation of motion for

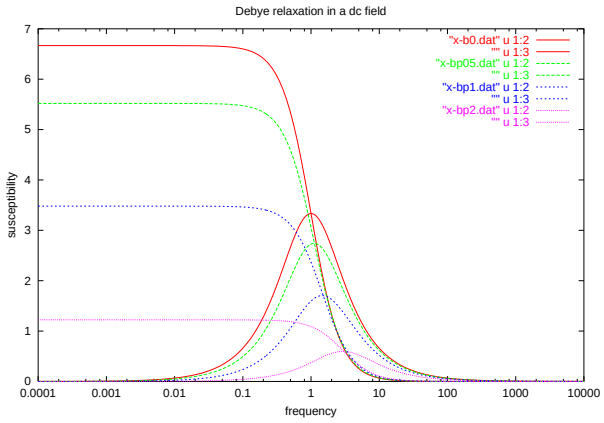

Figure 15: Dynamical susceptibility (polarisability) of a dipole in various static fields. The curves have been obtained with the continued-fraction method for $T = 0.1$ and the fields used are $pE/k_{\rm B}T = 0$, 1, 2 and 4.

the c_n reads [cf. Eq. (7.18)]

$$\frac{4\tau_{\rm D}}{2n+1}\dot{c}_n = \int {\rm d}z\, p_n \overbrace{\mathcal{L}_{\rm FP}\underbrace{P}_{\sum_m c_m p_m}}^{\partial_t P} = \sum_m \hat{Q}_{n,m} c_m \ , \quad \text{with} \quad Q_{n,m} = \int {\rm d}z\, p_n \mathcal{L}_{\rm FP} p_m \ .$$

To calculate $Q_{n,m}$ (which is not an operator here), we use relations obeyed by the Legendre polynomials[27] and an integration by parts:

$$\begin{aligned} Q_{n,m} &= \int {\rm d}z\, p_n \frac{\rm d}{{\rm d}z}\left[(1-z^2)\left(\frac{{\rm d}p_m}{{\rm d}z} - \alpha\, p_m\right)\right] \\ &= -m(m+1)\int {\rm d}z\, p_n p_m + \alpha \int {\rm d}z\, \frac{{\rm d}p_n}{{\rm d}z}(1-z^2)p_m \\ &= -n(n+1)\frac{2\delta_{nm}}{2n+1} + \alpha\, \frac{n(n+1)}{2n+1}\left(\frac{2\delta_{n-1,m}}{2n-1} - \frac{2\delta_{n+1,m}}{2n+3}\right) \ . \end{aligned}$$

[27] Specifically, the Legendre equation, and a sort of first integral:

$$\frac{\rm d}{{\rm d}z}\left[(1-z^2)\frac{{\rm d}p_n}{{\rm d}z}\right] + n(n+1)p_n = 0 \ , \qquad (1-z^2)\frac{{\rm d}p_n}{{\rm d}z} = \frac{n(n+1)}{2n+1}\,(p_{n-1} - p_{n+1}) \ .$$

Inserting this result into the equation for $\dot{c}_n$, we finally find

$$\boxed{\frac{2\tau_{\rm D}}{n(n+1)}\dot{c}_n = -c_n + \frac{pE}{k_{\rm B}T}\left(\frac{c_{n-1}}{2n-1} - \frac{c_{n+1}}{2n+3}\right) .} \tag{7.33}$$

which is a tridiagonal differential-recurrence relation for the dipole analogous to the Brinkman hierarchy (7.23).

Solving Eq. (7.33) by continued fractions, we can obtain any observable. For instance, we can compute the linear susceptibility in a static field (remember that the solved this problem in the absence of bias field), by setting $E = E_0 + \Delta E \cos(\omega t)$. The results are shown in Fig. 15, for various E_0. It is seen that the equilibrium susceptibility (real part at $\omega \to 0$) decreases with E_0, since the susceptibility measures the slope of the static polarization, which saturates at large fields. Besides the relaxation time (roughly, the inverse of the location of the maximum of the imaginary part) decreases with E_0, so that the relaxation is faster in a static field.

8 Derivation of Langevin equations in the bath-of-oscillators formalism

Now we shall discuss the foundations of the Langevin equations by showing how they can be "derived" from more or less microscopic description of a system coupled with its environment.

8.1 Dynamical approaches to the Langevin equations

The Langevin equations we have discussed for Brownian particles and spins, are phenomenological inasmuch as they constructed by augmenting known phenomenological equations by fluctuating terms. For subsequent reference, let us first rewrite these equations in the notation we shall employ later:
• *Brownian particle*

$$\left\{ \begin{array}{rcl} \dfrac{\mathrm{d}q}{\mathrm{d}t} & = & \dfrac{p}{m} \\ \dfrac{1}{m}\dfrac{\mathrm{d}p}{\mathrm{d}t} & = & -\dfrac{\partial V}{\partial q} - \gamma \dfrac{\mathrm{d}q}{\mathrm{d}t} + \xi(t) \ , \qquad \langle \xi(t)\xi(t')\rangle = 2(\gamma k_{\mathrm{B}}T/m)\delta(t-t') \ . \end{array} \right. \tag{8.1}$$

Note that the fluctuating terms only enter in the equation for the momentum.
• *Brownian spins and dipoles*

$$\frac{\mathrm{d}\vec{s}}{\mathrm{d}t} = \vec{s} \wedge \left[\vec{B} + \vec{\xi}(t) \right] - \lambda \, \vec{s} \wedge \left(\vec{s} \wedge \vec{B} \right) \ , \quad \langle \xi_k(t)\xi_\ell(t')\rangle = 2(\lambda k_{\mathrm{B}}T)\delta_{k\ell}\delta(t-t') \ . \tag{8.2}$$

This *stochastic Landau–Lifshitz equation* is equivalent to the Gilbert equation where the damping term is $\propto -\lambda\vec{s} \wedge (\mathrm{d}\vec{s}/\mathrm{d}t)$. Recall that Eq. (8.2) contains as a special case the stochastic dynamical equation for the electrical dipole.

Note that in both equations the fluctuating and dissipation terms are not independent: $D = \gamma k_{\mathrm{B}}T/m$ for the particle, which corresponds to $D = \lambda k_{\mathrm{B}}T$ for the spin. Besides, the force $F = -\partial V/\partial q$, corresponds to $\vec{B} = -\partial\mathcal{H}/\partial\vec{s}$, while the damping $-\gamma(\mathrm{d}q/\mathrm{d}t)$, corresponds to $-\lambda\vec{s}\wedge\left(\vec{s}\wedge\vec{B}\right)$, which as mentioned above is related with $\propto -\lambda\vec{s} \wedge (\mathrm{d}\vec{s}/\mathrm{d}t)$.

There have been several attempts to justify, starting from dynamical descriptions of a system coupled to its surroundings, these important Langevin equations. The effort was first directed to the Langevin equations for the

Brownian particle (translational problems) and then to rotational Brownian motion of spins and dipoles.

In most of those studies, the environment is represented as a set of independent harmonic oscillators. The oscillators are somehow "projected out" and an equation for the system variables is derived. The final equation has the form of a generalized Langevin equation (i.e., containing "memory" terms), whose fluctuating and dissipative terms naturally obey fluctuation-dissipation relations. For instance, for the particle problem one gets

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial \mathcal{H}}{\partial q} + f(t) - \int_{t_0}^{t} \mathrm{d}t' \, \mathcal{K}(t-t') \frac{\mathrm{d}q}{\mathrm{d}t}(t') \; , \qquad \langle f(t) f(t') \rangle = k_{\mathrm{B}} T \, \mathcal{K}(t-t') \; . \tag{8.3}$$

Thus the relaxation (damping) term, involves a memory integral taken along the past history of the system. Besides, the memory kernel $\mathcal{K}(t-t')$ is determined by the correlation properties of the fluctuating force $f(t)$. Thus, if the autocorrelation of the fluctuating terms is very short $\langle f(t) f(t') \rangle \propto \delta(t-t')$, the damping term reduces to minus the velocity $-(\mathrm{d}q/\mathrm{d}t)$ of the particle. Similar results are obtained for the spin and dipole problems.

In what follows we shall discuss the bath of oscillators formalism. First, since we shall use a Hamiltonian formalism throughout, we shall start with a brief review of the main results from Hamiltonian mechanics that we shall need. Subsequently, we shall introduce the model for the system coupled with its environment, deduce the corresponding dynamical equations, and finally discuss some examples.

8.2 Quick review of Hamiltonian dynamics

The dynamical equations for a system with Hamiltonian $\mathcal{H}(p,q)$ are

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial p} \; , \qquad \frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial \mathcal{H}}{\partial q} \; . \tag{8.4}$$

Then, the time evolution of an arbitrary dynamical variable of the system $A(p,q)$ (assumed not explicitly time dependent), is $\mathrm{d}A/\mathrm{d}t = (\partial A/\partial q)(\mathrm{d}q/\mathrm{d}t) + (\partial A/\partial p)(\mathrm{d}p/\mathrm{d}t)$. Then, using for $\mathrm{d}q/\mathrm{d}t$ and $\mathrm{d}p/\mathrm{d}t$ the Hamilton equations and introducing the *Poisson bracket* of two arbitrary dynamical variables

$$\big\{ A, B \big\} \equiv \frac{\partial A}{\partial q} \frac{\partial B}{\partial p} - \frac{\partial A}{\partial p} \frac{\partial B}{\partial q} \; , \tag{8.5}$$

we have for $\mathrm{d}A/\mathrm{d}t$ the basic Hamiltonian evolution equation

$$\boxed{\frac{\mathrm{d}A}{\mathrm{d}t} = \big\{A, \mathcal{H}\big\} \ .} \tag{8.6}$$

Finally, by using this equation for $A = q$, with $\partial q/\partial q = 1$, $\partial q/\partial p = 0$, and for $A = p$ with $\partial p/\partial q = 0$, and $\partial p/\partial p = 1$, we see that the Hamilton equations (8.4) are a particular case of Eq. (8.6). For a system with variables (p_a, q_a) $a = 1, \ldots, N$, the above results are the same, with the only change of introducing a sum over a in the definition of the Poisson bracket.

Two more results we shall need are the *product rule* of the Poisson bracket

$$\big\{A, BC\big\} = \big\{A, B\big\}C + B\big\{A, C\big\} \ , \tag{8.7}$$

and the *chain rule*

$$\big\{f, g\big\} = \sum_{i,k} \frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_k}\big\{x_i, x_k\big\} \ , \qquad x_i = x_i(p, q) \ , \tag{8.8}$$

which immediately follow from the ordinary differentiation rules.

Spin dynamics case. The equations for an isolated classical spin (not subjected to fluctuations and dissipation)

$$\frac{\mathrm{d}\vec{s}}{\mathrm{d}t} = \vec{s} \wedge \vec{B} \ , \qquad \vec{B} = -\frac{\partial \mathcal{H}}{\partial \vec{s}} \ . \tag{8.9}$$

can also be written in Hamiltonian form. To this end, let us write the formula for the gradient operator in spherical coordinates

$$\frac{\partial u}{\partial \vec{s}} = \hat{s}\frac{\partial u}{\partial s} + \hat{\vartheta}\frac{1}{s}\frac{\partial u}{\partial \vartheta} + \hat{\varphi}\frac{1}{s\sin\vartheta}\frac{\partial u}{\partial \varphi} \ , \tag{8.10}$$

where φ and ϑ are the azimuthal and polar angles of $\vec{s}$. Since the length of $\vec{s}$ is constant, the set vectorial equations (8.9), can be written as

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = -\frac{1}{s\sin\vartheta}\frac{\partial \mathcal{H}}{\partial \vartheta} \ , \qquad \frac{\mathrm{d}\vartheta}{\mathrm{d}t} = \frac{1}{s\sin\vartheta}\frac{\partial \mathcal{H}}{\partial \varphi} \ , \tag{8.11}$$

which correspond to the Hamilton equations (8.4) with the conjugate canonical variables

$$\boxed{q = \varphi \ , \qquad p = s_z \ .} \tag{8.12}$$

In terms of the variables (8.12) the Cartesian components of the spin are given by

$$s_x = \sqrt{s^2 - p^2}\cos q \ , \qquad s_y = \sqrt{s^2 - p^2}\sin q \ , \qquad s_z = p \ .$$

From these $s_i(p,q)$ and the definition of the Poisson bracket of two arbitrary dynamical variables [Eq. (8.5)], one can readily obtain the customary Poisson-bracket ("commutation") relations among the spin variables

$$\big\{s_i, s_j\big\} = \sum_k \epsilon_{ijk} s_k \ , \tag{8.13}$$

where ϵ_{ijk} is the Levi–Civita symbol.[28] In addition, on using the chain rule of the Poisson bracket [Eq. (8.8)], one gets the useful relation

$$\big\{s_i, W(\vec{s})\big\} = -\left(\vec{s} \wedge \frac{\partial W}{\partial \vec{s}}\right)_i \ , \tag{8.14}$$

which is valid for any function of the spin variables $W(\vec{s})$.[29]

8.3 Dynamical equations in the bath-of-oscillators formalism

We shall now study a classical system surrounded by an environment that can be represented by a set of independent classical harmonic oscillators. In spite of its academic appearance, those oscillators could correspond to the *normal modes* of an electromagnetic field, the lattice vibrations of a crystal (in the harmonic approximation), or they can be an effective low-energy description of a more general surrounding medium (Caldeira and Leggett, [1]).

[28] To illustrate, from

$$\begin{aligned} \partial s_x/\partial q &= -\big[s^2 - p^2\big]^{1/2}\sin q \ , & \partial s_x/\partial p &= -p\,\big[s^2 - p^2\big]^{-1/2}\cos q \ , \\ \partial s_y/\partial q &= \big[s^2 - p^2\big]^{1/2}\cos q \ , & \partial s_y/\partial p &= -p\,\big[s^2 - p^2\big]^{-1/2}\sin q \ , \end{aligned}$$

one gets $\big\{s_x, s_y\big\} = p\sin^2 q + p\cos^2 q = s_z$. Q.E.D.

[29] Note that one can conversely *postulate* the relations $\{s_i, s_j\} = \sum_k \epsilon_{ijk} s_k$, and then *derive* Eq. (8.9) starting from the basic Hamiltonian evolution equation $\mathrm{d}s_i/\mathrm{d}t = \{s_i, \mathcal{H}\}$ and using Eq. (8.14). This can be considered as a justification of the presence of the expression $\vec{B} = -\partial\mathcal{H}/\partial\vec{s}$ in the dynamical equations for a classical spin.

8.3.1 The system-environment Hamiltonian

The total system consisting of the "system of interest" plus the oscillators representing the environment forms a *closed* dynamical system that we shall describe by augmenting the isolated-system Hamiltonian as follows

$$\mathcal{H}_{\rm T} = \mathcal{H}(p,q) + \sum_\alpha \tfrac{1}{2}\Big\{ P_\alpha^2 + \omega_\alpha^2 \Big[Q_\alpha + \frac{\varepsilon}{\omega_\alpha^2} F_\alpha(p,q) \Big]^2 \Big\} \,. \tag{8.15}$$

Here, α is an oscillator index [e.g., the pair $(\vec{k}, s)$ formed by the wave-vector and branch index of a normal mode of the environment], and the coupling terms $F_\alpha(p,q)$ are arbitrary functions of the system variables. These terms may depend on the parameters of the oscillators ω_α, but not on their dynamical variables P_α, Q_α. On the other hand, we have introduced a system-environment coupling constant ε for the sake of convenience in keeping track of the orders of the various contributions.

The terms proportional to F_α^2, which emerge when squaring $Q_\alpha + (\varepsilon/\omega_\alpha^2)F_\alpha$, are "counter-terms" introduced to balance the coupling-induced renormalization of the Hamiltonian of the system. The formalism takes as previously considered whether such a renormalization actually occurs for a given interaction, so that $\mathcal{H}$ would already include it (whenever exists). An advantage of this convention is that one deals with the experimentally accessible energy of the system, instead of the "bare" one, which might be difficult to determine.

8.3.2 Dynamical equations

Let us first cast the Hamiltonian (8.15) into the form

$$\mathcal{H}_{\rm T} = \mathcal{H}^{(\rm m)}(p,q) + \sum_\alpha \tfrac{1}{2}\left(P_\alpha^2 + \omega_\alpha^2 Q_\alpha^2 \right) + \varepsilon \sum_\alpha Q_\alpha F_\alpha(p,q) \,, \tag{8.16}$$

where q and p are the canonical coordinate and conjugate momentum of a system with Hamiltonian $\mathcal{H}(p,q)$ and the "modified" system Hamiltonian $\mathcal{H}^{(\rm m)}$ augments $\mathcal{H}$ by the aforementioned counter-terms

$$\mathcal{H}^{(\rm m)} = \mathcal{H} + \frac{\varepsilon^2}{2} \sum_\alpha \frac{F_\alpha^2}{\omega_\alpha^2} \,. \tag{8.17}$$

Besides, in the above expression for $\mathcal{H}_{\rm T}$ the Hamiltonian of the oscillators is clearly recognised $\mathcal{H}_{\rm E} = \sum_\alpha \frac{1}{2}\left(P_\alpha^2 + \omega_\alpha^2 Q_\alpha^2\right)$, and the same for the coupling term $\mathcal{H}_{\rm int} = \varepsilon \sum_\alpha Q_\alpha F_\alpha(p,q)$.

The equation of motion for any dynamical variable C without explicit dependence on the time, $\partial C/\partial t \equiv 0$, is given by the basic Hamiltonian evolution equation (8.6) with $\mathcal{H} \to \mathcal{H}_{\rm T}$, with the whole Poisson bracket is given by

$$\big\{A, B\big\} \equiv \frac{\partial A}{\partial q}\frac{\partial B}{\partial p} - \frac{\partial A}{\partial p}\frac{\partial B}{\partial q} + \sum_\alpha \frac{\partial A}{\partial Q_\alpha}\frac{\partial B}{\partial P_\alpha} - \frac{\partial A}{\partial P_\alpha}\frac{\partial B}{\partial Q_\alpha} \, .$$

Therefore, the (coupled) equations of motion for *any* dynamical variable of the system $A(p,q)$ and the environment variables read ($C = A, P_\alpha$, and Q_α)

$$\frac{{\rm d}A}{{\rm d}t} = \big\{A, \mathcal{H}^{(\rm m)}\big\} + \varepsilon \sum_\alpha Q_\alpha \big\{A, F_\alpha\big\} \, , \tag{8.18}$$

$$\frac{{\rm d}Q_\alpha}{{\rm d}t} = P_\alpha \, , \qquad \frac{{\rm d}P_\alpha}{{\rm d}t} = -\omega_\alpha^2 Q_\alpha - \varepsilon F_\alpha \, , \tag{8.19}$$

where we have used $\big\{Q_\alpha, P_\alpha\big\} = 1$. The goal is to derive a dynamical equation for $A(p,q)$ involving the system variables only (*reduced* dynamical equation).

On considering that in Eqs. (8.19) the term $-\varepsilon F_\alpha(t) = -\varepsilon F_\alpha[p(t), q(t)]$ plays the rôle of a time-dependent forcing on the oscillators, those equations can be explicitly integrated, yielding

$$Q_\alpha(t) = Q_\alpha^{\rm h}(t) - \frac{\varepsilon}{\omega_\alpha}\int_{t_0}^{t} {\rm d}t' \, \sin[\omega_\alpha(t-t')] F_\alpha(t') \, , \tag{8.20}$$

where

$$Q_\alpha^{\rm h}(t) = Q_\alpha(t_0)\cos[\omega_\alpha(t-t_0)] + [P_\alpha(t_0)/\omega_\alpha]\sin[\omega_\alpha(t-t_0)] \, , \tag{8.21}$$

are the solutions of the *homogeneous* system of equations for the oscillators in the absence of the system-environment interaction (proper modes of the environment). Then, on integrating by parts in Eq. (8.20) one gets for the combination εQ_α that appears in Eq. (8.18)

$$\varepsilon Q_\alpha(t) = f_\alpha(t) - [\mathcal{K}_\alpha(t-t')F_\alpha(t')]_{t'=t_0}^{t'=t} + \int_{t_0}^{t} {\rm d}t' \, \mathcal{K}_\alpha(t-t')\frac{{\rm d}F_\alpha}{{\rm d}t}(t') \, , \tag{8.22}$$

where

$$f_\alpha(t) = \varepsilon Q_\alpha^{\rm h}(t) \, , \qquad \mathcal{K}_\alpha(\tau) = \frac{\varepsilon^2}{\omega_\alpha^2}\cos(\omega_\alpha \tau) \, . \tag{8.23}$$

Next, in order to eliminate the environment variables from the equation for $A(p,q)$, one substitutes Eq. (8.22) back into Eq. (8.18), getting

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \big\{A, \mathcal{H}^{(\mathrm{m})}\big\} \;-\; \sum_\alpha \big\{A, F_\alpha\big\}\mathcal{K}_\alpha(0)F_\alpha(t) + \sum_\alpha \big\{A, F_\alpha\big\}\mathcal{K}_\alpha(t-t_0)F_\alpha(t_0) \\ +\; \sum_\alpha \big\{A, F_\alpha\big\}\left[f_\alpha(t) + \int_{t_0}^t \mathrm{d}t'\,\mathcal{K}_\alpha(t-t')\frac{\mathrm{d}F_\alpha}{\mathrm{d}t}(t')\right].$$

The term $\sum_\alpha \big\{A, F_\alpha\big\}\mathcal{K}_\alpha(t-t_0)F_\alpha(t_0)$ depends on the initial state of the system $(p(t_0), q(t_0))$ and produces a transient response that can be ignored in the long-time dynamics (we shall return to this question below). The parallel term $-\sum_\alpha \big\{A, F_\alpha\big\}\mathcal{K}_\alpha(0)F_\alpha(t)$ is derivable from a Hamiltonian and balances exactly the term due to the counter-terms in $\big\{A, \mathcal{H}^{(\mathrm{m})}\big\}$. This can be shown by using

$$-\sum_\alpha \big\{A, F_\alpha\big\}\mathcal{K}_\alpha(0)F_\alpha = \big\{A, -\tfrac{1}{2}\sum_\alpha \mathcal{K}_\alpha(0)F_\alpha^2\big\}\;,$$

which follows from the product rule (8.7) of the Poisson bracket and then using $\mathcal{K}_\alpha(0) = \varepsilon^2/\omega_\alpha^2$ [see Eq. (8.23)]. Therefore, one is left with the *reduced* dynamical equation

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \big\{A, \mathcal{H}\big\} + \sum_\alpha \big\{A, F_\alpha\big\}\left[f_\alpha(t) + \int_{t_0}^t \mathrm{d}t'\,\mathcal{K}_\alpha(t-t')\frac{\mathrm{d}F_\alpha}{\mathrm{d}t}(t')\right], \qquad (8.24)$$

where the first term yields the free (conservative) time evolution of the system, whereas the second term incorporates the effects of the interaction of the system with its environment.

To conclude, let us decompose the coupling functions as

$$F_\alpha(p,q) = \sum_{\mathrm{a}} c_\alpha^{\mathrm{a}} W_{\mathrm{a}}(p,q)\;. \qquad (8.25)$$

Here "a" stands for a general index depending on the type of interaction. The idea is to split the part of the coupling $W_{\mathrm{a}}(p,q)$ which is common to all the modes, so that F_α is obtained multiplying that part by certain mode-dependent system-environment coupling constants c_α^{a}. For instance, if α is a mode of wave-vector $\vec{k}$, and $F_{\vec{k}} = \vec{k}\cdot\vec{r}$, then $\mathrm{a} = i$ (the Cartesian index)

with $c^i_{\vec{k}} = k_i$ and $W_i = r_i$. Introducing the above coupling function into Eq. (8.24), we have

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \big\{A, \mathcal{H}\big\} + \sum_\alpha \big\{A, \sum_{\mathrm{a}} c^{\mathrm{a}}_\alpha W_{\mathrm{a}}\big\} \left[f_\alpha(t) + \int_{t_0}^t \mathrm{d}t' \, \mathcal{K}_\alpha(t-t') \sum_{\mathrm{b}} c^{\mathrm{b}}_\alpha \frac{\mathrm{d}W_{\mathrm{b}}}{\mathrm{d}t}(t') \right] .$$

Therefore, we finally have

$$\boxed{\frac{\mathrm{d}A}{\mathrm{d}t} = \big\{A, \mathcal{H}\big\} + \sum_{\mathrm{a}} \big\{A, W_{\mathrm{a}}\big\} \left[f_{\mathrm{a}}(t) + \int_{t_0}^t \mathrm{d}t' \sum_{\mathrm{b}} \mathcal{K}_{\mathrm{ab}}(t-t') \frac{\mathrm{d}W_{\mathrm{b}}}{\mathrm{d}t}(t') \right]} \tag{8.26}$$

where

$$\boxed{f_{\mathrm{a}}(t) = \sum_\alpha c^{\mathrm{a}}_\alpha f_\alpha(t) \; , \qquad \mathcal{K}_{\mathrm{ab}}(\tau) = \sum_\alpha c^{\mathrm{a}}_\alpha c^{\mathrm{b}}_\alpha \mathcal{K}_\alpha(\tau) \; .} \tag{8.27}$$

The terms $f_{\mathrm{a}}(t)$ are customarily interpreted as *fluctuating* "forces" (or "fields"). Indeed $f_{\mathrm{a}}(t)$ is a sum of a large number of sinusoidal terms with different frequencies and phases; this can give to $f_{\mathrm{a}}(t)$ the form of a highly irregular function of t that is expected for a fluctuating term (see below).[30] The integral term keeps in general memory of the previous history of the system, and provides the *relaxation* due to the interaction with the surrounding medium.

The origin of both types of terms can be traced back as follows. Recall that in Eq. (8.20) the time evolution of the oscillators has formally been written as if they were driven by (time-dependent) forces $-\varepsilon F_\alpha[p(t'), q(t')]$ depending on the state of the system. Therefore, $Q_\alpha(t)$ consists of the sum of the proper (free) mode $Q^{\mathrm{h}}_\alpha(t)$ and the driven-type term, which naturally depends on the "forcing" (state of the system) at previous times. Then, the replacement of Q_α in the equation for the system variables by the driven-oscillator solution incorporates:

1. The time-dependent modulation due to the proper modes of the environment.

[30] Explicit expressions for the f_{a} and the kernels in terms of the proper modes are

$$f_{\mathrm{a}}(t) = \varepsilon \sum_\alpha c^{\mathrm{a}}_\alpha Q^{\mathrm{h}}_\alpha(t) \; , \qquad \mathcal{K}_{\mathrm{ab}}(\tau) = \varepsilon^2 \sum_\alpha \frac{c^{\mathrm{a}}_\alpha c^{\mathrm{b}}_\alpha}{\omega^2_\alpha} \cos(\omega_\alpha \tau) \tag{8.28}$$

2. The "back-reaction" on the system of its preceding action on the surrounding medium.

Thus, the formalism leads to a description in terms of a reduced number of dynamical variables at the expense of both explicitly time-dependent (fluctuating) terms and history-dependent (relaxation) terms.

8.3.3 Statistical properties of the fluctuating terms

In order to determine the statistical properties of the fluctuating sources $f_{\mathrm{a}}(t)$, one usually assumes that the environment was in thermodynamical equilibrium at the *initial* time (recall that no statistical assumption has been explicitly introduced until this point):

$$P_{\mathrm{eq}}(\mathbf{P}(t_0),\mathbf{Q}(t_0)) \propto \exp\left[-\beta\mathcal{H}_{\mathrm{E}}(t_0)\right] , \quad \mathcal{H}_{\mathrm{E}}(t_0) = \sum_{\alpha} \tfrac{1}{2}\left[P_\alpha(t_0)^2 + \omega_\alpha^2 Q_\alpha(t_0)^2\right] .$$

The initial distribution is therefore Gaussian and one has for the first two moments

$$\begin{array}{ccc} \langle Q_\alpha(t_0)\rangle = 0 , & & \langle P_\alpha(t_0)\rangle = 0 , \\ \langle Q_\alpha(t_0)Q_\beta(t_0)\rangle = \delta_{\alpha\beta}\dfrac{k_{\mathrm{B}}T}{\omega_\alpha^2} , & \langle Q_\alpha(t_0)P_\beta(t_0)\rangle = 0 , & \langle P_\alpha(t_0)P_\beta(t_0)\rangle = \delta_{\alpha\beta}k_{\mathrm{B}}T . \end{array}$$

From these results one readily gets the averages of the proper modes over initial states of the environment (ensemble averages):

$$\begin{aligned}
\left\langle Q^{\rm h}_\alpha(t)\right\rangle &= \underbrace{\langle Q_\alpha(t_0)\rangle}_{0}\cos[\omega_\alpha(t-t_0)] + \underbrace{\langle P_\alpha(t_0)\rangle}_{0}\frac{1}{\omega_\alpha}\sin[\omega_\alpha(t-t_0)]\ ,\\
\left\langle Q^{\rm h}_\alpha(t)Q^{\rm h}_\beta(t')\right\rangle &= \underbrace{\langle Q_\alpha(t_0)Q_\beta(t_0)\rangle}_{\delta_{\alpha\beta}k_{\rm B}T/\omega^2_\alpha}\cos[\omega_\alpha(t-t_0)]\cos[\omega_\beta(t'-t_0)]\\
&\quad + \underbrace{\langle Q_\alpha(t_0)P_\beta(t_0)\rangle}_{0}\frac{1}{m_b\omega_\beta}\cos[\omega_\alpha(t-t_0)]\sin[\omega_\beta(t'-t_0)]\\
&\quad + \underbrace{\langle P_\alpha(t_0)Q_\beta(t_0)\rangle}_{0}\frac{1}{\omega_\alpha}\sin[\omega_\alpha(t-t_0)]\cos[\omega_\beta(t'-t_0)]\\
&\quad + \underbrace{\langle P_\alpha(t_0)P_\beta(t_0)\rangle}_{\delta_{\alpha\beta}k_{\rm B}T}\frac{1}{\omega_\alpha m_b\omega_\beta}\sin[\omega_\alpha(t-t_0)]\sin[\omega_\beta(t'-t_0)]\\
&= k_{\rm B}T\frac{\delta_{\alpha\beta}}{\omega^2_\alpha}\{\cos[\omega_\alpha(t-t_0)]\cos[\omega_\alpha(t'-t_0)]\\
&\qquad + \sin[\omega_\alpha(t-t_0)]\sin[\omega_\alpha(t'-t_0)]\}\ ,
\end{aligned}$$

so that

$$\left\langle Q^{\rm h}_\alpha(t)\right\rangle = 0\ ,\qquad \left\langle Q^{\rm h}_\alpha(t)Q^{\rm h}_\beta(t')\right\rangle = k_{\rm B}T\frac{\delta_{\alpha\beta}}{\omega^2_\alpha}\cos[\omega_\alpha(t-t')]\ . \tag{8.29}$$

Then, since Eq. (8.28) says that $f_{\rm a}(t) = \varepsilon\sum_\alpha c^{\rm a}_\alpha Q^{\rm h}_\alpha(t)$ and $\mathcal{K}_{\rm ab}(\tau) = \varepsilon^2\sum_\alpha(c^{\rm a}_\alpha c^{\rm b}_\alpha/\omega^2_\alpha)\cos(\omega_\alpha\tau)$, the equations (8.29) give for the averages of the fluctuating terms $f_{\rm a}(t)$:

$$\boxed{\begin{aligned}\langle f_{\rm a}(t)\rangle &= 0\ ,\\ \langle f_{\rm a}(t)f_{\rm b}(t')\rangle &= k_{\rm B}T\,\mathcal{K}_{\rm ab}(t-t')\ .\end{aligned}} \tag{8.30}$$

The second equation relates the statistical time correlation of the fluctuating terms $f_{\rm a}(t)$ with the relaxation memory kernels $\mathcal{K}_{\rm ab}(\tau)$ occurring in the dynamical equations (*fluctuation-dissipation* relations). Short (long) correlation times of the fluctuating terms entail short-range (long-range) memory effects in the relaxation term, and vice versa. The emergence of this type of relations is not surprising in this context, since fluctuations and relaxation

arise as different manifestations of the *same* interaction of the system with the surrounding medium.[31] To conclude, we show in Fig. 16, the quantity $f(t) = \varepsilon \sum_k c_k \left[Q_k(t_0) \cos(\omega_k t) + [P_k(t_0)/\omega_k] \sin(\omega_k t) \right]$, with $c_k \propto k$, $\omega_k = ck$, and the $(P(t_0), Q(t_0))$ drawn from a Gaussian distribution. The graph shows that a quantity obtained by adding many sinusoidal terms with different frequencies and phases can actually be a highly irregular function of t.

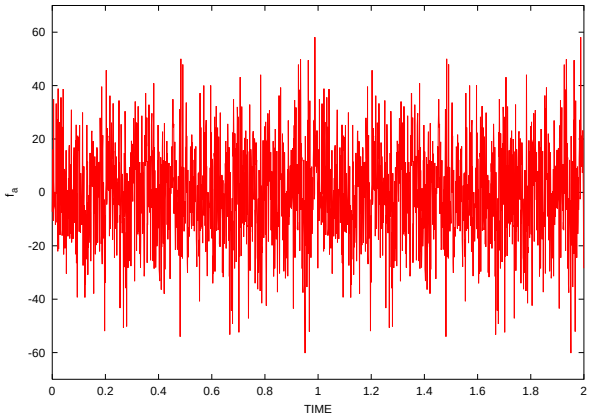

Figure 16: The quantity $f(t)$ obtained by summing over 1000 oscillators with initial conditions drawn from a Gaussian distribution.

8.3.4 Markovian regime

We shall now study the form that the dynamical equations derived exhibit in the absence of memory effects. This occurs when the memory kernels

[31] If one assumes that the environment is at $t = t_0$ in thermal equilibrium *in the presence of the system*, which is however taken as *fastened* in its initial state, the corresponding initial distribution of the environment variables is $P_{\rm eq} \propto \exp[-\mathcal{H}_{\rm SE}(t_0)/k_{\rm B}T]$, with

$$\mathcal{H}_{\rm SE}(t_0) = \sum_\alpha \tfrac{1}{2} \left\{ P_\alpha(t_0)^2 + \omega_\alpha^2 \left[Q_\alpha(t_0) + (\varepsilon/\omega_\alpha^2) F_\alpha(t_0) \right]^2 \right\} .$$

In this case, the dropped terms $\mathcal{K}_\alpha(t - t_0) F_\alpha(t_0)$, which for $F_\alpha = \sum_{\rm a} c_\alpha^{\rm a} W_{\rm a}$ lead to $\sum_{\rm b} \mathcal{K}_{\rm ab}(t - t_0) W_{\rm b}(t_0)$, are included into an alternative definition of the fluctuating sources, namely $\tilde{f}_{\rm a}(t) = f_{\rm a}(t) + \sum_{\rm b} \mathcal{K}_{\rm ab}(t - t_0) W_{\rm b}(t_0)$. The statistical properties of these terms, as determined by the above distribution, are given by expressions *identical* with Eqs. (8.30). Then, with both types of initial conditions one obtains the *same* Langevin equation after a time of the order of the width of the memory kernels $\mathcal{K}_{\rm ab}(\tau)$, which is the characteristic time for the "transient" terms $\sum_{\rm b} \mathcal{K}_{\rm ab}(t - t_0) W_{\rm b}(t_0)$ to die out.

are sharply peaked at $\tau = 0$, the remainder terms in the memory integrals change slowly enough in the relevant range, and the kernels enclose a finite non-zero algebraic area. Under these conditions, one can replace the kernels by Dirac deltas and no memory effects occur.

Doing this with the memory kernel (8.27), we write

$$\mathcal{K}_{\rm ab}(\tau) = 2\gamma_{\rm ab}\delta(\tau) \ , \tag{8.31}$$

where the $\gamma_{\rm ab}$ are *damping coefficients* related with the strength and characteristics of the coupling (see below). Then, on using $\int_0^\infty {\rm d}\tau\, \delta(\tau)h(\tau) = h(0)/2$, equation (8.26) reduces to

$$\boxed{\frac{{\rm d}A}{{\rm d}t} = \big\{A, \mathcal{H}\big\} + \sum_{\rm a} \big\{A, W_{\rm a}\big\}\left[f_{\rm a}(t) + \sum_{\rm b} \gamma_{\rm ab}\frac{{\rm d}W_{\rm b}}{{\rm d}t}\right] \ ,} \tag{8.32}$$

with

$$\langle f_{\rm a}(t)\rangle = 0 \ , \qquad \langle f_{\rm a}(t) f_{\rm b}(t')\rangle = 2\gamma_{\rm ab} k_{\rm B}T\, \delta(t-t') \ . \tag{8.33}$$

Inspecting Eq. (8.31), one sees that the damping coefficients can be obtained from the area enclosed by the memory kernels or, alternatively, by inserting the definition of the kernel (8.28) into the corresponding integral and then using $\int_0^\infty {\rm d}\tau\, \cos(\omega\tau) = \pi\delta(\omega)$:

$$\gamma_{\rm ab} = \int_0^\infty {\rm d}\tau\, \mathcal{K}_{\rm ab}(\tau) \ , \qquad \gamma_{\rm ab} = \pi\varepsilon^2 \sum_\alpha \frac{c_\alpha^{\rm a} c_\alpha^{\rm b}}{\omega_\alpha^2}\delta(\omega_\alpha) \ . \tag{8.34}$$

The area $\int_0^\infty {\rm d}\tau\, \mathcal{K}_{\rm ab}(\tau)$ must be: (i) *finite* and (ii) *different from zero*, for the Markovian approximation to work. The second expression gives the damping coefficients in terms of the distribution of normal modes and system-environment coupling constants, and could be useful in cases where it could be difficult to find the kernels exactly.

8.4 Examples: Brownian particles and spins

In order to particularize the general expressions to definite situations, we only need to specify the structure of the coupling terms $F_{\rm a}$.

Brownian particle. For instance, let us set $F_\alpha(p,q) = -c_\alpha q$ (bilinear coupling), and write down Eq. (8.24) for $A = q$ and $A = p$ with help from $\{p, B\} = -\partial B/\partial q$ and $\{q, B\} = \partial B/\partial p$. Then one gets $\mathrm{d}q/\mathrm{d}t = \partial\mathcal{H}/\partial p$ plus Eq. (8.3), which is the celebrated generalized Langevin equation for a "Brownian" particle. The fluctuating force is explicitly given by $f(t) = \sum_\alpha c_\alpha f_\alpha(t)$ and the memory kernel by $\mathcal{K}(\tau) = \sum_\alpha c_\alpha^2 \mathcal{K}_\alpha(\tau)$. Naturally, in the Markovian limit $\mathcal{K}(\tau) = 2m\gamma\delta(\tau)$ we have

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial\mathcal{H}}{\partial p}\,, \qquad \frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial\mathcal{H}}{\partial q} + f(t) - \gamma\frac{\mathrm{d}q}{\mathrm{d}t}\,, \tag{8.35}$$

whose relaxation term comprises minus the velocity $-(\mathrm{d}q/\mathrm{d}t)$ of the particle.

In general, when $\{A, F_\mathrm{a}\}$ in Eq. (8.32) is not constant, the fluctuating terms $f_\mathrm{a}(t)$ enter multiplying the system variables (*multiplicative* fluctuations). In this example, owing to the fact that $\{q, -c_\alpha q\} = 0$ and $\{p, -c_\alpha q\} = c_\alpha$, the fluctuations are *additive*.

Spin-dynamics. Let us now particularize the above results to the dynamics of a classical spin. To do so, we merely put $A = s_i$, $i = x, y, z$, in Eq. (8.24), and then use Eq. (8.14) to calculate the Poisson brackets required. Using also $\mathrm{d}W_\mathrm{b}/\mathrm{d}t = (\partial W_\mathrm{b}/\partial\vec{s}) \cdot (\mathrm{d}\vec{s}/\mathrm{d}t)$, we have

$$\frac{\mathrm{d}s_i}{\mathrm{d}t} = -\left(\vec{s}\wedge\frac{\partial\mathcal{H}}{\partial\vec{s}}\right)_i - \sum_\mathrm{a}\left(\vec{s}\wedge\frac{\partial W_\mathrm{a}}{\partial\vec{s}}\right)_i\left[f_\mathrm{a}(t) + \sum_\mathrm{b}\gamma_\mathrm{ab}\frac{\partial W_\mathrm{b}}{\partial\vec{s}}\cdot\frac{\mathrm{d}\vec{s}}{\mathrm{d}t}\right]\,,$$

On gathering these results for $i = x, y, z$ in vectorial form and recalling the definition of the effective field $\vec{B} = -\partial\mathcal{H}/\partial\vec{s}$, we arrive at

$$\frac{\mathrm{d}\vec{s}}{\mathrm{d}t} = \vec{s}\wedge\vec{B} - \vec{s}\wedge\left(\sum_\mathrm{a} f_\mathrm{a}(t)\frac{\partial W_\mathrm{a}}{\partial\vec{s}} + \left[\sum_\mathrm{ab}\gamma_\mathrm{ab}\frac{\partial W_\mathrm{a}}{\partial\vec{s}}\frac{\partial W_\mathrm{b}}{\partial\vec{s}}\right]\cdot\frac{\mathrm{d}\vec{s}}{\mathrm{d}t}\right)\,.$$

Then, defining the *fluctuating magnetic field*

$$\vec{\xi}(t) = -\sum_\mathrm{a} f_\mathrm{a}(t)\frac{\partial W_\mathrm{a}}{\partial\vec{s}}\,, \tag{8.36}$$

and the second-rank tensor $\hat{\Lambda}$ with elements

$$\Lambda_{ij} = \sum_\mathrm{a,b}\gamma_\mathrm{ab}\frac{\partial W_\mathrm{a}}{\partial s_i}\frac{\partial W_\mathrm{b}}{\partial s_j}\,, \tag{8.37}$$

we finally obtain the Langevin equation for the spin[32]

$$\boxed{\frac{\mathrm{d}\vec{s}}{\mathrm{d}t} = \vec{s} \wedge \left[\vec{B} + \vec{\xi}(t)\right] - \vec{s} \wedge \hat{\Lambda} \frac{\mathrm{d}\vec{s}}{\mathrm{d}t} \; .} \tag{8.38}$$

Equation (8.38) contains $\mathrm{d}\vec{s}/\mathrm{d}t$ on its right-hand side, so it will be referred to as a *Gilbert-type* equation. For $\varepsilon \ll 1$, on replacing perturbatively that derivative by its conservative part, $\mathrm{d}\vec{s}/\mathrm{d}t \simeq \vec{s} \wedge \vec{B}$, one gets the weak-coupling *Landau–Lifshitz-type* equation

$$\boxed{\frac{\mathrm{d}\vec{s}}{\mathrm{d}t} = \vec{s} \wedge \left[\vec{B} + \vec{\xi}(t)\right] - \vec{s} \wedge \hat{\Lambda} \left(\vec{s} \wedge \vec{B}\right) \; .} \tag{8.39}$$

which describes weakly damped precession. From the statistical properties (8.33) of the fluctuating sources $f_{\mathrm{a}}(t)$, one gets

$$\langle \xi_i(t) \rangle = 0 \; , \quad \langle \xi_i(t) \xi_j(t') \rangle = 2 \Lambda_{ij} k_{\mathrm{B}} T \delta(t - t') \; , \tag{8.40}$$

which relates the structure of the correlations of the fluctuating field and the relaxation tensor.

For a general form of the spin-environment interaction, due to the occurrence of the tensor $\hat{\Lambda}$ the structure of the relaxation terms in the above equations deviates from the forms proposed by Gilbert and Landau and Lifshitz. However, if the spin-environment interaction yields uncorrelated *and* isotropic fluctuations ($\Lambda_{ij} = \lambda \delta_{ij}$), one finds that: (i) the statistical properties (8.40) reduce to those in (8.2) and (ii) the Langevin equation (8.39) reduces to the stochastic Landau–Lifshitz equation (8.2).

We remark in closing that the occurrence of the vector *product* $\vec{s} \wedge \vec{\xi}$ in the dynamical equations entails that the fluctuating terms enter in a *multiplicative* way. In the spin-dynamics case, in analogy with the results obtained for mechanical rigid rotators, the multiplicative character of the fluctuations is an inevitable consequence of the Poisson bracket relations for angular-momentum-type dynamical variables $\left\{s_i, s_j\right\} = \sum_k \epsilon_{ijk} s_k$, which, even for F_{a} linear in $\vec{s}$, lead to non-constant $\left\{A, F_{\mathrm{a}}\right\}$ in Eq. (8.24). In our derivation this can straightly be traced back by virtue of the Hamiltonian formalism employed.

[32] Although we omit the symbol of scalar product, the action of a dyadic $\vec{A}\,\vec{B}$ on a vector $\vec{C}$ is the standard one: $(\vec{A}\,\vec{B})\vec{C} \equiv \vec{A}(\vec{B} \cdot \vec{C})$.

8.5 Discussion

We have seen that starting from a Hamiltonian description of a classical system interacting with the surrounding medium, one can derive generalized Langevin equations, which, in the Markovian approach, reduce to known phenomenological Langevin equations.

Note however that the presented derivation of the equations is formal in the sense that one must still investigate specific realizations of the system-plus-environment whole system, and then prove that the assumptions employed (mainly that of Markovian behavior) are at least approximately met. Let us give an example for a particle coupled to the elastic waves (phonons) of the substrate where it moves. The interaction would be proportional to the deformation tensor $\mathcal{H}_{\rm SE} \propto \partial u/\partial x$ in one dimension. Expanding the displacement field in normal modes $u(x) = \sum_k u_k \exp(\mathrm{i}kx)$, where the u_k are the coordinates of the environment variables (our Q_α), we have $\mathcal{H}_{\rm SE} \propto \sum_k \mathrm{i}k \exp(\mathrm{i}kx) u_k$, so that $c_k \propto \mathrm{i}k \exp(\mathrm{i}kx)$. If we had allowed complex c_α, the products c_α^2 would had been replaced by $|c_\alpha|^2$. Well, the corresponding memory kernel [Eq. (8.28)], then gives

$$\mathcal{K}(\tau) = \varepsilon^2 \sum_\alpha \frac{|c_\alpha|^2}{\omega_\alpha^2} \cos(\omega_\alpha \tau) \stackrel{\omega_k = ck}{\longrightarrow} \int_0^{k_{\rm D}} \mathrm{d}k \frac{\not{k}^2}{c^2\, \not{k}^2} \cos(ck\tau) \propto \frac{\sin(\omega_{\rm D}\tau)}{\tau} \,.$$

But, $\sin(\Omega\tau)/\tau$ plays the role of a Dirac delta for any process with time-scales much larger than $1/\Omega$. Thus, taking the Markovian limit is well justified in this case.

On the other hand, we have considered the classical regime of the environment and the system. A classical description of the environment is adequate, for example, for the coupling to low-frequency ($\hbar\omega_\alpha/k_{\rm B}T \ll 1$) normal modes. Nevertheless, the fully Hamiltonian formalism used, allows to guess the structure of the equations in the quantum case (just replacing Poisson brackets by commutators).

References

[1] A. O. Caldeira and A. J. Leggett, *Dissipation and quantum tunnelling*, Ann. Phys. (NY) **149** (1983), 374–456, Physica A **121**, 587 (1983).

[2] C. W. Gardiner, *Handbook of stochastic methods*, 2nd ed., Springer, Berlin, 1990.

[3] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes*, 2nd ed., Cambridge University Press, New York, 1992.

[4] H. Risken, *The Fokker–Planck equation*, 2nd ed., Springer, Berlin, 1989.

[5] N. G. van Kampen, *Stochastic processes in physics and chemistry*, 2nd ed., North-Holland, Amsterdam, 1992.

[6] H. S. Wio, *An introduction to stochastic processes and nonequilibrium statistical physics*, World Scientific, Singapore, 1994, Series on advances in statistical mechanics – vol. 10.

APPENDICES

Dynamical equations for the averages: macroscopic equation

From the master equation one can derive the dynamical equations for the averages of a Markov stochastic process. We shall write down the corresponding derivations directly in the multivariate case.

Let us first write the equation for the time evolution of an arbitrary function $\langle f(\mathbf{y})\rangle$.[33] First, one has

$$\frac{d}{dt}\langle f(\mathbf{y})\rangle = \frac{d}{dt}\int \mathrm{d}\mathbf{y}\, f(\mathbf{y})P(\mathbf{y},t) = \int \mathrm{d}\mathbf{y}\, f(\mathbf{y})\frac{\partial P(\mathbf{y},t)}{\partial t}\ .$$

Then, by using the master equation to express $\partial P/\partial t$, one has

$$\begin{aligned}
\frac{d}{dt}\langle f(\mathbf{y})\rangle &= \underbrace{\int \mathrm{d}\mathbf{y}\, f(\mathbf{y})\int \mathrm{d}\mathbf{y}'\, W(\mathbf{y}|\mathbf{y}')P(\mathbf{y}',t)}_{\mathbf{y}'\leftrightarrow\mathbf{y}} - \int \mathrm{d}\mathbf{y}\, f(\mathbf{y})\int \mathrm{d}\mathbf{y}'\, W(\mathbf{y}'|\mathbf{y})P(\mathbf{y},t) \\
&= \int \mathrm{d}\mathbf{y}'\, f(\mathbf{y}')\int \mathrm{d}\mathbf{y}\, W(\mathbf{y}'|\mathbf{y})P(\mathbf{y},t) - \int \mathrm{d}\mathbf{y}\, f(\mathbf{y})\int \mathrm{d}\mathbf{y}'\, W(\mathbf{y}'|\mathbf{y})P(\mathbf{y},t) \\
&= \int \mathrm{d}\mathbf{y}\, P(\mathbf{y},t)\int \mathrm{d}\mathbf{y}'\, [f(\mathbf{y}')-f(\mathbf{y})]W(\mathbf{y}'|\mathbf{y})\ . \qquad (.41)
\end{aligned}$$

On applying now this equation to $f(\mathbf{y}) = y_i$, and defining [cf. Eq. (4.2)]

$$a_i^{(1)}(\mathbf{y},t) = \int \mathrm{d}\mathbf{y}'\, (y_i' - y_i)W(\mathbf{y}'|\mathbf{y})\ , \qquad (.42)$$

the last line in Eq. (.41) is the average of $a_i^{(1)}(\mathbf{y},t)$, so we can finally write

$$\boxed{\frac{d}{dt}\langle y_i\rangle = \left\langle a_i^{(1)}(\mathbf{y},t)\right\rangle \qquad (i=1,2,\ldots)\ .} \qquad (.43)$$

This is an exact consequence of the master equation and therefore holds for any Markov process.

[33] Here we use the same notation for the stochastic process and its realisations.

Note that when $a_i^{(1)}$ is *linear* function of $\mathbf{y}$ one has $\left\langle a_i^{(1)}(\mathbf{y},t)\right\rangle = a_i^{(1)}(\langle\mathbf{y}\rangle,t)$, whence

$$\frac{d}{dt}\langle y_i\rangle = a_i^{(1)}(\langle\mathbf{y}\rangle,t)\ ,$$

which is a system of ordinary differential equations for $\langle\mathbf{y}\rangle$ and can be identified with the *macroscopic* equation of the system. For instance in the decay problem (4.9), since $W_{n',n}$ is non-zero for $n'=n-1$, we have

$$a^{(1)}(n,t) = \sum_{n'}(n'-n)\ \underbrace{W_{n',n}}_{\gamma n\delta_{n',n-1}} = [(\not{n}-1)-\not{n}]\,\gamma n = -\gamma n\ .$$

Therefore, from the general result (.43) we have $\mathrm{d}\langle n\rangle/\mathrm{d}t = \langle a^{(1)}(n,t)\rangle = -\langle\gamma n\rangle$, in agreement with Eq. (4.10).

On the other hand, when $a_i^{(1)}$ is a *non-linear* function of $\mathbf{y}$, Eq. (.43) is not a differential equation for $\langle y_i\rangle$. Then Eq. (.43) is not a closed equation for $\langle y_i\rangle$ but higher-order moments enter as well. Thus, the evolution of $\langle y_i\rangle$ in the course of time is not determined by $\langle y_i\rangle$ itself, but is also influenced by the fluctuations around this average. To get equations for higher-order moments we proceed analogously. For instance, for $\langle y_i(t)y_j(t)\rangle$, we can use Eq. (.41) with $f(\mathbf{y}) = y_iy_j$. Writting $(y_i'y_j'-y_iy_j) = (y_i'-y_i)(y_j'-y_j)+y_i(y_j'-y_j)+y_j(y_i'-y_i)$, and defining, analogously to Eqs. (4.2) and (.42),

$$a_{ij}^{(2)}(\mathbf{y},t) = \int\mathrm{d}\mathbf{y}'\,(y_i'-y_i)(y_j'-y_j)W(\mathbf{y}'|\mathbf{y})\ . \qquad (.44)$$

we finally have

$$\boxed{\frac{d}{dt}\langle y_iy_j\rangle = \left\langle a_{ij}^{(2)}(\mathbf{y},t)\right\rangle + \left\langle y_ia_j^{(1)}(\mathbf{y},t)\right\rangle + \left\langle y_ja_i^{(1)}(\mathbf{y},t)\right\rangle\ .} \qquad (.45)$$

which is also an exact consequence of the master equation.[34] However, if $a_{ij}^{(2)}$ is a non-linear function of $\mathbf{y}$, the equation involves even higher order moments

[34] Note that for one variable (or for $i=j$) Eq. (.45) reduces to

$$\frac{d}{dt}\langle y^2\rangle = \left\langle a^{(2)}(y,t)\right\rangle + 2\left\langle ya^{(1)}(y,t)\right\rangle\ ,$$

where [cf. Eq. (4.2)]

$$a^{(1)}(y,t) = \int\mathrm{d}y'\,(y'-y)W(y'|y)\ ,\qquad a^{(2)}(y,t) = \int\mathrm{d}y'\,(y'-y)^2W(y'|y)\ , \qquad (.46)$$

are the one-variable counterparts of Eqs. (.42) and (.44), respectively.

$\langle y_i y_j y_k \rangle$, so what we have is an infinite hierarchy of coupled equations for the moments.

The Langevin process $\xi(t)$ as the derivative of the Wiener–Lévy process

Let us formally write $\mathrm{d}W/\mathrm{d}t = \xi(t)$, and see which are the properties of the $W(t)$ so defined. On integrating over an interval τ, we have

$$w(\tau) \equiv \Delta W(\tau) \equiv W(t+\tau) - W(t) = \int_t^{t+\tau} \xi(s) ds \; . \qquad (.47)$$

Let us show that this $w(\tau)$ is indeed a Wiener–Lévy process. Firstly, $w(\tau)$ is Gaussian because $\xi(t)$ is so. Furthermore, on using the statistical properties (5.2) one gets $(\tau, \tau_1, \tau_2 \geq 0)$

$$w(0) = 0 \; , \qquad \langle w(\tau) \rangle = 0 \; , \qquad \langle w(\tau_1) w(\tau_2) \rangle = 2D \min(\tau_1, \tau_2) \; . \qquad (.48)$$

Proof: $w(0) = 0$ follows immediately from the definition (.47), while for the average $\langle w(\tau) \rangle$, one gets $\langle w(\tau) \rangle = \int_t^{t+\tau} \underbrace{\langle \xi(s) \rangle}_{0} ds = 0$. On the other hand, for $\langle w(\tau_1) w(\tau_2) \rangle$, one finds

$$\begin{aligned} \langle w(\tau_1) w(\tau_2) \rangle &= \int_t^{t+\tau_1} \int_t^{t+\tau_2} \overbrace{\langle \xi(s) \xi(s') \rangle}^{2D\delta(s-s')} ds' ds \\ &= 2D \int_t^{t+\min(\tau_1,\tau_2)} \overbrace{\int_t^{t+\max(\tau_1,\tau_2)} \delta(s-s') ds'}^{1} \, ds \\ &= 2D \min(\tau_1, \tau_2) \; , \end{aligned}$$

where we have sorted the integrals to ensure that, when using the Dirac delta to take one of them, the location of the "maximum" of the delta is inside the corresponding integration interval, and the result is therefore unity.

Now on comparing these results with those for the *increment* of the Wiener–Lévy process, whose average is zero since that of the Wiener–Lévy process is zero and the second moment is given by Eqs. (3.12), one realises that the process defined by Eq. (.47) coincides with the increment of a Wiener–Lévy process.[35]

[35] They exactly coincide if $w(\tau)$ is multiplied by $1/\sqrt{2D}$.

Proof of the convergence of the Heun scheme

We shall check that the Heun scheme correctly generates the Kramers–Moyal coefficients, by carrying out the Taylor expansion of Eq. (7.4), accounting for Eq. (7.5). Concerning the terms involving A_i, one has

$$\begin{aligned}
&\tfrac{1}{2}\left\{A_i(\tilde{\mathbf{y}},t+\Delta t)+A_i[\mathbf{y}(t),t]\right\}\Delta t \\
&\qquad = \tfrac{1}{2}\left\{A_i[\mathbf{y}(t),t]+\frac{\partial A_i}{\partial t}\Delta t+\sum_j \frac{\partial A_i}{\partial y_j}\overbrace{[\tilde{y}_j-y_j(t)]}^{A_j\Delta t+\sum_k B_{jk}\Delta W_k}+\cdots+A_i[\mathbf{y}(t),t]\right\}\Delta t \\
&\qquad = A_i[\mathbf{y}(t),t]\Delta t+\mathcal{O}[(\Delta t)^{3/2}]\;,
\end{aligned}$$

whereas, the terms involving B_{ik} can be expanded as

$$\begin{aligned}
&\tfrac{1}{2}\left\{B_{ik}(\tilde{\mathbf{y}},t+\Delta t)+B_{ik}[\mathbf{y}(t),t]\right\}\Delta W_k \\
&\quad = \tfrac{1}{2}\left\{B_{ik}[\mathbf{y}(t),t]+\frac{\partial B_{ik}}{\partial t}\Delta t+\sum_k \frac{\partial B_{ik}}{\partial y_j}\overbrace{[\tilde{y}_j-y_j(t)]}^{A_j\Delta t+\sum_\ell B_{j\ell}\Delta W_\ell}+\cdots+B_{ik}[\mathbf{y}(t),t]\right\}\Delta W_k \\
&\quad = B_{ik}[\mathbf{y}(t),t]\Delta W_k+\sum_k \frac{\partial B_{ik}}{\partial y_j}[\mathbf{y}(t),t]\sum_\ell B_{j\ell}[\mathbf{y}(t),t]\Delta W_\ell\Delta W_k+\mathcal{O}[(\Delta t)^{3/2}]\;.
\end{aligned}$$

In this case we have retained in $\tilde{y}_j-y_j(t)$ terms up to order $(\Delta t)^{1/2}$, which in the corresponding expansion of A_i are omitted since they yield terms of order $(\Delta t)^{3/2}$. Finally, on inserting these expansions in Eq. (7.4), on gets

$$\begin{aligned}
y_i(t+\Delta t) \simeq\; & y_i(t)+A_i[\mathbf{y}(t),t]\Delta t+\sum_k B_{ik}[\mathbf{y}(t),t]\Delta W_k \\
& +\frac{1}{2}\sum_{k\ell}\left\{\sum_j B_{j\ell}[\mathbf{y}(t),t]\frac{\partial B_{ik}}{\partial y_j}[\mathbf{y}(t),t]\right\}\Delta W_k\Delta W_\ell\;, \qquad (.49)
\end{aligned}$$

which corresponds to Eq. (2.8) of Ramírez-Piscina, Sancho and Hernández-Machado. Finally, to obtain the Kramers–Moyal coefficients, we have to average Eq. (.49) for fixed initial values $\mathbf{y}(t)$ (conditional average). To do so, one can use $\langle\Delta W_k\rangle=0$ and $\langle\Delta W_k\Delta W_\ell\rangle=(2D\Delta t)\delta_{k\ell}$, to get

$$\begin{aligned}
\left\langle\sum_k B_{ik}\Delta W_k\right\rangle &= 0\;, \\
\left\langle\frac{1}{2}\sum_{jk\ell}B_{j\ell}\frac{\partial B_{ik}}{\partial y_j}\Delta W_k\Delta W_\ell\right\rangle &= D\left(\sum_{jk}B_{jk}\frac{\partial B_{ik}}{\partial y_j}\right)\Delta t\;, \\
\left\langle\sum_k B_{ik}\Delta W_k\sum_\ell B_{j\ell}\Delta W_\ell\right\rangle &= 2D\left(\sum_k B_{ik}B_{jk}\right)\Delta t\;.
\end{aligned}$$

Therefore, from Eq. (.49) one obtains

$$\langle y_i(t+\Delta t) - y_i(t)\rangle = \left(A_i + D\sum_{jk} B_{jk}\frac{\partial B_{ik}}{\partial y_j}\right)\Delta t + \mathcal{O}[(\Delta t)^{3/2}]$$

$$\langle [y_i(t+\Delta t) - y_i(t)]\,[y_j(t+\Delta t) - y_j(t)]\rangle = 2D\left(\sum_k B_{ik}B_{jk}\right)\Delta t + \mathcal{O}[(\Delta t)^{3/2}]\ ,$$

which lead to the Kramers–Moyal coefficients (5.15) via Eq. (4.21). Q.E.D.

Proof of the Box–Muller algorithm.

We can verify that the transformation (7.6) leads to a pair of independent Gaussian random numbers as an exercise of *transformation of variables* as introduced in Sec. 2.4:

$$P_{W_1,W_2}(w_1,w_2) = \int_0^1 \mathrm{d}r_1 \int_0^1 \mathrm{d}r_2\, \delta[w_1 - \sqrt{-2\ln(r_1)}\cos(2\pi r_2)] \times \delta[w_2 - \sqrt{-2\ln(r_1)}\sin(2\pi r_2)] \overbrace{P_{R_1,R_2}(r_1,r_2)}^{1 \text{ by hypothesis}} .$$

Let us now introduce the substitution

$$u_1(r_1,r_2) = \sqrt{-2\ln(r_1)}\cos(2\pi r_2) , \quad u_2(r_1,r_2) = \sqrt{-2\ln(r_1)}\sin(2\pi r_2) ,$$

the Jacobi matrix of which reads

$$\begin{pmatrix} \frac{\partial u_1}{\partial r_1} & \frac{\partial u_1}{\partial r_2} \\ \frac{\partial u_2}{\partial r_1} & \frac{\partial u_2}{\partial r_2} \end{pmatrix} = \begin{pmatrix} -\frac{1}{r_1}\frac{1}{\sqrt{-2\ln(r_1)}}\cos(2\pi r_2) & -2\pi\sqrt{-2\ln(r_1)}\sin(2\pi r_2) \\ -\frac{1}{r_1}\frac{1}{\sqrt{-2\ln(r_1)}}\sin(2\pi r_2) & 2\pi\sqrt{-2\ln(r_1)}\cos(2\pi r_2) \end{pmatrix}$$

and the corresponding Jacobian (the determinant of this matrix) is given by $\partial(u_1,u_2)/\partial(r_1,r_2) = -2\pi/r_1$. Nevertheless, when changing the variables in the above integrals one needs the *absolute value* of the Jacobian of the *inverse* transformation, which is given by $|\partial(r_1,r_2)/\partial(u_1,u_2)| = r_1/2\pi$. Besides, $r_1(u_1,u_2)$ can be obtained from the above transformation: $-2\ln(r_1) = u_1^2 + u_2^2 \Rightarrow r_1 = \exp[-\frac{1}{2}(u_1^2+u_2^2)]$. On using all these results the probability distribution of (w_1,w_2) is finally given by

$$\begin{aligned} P_{W_1,W_2}(w_1,w_2) &= \int_{-\infty}^{\infty}\mathrm{d}u_1 \int_{-\infty}^{\infty}\mathrm{d}u_2\, \delta(w_1-u_1)\delta(w_2-u_2)\frac{1}{2\pi}\exp\left[-\tfrac{1}{2}\left(u_1^2+u_2^2\right)\right] \\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}w_1^2\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}w_2^2\right) . \end{aligned}$$

This expression demonstrates that when r_1 and r_2 are independent random numbers uniformly distributed in the interval $(0,1)$, the random variables w_1 and w_2 given by the transformation (7.6) are indeed independent and Gaussian-distributed with zero mean and variance unity.