

ADVANCED STATISTICAL PHYSICS

Master 2 iCFP - Soft Matter

Christophe Texier



Contents

PART 1 : Out-of-equilibrium statistical physics	3
0 Introduction	3
1 Stochastic processes (1) : the Langevin equation for a particle in a fluid	3
1.1 Fluctuations and Langevin force	3
1.2 A fluctuation-dissipation relation	5
1.3 Diffusion	6
1.4 Large scale properties and the overdamped regime	7
1.5 The free Brownian motion (the Wiener process)	8
Appendix : Microscopic foundations of the Langevin equation	10
Additional exercises	15
2 Stochastic processes (2) : Markov processes and master equation	16
2.1 Generalities : joint probabilities and conditional probabilities	16
2.2 Markov processes	18
2.3 Master equation	21
2.4 Classification of Markov processes	28
2.5 Spectral decomposition - Relaxation	30
2.6 Markov chains	32
2.7 Fourier analysis of stochastic processes – Wiener-Khintchin theorem	37
Additional exercises	39
3 Stochastic processes (3) : stochastic differential equations	41
3.1 SDE with drift and additive noise	41
3.2 SDE with multiplicative noise : Itô or Stratonovich ?	43
Historical note on Döblin-Itô calculus	52
Appendix : Stochastic integrals	53
Additional exercises	54
4 Stochastic processes (4) : the Fokker-Planck approach	56
4.1 Interpretation of the Fokker-Planck equation	57
4.2 The generator of the diffusion and the stationary state	59
4.3 Conditional probability, backward FPE and a useful relation	62
4.4 Relaxation and spectral analysis	64
4.5 Boundary conditions for the FPE	66
4.6 First passage and exit problem (in 1D)	68
Additional exercises	74

5	Linear response theory	77
5.1	Theory of thermodynamic fluctuations	77
5.2	Thermodynamics of irreversible processes	81
5.3	Correlation functions and response functions	84
5.4	Once more the fluctuation-dissipation theorem	88
5.5	Causality and Kramers-Kronig relations	89
	PART 2 : Introduction to phase transitions and critical phenomena	92
6	Mean field	92
6.1	Introduction : the liquid-gas transition	92
6.2	The phenomenological Landau approach	103
6.3	The case of inhomogeneous systems : Ginzburg-Landau's approach	115
7	Beyond mean field	122
7.1	Fluctuation-response relation	122
7.2	The Gaussian approximation and the "one-loop" correction	123
7.3	The Ginzburg criterion	127
7.4	Scaling laws	130
8	Introduction to the renormalization group	137
8.1	The Ising chain	138
8.2	Fixed points	140
8.3	RG flow	144
8.4	General discussion	153
A	Mathematical tools	156
A.1	Linear differential equations with source terms	156
A.2	Poisson formulae	158
A.3	Asymptotics of integrals	158
A.4	Special functions	160
B	Solutions of some of the exercises	162
	References	191
	Index	193

Prerequisites :

- equilibrium statistical physics (ensemble theory)
- probability : moments, cumulants (variance, etc), generating function, central limit theorem,...
- Gaussian integrals, saddle-point (steepest descent)

PART 1 : Out-of-equilibrium statistical physics

0 Introduction

This first set of lectures is devoted to out-of-equilibrium statistical physics. Equilibrium statistical physics provides a well defined procedure to study the thermodynamic properties of systems with complex dynamics. The main idea is to replace the study of the complex dynamics of the system, i.e. how its state $\vec{\Gamma}(t)$ evolves in time (here $\vec{\Gamma}$ represents a point in phase space), by some statistical information, i.e. the probability $\rho(\vec{\Gamma})$ to find the system in a given state. The first approach would require to solve a macroscopic number of differential equations, while the beauty of the second approach lies on the fact that the determination of the probability density relies on very few information, what can be understood as a result of a maximum entropy principle. The choice of the distribution, microcanonical, canonical, grand canonical, etc, is driven by physical considerations or simply by convenience.

Out-of-equilibrium statistical physics requires a statistical treatment of the dynamics, which can be achieved by various approaches, phenomenological or microscopic. On the more phenomenological side : the Langevin equation, the master equation and the Fokker-Planck equation provide different approaches for the analysis of stochastic processes. On the more microscopic side : kinetic equations (BBGKY hierarchy, Boltzmann equation, Vlasov equation, hydrodynamic equations,...). Note that the frontier between phenomenological and microscopic is not so sharp, as we will see by deriving a Langevin equation from a microscopic model (§ 1.5 page 10).

1 Stochastic processes (1) : the Langevin equation for a particle in a fluid

The aim of this introductory chapter is to start the discussion of stochastic processes with a concrete and simple example, and introduce several general and important ideas. Consider a particle in a fluid, submitted to a friction force. The usual *phenomenological* model is friction proportional to the velocity (Stokes regime) :

$$F_f = -\gamma v, \quad (1.1)$$

where γ is the friction coefficient. For a spherical particle of radius R , fluid mechanics gives $\gamma = 6\pi\eta R$ where η is the viscosity of the fluid (for example, $\eta \simeq 10^{-3} \text{ kg.m}^{-1}.\text{s}^{-1}$ for water at $T = 20^\circ\text{C}$). In the absence of any other external force, the Newton equation of motion takes the form $m\dot{v} = -\gamma v$. The friction coefficient has dimension of a mass divided by a time, hence we can write

$$\gamma = \frac{m}{\tau} \quad (1.2)$$

where τ is the *relaxation time* for the velocity.

1.1 Fluctuations and Langevin force

In 1827, the scottish botanist Robert Brown observed with a microscope that pollen grains at the surface of water move erratically. ¹ It was understood later that this observation supports the atomist description of matter as it is the manifestation of the *fluctuations* in the fluid (erratic

¹Brown was not the first one to observe this phenomenon, but he was the first to study it systematically (varying parameters of the experiment). You can find some historical perspectives in the excellent article of Bertrand Duplantier [10].

motion of the molecules). A clear description of the phenomenon was given much later by Albert Einstein in 1905. If the particle (the pollen grain) is small, it is not only submitted to the friction force but it is also sensitive to the *fluctuations in the fluid*, i.e. the collisions with molecules. The typical collision time between molecules in a fluid is $\tau_{\text{coll}} \sim 10^{-15}$ s, thus we expect that the Brownian particle experiences collisions with the rate $1/\tau_{\text{coll}}$ and which can be considered as *independent*. The friction force is due to the effect of these collisions over a much larger time scale. Additionally to the friction force, we model the frequent collisions by introducing in the Newton equation a force $\xi(t)$ fluctuating in time, called the “**Langevin force**” :

$$m \frac{dv(t)}{dt} = -\gamma v(t) + \xi(t) \quad (1.3)$$

$$\frac{dx(t)}{dt} = v(t) \quad (1.4)$$

Because the collisions are exerted at random along all directions, we expect that

$$\langle \xi(t) \rangle = 0 \quad (1.5)$$

where $\langle \dots \rangle$ denotes *statistical averaging*² (it is also true if we consider averaging over time for a single history). As the Langevin force models the force exerted on the particle by the molecules, it is natural to assume short time correlations $\langle \xi(t)\xi(t') \rangle = \frac{C}{\tau_{\text{coll}}} \varphi((t-t')/\tau_{\text{coll}})$ where φ is normalised function of width ~ 1 centered on the origin (like $(1/2)e^{-|x|}$ or $\pi^{-1/2}e^{-x^2}$) and C the strength of the fluctuations. As we are interested in the dynamics of the Brownian particle over time $\gg \tau_{\text{coll}}$ we can simply consider

$$\langle \xi(t)\xi(t') \rangle = C \delta(t-t') \quad (1.6)$$

(which corresponds formally to $\tau_{\text{coll}} \rightarrow 0$). A random function characterised by such local correlations is called a “*white noise*”.



Figure 1.1: *Robert Brown (1773-1858), Albert Einstein (1879-1955), Paul Langevin (1872-1946) and Jean Perrin (1870-1942).*

This model was introduced by Paul Langevin [20]. Why studying a model for the motion of a pollen grain at the surface of a fluid (or more generally a “colloid” in a fluid) is an important problem ? The reason is that several ideas of the Langevin model have a much broader application in out-of-equilibrium statistical physics.

We now analyse the statistical properties of the particle. Taking advantage that the equation of motion is *linear*, its integration gives

$$v(t) = v(0) e^{-t/\tau} + \frac{1}{m} \int_0^t dt' \xi(t') e^{-(t-t')/\tau} . \quad (1.7)$$

²Statistical averaging corresponds to average over **different histories** of the particle, with same initial conditions but different realisations of the Langevin force. This is the procedure followed by Jean Perrin (Nobel prize in 1926) in his experiments [37] ; cf. Fig. 1.3. The difference between statistical averaging and time averaging is described more concretely in Exercise 1.6.

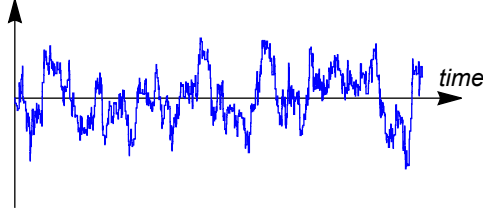


Figure 1.2: A realization of the Ornstein-Uhlenbeck process $v(t)$ defined by (1.3) when $\xi(t)$ is a Gaussian white noise.

This representation makes easy to deduce the statistical properties of $v(t)$ from those of $\xi(t)$. If the initial velocity is non random, we have

$$\langle v(t) \rangle = v(0) e^{-t/\tau} . \quad (1.8)$$

After a time larger than $\tau = m/\gamma$, the memory of the initial velocity is lost and the velocity is independent of $v(0)$. We also get the correlator $\langle v(t)v(t') \rangle_c \stackrel{\text{def}}{=} \langle v(t)v(t') \rangle - \langle v(t) \rangle \langle v(t') \rangle$:

$$\langle v(t)v(t') \rangle_c = \frac{1}{m^2} \int_0^t dt_1 e^{-(t-t_1)/\tau} \int_0^{t'} dt_2 e^{-(t'-t_2)/\tau} \langle \xi(t_1) \xi(t_2) \rangle = \frac{C e^{-(t+t')/\tau}}{m^2} \int_0^{\min(t,t')} dt_1 e^{2t_1/\tau} \quad (1.9)$$

thus

$$\langle v(t)v(t') \rangle_c = \frac{C \tau}{2m^2} \left(e^{-|t-t'|/\tau} - e^{-(t+t')/\tau} \right) . \quad (1.10)$$

The correlations decay in time over the same time scale τ like the average time. The decorrelation of the velocity can be understood from the fact that the velocity decorrelates from the noise on the same time scale : we check easily that

$$\langle v(t)\xi(t') \rangle = \frac{C}{m} \theta_H(t-t') e^{-(t-t')/\tau} . \quad (1.11)$$

The stochastic process described by (1.3) where $\xi(t)$ is a Gaussian white noise is known as the “Ornstein-Uhlenbeck process”.

1.2 A fluctuation-dissipation relation

After a sufficient long time, we expect that the particle is at thermal equilibrium with the fluid (which should be at thermal equilibrium), hence $\langle v(t)^2 \rangle = k_B T / m$ (equipartition theorem). On the other hand, the outcome of the Langevin analysis is $\langle v(t)^2 \rangle = C / (2m\gamma)$. Consistency of the previous analysis with equilibrium statistical physics imposes a constraint between the strength C of the Langevin noise (the fluctuations in the fluid), the friction coefficient γ and the temperature :

$$C = 2 \overset{\substack{\text{dissipation} \\ \uparrow}}{\gamma} k_B T \quad (1.12)$$

$\downarrow$
fluctuations

this is a “*fluctuation-dissipation relation*” and a first formulation of the **fluctuation-dissipation theorem** (FDT). The mass has disappeared, which emphasizes that this is a property of the thermal bath (the fluid). We could write the correlator of the noise

$$\langle \xi(t)\xi(t') \rangle = 2\gamma k_B T \delta(t-t') . \quad (1.13)$$

The phenomenological coefficients C , the strength of the Langevin force, and γ , the friction coefficient, are *not* two independent parameters (at least when thermal equilibrium holds). This reminds that friction and fluctuations have the same physical origin : the interaction of the particle with the molecules of the fluid. Below (§ p. 10), we will introduce a microscopic model of friction and try to clarify the origin of this relation.

✎ **Exercise 1.1 Langevin equation for random initial velocity:** The correlator (1.10) corresponds to a fixed initial velocity. Consider now the case where the initial velocity is random, distributed according to $P(v_0) \propto \exp \left\{ -\frac{mv_0^2}{2k_B T} \right\}$.

- Compute the new correlator, denoted $\langle v(t)v(t') \rangle^{\text{equil}}$.
- Can we compare the two correlators ?

✎ **Exercise 1.2 Stationary measure of the Ornstein-Uhlenbeck process:** In the stationary regime, compare the correlator with the one obtained in Exercise ?? . Deduce what is the measure of the Ornstein-Uhlenbeck process.

In Eq. (1.10), the second term is a transient term coming from the initial condition $v(0) = 0$. Hence for times $t, t' \gg \tau$ we can write

$$\langle v(t)v(t') \rangle = \frac{k_B T}{m} e^{-|t-t'|/\tau} \quad (1.14)$$

The correlations (1.11) between velocity and Langevin force can also be rewritten in terms of the temperature

$$\langle v(t)\xi(t') \rangle = \frac{2k_B T}{\tau} \theta_H(t-t') e^{-(t-t')/\tau} . \quad (1.15)$$

1.3 Diffusion

Now that we have obtained the statistical properties of the velocity, let us analyze the statistical properties of the position. The average position is simply

$$\langle x(t) \rangle = v(0)\tau \left(1 - e^{-t/\tau} \right) \quad (1.16)$$

On average, after a time $t \gg \tau$, the particule covers a distance $v(0)\tau$, which is not surprising as the memory of the initial velocity $v(0)$ is lost after time τ . What about the fluctuations ? In the stationary regime the correlator of the speed is a "narrow function" of width τ , with weight

$$\int_{-\infty}^{+\infty} dt \langle v(t)v(t') \rangle_c^{\text{equil}} = \frac{2k_B T}{\gamma} . \quad (1.17)$$

As a result, neglecting the transient regime at short times, we can write

$$\langle x(t)^2 \rangle_c = \int_0^t dt_1 \int_0^t dt_2 \langle v(t_1)v(t_2) \rangle_c \simeq t \int_{-\infty}^{+\infty} dt_1 \langle v(t_1)v(t_2) \rangle_c^{\text{equil}} = \frac{2k_B T}{\gamma} t \quad (1.18)$$

The linear behaviour characterizes **diffusive motion**. We introduce the *diffusion constant*

$$D \stackrel{\text{def}}{=} \lim_{t \rightarrow \infty} \frac{\langle x(t)^2 \rangle_c}{2t} \quad (1.19)$$

In other terms, we have obtained above a general relation between the velocity correlator and the diffusion constant

$$D \stackrel{\text{def}}{=} \int_0^\infty dt \langle v(t)v(0) \rangle \quad (1.20)$$

Coming back to Langevin model, Eqs. (1.3,1.4), we get the expression

$$D = \frac{k_B T}{\gamma} \quad (1.21)$$

which is known as the “*Einstein relation*”. It was obtained in Einstein 1905’s article on the Brownian motion.³ It is also known as the “Einstein-Stokes law” or “Einstein-Stokes-Sutherland law”. This is another formulation of the FDT, relating three different physical quantities, the diffusion constant characterizing the fluctuations of the motion, the friction coefficient characterizing the dissipation and the temperature.

✎ **Exercise 1.3** : Choose initial conditions for a fixed initial velocity $x(0) = 0$ and $v(0) = v_0$. Compute $\langle x(t) \rangle$. Study precisely $\langle x(t)^2 \rangle_c$. Analyze the crossover between the short time and large time $\langle x(t)^2 \rangle_c \propto t$ diffusive behaviour.

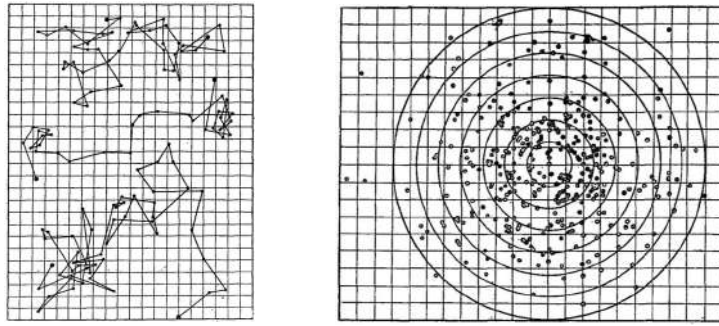


Figure 1.3: Measurements of Jean Perrin (1908) ; from [37]. Left : few examples of trajectories. Right : final points after the several histories (for a fixed time).

1.4 Large scale properties and the overdamped regime

Over large time scales ($\gg \tau$), the correlator (1.14) seems a narrow function which can be replaced by a delta function

$$\langle v(t)v(t') \rangle_c^{\text{equil}} = \frac{C}{\gamma^2} \frac{e^{-|t-t'|/\tau}}{2\tau} \stackrel{\text{large scale}}{\approx} \frac{C}{\gamma^2} \delta(t-t') = \frac{1}{\gamma^2} \langle \xi(t)\xi(t') \rangle . \quad (1.22)$$

Furthermore, assuming that $\xi(t)$ is Gaussian,⁴ this corresponds to write

$$v(t) \stackrel{\text{large scale}}{\approx} \frac{1}{\gamma} \xi(t) \quad (1.23)$$

i.e. to neglect the acceleration term in Newton’s equation :

$$0 \approx -\gamma v(t) + \xi(t) \quad (\text{overdamped regime}) . \quad (1.24)$$

This approximation is called the “*overdamped regime*”, which is achieved either by studying the process over large time scales, $t \gg \tau$, or by formally considering the limit of strong damping, $\gamma \rightarrow \infty$. As a result we obtain that the **velocity equals the force**.

³A. Einstein, “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen”, Annalen der Physik **322**(8), 549–560 (1905).

⁴Gaussian processes have only two finite cumulants κ_1 and κ_2 , all others being zero. Equality of averages and variances of two Gaussian processes implies that they have the same statistical properties.

Overdamped regime : "Aristotle equation".— The pre-Galileo-Newtonian postulate, proposed by Aristotle, asserted that velocity is proportional to force.⁵ The above discussion has shown that this makes sense for the motion of a particle in a viscous fluid. If an additional (conservative) force $F(x)$ is introduced in the equation of motion for the overdamped regime (1.24), we have

$$\boxed{\frac{dx(t)}{dt} \approx \frac{1}{\gamma} [F(x(t)) + \xi(t)]} \quad (\text{overdamped regime}) . \quad (1.25)$$

We can call the equation (velocity) = (force)/(friction) the "*Aristotle equation*", although this is a bit anachronistic (differential calculus was invented only at the end of XVIIIth century by Leibniz and Newton).

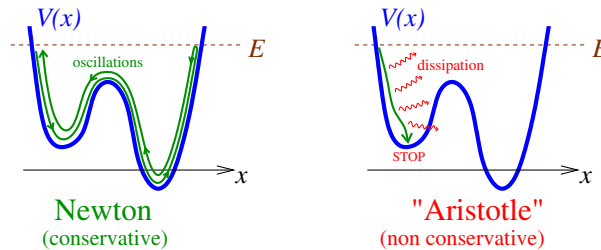


Figure 1.4: *The dynamics described by the Newton equation conserves the energy, whereas the energy is dissipated during the "Aristotle dynamics".*

A last comment about the physical content of the overdamped regime, which we rewrite

$$\frac{dx(t)}{dt} = -\frac{1}{\gamma} \frac{\partial V(x)}{\partial x} + \sqrt{2D} \eta(t) \quad (1.26)$$

where $V(x)$ is the potential, such that $F(x) = -V'(x)$, and $\eta(t)$ a normalised Gaussian white noise, i.e. with $\langle \eta(t) \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$. Contrary to the Newton equation, which describes a *conservative* dynamics, the Aristotle equation describes a *non conservative* dynamics, along which energy is dissipated (cf. Exercise 1.8). In particular, in the $D \rightarrow 0$ limit, the trajectory converges towards the first stationary point (Fig. 1.4). If the first stationary point is a local minimum like on the figure, only a fluctuation (finite D) can untrap the particle in order to visit another neighbouring (local or global) minimum.

We will come back later to a general analysis of this stochastic differential equation.

1.5 The free Brownian motion (the Wiener process)

In the overdamped regime, in the absence of the external force $F(x)$, the position is just the integral of a Gaussian white noise

$$x(t) = \underbrace{x(0)}_{=0} + \frac{1}{\gamma} \int_0^t du \xi(u) \quad (1.27)$$

⁵In the book "Aristotle's Physics", it is claimed that the velocity of a falling body is proportional to its weight, i.e. the velocity is proportional to the force. In other terms the motion requires a force for Aristotle (384-322 BC). It required almost two thousand years to understand that force induces *acceleration*, not motion : motion can exist without force. This was the achievement of Galileo Galilei (1564-1642 AC).

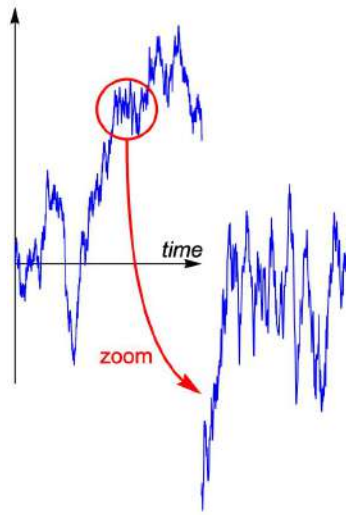


Figure 1.5: A typical Brownian trajectory : $W(t)$ as a function of t . If one zooms, trajectory looks the same (scale invariance).

Let us simplify the notations and introduce a normalised Gaussian white noise $\eta(t) = \xi(t)/\sqrt{C}$ so that

$$\langle \eta(t) \rangle = 0 \quad \text{and} \quad \langle \eta(t)\eta(t') \rangle = \delta(t - t'). \quad (1.28)$$

We now consider the normalised free Brownian motion, with no external force (the “Wiener process”) by setting $x(t) = \frac{\sqrt{C}}{\gamma} W(t)$. Obviously

$$\frac{dW(t)}{dt} = \eta(t) \quad \Rightarrow \quad W(t) = \int_0^t du \eta(u) \quad (1.29)$$

thus

$$\langle W(t)W(t') \rangle = \int_0^t du \int_0^{t'} dv \delta(u - v) = \int_0^{\min(t, t')} du \quad (1.30)$$

Finally

$$\boxed{\langle W(t)W(t') \rangle = \min(t, t')} \quad (1.31)$$

Interpretation : consider the case $t < t'$, we have $\langle W(t)W(t') \rangle = \langle [W(t') - W(t)]W(t) \rangle + \langle W(t)^2 \rangle$. The second term is t (diffusion) ; the first term vanishes as the two increments $W(t') - W(t) = \int_t^{t'} du \eta(u)$ and $W(t) = \int_0^t du \eta(u)$ are independent.

✎ **Exercise 1.4** : Check that the increment depends only on the time difference

$$\langle [W(t) - W(t')]^2 \rangle = |t - t'| \quad (1.32)$$

Continuity.— In the limit $t \rightarrow t'$ we get $\lim_{t \rightarrow t'} \langle [W(t) - W(t')]^2 \rangle = 0$. This last equation suggests that the curve $W(t)$ is **continuous**. As a matter of fact the two statements are disconnected : the fact that this limit is zero does not imply continuity (cf. Exercise 2.5)!

In order to prove continuity, we should consider the probability

$$\text{Proba}\{|W(t) - W(t')| < \epsilon\} = \text{erf}\left(\frac{\epsilon}{\sqrt{2|t - t'|}}\right). \quad (1.33)$$

This expression simply follows from the fact that the increment between times t and t' is Gaussian, with variance $|t - t'|$, hence the cumulative distribution related to the error function $\text{erf}(x)$ (see Appendix A). This expression shows that $\forall \epsilon > 0$, there exists a sufficiently small time difference, choose $|t - t'| \ll \epsilon^2$, so that we are sure that the distance is smaller than ϵ : $\text{Proba}\{|W(t) - W(t')| < \epsilon\} \simeq 1 - \mathcal{O}(e^{-\epsilon^2/(2|t-t'|)})$. Because we can choose ϵ as small as we want, this corresponds to continuity of the process.

Continuity can also be understood in a simpler manner by constructing the Brownian motion from the discrete random walk, $W(t) = \sum_{\tau=1}^{t/\epsilon} \delta W_\tau$ where $\langle \delta W_\tau \rangle = 0$ and $\langle \delta W_\tau \delta W_{\tau'} \rangle = \epsilon \delta_{\tau, \tau'}$, and taking the continuum limit of small jumps $\delta W_\tau \sim \sqrt{\epsilon} \rightarrow 0$ occuring with high rate $1/\epsilon \rightarrow \infty$.

Non differentiability Another consequence of (1.32) is that

$$\left\langle \left(\frac{W(t) - W(t')}{t - t'} \right)^2 \right\rangle = \frac{1}{|t - t'|} \quad (1.34)$$

which goes to infinity when $t \rightarrow t'$: the derivative of typical curves is infinite, i.e. $W(t)$ is *non differentiable*. We can see on the plot that the curve is indeed extremely irregular (see Fig. 1.5).

Scale invariance.— This irregularity is related to *scale invariance* and fractal behaviour (with fractal dimension 1/2). For a scaling factor $\alpha > 0$, it is clear that

$$\langle W(\alpha t) W(\alpha t') \rangle = \alpha \min(t, t') = \alpha \langle W(t) W(t') \rangle \quad (1.35)$$

Because $W(t)$ is Gaussian, all statistical information is encoded in the two point function, hence this equality means that we can identify the statistical properties of $W(\alpha t)$ with those of $\sqrt{\alpha} W(t)$. Mathematicians express this through and “*equality in law*” (or “*equality in distribution*”)

$$\boxed{W(\alpha t) \stackrel{(\text{law})}{=} \sqrt{\alpha} W(t)} \quad (1.36)$$

(the “laws” of the two sides are equal). The same argument applies to the Gaussian white noise gives

$$\eta(\alpha t) \stackrel{(\text{law})}{=} \frac{1}{\sqrt{\alpha}} \eta(t). \quad (1.37)$$

☺ Important points

- Master the analysis of the linear Langevin equation (1.3) (integrate, average, etc)
- Fluctuation-dissipation relation (different forms) : Langevin force and damping force have the same origin, hence the relation.
- Wiener process : main properties.

APPENDIX : Microscopic foundations of the Langevin equation

The aim of this paragraph is to go beyond the phenomenological Langevin model and clarify the physical origin of the Langevin equation from a microscopic description. We introduce a model with a purely Hamiltonian *conservative* dynamics, within which will emerge the effective *dissipative* dynamics described by the Langevin equation. This will allow to identify the microscopic origin of the dissipation.

Model.— One could be tempted to model the collisions in the fluid, however the microscopic dynamics would be difficult to analyze. Instead, we consider a particle coupled to a macroscopic number of uncoupled harmonic oscillators modelling the “environment” (also called the “bath”). In this model, the oscillators represent the eigen-modes of the macroscopic system (like the phonon modes in a fluid). We now study the deterministic dynamics governed by the Hamiltonian

$$H = \frac{p^2}{2m} + V(x) + \sum_n \left[\frac{p_n^2}{2} + \frac{1}{2} \omega_n^2 \left(q_n - \frac{c_n x}{\omega_n^2} \right)^2 \right] \quad (1.38)$$

i.e.

$$H_{\text{sys}}(x, p) = \frac{p^2}{2m} + V(x) \quad (1.39)$$

$$H_{\text{env}}(\{q_n, p_n\}) = \sum_n \left[\frac{p_n^2}{2} + \frac{1}{2} \omega_n^2 q_n^2 \right] \quad (1.40)$$

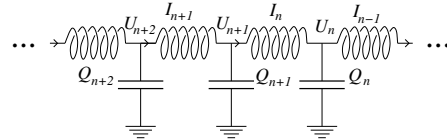
The crucial assumption here is that the coupling between the system and the bath is *linear*

$$H_{\text{int}} = -x \sum_n c_n q_n + \frac{1}{2} x^2 \sum_n \frac{c_n^2}{\omega_n^2} \quad (1.41)$$

where c_n are coupling constants.

Physical realizations are : an electron in an atom, coupled to the electromagnetic modes, or an electric device coupled to a L-C line, etc.

✎ Exercice 1.5 Dissipation in a transmission line : A perfect transmission line (a coaxial cable) is characterised by an inductance and a capacitance per length. A possible discrete model is a series of discrete capacitive and inductive elements (without resistance), i.e. only non-dissipative elements.



We consider harmonic solutions $I_n(t) = \tilde{I}_n e^{-i\omega t}$. We recall that the impedance of the capacitance is $Z_C = 1/(-i\omega C)$ and that of the inductance $Z_L = -i\omega L$, where ω is the frequency.

1/ We first study the eigenmodes of the infinite line. Using Kirchhoff laws, write the equations satisfied by the currents $\tilde{I}_n$.

2/ Propagative modes.— Show that the modes $I_n(t) = e^{iqn - i\omega(q)t}$ only exist in a finite bandwidth $\omega \in [0, \omega_0]$ where $\omega_0 \stackrel{\text{def}}{=} 2/\sqrt{LC}$. Give the dispersion relation.

3/ Evanescent modes.— Study solutions of the form $I_n(t) = (-1)^n e^{qn - i\omega(q)t}$. Over what distance can propagate such modes ?

4/ Impedance of semi-infinite line.— We denote by Z_n the impedance of a finite line involving n couples of $L - C$ elements. Give the recurrence between Z_n and Z_{n+1} . Deduce the impedance of the semi-infinite line $Z_\infty \equiv Z(\omega)$. Plot $\text{Re } Z(\omega)$ and $\text{Im } Z(\omega)$. Discuss the fact that $\text{Re } Z(\omega) \neq 0$ for a certain interval of frequencies (comment this at the light of question 1).

Integration of the bath equations of motion.— We first derive the equations of motion

$$\begin{cases} m \ddot{x} = F(x) - x \sum_n \frac{c_n^2}{\omega_n^2} + \sum_n c_n q_n \\ \ddot{q}_n = -\omega_n^2 q_n + c_n x \end{cases} \quad (1.42)$$

where $F(x) = -V'(x)$. Let us integrate the equations of motion for the bath, which is possible thanks to the linearity. We can use that the retarded Green's function for the harmonic oscillator, i.e. the causal solution of $\ddot{G}^R(t) + \omega_n^2 G^R(t) = \delta(t)$ is $G^R(t) = \theta_H(t) \frac{\sin(\omega_n t)}{\omega_n}$. Thus we can solve the equation of motion for the oscillators

$$q_n(t) = q_n(0) \cos(\omega_n t) + \dot{q}_n(0) \frac{\sin(\omega_n t)}{\omega_n} + c_n \int_0^t dt' \frac{\sin(\omega_n(t-t'))}{\omega_n} x(t') \quad (1.43)$$

We split the source term in Eq. (1.42) in two parts

$$\sum_n c_n q_n(t) = \overbrace{\sum_n c_n \left(q_n(0) \cos(\omega_n t) + \dot{q}_n(0) \frac{\sin(\omega_n t)}{\omega_n} \right)}^{\stackrel{\text{def}}{=} \xi(t)} + \int_0^t dt' \Gamma(t-t') x(t') \quad (1.44)$$

where

$$\Gamma(t) \stackrel{\text{def}}{=} \sum_n c_n^2 \frac{\sin(\omega_n t)}{\omega_n} \quad (1.45)$$

is a function depending of the microscopic parameters of the model. We denote $\xi(t)$ “the noise”, which is controlled by the dynamical variables of the bath, i.e. a macroscopic number of degrees of freedom. For this reason, it is expected to exhibit a complex dynamics. Using $\int_0^\infty dt e^{i\omega t} = \frac{1}{0^+ - i\omega}$ we remark that

$$\int_0^\infty dt \Gamma(t) = \sum_n \frac{c_n^2}{\omega_n^2} \quad (1.46)$$

which appears in the equation of motion above. With these definitions, we can rewrite the effective equation for the particle as

$$m \ddot{x}(t) = F(x(t)) - x(t) \int_0^\infty d\tau \Gamma(\tau) + \int_0^t d\tau \Gamma(\tau) x(t-\tau) + \xi(t) \quad (1.47)$$

Integration over the bath degrees of freedom is responsible for both the integral term and the “noise” term.

Noise and spectral function.— Because the bath involves a macroscopic number of degrees of freedom, it is natural to assume **thermal equilibrium for the bath**, say at $t = 0$, for the bath variables

$$P(\{q_n, p_n\}) \propto e^{-\beta H_{\text{env}}} \quad (1.48)$$

so that

$$\langle q_n(0) q_m(0) \rangle = \delta_{n,m} \frac{k_B T}{\omega_n^2} \quad (1.49)$$

$$\langle \dot{q}_n(0) \dot{q}_m(0) \rangle = \delta_{n,m} k_B T \quad (1.50)$$

Then the noise correlator is

$$C(t-t') = \langle \xi(t) \xi(t') \rangle = k_B T \sum_n \frac{c_n^2 \cos(\omega_n(t-t'))}{\omega_n^2} \quad (1.51)$$

At this stage it is useful to define the spectral function

$$J(\omega) \stackrel{\text{def}}{=} \pi \sum_n \frac{c_n^2}{2\omega_n} \delta(\omega - \omega_n) \quad (1.52)$$

which depends on the distribution of frequencies and coupling constants. We can write the function

$$\Gamma(t) = \frac{2}{\pi} \int_0^\infty d\omega J(\omega) \sin(\omega t) \quad (1.53)$$

and the correlator

$$C(t) = \frac{2k_B T}{\pi} \int_0^\infty d\omega \frac{J(\omega)}{\omega} \cos(\omega t) \quad (1.54)$$

in terms of the spectral function. In particular $C'(t) = -k_B T \Gamma(t)$. Two remarks :

- In practice, we expect a dense spectrum of oscillators for frequencies $\omega \geq 0$ (it is natural to assume that the spectrum of eigenmodes start at $\omega = 0$ since there exist low frequency excitations ususalley).

What kind of behaviour can we expect for $J(\omega)$? Imagine that coupling constant is a smooth function of the frequency $c_n^2 = g(\omega_n)$. Then $J(\omega) = \frac{\pi}{2\omega} \sum_n g(\omega_n) \delta(\omega - \omega_n) \simeq \frac{\pi}{2\omega} g(\omega) \rho(\omega)$ for $\omega \rightarrow 0$, where $\rho(\omega)$ is the spectral density. For a linear spectrum (like photons, or phonons) we have $\rho(\omega) \sim \omega^{d-1}$ and thus we expect a power law $J(\omega) \sim g(\omega) \omega^{d-2}$ at low frequency. A simple assumption is $g(0) = \text{cste}$. In dimension $d = 3$, $g(0) = \text{cste}$ gives $J(\omega) \propto \omega$ for $\omega \rightarrow 0$.

- The spectrum of eigen-frequencies is usually cut off at a frequency ω_D related to the microscopic scale (for phonon modes in a crystal, ω_D is the Debye frequency, related to the lattice spacing).

The Ohmic case, $J(\omega) \propto \omega$ for small frequency : a concrete example.— Assuming a broad spectrum of frequencies, of width ω_D of the form

$$J(\omega) = \gamma_0 \omega \frac{\omega_D^2}{\omega^2 + \omega_D^2} \quad (1.55)$$

gives

$$C(t) = k_B T \gamma_0 \omega_D e^{-\omega_D |t|} . \quad (1.56)$$

Its integral is

$$\int_{-\infty}^{+\infty} dt C(t) = 2\gamma_0 k_B T \quad (1.57)$$

which recall us something...

Effective equation of motion.— Let us come back to the analysis of the effective equation of motion (1.47). If the spectral function $J(\omega)$ is broad (width $\sim \omega_D$), we expect the function $\Gamma(t)$ to be narrow in time (width $\sim 1/\omega_D$). For future convenience, we introduce

$$\gamma(t) = \int_t^\infty dt' \Gamma(t') \quad (1.58)$$

which also decays rapidly over the scale $\sim 1/\omega_D$. We introduce a Heaviside function in its definition to make it causal

$$\gamma(t) = \theta_H(t) \sum_n \frac{c_n^2 \cos(\omega_n t)}{\omega_n^2} = \frac{2\theta_H(t)}{\pi} \int_0^\infty d\omega \frac{J(\omega)}{\omega} \cos(\omega t) \quad (1.59)$$

is a "narrow function" of width $1/\omega_D$.

An integration by parts gives

$$\int_0^t d\tau \Gamma(\tau) x(t - \tau) = \gamma(0) x(t) - \gamma(t) x(0) - \int_0^t d\tau \gamma(\tau) \dot{x}(t - \tau) \quad (1.60)$$

Considering times $t \gg 1/\omega_D$, we drop the term $\gamma(t) x(0)$. We end with the effective equation of motion

$$\boxed{m \ddot{x}(t) = F(x(t)) - \int_0^t d\tau \gamma(\tau) \dot{x}(t - \tau) + \xi(t)} \quad (1.61)$$

After "integration over the bath degree of freedom", we have obtained that the evolution the system is described by a integro-differential equation. This form makes clear the physical interpretation of the integral term as a *friction* term. The nonlocality of the friction term is not a surprise : from a microscopic perspective, friction results from the multiple interactions between the system and the bath, hence damping needs some time to establish.

FDT.— Finally we have the relation between the correlator of the noise and the friction

$$\boxed{C(\tau) = k_B T \gamma(\tau) \quad \text{for } \tau > 0} \quad (\text{FDT}) \quad (1.62)$$

which relates the correlator of the noise to the damping (friction) function.

In the microscopic model, the damping term is an integral term. The relation between the damping and the strength of the noise results from the integration of the microscopic equation of motions, assuming equilibrium for the bath *only* (not for the particle, like in the phenomenological Langevin approach). We can compare the two approaches

- In the § 1 we have introduced two terms in the Langevin equation : the friction controlled by γ_0 and the noise controlled by the strength C . We have then assumed that the *particle* is at canonical equilibrium $P_{\text{sys}}(x, p) \propto \exp[-\frac{\beta}{2}mv^2]$. Comparing with the statistical properties of the solution of the Langevin equation, we have deduced that the two parameters of the model cannot be independent but must be related by $C = 2\gamma_0 k_B T$. To some extent, this relation was *assumed* for consistency.
- Here, we have only assumed that, being macroscopic, the *bath* is at thermal equilibrium $P_{\text{bath}}(x, p) \propto e^{-\beta H_{\text{bath}}}$. As a result of the integration of the conservative dynamics, we have deduced the relation $C = 2\gamma_0 k_B T$. A by-product is that if we study the statistic for the particle, one can show that it is described by canonical equilibrium (the particle reaches equilibrium because it interacts with the bath).

Quantum model : a very similar analysis can be performed within a quantum frame. Mainly, the correlator of the noise (i.e. of the initial bath variables) involves a different function and one is led to a "quantum Langevin equation" (cf. [13] or [47]).

Energetic considerations : We now study the energy of the system

$$\frac{d}{dt} H_{\text{sys}} = \dot{x} [m\ddot{x} - F(x)] = -v(t) \int_0^t d\tau \gamma(\tau) v(t - \tau) + v(t) \xi(t) \quad (1.63)$$

Clearly, the second term corresponds to the work of the Langevin force

$$\frac{dW}{dt} = v(t) \xi(t) \quad (1.64)$$

hence the first term should be interpreted as the heat received by the system

$$\frac{dQ}{dt} = -v(t) \int_0^t d\tau \gamma(\tau) v(t - \tau). \quad (1.65)$$

One can consider the model with $\gamma(\tau) = \gamma_0 \omega_D e^{-\omega_D \tau}$. Assuming $1/\omega_D \ll \tau = m/\gamma_0$, we expect that $v(t)$ is smooth on the scale $1/\omega_D$ so that we can treat $\xi(t)$ as a white noise. Hence $v(t) \simeq \frac{1}{m} \int_0^t dt' \xi(t') e^{-(t-t')/\tau}$. We can estimate the average work of the Langevin force

$$\frac{\langle \dot{W} \rangle}{dt} = \langle v(t) \xi(t) \rangle = \frac{2k_B T}{\tau} \theta_H(0) = \frac{k_B T}{\tau} \quad (1.66)$$

where $\langle v(t) \xi(t) \rangle$ was computed in Chapter 1, cf. Eq. (1.15). We have used that $\theta_H(0) = 1/2$ (this is consistent with a symmetric regularised δ function). The particle receives work $\langle \dot{W} \rangle > 0$ on average. The averaged heat is

$$\frac{\langle \dot{Q} \rangle}{dt} = - \int_0^t dt' \gamma(t-t') \langle v(t)v(t') \rangle \quad (1.67)$$

Because the correlator $\langle v(t)v(t') \rangle$ decays much slower than the damping function (we have assumed $\omega_D \tau \gg 1$) we can write $\gamma(t-t') \langle v(t)v(t') \rangle \simeq \gamma(t-t') \langle v(t)^2 \rangle$, hence

$$\frac{\langle \dot{Q} \rangle}{dt} \simeq - \langle v(t)^2 \rangle \int_0^\infty dt'' \gamma(t'') = - \frac{k_B T}{m} \gamma_0 = - \frac{k_B T}{\tau} \quad (1.68)$$

i.e. the bath receives some heat from the particle. As it should the total energy is conserved on average

$$\langle \dot{W} \rangle + \langle \dot{Q} \rangle = 0 \quad (1.69)$$

On average, the Langevin force gives some work to the particle and the bath receives the heat which is dissipated. Fluctuations allow that the opposite occurs.

The bath receives the entropy $dS_{\text{bath}} = -\dot{Q}/T$, thus the dissipation corresponds to the entropy production with rate

$$\boxed{\frac{d \langle S_{\text{bath}} \rangle}{dt} = + \frac{k_B}{\tau}} \quad (1.70)$$

Stochastic thermodynamics.— Here, I have applied some concepts of thermodynamics to a single particle. This type of question has attracted a lot of attention for ~ 25 years and is the subject of the field of “stochastic thermodynamics”. If you are interested you can have a look to the reviews [22, 9, 26, 44, 43, 24] or to the lectures of Bernard Derrida at coll ge de France (2015-2016), <https://www.college-de-france.fr/site/bernard-derrida/>

Additional exercises

 **Exercice 1.6 Comparison between time average and statistical average:** One considers the random “function” given by the sum of impulses $\xi(t) = \sum_{n=1}^N \kappa_n \delta(t-t_n)$ defined over the interval $[0, T]$, where

- the t_n ’s are independent and identically distributed (i.i.d) random times uniformly distributed over $[0, T]$ (i.e. one t_n has distribution $p(t_n) = 1/T$). We denote by $\lambda = N/T$ (for $N \rightarrow \infty$ and $T \rightarrow \infty$) the rate of occurrence of the random times.
- The κ_n ’s are i.i.d random variables with common distribution $w(\kappa)$ with finite $\langle \kappa_n^2 \rangle$.

a) Compute the time average of $\overline{\xi(t)}$, over the time interval $[0, T]$. Compare with the statistical average (over t_n ’s and κ_n ’s).

b) What is the condition on the random function $\xi(t)$ allowing to define a time averaged correlator $\tilde{C}(t-t') = \overline{\xi(t)\xi(t')^c} = \overline{\xi(t)\xi(t')} - \overline{\xi(t)} \overline{\xi(t')}$? Compare to $C(t-t') = \langle \xi(t)\xi(t') \rangle_c = \langle \xi(t)\xi(t') \rangle - \langle \xi(t) \rangle \langle \xi(t') \rangle$.

✎ **Exercice 1.7 Mean square displacement from the Langevin equation:** Our aim is to compute the mean square displacement $\langle x(t)^2 \rangle$ of a Brownian particle in a fluid described by (1.3,1.4). We assume that $x(0) = 0$ and that the particle is initially at equilibrium with the fluid.

- a) Recover the correlator $\langle v(t)\xi(t') \rangle$ and deduce $\langle x(t)\xi(t') \rangle$. What is $\langle x(t)\xi(t) \rangle$?
- b) Prove that $\frac{d^2}{dt^2}x(t)^2 + \frac{1}{\tau}\frac{d}{dt}x(t)^2 = 2v(t)^2 + \frac{2}{m}x(t)\xi(t)$.
- c) What is $\langle v(t)^2 \rangle$?
- d) Argue that $\frac{d}{dt}\langle x(t)^2 \rangle|_{t=0} = 0$ and deduce that

$$\langle x(t)^2 \rangle = 2D \left[t - \tau \left(1 - e^{-t/\tau} \right) \right] \quad (1.71)$$

Analyze carefully the limiting behaviours (interpret the $t \rightarrow 0$ behaviour) and plot the function.

✎ **Exercice 1.8 Dissipation of the energy in the Langevin equation:** We consider the Langevin equation $m\frac{d}{dt}v(t) = -\gamma v(t) + \xi(t)$ for the velocity of a particle in a fluid where $\xi(t)$ is a Gaussian white noise such that $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = 2\gamma k_B T \delta(t - t')$.

- a) Solve the Langevin equation. Deduce the correlator between the velocity and the Langevin force $\langle v(t)\xi(t') \rangle$.
- b) Consider now the kinetic energy $E_{\text{kin}}(t) = \frac{1}{2}mv(t)^2$. Using the Langevin equation, deduce $\frac{dE_{\text{kin}}}{dt}$ and interpret the two terms.
- c) Deduce a differential equation for $\langle E_{\text{kin}}(t) \rangle$.
Hint : you should need $\langle v(t)\xi(t) \rangle$ at coinciding times, which requires $\theta_H(0) = 1/2$ (Heaviside function at the origin).
- d) Solve the differential equation for $\langle E_{\text{kin}}(t) \rangle$ and discuss the result.

✎ **Exercice 1.9 From the Wiener process to the Ornstein-Uhlenbeck process:** We consider the Wiener process described by the equation $\frac{dW(u)}{du} = \eta(u)$, where $\eta(u)$ is a normalised Gaussian white noise.

- a) Consider $\varphi(u)$ a monotonous function. Argue that

$$\eta(\varphi(u)) \stackrel{(\text{law})}{=} \frac{1}{\sqrt{|\varphi'(u)|}} \eta(u) \quad (1.72)$$

- b) Deduce the stochastic differential equation for

$$x(t) = \frac{W(u)}{\sqrt{u}} \quad \text{with} \quad u = u_0 e^{2\gamma t} \quad (1.73)$$

2 Stochastic processes (2) : Markov processes and master equation

We have discussed above few simple stochastic processes (the Wiener process and the Ornstein-Uhlenbeck process). Let us now introduce some general ideas and concepts (and the vocabulary) allowing for a general analysis of stochastic processes.

2.1 Generalities : joint probabilities and conditional probabilities

The aim of the section is to introduce some useful tools and concepts needed to describe random processes. In general, the characterization of the distribution of a random process $X(t)$ defined on an interval $[0, T]$, i.e. a random function of time, should require to introduce a probability



Figure 2.1: *Marian von Smoluchowski (1872-1917) is considered as the father of the theory of stochastic processes.*

weight taking the form of a *functional* $P[x(t)] \sim \text{Proba}\{X(t) = x(t), \forall t \in [0, T]\}$, which is rather tricky to handle. However if we are interested in an average value, formally written as a functional integral $\langle X(t) \rangle = \int \mathcal{D}x P[x(t)] x(t)$, we do not need all the information about the process at all times, but only at time t . In this chapter and the following two chapters, we will restrict ourselves to simple questions which only require to know the probability that the process $X(t)$ takes some values at a *discrete* set of times (at t for $\langle X(t) \rangle$, or at t_1 and t_2 for $\langle X(t_1)X(t_2) \rangle$, etc). This leads to introduce the joint distribution of the values taken by the process at those times only.

Joint probability : In order to characterize the statistical properties of the random process, we introduce the joint probability or “the n -point function”

$$P_n(\underbrace{x_n, t_n; \dots; x_2, t_2; x_1, t_1}_{\substack{\longleftarrow \\ \text{time}}}) = \langle \delta(x_n - X(t_n)) \dots \delta(x_1 - X(t_1)) \rangle \quad (2.1)$$

corresponding to the probability (density) for the process to be equal to $x_1, \dots, x_n$ at times $t_1, \dots, t_n$. We can also write

$$\begin{aligned} & P_n(x_n, t_n; \dots; x_2, t_2; x_1, t_1) dx_1 \dots dx_n \\ &= \text{Proba}\{X(t_1) \in [x_1, x_1 + dx_1] \& \dots \& X(t_n) \in [x_n, x_n + dx_n]\} \end{aligned} \quad (2.2)$$

From the definition, it is clear that one integration relates the n -point to the $n-1$ -point function

$$\int dx_k P_n(x_n, t_n; \dots; x_{k+1}, t_{k+1}; \underline{x_k, t_k}; x_{k-1}, t_{k-1}; \dots; x_1, t_1) \quad (2.3)$$

$$= P_{n-1}(x_n, t_n; \dots; x_{k+1}, t_{k+1}; x_{k-1}, t_{k-1}; \dots; x_1, t_1) \quad (2.4)$$

The main interest of the joint distribution with n arguments is to allow us to express the average and correlations of the process. Eq. (2.1) makes clear that

$$\langle X(t_1) \rangle = \int dx_1 P_1(x_1, t_1) x_1 \quad (2.5)$$

$$\langle X(t_2)X(t_1) \rangle = \int dx_1 dx_2 P_2(x_2, t_2; x_1, t_1) x_2 x_1 \quad (2.6)$$

etc.

Conditional probability : Another important concept is the one of conditional probability corresponding to the probability for the process to pass through $x_1, \dots, x_n$ at successive times $t_1, \dots, t_n$, *given* that it has previously passed through $y_1, \dots, y_n$ at successive times $\tau_1, \dots, \tau_n$:

$$P_{n|m}(x_n, t_n; \dots; x_1, t_1 | y_m, \tau_m; \dots; y_1, \tau_1) = \frac{P_{n+m}(x_n, t_n; \dots; x_1, t_1; y_m, \tau_m; \dots; y_1, \tau_1)}{P_m(y_m, \tau_m; \dots; y_1, \tau_1)} \quad (2.7)$$

In particular, the conditional probability is useful if we are interested in conditional averages. Examples :

$$\langle X(t_1) | X(0) = x_0 \rangle = \int dx_1 x_1 P_{1|1}(x_1, t_1 | x_0, 0) \quad (2.8)$$

$$\langle X(t_2)X(t_1) | X(0) = x_0 \rangle = \int dx_1 dx_2 x_2 x_1 P_{2|1}(x_2, t_2; x_1, t_1 | x_0, 0), \quad (2.9)$$

etc, which involve averaging over trajectories which all start from the same point $X(0) = x_0$.

Propagator : Integrating the two point-point function $P_2(x, t; x_0, t_0) = P_{1|1}(x, t | x_0, t_0) P_1(x_0, t_0)$ over the initial coordinate, we get

$$P_1(x, t) = \int dx_0 P_{1|1}(x, t | x_0, t_0) P_1(x_0, t_0). \quad (2.10)$$

The equation shows that the conditional probability plays the role of a “*time evolution operator*” for the distribution $P_1(x, t)$: it “propagates” the stochastic evolution. For this reason, it is also called the “*propagator*”.

2.2 Markov processes

A very important class of random processes are *Markov processes*. A Markov process is a random process whose evolution after time t_0 only depends on its value at time t_0 , and *not* on its past history before t_0 . Mathematically, this property takes the form

$$P_{n|m}(x_n, t_n; \dots; x_1, t_1 | y_m, \tau_m; \underbrace{\dots; y_1, \tau_1}_{\text{past history}}) = P_{n|1}(x_n, t_n; \dots; x_1, t_1 | y_m, \tau_m) \quad (2.11)$$

which expresses that history prior to τ_m does not matter : only the *last* position y_m at time τ_m determines the evolution at times $t_1, \dots, t_n$.



Figure 2.2: Andreï Andreïevich Markov (1856-1922).

Let us examine the consequences of this fundamental assumption. Consider for example the three point function :

$$P_3(x_3, t_3; x_2, t_2; x_1, t_1) = P_{1|2}(x_3, t_3 | x_2, t_2; x_1, t_1) P_2(x_2, t_2; x_1, t_1) \quad (2.12)$$

$$= P_{1|2}(x_3, t_3 | x_2, t_2; x_1, t_1) P_{1|1}(x_2, t_2 | x_1, t_1) P_1(x_1, t_1) \quad (2.13)$$

$$\stackrel{\text{Markov}}{=} P_{1|1}(x_3, t_3 | x_2, t_2) P_{1|1}(x_2, t_2 | x_1, t_1) P_1(x_1, t_1) \quad (2.14)$$

We can generalize this to any joint distribution. We simplify the notation as $P_{1|1}(x, t | y, t_0) \equiv P(x, t | y, t_0)$ and $P_1(x, t) \equiv P(x, t)$ and we conclude that

A Markov process is fully characterized by $P(x, t | x_0, t_0)$ and $P(x, t)$ only.

In practice, the following property is useful

$$P_{n|1}(x_n, t_n; \dots; x_1, t_1 | x_0, t_0) = \prod_{k=1}^n P(x_k, t_k | x_{k-1}, t_{k-1}) \quad (2.15)$$

as it allows to compute any average $\langle X(t_n) \dots X(t_1) X(t_0) \rangle$ or $\langle X(t_n) \dots X(t_1) | X(t_0) = x_0 \rangle$.

Chapman-Kolmogorov equation : Start from the general property

$$\int dx_2 P_3(x_3, t_3; x_2, t_2; x_1, t_1) = P_2(x_3, t_3; x_1, t_1) . \quad (2.16)$$

For a Markov process, using (2.14), one gets the Chapman-Kolmogorov equation

$$\int dx_2 P(x_3, t_3 | x_2, t_2) P(x_2, t_2 | x_1, t_1) = P(x_3, t_3 | x_1, t_1) \quad (2.17)$$

The probability to go from x_1 to x_3 is the sum over x_2 of the probabilities conditioned to passed through x_2 .



Figure 2.3: *Andreï Nikolaïevitch Kolmogorov (1903-1987), well-known by physicists for his major contributions to the theory of dynamical systems and probability.*

Homogeneous Markov processes : In the following, in order to simplify the discussion, we will restrict ourselves to Markov processes such that the transition probability is invariant under time translation

$$P(x_2, t_2 | x_1, t_1) = P(x_2, t_2 - t_1 | x_1, 0) \quad (2.18)$$

Such random processes are denoted “homogeneous”. Since it only depends on the time difference, we will also write the propagator as $P_t(x | x_0)$.

a) A first example of Markov process

We can come back to the Langevin equation (1.3) for the velocity. (i) The equation for $v(t)$ is first order, hence the evolution at $t > 0$ is fully determined by $v(0) = v_0$. (ii) The Langevin force is a white noise (ξ uncorrelated in time), thus no time correlations are hidden in the noise. From (i) & (ii) we conclude that the process is Markovian.

Let us now discuss the related fundamental probabilities $P(v, t)$ and $P(v, t|v_0, 0)$, which can be easily obtained if we assume furthermore that the Langevin noise is **Gaussian**.⁶ We have obtained above the two first moments, cf. (1.8, 1.10), which are conditioned averages now denoted

$$\langle v(t) | v(0) = v_0 \rangle = v_0 e^{-t/\tau} \quad (2.19)$$

$$\text{Var}[v(t) | v(0) = v_0] = \frac{k_B T}{m} \left(1 - e^{-2t/\tau} \right) \quad (2.20)$$

The process $v(t)$ is a convolution of the Gaussian Langevin force $\xi(t)$, hence it is also Gaussian (a sum of Gaussian variables is also Gaussian). The knowledge of these two moments is sufficient to characterize the full distribution :⁷

$$P_t(v|v_0) = \sqrt{\frac{m}{2\pi k_B T (1 - e^{-2t/\tau})}} \exp \left\{ -\frac{m (v - v_0 e^{-t/\tau})^2}{2k_B T (1 - e^{-2t/\tau})} \right\}. \quad (2.21)$$

At large time, the conditional probability converges toward the equilibrium distribution

$$P_t(v|v_0) \xrightarrow{t \rightarrow \infty} P(v) = \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{m}{2k_B T} v^2} \quad (2.22)$$

the Maxwell distribution (or Gibbs distribution). Here, the “one point distribution” $P(v)$ is independent of the time due to the existence of a stationary state (this is not always the case). This is also why the conditional probability rapidly converges (exponentially fast) toward the equilibrium distribution. The process described by equation (1.3), or the conditional probability (2.21), is known as the **Ornstein-Uhlenbeck process**. It is the subject of a theorem by Doob (1942) : the only homogeneous Gaussian stationary random process is the Ornstein-Uhlenbeck process.⁸

✎ Exercice 2.1 : Express the correlator $\langle v(t)v(t') | v(0) = v_0 \rangle$ as a double integral involving the conditional probability $P_t(v|v_0)$. Check that one recovers Eq. (1.10) by using (2.21). Same question for the correlator $\langle v(t)v(t') \rangle_c$ corresponding to an initially random velocity v_0 .

b) Markovian or non-Markovian ? Few examples

Random walk.— After each step, a random walker forgets his past history (where he is coming from) : this is a Markov process.

Choosing clothes every morning.— Consider choosing clothes from a closet every morning without looking at the weather forecast, as a stochastic process. The selection depends on the weather of the previous day : there is a memory effect hence the process is non Markovian.

⁶The distribution of the noise is a Gaussian : $P[\xi] \propto \exp \left\{ -\frac{1}{2C} \int dt \xi(t)^2 \right\}$. One can deduce from this that $\langle \xi(t)\xi(t') \rangle = C \delta(t - t')$ [hint : discretize the time to check this].

⁷This argument was used in the historical paper : G. E. Uhlenbeck and L. S. Ornstein, “On the Theory of the Brownian Motion”, Phys. Rev. **36**(5), 823–841 (1930).

⁸The Ornstein-Uhlenbeck process is more frequently introduced as a model for a particle attached to a spring in the overdamped regime : $\dot{x}(t) = -k x(t) + \xi(t)$ (I forget the friction coefficient γ).

Velocity $v(t)$: a Markov process.— We have pointed out that the Langevin equation (1.3) describes a Markov process because

- (i) the differential equation is first order, hence the solution for $t > 0$ depends only on some initial value $v(0)$;
- (ii) the noise is δ -correlated, hence there is no memory hidden in the Langevin force.

Position $x(t) = \int_0^t dt' v(t')$: a non Markovian process.— The position $x(t)$ of the particle is a non Markovian process. There are two points of view leading to this conclusion.

(1) it obeys the stochastic differential equation (SDE) $m \ddot{x}(t) = -\gamma \dot{x}(t) + \xi(t)$, which is *second* order. The noise is still δ -correlated, however, the solution of the differential equation is determined both by the initial position $x(0)$ *and* the initial velocity $\dot{x}(0)$, which depends on the history before $t = 0$ (velocity indicates where the particle comes from). We break point (i) while point (ii) holds. The process $x(t)$ is non Markovian.

(2) We could also argue that $x(t)$ obeys a first order SDE, $\dot{x}(t) = v(t)$, whose solution is determined only by $x(0)$, hence point (i) holds. The velocity now plays the role of the noise. This “noise” is now characterized by a finite correlation time (memory time), $\langle v(t)v(t') \rangle = \frac{k_B T}{m} e^{-|t-t'|/\tau}$ (one says that the SDE for $x(t)$ involves a “colored noise”). Within this point of view we break point (ii), which leads to the same conclusion, as it should, that $x(t)$ is non Markovian.

A 2D Markov process : $(x(t), v(t))$.— $x(t)$ alone is a non Markovian process, however it can be considered as the first component of a two-dimensional Markov process $\vec{X}(t) = (x(t), v(t))$. Indeed, the system of differential equations for $(x(t), v(t))$ is *first* order and involves a δ -correlated noise. It presents the more general form of a *multidimensional Langevin equation*

$$\dot{X}_i = A_i(\vec{X}) + B_{ij} \Xi_j(t) \quad (2.23)$$

where here $\vec{A} = (v, F(x) - \gamma v)$ is the drift (we have added a conservative force). The noise $\vec{\Xi}(t) = (\xi_x(t), \xi(t))$ is uncorrelated in time. The matrix is $B_{xx} = B_{xv} = B_{vx} = 0$ and $B_{vv} = 1$. Hence it is a 2D Markovian process. We will see that the joint distribution $P_t(x, v)$ obeys the “Kramers equation” $\partial_t P_t = \left[-\partial_x v - \frac{1}{m} \partial_v (F(x) - \gamma v) + \frac{1}{m^2} \gamma k_B T \partial_v^2 \right] P_t$.

Conclusion : The analysis of this example shows that the identification of a Markov process is sometimes a question of perspective, and also illustrates that Markov processes are elementary building blocks. Many non Markovian processes can be constructed from Markov processes.

2.3 Master equation

a) Master equation for continuous processes

Let us start with the case of continuous processes, which is more general. As we have seen, the time evolution of the distribution $P(x, t)$ of a Markov process can be represented in terms of the conditional probability, Eq. (2.10). However, this equation and (2.17) are not of great help to determine the two fundamental functions $P(x, t)$ and $P(x, t|x_0, t_0)$. The distribution is more conveniently obtained by solving an evolution equation for an infinitesimal time step : such an evolution equation can be related to the above integral equation by considering the evolution during an infinitesimal time $\delta t \rightarrow 0$. In this case we expect

$$P(x, t + \delta t|x_0, t) \simeq \delta(x - x_0) + \delta t W_t(x|x_0) + \mathcal{O}(\delta t^2) \quad (2.24)$$

The linear correction follows from the Markov assumption : at short time, the transition probability is linear with time and involves a transition rate $W_t(x|x_0)dx$ for performing the transition from x_0 to $[x, x + dx]$, which is the elementary information defining the process.

If we restrict ourselves to homogeneous processes (time translation invariant) we have

$$W_t(x|x_0) \rightarrow W(x|x_0) \quad (\text{homogeneous process}) \quad (2.25)$$

independent of time. For homogeneous processes, we find the differential equation (in time)

$$\boxed{\frac{\partial P(x, t)}{\partial t} = \int dx' W(x|x') P(x', t)} \quad (2.26)$$

This is the general form of the Master equation for a continuous process. The fact that the differential equation in time is *first order* means that the future evolution of the probability density for times $t > 0$ depends only on the distribution $P(x, 0)$ and not on the past history (Markov property). This is the (*a posteriori*) justification for the expansion (2.24). Note that the conservation of probability requires that

$$\int dx_f W(x_f|x_i) = 0 \quad \Rightarrow \quad \frac{\partial}{\partial t} \int dx P(x, t) = 0 \quad \forall t \quad (2.27)$$

so that probability $\int dx P(x, t) = 1$ is conserved. This condition follows from the normalization condition of the conditional probability, in the expansion (2.24).

Note that, being related to a transition rate, the transition kernel should be positive $W(x|x_0) \geq 0$ for $x \neq x_0$. As a result, it must contain a singular negative contribution of the form $-\lambda(x) \delta(x - x_0)$ in order to fulfill the condition (2.27), where $\lambda(x)$ is therefore the probability rate for leaving point x (see examples below in this chapter).

Obviously, the conditional probability obeys the same equation

$$\frac{\partial P_t(x|x_0)}{\partial t} = \int dx' W(x|x') P_t(x'|x_0) \quad \text{for initial condition } P_0(x|x_0) = \delta(x - x_0). \quad (2.28)$$

In a specific problem, the transition “rates” $W(x|x_0)$ encode the elementary information about the stochastic process. The aim is to solve the master equation (2.26), or (2.28), given the integral kernel $W(x|x_0)$.

In the most general case, a Markov process can combine

- a diffusion : in this case the integral kernel $W(x|x')$ is replaced by a second order differential operator
- jumps : leading to an integral term in the master equation, like in (2.26).

Below, we will give concrete examples. We will come back on this discussion in chapter 4 (§ ??).

b) Master equation for discrete processes

For simplicity, let us first consider a random process which takes discrete values $X(t) \in \{x_1, \dots, x_N\}$ and denote $P_n(t) = \text{Proba}\{X(t) = x_n\}$ (in certain situations the total number of states can be infinite, $N \rightarrow \infty$). The Markovian nature of the process implies that what happens on $[t, t + \delta t]$ is independent to what has occurred on $[0, t]$. Hence the probability $P_n(t + \delta t)$ can be expressed in terms of products of probabilities, one characterizing the evolution up to time t , i.e. the probabilities $P_m(t)$ ’s, and the probabilities for transitions between states during time interval

δt . We denote by $W_{n,m} \geq 0$ the rate for a transition from m to n , with $n \neq m$, meaning that the probability to make a transition $m \rightarrow n$ in the time interval $[t, t + \delta t]$ is $W_{n,m} \delta t$. Therefore

$$P_n(t + \delta t) \simeq \underbrace{\left(1 - \sum_{m(\neq n)} W_{m,n} \delta t\right)}_{\text{proba to leave } n} \overset{\text{Markov}}{\times} P_n(t) + \sum_{m(\neq n)} \underbrace{W_{n,m} \delta t}_{\text{proba for } m \rightarrow n} \overset{\text{Markov}}{\times} P_m(t) \quad (2.29)$$

Letting $\delta t \rightarrow 0$ we obtain the *first order* differential equation

$$\frac{d}{dt} P_n(t) = \sum_{m(\neq n)} [W_{n,m} P_m(t) - W_{m,n} P_n(t)] \quad (2.30)$$

(we can also replace $\sum_{m(\neq n)} \rightarrow \sum_m$). Eq. (2.30) is the *master equation* for a discrete process.

The previous equation does not depend on $W_{n,n}$, which was not defined so far, however it is convenient to introduce $W_{n,n} \stackrel{\text{def}}{=} -\sum_{m(\neq n)} W_{m,n} < 0$, which is the probability rate for leaving state n , up to the sign. We deduce the form

$$\boxed{\frac{dP_n(t)}{dt} = \sum_m W_{n,m} P_m(t)} \quad (2.31)$$

The coefficients $W_{n,m}$ form a $N \times N$ matrix W , which satisfies the condition

$$\boxed{\sum_n W_{n,m} = 0} \quad (2.32)$$

ensuring the conservation of probability $\sum_n P_n(t) = 1 \forall t$. Note that this condition is related to an asymmetry of the matrix ($W_{n,m} \neq W_{m,n}$ in general) : the sum of all elements of a column is zero, however summing elements along a line is non zero in general.

We now discuss several examples of Markov processes (discrete or continuous).

c) The random telegraph process ($N = 2$)

Let us first consider the *most simple Markov process* $X(t)$, taking only two possible values X_1 or X_2 (this is a “two level system” for stochastic processes). A typical realization of the process is represented on Fig. 2.4.

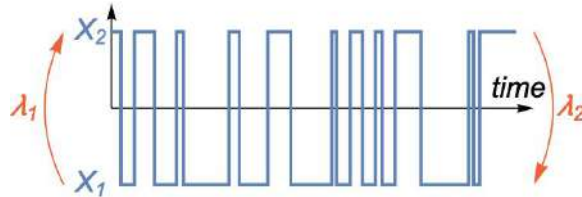


Figure 2.4: *Random telegraph process with $X_1 = -1$ and $X_2 = +1$.*

There are many different realizations of the random telegraph process (according to wikipedia, it is also known as the Kac process, from the name of the famous Mathematician Mark Kac, or random telegraph noise, or burst noise, or popcorn noise). It may occur any time we deal with a two state system with a random switching mechanism. For example, since the 1990s, there has been a huge interest in micrometric structures known as “*quantum dots*” (QD), which can trap a tunable number of electrons (see for example the review [52]). Such a device is shown in

Fig. 2.5 (see also Fig. 2.6 p. 25) : a small island is patterned in a two dimensional electron gas at an interface GaAs/GaAlAs. Due to Coulomb interaction, the addition of an extra electron in the island has a high energy cost, however, playing with some external gate voltages, the QD can be prepared in a configuration such that the two states, the one with n electrons and the one with $n + 1$ electrons, are degenerate. Then, the QD is switching between the two states. Moreover, the figure 2.5 shows that the switching occurs with a rather low frequency, on a millisecond scale, which is a macroscopic time for the electrons.

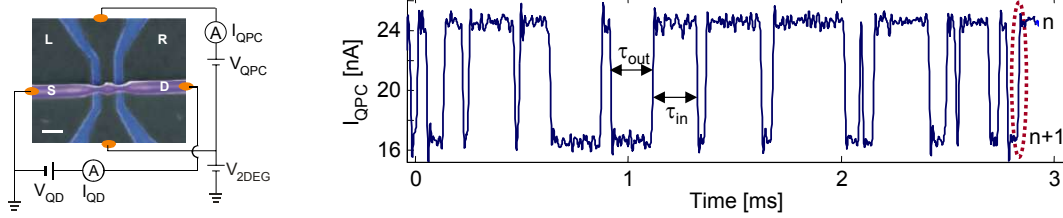


Figure 2.5: A physical realization of the random telegraph process : the charge inside a quantum dot is measured as a function of the time with a neighbouring probe. The current I_{QPC} through the probe is proportional to the number of electrons inside the central island, which fluctuates by one unit (one electron). From : S. Gustavsson, I. Shorubalko, R. Leturcq, S. Schön, and K. Ensslin, “Measuring current by counting electrons in a nanowire quantum dot”, Appl. Phys. Lett. **92**, 152101 (2008)

We introduce two probabilities $P_i(t) = \text{Proba}\{X(t) = X_i\}$ with $i \in \{1, 2\}$ and in general, there are two different transition rates $\lambda_1 = \text{rate}(X_1 \rightarrow X_2)$ and $\lambda_2 = \text{rate}(X_2 \rightarrow X_1)$. The matrix W is 2×2

$$W = \begin{pmatrix} -\lambda_1 & \lambda_2 \\ \lambda_1 & -\lambda_2 \end{pmatrix}. \quad (2.33)$$

The two probabilities obey the set of coupled differential equations

$$\frac{dP_1(t)}{dt} = -\lambda_1 P_1(t) + \lambda_2 P_2(t) \quad (2.34)$$

$$\frac{dP_2(t)}{dt} = +\lambda_1 P_1(t) - \lambda_2 P_2(t) \quad (2.35)$$

in each equation, one term populates the state and the other empties the state.

In the following exercise, we solve this system of equations.

✎ Exercise 2.2 Random telegraph process :

a) Derive the set of differential equations for $P_1(t)$ and $P_2(t)$ from the definitions of the rates. Deduce a matricial form $\frac{d}{dt}P(t) = W P(t)$, where $P = (P_1 \ P_2)^T$ is the column vector (T denotes transposition).

b) Find the stationary solution, denoted by P_i^* , and give the general solution of the master equation.

c) Determine the conditional probability $P_t(i|j)$. Compute $\sum_j P_t(n|j) P_j^*$ and interpret. Discuss detailed balance.

d) Express $\langle X(t) \rangle$ and $\langle X(t)X(t') \rangle$ in the stationary regime. For simplicity, choose $X_1 = 0$ and $X_2 = 1$. Show that the correlator $C(t - t') = \langle X(t)X(t') \rangle - \langle X(t) \rangle \langle X(t') \rangle$ is

$$C(t - t') = \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2} e^{-(\lambda_1 + \lambda_2)|t - t'|}. \quad (2.36)$$

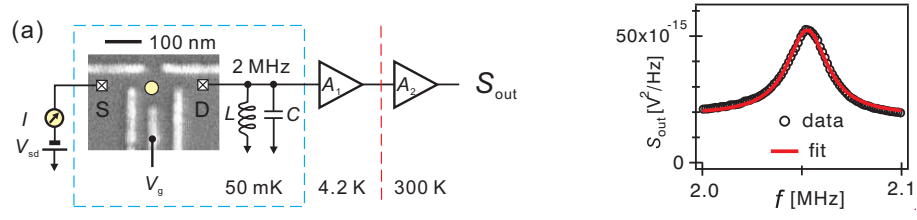


Figure 2.6: Left : *The microstructure studied*. Right : *Power spectrum*. From : Y. Okazaki, S. Sasaki and K. Muraki, *Shot noise spectroscopy on a semiconductor quantum dot in the elastic and inelastic cotunneling regimes*, Phys. Rev. B **87**, 041302(R) (2013).

e) **Electric noise power spectrum** : in the experiments, the fluctuations of the charge in the QD can be characterized by measuring the power spectrum $S(\omega)$ (see the Wiener-Khinchin theorem discussed below, § 2.7, and the relation with the correlation function $C(t)$). Deduce $S(\omega)$ for the random telegraph process. The measurement is reported in another article, Fig. 2.6. Compare your result with the data.

d) Birth and death processes

A subclass of discrete processes are “birth and death processes”. They correspond to the case where the transition matrix is tridiagonal, i.e. allows only transitions between nearest neighbour states. The master equation has the form

$$\frac{dP_n(t)}{dt} = d_{n+1} P_{n+1}(t) + b_{n-1} P_{n-1}(t) - (d_n + b_n) P_n(t) \quad (2.37)$$

where $d_n > 0$ and $b_n > 0$ are death and birth rates, respectively (cf. Fig. 2.7). A simple example is the Poisson process studied below in great detail (in this case the death rates are zero, $d_n = 0$ and birth rates are all equal, $b_n = \lambda \forall n$). Birth and death processes have some applications in queueing theory (Exercise ??), in biology, etc. Another example will be studied in Exercise 2.17.

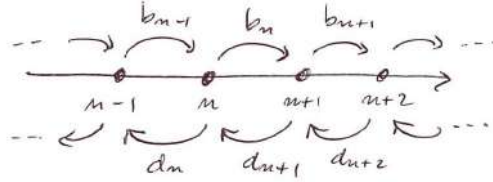


Figure 2.7: *Transition rates in the birth and death process*.

e) The Poisson process (statistics of uncorrelated events)

The Poisson process takes integer values $\mathcal{N}(t) \in \mathbb{N}$ and starts at initial time from $\mathcal{N}(0) = 0$. With probability rate λ , the process is incremented by one, i.e. during an interval of time of duration dt , the process increases by one with probability λdt . We denote $P_n(t) = \text{Proba}\{\mathcal{N}(t) = n\}$ its probability.

The Poisson process (PP) counts the occurrences of *independent events*. For instance the number of drops of rain falling on the floor during a time interval t . Or the number of disintegrations in a radioactive material during a time t .

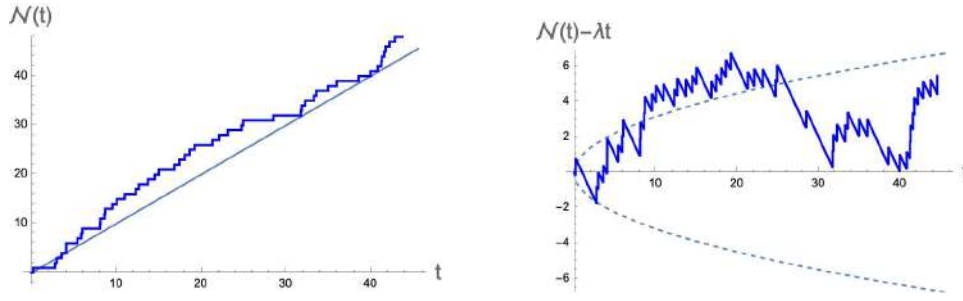


Figure 2.8: A typical realization of the Poisson process : $\mathcal{N}(t)$ as a function of t . On the right plot, the average is removed. Typical fluctuations grow as $\sim \sqrt{t}$.

The master equation of the PP is

$$\begin{cases} \frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) - \lambda P_n(t) & \text{for } n > 0 \\ \frac{dP_0(t)}{dt} = -\lambda P_0(t) \end{cases} \quad (2.38)$$

In other terms, the rate “matrix” has elements on the diagonal and just below the diagonal $W_{n,m} = \lambda (-\delta_{m,n} + \delta_{m,n-1})$.

✎ Exercice 2.3 Master equation for the PP :

- Show that the master equation for the Poisson process is (2.38).
- Introduce the generating function $G(z; t) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} z^n P_n(t)$. What is the value of $G(z; 0)$? Get a differential equation for $G(z; t)$ and solve it.
- Deduce that

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (2.39)$$

- Determine the cumulants $\langle \mathcal{N}(t)^k \rangle_c$ of the Poisson process.
- Give the distribution $q(\tau)$ of the time separating two successive events [indication : relate $q(\tau)$ and $P_0(t)$].

Note that (2.39) was obtained by choosing the initial condition $\mathcal{N}(0) = 0$, hence, (2.39) corresponds to the conditional probability $P_t(n|0)$. Using translation invariance of the Poisson process, we deduce the general expression of the conditional probability

$$P_t(n|m) = \begin{cases} \frac{(\lambda t)^{n-m}}{(n-m)!} e^{-\lambda t} & \text{for } n \geq m \\ 0 & \text{for } n < m \end{cases} \quad (2.40)$$

The Poisson process is both *translation invariant* in time ($W_{n,m}$ does not depend on time) and “space”

$$W_{n,m} = \lambda (-\delta_{m-n,0} + \delta_{n-m,1}) = \text{function}(n-m). \quad (2.41)$$

✎ Exercice 2.4 Conditional probability for the PP : Consider the Poisson process with $\mathcal{N}(0) = 0$. Check that $\sum_m P_t(n|m) P_{t'}(m|n_0) = P_{t+t'}(n|n_0)$.

✎ Exercice 2.5 Two-point correlator of the PP :

- For $\mathcal{N}(0) = 0$, express $\langle \mathcal{N}(t) \rangle$ in terms of the probability. Compute it.
- Express $\langle \mathcal{N}(t)\mathcal{N}(t') \rangle$ as a double sum and compute it when $t' < t$. Deduce the correlator $\langle \mathcal{N}(t)\mathcal{N}(t') \rangle_c = \langle \mathcal{N}(t)\mathcal{N}(t') \rangle - \langle \mathcal{N}(t) \rangle \langle \mathcal{N}(t') \rangle$. Compare with the Wiener process.
- Deduce $\left\langle [\mathcal{N}(t) - \mathcal{N}(t')]^2 \right\rangle_c$. Comment on the limit $t \rightarrow t'$.

f) The compound Poisson process (or “Poisson jump process”)

A natural generalization of the Poisson process is the *compound Poisson process* (CPP) ⁹ : we consider now that the process $X(t)$ makes random jumps

$$X(t_n^+) = X(t_n^-) + \eta_n \quad (2.42)$$

where the η_n 's are i.i.d., distributed according to a distribution $w(\eta)$. As for the Poisson process, the jumps occur at random times t_n with rate λ (for the PP, we have $\mathcal{N}(t_n^+) = \mathcal{N}(t_n^-) + 1$). After a time t , the number of jumps $\mathcal{N}(t)$ is random (it is a PP), hence we can represent the CPP in terms of a PP as

$$X(t) = \sum_{n=0}^{\mathcal{N}(t)} \eta_n \quad \text{with } X(0) = \eta_0 = 0. \quad (2.43)$$

Numerics : this form shows that it is very simple to simulate the CPP with a computer. One can generate two independent sequences of random numbers : the random time intervals $\tau_n = t_n - t_{n-1} > 0$ with exponential distribution $q(\tau) = \lambda e^{-\lambda\tau}$ (cf. Exercise 2.3) and the random amplitudes η_n 's with distribution $w(\eta)$.

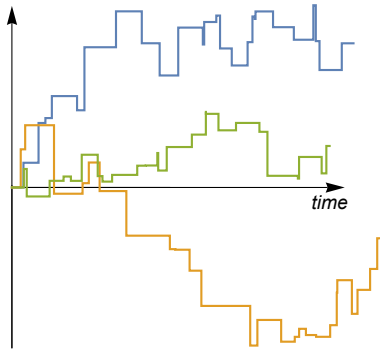


Figure 2.9: Three realizations of the Compound Poisson Process $X(t)$ for Gaussian jumps.

Master equation : The master equation of the CPP is

$$\frac{\partial P(x, t)}{\partial t} = \lambda \int d\eta w(\eta) [P(x - \eta, t) - P(x, t)] \quad (2.44)$$

i.e. of the form (2.26) for

$$W(x|x_0) = \lambda [w(x - x_0) - \delta(x - x_0)] . \quad (2.45)$$

Note that here, the transition kernel is (time and space) translation invariant, as for the PP.

✎ Exercice 2.6 Master equation for the CPP :

a) Show that the master equation for the CPP is (2.44).

b) *Continuum limit.*— Study the limit $\lambda \rightarrow \infty$ with $\eta_n \rightarrow 0$ such that $a = \lambda \langle \eta_n \rangle$ and $b = \lambda \langle \eta_n^2 \rangle$ are kept finite (argue that, in this limit, $\lambda \langle \eta_n^k \rangle \rightarrow 0$ for $k > 2$).

c) Introducing the Fourier transforms $\hat{P}(k, t) = \int dx e^{-ikx} P(x; t)$ and $\hat{w}(k) = \int d\eta e^{-ik\eta} w(\eta)$, show that the solution is

$$P(x, t) = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} e^{\lambda t [\hat{w}(k) - 1] + ikx} . \quad (2.46)$$

⁹in French: “processus de Poisson composé”.

Discuss the continuum limit.

d) When $\langle \eta_n^2 \rangle = \infty$, the process belongs to the class of **Lévy flights**. For example, if $w(\eta) \simeq c/\eta^2$ for $\eta \rightarrow \pm\infty$ we have $\hat{w}(k) \simeq 1 - \tilde{c}|k|$ for $k \rightarrow 0$. Deduce $P(x, t)$ over large scales. Discuss also the more general case where $\hat{w}(k) \simeq 1 - \tilde{c}|k|^\mu$ for $k \rightarrow 0$, with $\mu \in]0, 2[$.

2.4 Classification of Markov processes

We now discuss the different scenarii which might occur when studying the master equation (2.31).

Notations for vectors and matrices : In order to use unified notations, “vectors” in $\mathbb{R}^d$ will be column vectors ($d \times 1$ matrices) :

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \quad \text{and} \quad X^T = (x_1 \quad \cdots \quad x_d) \quad (2.47)$$

is a line vector (a $1 \times d$ matrix). We denote by $(\cdot)^T$ the transposition of a matrix. As a result, if X and Y are two (column) vectors, the scalar product is $X^T Y$, while $X Y^T$ represents a matrix (a projector). Explicitly (for $d = 2$) :

$$(x_1 \quad x_2) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 y_1 + x_2 y_2 \quad \text{and} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (y_1 \quad y_2) = \begin{pmatrix} x_1 y_1 & x_1 y_2 \\ x_2 y_1 & x_2 y_2 \end{pmatrix}. \quad (2.48)$$

a) Stationary state

The equation (2.32), which guarantees the conservation of the total probability, has a clear linear algebra interpretation : it expresses that the (column) vector with unit components

$$L^{(0)} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad (2.49)$$

is a *left* eigenvector with eigenvalue $\lambda_0 = 0$:

$$L^{(0)T} W = 0 \quad \text{or} \quad W^T L^{(0)} = 0. \quad (2.50)$$

Because the $N \times N$ matrix W is non-symmetric in general, the eigenvalue is associated with a pair of left and right eigenvectors which differ. Let us denote $R^{(0)}$ the corresponding right eigenvector, i.e. $W R^{(0)} = 0$, with components $R^{(0)} = (P_1^*, \dots, P_N^*)^T$. Using (2.32), it is convenient for future discussion to rewrite the equation under the form

$$\sum_{m (\neq n)} [W_{nm} P_m^* - W_{mn} P_n^*] = 0 \quad \forall n. \quad (2.51)$$

For matrices W with $W_{n,m} \geq 0$ and constraint $\sum_n W_{n,m} = 0$, there exists a theorem in Mathematics (the Perron-Fröbenius theorem discussed below) which guarantees that (i) the eigenvalue $\lambda_0 = 0$ is the largest one, (ii) it is non degenerate, (iii) it is associated with a right eigenvector with positive components $R^{(0)} = (P_1^*, \dots, P_N^*)^T$, which can therefore be interpreted as the stationary distribution.¹⁰

¹⁰The Perron-Fröbenius theorem applies to “stochastic matrices”, i.e. matrices with positive entries and such that $\sum_n M_{nm} = 1 \forall m$. Here, the matrix W is a stochastic matrix shifted by the unity matrix : $M = \mathbf{1}_N + \mu W$ where $\mathbf{1}_N$ is the identity matrix and μ is a real number chosen such that all diagonal elements of M are positive.

b) Three types of stochastic processes

The nature of the different stochastic processes can be related to the properties of the stationary state. Depending on the matrix W , which *defines* the stochastic process, one encounters one of the three following situations.

- (i) **Equilibrium.**— Often, the existence of a stationary solution is ensured by a condition *stronger* than (2.51), called the **detailed balance condition**

$$\boxed{W_{nm} P_m^* = W_{mn} P_n^* \quad \forall (n, m)} \quad (\text{detailed balance} \equiv \text{equilibrium}) \quad (2.52)$$

i.e. ratio of occupations coincides with the ratio of transition rates

$$\frac{W_{nm}}{W_{mn}} = \frac{P_n^*}{P_m^*}. \quad (2.53)$$

If detailed balance is fulfilled, one says that the distribution P_n^* describes an *equilibrium state*. This is a *probabilistic definition of equilibrium*.

The detailed balance condition defines a first class of stochastic processes, those eventually reaching an equilibrium. From the point of view of the $N \times N$ matrix W , there are $\sim N^2$ conditions [there are $\sim N^2$ equations (2.52)], which is rather restrictive.

- (ii) **Non-equilibrium stationary state (NESS).**— If the detailed balance condition (2.52) is *not* fulfilled (at least for some couples (n, m)) but the condition

$$\sum_{m(\neq n)} \underbrace{[W_{nm} P_m^* - W_{mn} P_n^*]}_{\neq 0} = 0 \quad (\text{stationarity}) \quad (2.54)$$

holds ($\forall n$) for a normalisable distribution, one says that the stationary state is a *non-equilibrium steady state*.

This property defines a second class of stochastic processes, those reaching a NESS at large time. Note that the existence of the NESS involves only a set of $\sim N$ conditions ¹¹

- (iii) **Transient process.**— When the size N of the matrix W is finite, the existence of a stationary state follows from the definition of the matrix (i.e. from the Perron-Fröbenius theorem). However, when $N \rightarrow \infty$, the eigenvector $R^{(0)} = (\dots, P_n^*, \dots)^T$ can be normalisable [cases (i) or (ii)], or not. In this latter case, when $\sum_n P_n^* = \infty$, the system never reaches a stationary state and the process is said to be *transient*. This defines the last family of stochastic processes.

Current interpretation : Detailed balance condition has a simple interpretation in terms of *probability currents* : for $n \neq m$, the matrix element W_{nm} being a rate, $J^*(m \rightarrow n) = W_{nm} P_m^*$ is the probability current from state m to state n . In the stationary state, currents between states should compensate globally, if probability is conserved : (2.54) expresses the vanishing of the sum of all currents arriving and leaving state n .

Detailed balance (2.52) means that probability currents equilibrate, $J^(m \rightarrow n) = J^*(n \rightarrow m)$, for all pairs of states.* There is no circulating probability current between more than two states (Fig. 2.10).

On the other hand, if detailed balance does not hold, the NESS is characterized by the existence of circulating probability currents around loops.

¹¹The definition of the matrix W involves N constraints $\sum_n W_{nm} = 0$ for $m = 1, \dots, N$; those conditions are directly related to the N equations (2.54).

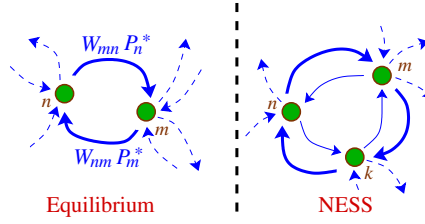


Figure 2.10: **Equilibrium state versus NESS.**— Denote by $J^*(m \rightarrow n) = W_{nm} P_m^*$ the probability current between two states, in the stationary state. At equilibrium, currents equilibrate for all pairs of states $J^*(n \rightarrow m) = J^*(m \rightarrow n)$. In the NESS, there exist loop(s) in which probability current(s) circulate : here, due to an imbalance $J^*(n \rightarrow m) > J^*(m \rightarrow n)$, there is a net circulation $n \rightarrow m \rightarrow k \rightarrow n$.

c) Simple examples :

Molecular vapour at thermal equilibrium : consider a vapour of molecules at thermal equilibrium. Each molecule has energy levels ε_n , expected to be occupied according to canonical weights $P_n^* \propto e^{-\beta \varepsilon_n}$. The molecule in an excited state falls in a state with lower energy by emission. Equilibrium and detailed balance imply that the absorption and emission rates between two levels fulfill the relation

$$\frac{\Gamma_{n \leftarrow m}}{\Gamma_{m \leftarrow n}} = \frac{P_n^*}{P_m^*} = e^{-\beta(\varepsilon_n - \varepsilon_m)} \quad (2.55)$$

i.e. emission is more probable than absorption. This observation (Gibbs equilibrium is compatible with an imbalance between emission and absorption) is the key point of Einstein first theory of spontaneous emission (at the origin of the difference between emission and absorption).¹²

Equilibrium statistical physics : more generally, all situations described by equilibrium statistical physics and ensemble theory [50, 51] rely on the existence of transition rates which satisfy detailed balance (dynamical aspects, like transitions between microstates, are usually not deeply described in introductory courses on statistical physics ; however transitions are fundamental to ensure that the system reaches an equilibrium state).

Other simple examples : Exercise 2.10 discusses an extremely simple model of biased random walk in a ring, which illustrates the three situations. Another illustration is given in Exercise 2.17, where a richer model of random walk on $\mathbb{Z}$ for non uniform transition rates modeling the presence of a potential is studied,.

2.5 Spectral decomposition - Relaxation

In the previous paragraph, we have discussed the solution of the master equation (2.31) in the $t \rightarrow \infty$ limit. What can we say about the solution for finite t ? The master equation is a first order differential equation of the form $\frac{d}{dt} \text{Vector}(t) = \text{Matrix} \cdot \text{Vector}(t)$. Obviously such equation is conveniently solved by using the information on the spectrum of eigenvalues and eigenvectors of the matrix. The main difficulty is here related to the fact that the matrix is non symmetric in general.

¹²Albert Einstein, “Zur Quantentheorie der Strahlung”, Physikalische Zeitschrift **18**, 121–128 (1917).

Einstein’s article is reproduced (and translated in french) in : A. Einstein, *Œuvres choisies. 1. Quanta*, Seuil (1989), textes choisis et présentés par F. Balibar, O. Darrigol & B. Jech.

🔪 **Exercise 2.7 Warm up : right and left eigenvectors of a non-symmetric 2×2 matrix :** We consider the real non-symmetric matrix

$$M = \begin{pmatrix} 0 & e^{-h} \\ e^h & 0 \end{pmatrix} \quad (2.56)$$

Show that it is diagonalisable. Find the right and left eigenvectors and discuss the normalisation.

In the master equation, the matrix W is not symmetric in general, $W^T \neq W$. We have seen above that its eigenvalue $\lambda_0 = 0$ is associated with a couple of left and right eigenvectors $L^{(0)}$ and $R^{(0)}$. If W is diagonalisable, its eigenvalues are associated with a biorthogonal set of left and right eigenvectors $L^{(n)}$ and $R^{(n)}$. Let us write for convenience the equation for the eigenvalue as

$$WR^{(n)} = -\lambda_n R^{(n)} \quad \text{and} \quad W^T L^{(n)} = -\lambda_n L^{(n)} \quad (2.57)$$

where $\lambda_0 = 0 < \lambda_1 < \lambda_2 < \dots$ (the Perron-Fröbenius theorem discussed below guarantees this property).

Let us prove now that two left and right eigenvectors for different eigenvalues are orthogonal (but not the right eigenvectors among themselves, nor the left eigenvectors in general). We write the two equations (2.57) for two eigenvalues, $WR^{(n)} = -\lambda_n R^{(n)}$ and $L^{(m)T}W = -\lambda_m L^{(m)T}$ and multiply by the left and right eigenvectors, respectively,

$$\begin{cases} L^{(m)T}WR^{(n)} = -\lambda_n L^{(m)T}R^{(n)} \\ L^{(m)T}WR^{(n)} = -\lambda_m L^{(m)T}R^{(n)} \end{cases} \xrightarrow{\text{difference}} 0 = (\lambda_n - \lambda_m) L^{(m)T}R^{(n)}.$$

Hence, $\lambda_n \neq \lambda_m$ implies orthogonality $L^{(m)T}R^{(n)} = 0$. The choice of normalization is arbitrary. A convenient one is

$$L^{(n)T}R^{(m)} = \delta_{n,m}, \quad (2.58)$$

which ensures that the identity can be written

$$\sum_n R^{(n)} L^{(n)T} = \mathbf{1}_N \quad (2.59)$$

where $R^{(n)} L^{(n)T}$ is the projector on the eigenvector $R^{(n)}$. More importantly for the analysis, this leads to the **spectral representation** of the matrix

$$W = - \sum_n \lambda_n R^{(n)} L^{(n)T}. \quad (2.60)$$

This is very useful in order to solve the master equation. Denoting by $P(0) = (P_1(0), \dots, P_N(0))^T$ the initial state, we can write the solution of (2.31) in a matrix form and then get the probabilities :

$$P(t) = e^{tW} P(0) \quad \text{i.e.} \quad P_n(t) = \sum_j e^{-\lambda_j t} R_n^{(j)} \underbrace{L^{(j)T} P(0)}_{c_j \stackrel{\text{def}}{=}} = \sum_j c_j e^{-\lambda_j t} R_n^{(j)}. \quad (2.61)$$

c_j 's are the coefficients of the initial vector in the basis of eigenvectors $P(0) = \sum_j c_j R^{(j)}$. Note that $c_0 = L^{(0)T} P(0) = \sum_j P_j(0) = 1$ carries all the normalisation (orthonormalisation indeed implies that $L^{(0)T} R^{(j)} = \sum_n R_n^{(j)} = 0$ for $j \neq 0$).

Relaxation towards the stationary state.— Now let us discuss the large time behaviour. Using that $\lambda_0 = 0 < \lambda_1 < \lambda_2 < \dots$, the large time behaviour takes the form

$$P_n(t) \underset{t \rightarrow \infty}{\simeq} \underbrace{P_n^*}_{\equiv R_n^{(0)}} + \underbrace{c_1 e^{-\lambda_1 t} R_n^{(1)}}_{\xrightarrow{t \rightarrow \infty} 0} \quad (2.62)$$

where we have used $c_0 = 1$ (normalisation). This shows that $1/\tau_{\text{relax}} = \lambda_1$ is the relaxation rate towards the stationary state. Relaxation is usually exponentially fast, unless the gap in the spectrum vanishes and the spectrum is continuous.

Remark 1 : apart $\lambda_0 = 0$, the eigenvalues are not real in general. However, if some eigenvalues are complex, eigenvalues should come in conjugate pairs since M is a real matrix (and the same for eigenvectors). Thus, in the general case, the rate of relaxation towards stationary state is

$$\boxed{\frac{1}{\tau_{\text{relax}}} = \text{Re}(\lambda_1)} \quad (2.63)$$

and the ranking of e.v. is

$$\lambda_0 < \text{Re}(\lambda_1) < \text{Re}(\lambda_2) < \dots \quad (2.64)$$

Remark 2 : a non symmetric is not always diagonalisable. However, W can always be decomposed in terms of Jordan blocks associated to each degenerate eigenvalue. An example will be discussed in Exercise ??.

Conditional probability.— If we choose the initial condition $P_n(0) = \delta_{nm}$, the coefficients are $c_j = L_m^{(j)}$. This solution of the master equation is the conditional probability $P_n(t) \equiv P_t(n|m) = \sum_j e^{-\lambda_j t} R_n^{(j)} L_m^{(j)}$, which corresponds to the spectral decomposition of e^{tW} . This shows that the conditional probability can be written as

$$\boxed{P_t(n|m) = (e^{tW})_{nm}} \quad (2.65)$$

✎ **Exercise 2.8 :** Check that

$$\sum_m P_t(n|m) P_m^* = P_n^* \quad (2.66)$$

2.6 Markov chains

An important class of stochastic processes are *Markov chains*, which are homogeneous stochastic processes, discrete with respect to both the time and the state. This makes such processes rather convenient for numerical analysis.

a) Stochastic matrix

We consider a stochastic process $X(t) \in \{1, \dots, N\}$ and denote $P_n(t) = \text{Proba}\{X(t) = n\}$. The master equation (2.31) introduced above involves transitions at random times. For Markov chain, the jumps occur at regular discrete times, thus the master equation takes the form

$$P_n(t+1) = \sum_m M_{nm} P_m(t) \quad (2.67)$$

where

$$M_{nm} = \text{Proba}\{m \rightarrow n\} \in [0, 1] \quad (2.68)$$

is the $N \times N$ matrix of transition probabilities at each elementary time step. M is called a “*stochastic matrix*”. It has non negative elements and satisfies

$$\boxed{\sum_n M_{nm} = 1} \quad (2.69)$$

Example of Markov chain : the biased RW.— A simple example is the case of the random walk on the line, where, at each time step, the walker jumps to the left with probability q or to the right with probability p . Then

$$M_{nm} = p \delta_{m,n-1} + q \delta_{m,n+1} \quad (2.70)$$

(with $p + q = 1$). The problem is studied in detail in Exercise 2.10 below.

For the following, it is useful to rewrite the master equation (2.67) in a form closer to the differential equation (2.30) by using (2.69)

$$P_i(t+1) - P_i(t) = \sum_{j(\neq i)} [M_{ij} P_j(t) - M_{ji} P_i(t)] \quad (2.71)$$

✎ **Exercise 2.9 Continuum limit of the Markov chain:** Consider a Markov chain with jumps occurring every δt . Argue that the master equation (2.31) is recovered by considering the continuum limit

$$M_{ij} = \delta_{ij} + \delta t W_{ij} \quad \text{with } \delta t \rightarrow 0. \quad (2.72)$$

If M is a stochastic matrix, what is the constraint on the matrix W ?

b) The Perron-Fröbenius theorem and the stationary state

Discussion is similar as the one of previous paragraph. We can interpret the condition (2.69) as the existence of a left eigenvector $L^{(0)} = (1, \dots, 1)^T$ for eigenvalue $\Lambda_0 = 1$:

$$L^{(0)T} M = L^{(0)T} \quad \text{or} \quad M^T L^{(0)} = L^{(0)}, \quad (2.73)$$

The *Perron-Fröbenius theorem* states that, if M is a stochastic matrix,

- (i) $\Lambda_0 = 1$ is non-degenerate,¹³
- (ii) Λ_0 is the largest eigenvalue (and the spectral radius of M is one, i.e. $|\Lambda_n| \leq 1 \forall n$),
- (iii) the related right eigenvector

$$M R^{(0)} = R^{(0)}, \quad (2.74)$$

has positive components. For a finite number of states N , this corresponds to the **stationary solution**, $R^{(0)} = (P_1^*, \dots, P_N^*)^T$. Normalization condition reads $L^{(0)T} R^{(0)} = \sum_i P_i^* = 1$ (the scalar product is the product of the line and the column vectors). We can rewrite the equation (2.74) as

$$\sum_{j(\neq i)} [M_{ij} P_j^* - M_{ji} P_i^*] = 0. \quad (2.75)$$

¹³strictly speaking, $\Lambda_0 = 1$ is non-degenerate only if the states are not disconnected in the stationary state, for trivial reason(s). An example where this occurs is $M = \begin{pmatrix} 1 & 0 & p \\ 0 & 1 & q \\ 0 & 0 & 0 \end{pmatrix}$ with $p + q = 1$, which describes the situation where the system ends either on 1 or on 2, where it remains trapped forever. Because the stationary state describes two isolated islands, $\Lambda_0 = 1$ has degeneracy two. The Perron-Fröbenius theorem requires that the matrix M is *irreducible*, so that this situation does not happen.

c) Spectral analysis

We can solve (2.67) by using the spectral information on the matrix M . In the simplest situation, the matrix M is diagonalisable and has a biorthogonal set of right and left eigenvectors $R^{(n)}$, $L^{(n)}$ for eigenvalue Λ_n (with $|\Lambda_n| \leq 1$ according to the Perron-Fröbenius theorem). As a result, the solution of the master equation (2.67) is

$$P(t) = M^t P(0) = \sum_j c_j \Lambda_j^t R^{(j)} \underset{t \rightarrow \infty}{\simeq} R^{(0)} + c_1 \Lambda_1^t R^{(1)} + \dots \quad (2.76)$$

where $c_j = L^{(j)T} P(0)$ is the component of the initial vector on the eigenvector $R^{(j)}$; in particular $c_0 = 1$. In this case, the relaxation rate is $1/\tau_{\text{relax}} = -\ln |\Lambda_1|$.

Example : Let us illustrate the spectral analysis on a concrete example : I have constructed a random $N \times N$ stochastic matrix as follows : on each column, the number to matrix elements is taken from a binomial distribution $c = B(N, p)$ (there are c non zero matrix elements $M_{\cdot, m}$, i.e. there are c possible transitions from the state m to other states). The c matrix elements are then picked at random so that their sum is one (i.e. the non zero matrix elements are $\sim 1/c \sim 1/(Np)$). Here is an example of matrix obtained in this manner (for $N = 10$ and $p = 1/3$) :

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.068 & 0.125 & 0 \\ 0 & 0.156 & 0 & 0.322 & 0.400 & 0.060 & 0.641 & 0 & 0 & 0 \\ 0.452 & 0 & 0 & 0 & 0.220 & 0.137 & 0.054 & 0 & 0 & 0 \\ 0.087 & 0 & 0 & 0 & 0 & 0.329 & 0 & 0 & 0 & 0 \\ 0.075 & 0.844 & 0.419 & 0 & 0.157 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.373 & 0.071 & 0 & 0.067 & 0 & 0 & 0 & 0.556 \\ 0 & 0 & 0 & 0.071 & 0 & 0.155 & 0 & 0 & 0.396 & 0 \\ 0.048 & 0 & 0 & 0.536 & 0 & 0.112 & 0 & 0 & 0.268 & 0.366 \\ 0.202 & 0 & 0.208 & 0 & 0 & 0.140 & 0.305 & 0 & 0.116 & 0.078 \\ 0.136 & 0 & 0 & 0 & 0.223 & 0 & 0 & 0.932 & 0.093 & 0 \end{pmatrix} \quad (2.77)$$

On can check that the sum of probabilities on each column is one. Since some transitions are forbidden, this matrix describes an oriented graph. The possible connections are represented in Figure 2.11 : for example, the first column has $c = 6$ non zero elements corresponding to the possible transitions $1 \rightarrow 3$, $1 \rightarrow 4$, $1 \rightarrow 5$ and $1 \rightarrow 8$, $1 \rightarrow 9$, $1 \rightarrow 10$.

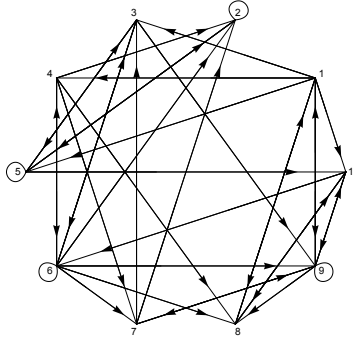


Figure 2.11: $N = 10$ states : connections described by the matrix (2.77).

The spectrum of eigenvalues of such matrix is complex. Examples of spectra are plotted in Fig. 2.12 for $N = 500$ and various values of p .

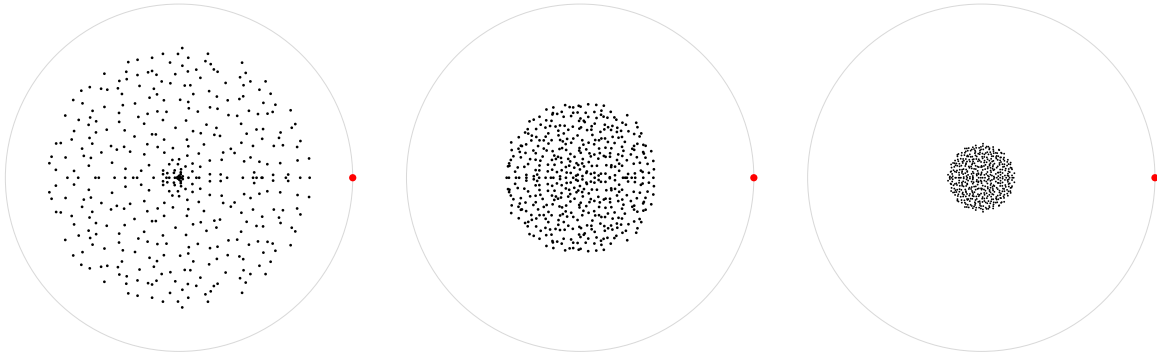


Figure 2.12: *Examples of spectra of some random stochastic $N \times N$ matrices in the complex plane. Here $N = 500$ and $p = 1/250$ (left), $p = 1/50$ (center) and $p = 1/10$ (right). The red dot is $\Lambda_0 = 1$, all others e.v. being inside the circle of unit radius. The symmetry with respect to the real axis is visible (e.v. come in c.c. pairs). The absence of an accumulation of e.v. close to $\Lambda_0 = 1$ ensures the existence of the stationary state.*

Conditional probability.— An example of initial condition is $P_n(0) = \delta_{nm}$ i.e. coefficients $c_j = L_m^{(j)}$. Then, the solution of the master equation (2.67) is the conditional probability

$$P_t(n|m) = (M^t)_{nm} \quad (2.78)$$

which is analogous to (2.65).

Exercise 2.10 Biased random walk in a ring: Consider the random walk in a ring with L sites, such that with $M_{nm} = p\delta_{n,m+1} + q\delta_{n,m-1}$ for $n, m \in \{1, \dots, L\}$. Periodic boundary conditions are $M_{1L} = p$ and $M_{L1} = q$.

- Argue that the stationary state is an equilibrium state when $p = q = 1/2$ and a NESS for $p \neq q$.
- Give the spectrum of eigenvalues Λ_k and eigenvectors (left/right) of the stochastic matrix M . Write $p = \frac{1+v}{2}$ and $q = \frac{1-v}{2}$ with $v \in [-1, +1]$. Check that the “spectral radius” is unity, i.e. $|\Lambda_k| \leq 1 \forall k$.
- Decompose the conditional probability $P_t(n|m)$ over the eigenvalues and the eigenvectors.
- Consider the limit $L \rightarrow \infty$ and discuss the bottom of the spectrum. Compute $P_t(n|m)$. Discuss the case $v = \pm 1$.

d) Detailed balance, reversibility and ergodicity

Let us consider a Markov chain with master equation (2.67), such that detailed balance is fulfilled. We denote by P_i^* the *equilibrium* solution. Define

$$\mathcal{D}_t \stackrel{\text{def}}{=} \sum_i \frac{(P_i(t) - P_i^*)^2}{P_i^*} = \sum_i \frac{P_i(t)^2}{P_i^*} - 1 \geq 0 \quad (2.79)$$

which measures the distance of the distribution at time t to the equilibrium distribution. One can study the evolution of the quantity by considering $\Delta \mathcal{D}_t = \mathcal{D}_{t+1} - \mathcal{D}_t$. Some algebra making use of detailed balance condition $M_{ij}P_j^* = M_{ji}P_i^*$ leads to

$$\Delta \mathcal{D}_t = -\frac{1}{2} \sum_{i,j,k} M_{ji}M_{ki}P_i^* \left(\frac{P_j(t)}{P_j^*} - \frac{P_k(t)}{P_k^*} \right)^2 \leq 0 \quad (2.80)$$

Conclusion :

- $\mathcal{D}_t \geq 0$
- $\Delta \mathcal{D}_t \leq 0$
- We conclude that $\mathcal{D}_t \searrow$ and thus $P_i(t) \rightarrow P_i^*$.

This shows that detailed balance ensures that the system reaches equilibrium. This remark is borrowed from [23].

e) A practical (and important) application of Markov chains : the Monte Carlo method

Consider a physical observable $\mathcal{O}$. At thermal equilibrium, the probability of a microstate is $P_\ell \propto e^{-\beta E_\ell}$. If the number of states N_{state} is too large, it might be difficult to compute numerically the sum

$$\langle \mathcal{O} \rangle_{\text{eq}} = \sum_{\ell=1}^{N_{\text{state}}} P_\ell \mathcal{O}_\ell \quad (2.81)$$

For example, if one considers N Ising spins, the sum runs over $N_{\text{state}} = 2^N$ microstates, which becomes rapidly untrackable if N is large (a square of 10×10 spins 1/2 has $N_{\text{state}} \sim 10^{30}$ microstates).

The central idea of equilibrium statistical physics is to replace the study of the microscopic (deterministic) dynamics by a probabilistic description. The main idea of Monte Carlo method is to replace the probabilistic description by a stochastic dynamics which explores reasonably well the states of the system, and perform a time averaging over the stochastic dynamics involving N_{step}

$$\overline{\mathcal{O}(t)} = \frac{1}{N_{\text{step}}} \sum_{t=1}^{N_{\text{step}}} \mathcal{O}(t). \quad (2.82)$$

The number of steps N_{step} can be chosen orders of magnitude smaller than N_{state} , still large enough in order to ensure some ergodicity (see the book [18] for a detailed discussion). There are several possible such dynamics.



Figure 2.13: *Nicholas Metropolis (1915-1999).*

Metropolis-Hastings algorithm.— if the system is in state $|i\rangle$ at time t , a move to another state $|f\rangle$ chosen randomly is made with probability

$$\text{Proba}\{\text{move } i \rightarrow f\} \equiv M_{fi} = \min \left(1, e^{-\beta(E_f - E_i)} \right). \quad (2.83)$$

For example, in a spin system, one chooses a spin randomly and flip it, thus the difference of energy $E_f - E_i$ is due to a local change, and the energy difference is very easy to compute.

This means that the matrix M_{nm} changes randomly at each elementary time step. Assuming $E_f > E_i$, the stochastic matrix has the form

$$M = \begin{pmatrix} \ddots & & & \\ & 1 - e^{-\beta(E_f - E_i)} & 1 & \\ & e^{-\beta(E_f - E_i)} & 0 & \\ & & & \ddots \end{pmatrix} \begin{matrix} \leftarrow |i\rangle \\ \leftarrow |f\rangle \end{matrix} \quad (2.84)$$

All other diagonal matrix elements are equal to one and all other non diagonal matrix elements equal to zero. Because $M_{fi}/M_{if} = e^{-\beta(E_f - E_i)}$, such dynamics converges towards the Gibbs equilibrium $P_n^* \propto e^{-\beta E_n}$.

Glauber dynamics.— Another popular dynamics is the Glauber dynamics, where

$$\text{Proba}\{\text{move } i \rightarrow f\} \equiv M_{fi} = \frac{1}{1 + e^{\beta(E_f - E_i)}} \quad (2.85)$$

while the probability for no move is $M_{ii} = e^{\beta(E_f - E_i)} / (1 + e^{\beta(E_f - E_i)})$. Selecting a pair of states, the stochastic matrix has a 2×2 block with these two probabilities and $M_{if} = M_{ii}$ and $M_{ff} = M_{fi}$. One can also check that Glauber dynamics satisfies detailed balance, hence we expect convergence towards the Gibbs equilibrium as well.

Conclusion.— Going from the microscopic dynamics to the Monte Carlo method, the scheme is the following :

<u>Classical mechanics</u>		<u>Equilib. statistical physics</u>		<u>Monte Carlo method</u>
deterministic	$\longrightarrow$	probabilistic	$\longrightarrow$	stochastic
evolution		description		dynamics

2.7 Fourier analysis of stochastic processes – Wiener-Khintchin theorem

Let us close the chapter on general properties of stochastic processes with a somehow disconnected problem : the Fourier analysis of stochastic processes.

Consider a homogeneous (time translation invariant) and **stationary** random process $x(t)$ defined on the interval $t \in [0, T]$, where T is the observation time. It is characterised by the correlation function $C_{xx}(\tau) \stackrel{\text{def}}{=} \langle x(t) x(t + \tau) \rangle$, assumed rapidly decreasing (assume $\langle x(t) \rangle = 0$ for simplicity). Because the process is stationary, we prefer to consider its discrete Fourier transform (cf. appendix page 156) ¹⁴

$$\tilde{x}_n = \int_0^T \frac{dt}{T} x(t) e^{+i\omega_n t} \quad \text{et} \quad x(t) = \sum_n \tilde{x}_n e^{-i\omega_n t} \quad \text{où} \quad \omega_n = \frac{2n\pi}{T} \quad \text{avec} \quad n \in \mathbb{Z}. \quad (2.86)$$

Noise spectrum.— Let $\Delta\omega$ be the bandwidth of the apparatus (with $\Delta\omega \gg 1/T$). We define the *noise spectrum* as the average of the square modulus of the Fourier components in the bandwidth, i.e. in the interval $[\omega, \omega + \Delta\omega]$:

$$\mathcal{S}(\omega) \stackrel{\text{def}}{=} \frac{1}{\Delta\omega} \sum_{\omega_n \in [\omega, \omega + \Delta\omega]} \langle |\tilde{x}_n|^2 \rangle \quad (2.87)$$

¹⁴Later, we will define the Fourier transform in space as $f_q = \int_V dr f(r) e^{-iqr}$, where q is quantized if the volume is finite, and $f(r) = \frac{1}{V} \sum_q f_q e^{+iqr} \rightarrow \int \frac{dq}{(2\pi)^d} f_q e^{+iqr}$.

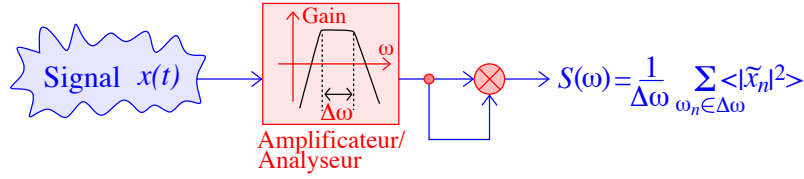


Figure 2.14: *Measure of the noise : the signal is amplified, duplicated and multiplied by itself. The result is averaged over a long time T .*

This is precisely the outcome of the device represented in figure 2.14 : sample $\rightarrow$ ampli/filter $\rightarrow$ multiplier $\rightarrow$ measurement. It measures the fluctuations of the Fourier components belonging to the bandwidth.



Figure 2.15: *Norbert Wiener (1894-1964) & Aleksandr Yakovlevich Khinchin (1894-1959).*

Wiener-Khintchin theorem.— From the above hypothesis, one can verify that :

$$\langle \tilde{x}_n \tilde{x}_m^* \rangle = \frac{1}{T} \delta_{n,m} \tilde{C}_{xx}(\omega_n) \quad \text{where } \tilde{C}_{xx}(\omega) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} d\tau C_{xx}(\tau) e^{i\omega\tau} \quad (2.88)$$

is the Fourier transform of the correlation function. Only components corresponding to opposite frequencies ω_n and ω_{-n} are correlated¹⁵. Thus one has : $\sum_{\omega_n \in [\omega, \omega + \Delta\omega]} \langle |\tilde{x}_n|^2 \rangle = \frac{\mathcal{N}_{\Delta\omega}}{T} \tilde{C}_{xx}(\omega)$ where $\mathcal{N}_{\Delta\omega} = \Delta\omega T / 2\pi$ is the number of frequencies ω_n in the bandwidth. Finally one gets

$$\boxed{\mathcal{S}(\omega) = \frac{\tilde{C}_{xx}(\omega)}{2\pi}} \quad (2.89)$$

i.e. a relation between the noise spectrum (fluctuations at frequency ω) and the correlations. A random process characterized by short time correlations thus corresponds to a broad noise spectrum. The limit of correlation with zero range is called a “white noise” (flat spectrum).

Details of derivation of (2.88) : Write

$$\langle \tilde{x}_n \tilde{x}_m^* \rangle = \int_0^T \frac{dt}{T} \int_0^T \frac{dt'}{T} e^{i\omega_n t - i\omega_m t'} C_{xx}(t - t') = \frac{1}{T} \int_0^T \frac{dt}{T} e^{i(\omega_n - i\omega_m)t} \int_{t-T}^t du e^{i\omega_n u} C_{xx}(u)$$

where we have performed the change of variable $(t, t') \rightarrow (t, u = t - t')$ (with Jacobian one). Short range correlation assumption allows to write $\int_{t-T}^t du \dots \rightarrow \int_{-\infty}^{+\infty} du \dots$, leading to (2.88). QED.

Note that we have considered the discrete Fourier transform of the process (due to translation invariance in time), however the calculation involves the Fourier transform of the correlation function on $\mathbb{R}$, as it decays fast.

¹⁵One can as well consider a process defined on $\mathbb{R}$ by writing $T \rightarrow \infty$. The Fourier transform is then defined as $\tilde{x}(\omega) = \int dt x(t) e^{i\omega t}$ and one can show that $\langle \tilde{x}(\omega) \tilde{x}(\omega') \rangle = 2\pi \delta(\omega + \omega') \tilde{C}_{xx}(\omega)$. Correspondence between the two formulations is ensured by the substitutions $\tilde{x}(\omega) \leftrightarrow T \tilde{x}_n$ and $2\pi \delta(\omega - \omega') \leftrightarrow T \delta_{n,n'}$.

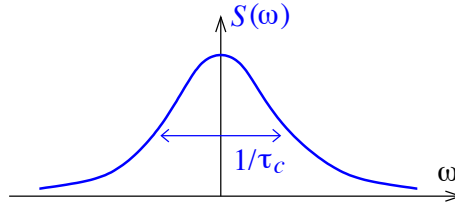


Figure 2.16: **Wiener-Khintchin theorem** : width of noise spectrum is inversely proportional to the correlation time τ_c of the process (τ_c is the width of $C(\tau)$).

✎ **Exercice 2.11** : As a simple application of the Wiener-Khintchin theorem, we analyze the correlation of the velocity for the process defined by the phenomenological Langevin equation

$$\frac{dv(t)}{dt} = - \int dt' \gamma(t-t') v(t') + \xi(t) \quad (2.90)$$

where $\xi(t)$ is the Langevin force (assumed to be a stationary random process with short time correlations). Here the friction is nonlocal in time, controlled by a causal function $\gamma(t)$, with finite width τ_m . Show that

$$C_{vv}(\tau) = \int_{-\infty}^{+\infty} d\omega \underbrace{\frac{\tilde{C}_{\xi\xi}(\omega)}{2\pi}}_{\mathcal{S}_{\text{Force}}(\omega)} \frac{e^{-i\omega\tau}}{|\tilde{\gamma}(\omega) - i\omega|^2} \quad (2.91)$$

Consider the limit $C_{\xi\xi}(\tau) = 2D\gamma^2\delta(\tau)$ et $\gamma(t) = \gamma\delta(t)$ and compute explicitly the correlator.

✎ **Exercice 2.12** : Consider now the case of a Langevin force correlated over the finite time τ_c (a microscopic time) : $C_{FF}(t) = 2D\gamma^2 \frac{1}{2\tau_c} e^{-|t|/\tau_c}$. We expect the function $\gamma(t)$ to be of finite width τ_m ; Assume $\gamma(t) = \gamma\theta(t) \frac{1}{\tau_m} e^{-t/\tau_m}$. The three time scales fulfill : $\tau_c \lesssim \tau_m \ll 1/\gamma$. Analyze the residus of $|\tilde{\gamma}(\omega) - i\omega|^{-2}$ justify that one can consider $\tau_m \rightarrow 0$ while keeping a finite τ_c , what simplifies the evaluation of the integral. Show then that $C_{vv}(\tau) = \frac{D\gamma}{1-(\gamma\tau_c)^2} [e^{-\gamma|\tau|} - \gamma\tau_c e^{-|\tau|/\tau_c}]$. Analyze the behaviour at short time as well.

☺ Important points

- Markov process (definition).
- Be familiar with the various forms of the master equation (continuous/discrete ; Markov chain).
- A good exercise : recover the properties of the Poisson process (and the CPP).
- Definition of the stochastic matrix. Use of spectral information to solve the master equation.
- Detailed balance and the classification of Markov processes.
- Wiener-Khintchin theorem : relation between the correlation function of an homogeneous process and its noise spectrum.

Additional exercises

✎ **Exercice 2.13 Solution of the Kramers equation (free case)** : We consider a particle in a fluid described by the two differential equations (1.3,1.4), where we recall that the Langevin force is a Gaussian white noise such that $\langle \xi(t)\xi(t') \rangle = 2\gamma k_B T \delta(t-t')$. The Kramers equation is the related equation for the joint distribution $P_t(x, v)$, which was discussed in this chapter and is also given in the next chapter, Eq. (3.34). Here we consider fixed initial conditions $x(0) = x_0$ and $v(0) = v_0$.

- a) What is the form of the conditional joint probability $P_t(x, v | x_0, v_0)$?
- b) Collecting the results obtained in chapter 1 (lecture and exercises), give $\langle x(t) \rangle$, $\langle v(t) \rangle$, $\langle x(t)^2 \rangle_c$ and $\langle v(t)^2 \rangle_c$. Compute also $\langle x(t)v(t) \rangle_c$.
- c) Deduce $P_t(x, v | x_0, v_0)$.
- d) Discuss the marginal distribution of the position $W_t(x | x_0, v_0) \stackrel{\text{def}}{=} \int dv P_t(x, v | x_0, v_0)$ for fixed initial velocity and also when the particle is initially at equilibrium, i.e. $Q_t(x | x_0) \stackrel{\text{def}}{=} \int dv dv_0 P_t(x, v | x_0, v_0) P_{\text{eq}}(v_0)$.
- e) Compute $\int dx_1 Q_t(x | x_1) Q_{t'}(x_1 | x_0)$ and discuss the result.

✎ **Exercise 2.14 Poisson noise (derivative of the PP):** We consider the noise $\xi(t) = \sum_n \delta(t - t_n)$, where the times are i.i.d. for a uniform density λ . I.e., when they are ordered, the events occur randomly and independently with rates λ . In other terms, the noise is the derivative of the Poisson process introduced above $\xi(t) = \mathcal{N}'(t)$.

We introduce the generating function of the noise $G[h] \stackrel{\text{def}}{=} \langle \exp \int dt h(t) \xi(t) \rangle$, where $\langle \bullet \rangle$ is the averaging over the random times t_n 's.

- a) Show that $G[h] = \exp \left\{ \lambda \int dt (e^{h(t)} - 1) \right\}$.

Hint: Consider that the N times are not ordered, distributed over $[0, T]^N$ with measure $dt_1 \cdots dt_N / T^N$.

- b) Deduce the connex correlation functions (cumulants) : $\langle \xi(t) \rangle = \lambda$ and $\langle \xi(t_1) \cdots \xi(t_n) \rangle_c = \lambda \delta(t_1 - t_2) \cdots \delta(t_1 - t_n)$.

Hint: Consider functional derivatives of $\ln G[h]$.

In conclusion, $\xi(t) = \mathcal{N}'(t)$ is a **non Gaussian white noise**.

✎ **Exercise 2.15 n-point correlations of the PP:** Using the result of Exercise 2.14, show that the n -point correlation function is

$$\langle \mathcal{N}(t_1) \mathcal{N}(t_2) \cdots \mathcal{N}(t_n) \rangle_c = \lambda \min(t_1, t_2, \dots, t_n). \quad (2.92)$$

✎ **Exercise 2.16 Derivative of the CPP - a non Gaussian white noise:**

- a) Using the representation $\xi(t) = X'(t) = \sum_n \eta_n \delta(t - t_n)$, where both the times t_n 's and the coefficients η_n 's are random, derive the connex correlation function of the noise $\langle \xi(t_1) \cdots \xi(t_n) \rangle_c$ (the analogous of a cumulant).

Hint: follow the same steps as in Exercise 2.14.

- b) Show that the noise becomes a Gaussian white noise in a certain limit.

✎ **Exercise 2.17 Random walk on $\mathbb{Z}$ with a potential : the 3 scenarii of stochastic processes:** Let us consider the master equation describing the one dimensional diffusion on $\mathbb{Z}$ with transitions between nearest neighbour sites

$$\frac{dP_n(t)}{dt} = W_{n,n-1} P_{n-1}(t) + W_{n,n+1} P_{n+1}(t) - (W_{n-1,n} + W_{n+1,n}) P_n(t) \quad (2.93)$$

i.e. $W_{n,m}$ is a tridiagonal matrix with $W_{n,n} = -W_{n-1,n} - W_{n+1,n}$ (if $n \in \mathbb{Z}$, the matrix has an infinite size). Hence, this is another example of birth and death process.

- a) *Current* : check that the master equation can be rewritten under the form

$$\frac{dP_n}{dt} = -J_n + J_{n-1} \quad (2.94)$$

and express the probability current $J_n(t)$ related to the distribution $P_n(t)$ (J_n measures the current at time t between sites n and $n+1$).

We now choose the matrix such that

$$W_{n,m} = e^{[V(m)-V(n)]/2} \quad (2.95)$$

where $V(x)$ is a known function.

b) *Equilibrium* ($J = 0$).— Show that

$$P_n^* = C e^{-V(n)} \quad (2.96)$$

is a stationary solution corresponding to a vanishing probability current. Discuss the normalisability. Give an example of $V(n)$ leading to an equilibrium.

c) *NESS* ($J \neq 0$).— Find the stationary solution corresponding to $J_n = J \forall n$. Show that it is

$$P_n^* = J e^{-V(n)} \sum_{m=n}^{\infty} e^{[V(m+1)+V(m)]/2} \quad (2.97)$$

Discuss the normalisability (consider the continuum limit for simplicity). Give an example of $V(n)$ leading to a NESS.

d) Provide an example where there is no stationary state.

3 Stochastic processes (3) : stochastic differential equations

In chapter 1, we have discussed a specific case of stochastic differential equation (SDE), the Langevin equation $m \frac{d}{dt} v(t) = -\gamma v(t) + \xi(t)$ involving a δ -correlated Langevin force. We took advantage of the linearity to obtain an integral representation of the solution, which makes easy the analysis of the statistical properties of the solution. The aim of this paragraph is to consider a more general situation and consider SDE of the form $\frac{d}{dt} X(t) = F(X(t)) + \sqrt{2D(X(t))} \eta(t)$, where $\eta(t)$ is a normalised Gaussian white noise :

$$\boxed{\langle \eta(t) \rangle = 0 \quad \text{and} \quad \langle \eta(t) \eta(t') \rangle = \delta(t - t')} \quad (3.1)$$

SDE are particularly well suited for numerical simulations (it is easy to generate many realizations of such processes). Here, the aim is to introduce some tools which pave the route for a statistical analysis of the solution. Finally, let us stress that by considering that $\eta(t)$ is a Gaussian white noise in this chapter, we restrict ourselves to the study of **continuous Markov processes** (with no jump).¹⁶ There will be only one exception in b) of § 3.2.

3.1 SDE with drift and additive noise

Let us come back to the analysis of the stochastic process described by Eq. (1.25). We write

$$\frac{dX(t)}{dt} = F(X(t)) + \sqrt{2D} \eta(t) \quad (3.2)$$

¹⁶Remember the end of chapter 1 : the Wiener process $W(t) = \int_0^t du \eta(u)$ is continuous but not differentiable. The solution $X(t)$ of the SDE has the same regularity.

We could also write a differential equation for a Markov process with jumps by considering that a noise $\eta(t)$ such that $\int_0^t du \eta(u)$ is a Poisson process or a compound Poisson process, i.e. $\eta(t) = \sum_n \eta_n \delta(t - t_n)$ is a *non* Gaussian white noise. The analysis would be more complicated because the equation for $P_t(x)$ would then involve an *integral* operator instead of a *differential* operator, like in the Fokker-Planck equation (e.g. the equation (2.44) describing the simple case of the CPP involves an integral operator).

An analysis similar to the one of Chapter 1 is not possible due to the nonlinear character of the equation (the method of Chapter 1 only applies to the linear force $F(x) = a + bx$). Let us introduce the Wiener process

$$W(t) = \int_0^t dt' \eta(t') \quad (3.3)$$

what allows to rewrite the SDE in a form popular in the mathematical literature :

$$dX(t) = F(X(t)) dt + \sqrt{2D} dW(t) . \quad (3.4)$$

This emphasizes that $X(t)$ has the same regularity as the Wiener process $W(t)$: *continuous* but *non differentiable*. One of the question raised in this chapter is precisely to give a rigorous meaning to the differential equation for non differentiable functions !

Being interested in statistical properties of the solution, it is natural to consider its distribution, or at least to build an equation for it, the Fokker-Planck equation. Below we show that the corresponding FPE is

$$\frac{\partial P_t(x)}{\partial t} = -\frac{\partial}{\partial x} [F(x) P_t(x)] + D \frac{\partial^2}{\partial x^2} P_t(x) \quad (3.5)$$

where $P_t(x) = \langle \delta(x - X(t)) \rangle$ is the distribution of $X(t)$. In the next section, we will further discuss how to solve this equation.

Proof : I define the increment $\delta W(t) = W(t + \delta t) - W(t)$. The most important observation is the **independence of the increments** (chapter 1) and the property

$$\boxed{\langle \delta W(t)^2 \rangle = \delta t \quad \text{i.e.} \quad \delta W(t) \sim \mathcal{O}(\sqrt{\delta t})} \quad (3.6)$$

see above, Eq. (1.32). The distribution of $\delta W(t)$ does not depend on time t , but only on the time difference δt . We now introduce the increment of the process

$$\delta X(t) \stackrel{\text{def}}{=} X(t + \delta t) - X(t) \simeq F(X(t)) \delta t + \sqrt{2D} \delta W(t) \quad (3.7)$$

Consider a test function $\varphi(x)$. We study the evolution of $\langle \varphi(X(t)) \rangle$.

$$\begin{aligned} & \langle \varphi(X(t + \delta t)) \rangle - \langle \varphi(X(t)) \rangle \\ &= \left\langle \varphi'(X) \left[F(X) \delta t + \sqrt{2D} \delta W \right] + \frac{1}{2} \varphi''(X) \left[F(X) \delta t + \sqrt{2D} \delta W \right]^2 + \dots \right\rangle \end{aligned} \quad (3.8)$$

$$= \langle \varphi'(X(t)) F(X(t)) \rangle \delta t + \sqrt{2D} \langle \varphi'(X(t)) \delta W(t) \rangle + D \langle \varphi''(X(t)) \rangle \delta t + \dots \quad (3.9)$$

where we have kept terms $\mathcal{O}(\delta t)$. Because $X(t)$ is only correlated with the increment $\delta W(t')$ for $t > t'$, we see that $X(t)$ and $\delta W(t)$ are uncorrelated, thus

$$\langle \varphi'(X(t)) \delta W(t) \rangle = \langle \varphi'(X(t)) \rangle \langle \delta W(t) \rangle = 0 . \quad (3.10)$$

Finally

$$\frac{d}{dt} \langle \varphi(X(t)) \rangle = \langle \varphi'(X(t)) F(X(t)) \rangle + D \langle \varphi''(X(t)) \rangle \quad (3.11)$$

The key point was that the expansion was performed up to second order in $\delta X(t)$ because the second order term in δW gives some first order contribution in δt , due to (3.6). We can now rewrite the equation in terms of the distribution

$$\frac{\partial}{\partial t} \int dx P_t(x) \varphi(x) = \int dx P_t(x) [\varphi'(x) F(x) + D \varphi''(x)] \quad (3.12)$$

$$= \int dx \varphi(x) \left(-\frac{\partial}{\partial x} [F(x) P_t(x)] + D \frac{\partial^2 P_t(x)}{\partial x^2} \right) \quad (3.13)$$

Because the equation is valid $\forall \varphi$, we can remove the integral, hence (3.5). QED.

3.2 SDE with multiplicative noise : Itô or Stratonovich ?

The SDE (3.2) is not the most general form of stochastic differential equation as it corresponds to the case where the diffusion constant is independent of x . The aim of the paragraph is to discuss the case of SDE of the form

$$\frac{dX(t)}{dt} = a(X(t)) + b(X(t))\eta(t) \quad (\text{not well defined!}) \quad (3.14)$$

where $b(x) = \sqrt{2D(x)}$ can be related to a x -dependent diffusion constant (note that the sign of $b(x)$ plays no role because $\eta(t) \stackrel{(\text{law})}{=} -\eta(t)$). As we explain now, this form is however not well defined.

The noise is here multiplied by a function of the process : one says that the **noise is multiplicative**, whereas it is said additive in the SDE (3.2). For a multiplicative white noise, with singular correlations $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$, the differential equation (3.14) is ambiguous : something is missing in order to define precisely the solution of this differential equation. This is not surprising : if $\eta(t)$ is a Gaussian white noise, $X(t)$ has the same regularity as the Brownian motion, i.e. is continuous but not differentiable. The existence of a difficulty comes from the fact that we manipulate a differential equation involving objects which are not differentiable in the sense of functions.

If the difficulty comes from the singular nature of the white noise (δ -correlations), one could argue that we should rather study a different model for a more regular (non white) noise. However, that would be very painful ! The white noise is the simplest model which ensures that the process defined by the SDE is Markovian. The price to pay for this simplicity is that we have to add a prescription to interpret precisely the SDE (3.14).

a) Discretization (for numerical simulations)

A first approach to give a precise meaning to (3.14) could be to discretize time (this is natural for numerical implementation of the stochastic differential equation). We introduce $X_n = X(t = n\delta t)$ together with i.i.d. Gaussian variables δW_n such that $\langle \delta W_n \rangle = 0$ and $\langle \delta W_n \delta W_m \rangle = \delta t \delta_{n,m}$. The discretized SDE is

$$X_{n+1} = X_n + a(X_n)\delta t + b(X_n)\delta W_n. \quad (3.15)$$

Note that by construction X_n and δW_m are independent for $n \leq m$ (the process does not depend on the noise in the future or *at equal time*). X_n only depends on the noise δW_m in the past, for $m < n$.

This discretization procedure is perfectly fine (the recurrence (3.15) is well defined), however it turns out that in the limit $\delta t \rightarrow 0$, the resulting continuous function $X(t) = X_n$ does not obey the usual rules of differential calculus (for regular functions), as we will see below. This is not necessarily a problem, however this deserves a clarification.

b) Origin of the ambiguity (the case of a non Gaussian white noise)

To clarify this point, we come back to the continuous description and consider $\frac{dX}{dt} = a(X) + b(X)\xi(t)$ for a different type of noise, made of δ -peaks at random times

$$\xi(t) = \sum_n \delta(t - t_n). \quad (3.16)$$

If the times t_n 's occur with a finite rate λ , the noise $\xi(t)$ is a white noise since $\langle \xi(t)\xi(t') \rangle_c = \lambda \delta(t - t')$, however it is of *non Gaussian* nature because higher cumulants are non zero (cf. Exercise 2.14, page 40). Note that the effect of the non-zero average noise $\langle \xi(t) \rangle = \lambda$ can be

removed by changing the drift as $a(X) \rightarrow a(X) - \lambda b(X)$ (we could also add some random weights, cf. Exercise ??, which would not change the following discussion).

In a close enough neighbourhood of time t_n , we can focus on the evolution due to a single δ -peak of the noise

$$\frac{dX(t)}{dt} = \dots + b(X(t)) \delta(t - t_n) \quad \text{for } t \sim t_n \quad (3.17)$$

This implies that $X(t)$ is discontinuous at t_n . When dealing with a continuous function $\psi(t)$ we can write $\psi(t) \delta(t - t_n) = \psi(t_n) \delta(t - t_n)$. The product is however not clearly defined when the function $\psi(t)$ is discontinuous at t_n : should we replace $\psi(t)$ by $\psi(t_n^-)$, $\psi(t_n^+)$ or $\frac{1}{2}[\psi(t_n^-) + \psi(t_n^+)]$, ... or something else ? Here, we are faced to this problem, which makes the time evolution ambiguous : Eq. (3.17) shows that $X(t)$ makes a jump whose amplitude is $b(X(t))$, i.e. depends on the process at a time where the process is discontinuous and still unknown ! How to choose this time ? At this point, we must specify a choice, i.e. add a prescription to define the evolution. Let us propose two possible interpretations of the evolution (3.17) :

- (i) **Proposal 1 ($\leftrightarrow$ Itô) :** Interpret the equation as $\frac{d}{dt}X(t) = \dots + b(X(t_n^-)) \delta(t - t_n)$, then

$$X(t_n^+) = X(t_n^-) + b(X(t_n^-)) \quad (3.18)$$

This is a natural choice for numerics. This is analogous to (3.15) : the process is not only independent of the noise in the future (causality) but also at equal time.

- (ii) **Proposal 2 ($\leftrightarrow$ Stratonovich) :** A physicist rather considers that the δ -peak is a mathematical model for a regular narrow function of finite width $\delta(t) \rightarrow \delta^\epsilon(t)$, for example $\delta^\epsilon(t) = \frac{1}{2\epsilon} e^{-|t|/\epsilon}$ or $\delta^\epsilon(t) = \frac{1}{\sqrt{2\pi\epsilon}} e^{-t^2/2\epsilon^2}$, etc. Then starting from (3.17) one writes $dX(t)/b(X(t)) \simeq \delta^\epsilon(t - t_n) dt$ and integrate around the δ^ϵ , eventually taking the limit $\epsilon \rightarrow 0^+$. One gets

$$\int_{X(t_n^-)}^{X(t_n^+)} \frac{dx}{b(x)} = 1 \quad (3.19)$$

which obviously differs from (3.18) (unless b is independent of x , i.e. for additive noise).

✎ **Exercise 3.1 :** Consider a multiplicative noise with $b(x) = \sigma x$ where σ is a positive constant. Compare the two evolutions (3.18) and (3.19) in this case.

The choice of the prescription, i.e. the precise meaning to give to the multiplicative noise term, determines the evolution and contribute to define the stochastic process with the SDE. We stress that given *two different* interpretations of the *same* equation (3.17) leads to two different evolutions, (3.18) or (3.19), i.e. define *two different processes*. An example is shown in Fig. 3.1.

A similar problem occurs with the SDE (3.14) where $\eta(t)$ is a Gaussian white noise. Several interpretations can be given to the multiplicative noise term.

The two evolutions (3.18) and (3.19) will be further studied in Exercise ?? below. In the following, we come back to the study of SDE involving Gaussian white noise.

Remark : as the previous discussion has made clear, the problem raised by the definition of the differential equation (3.14) is due the *singular nature* of the noise, but not its stochastic nature. The very same discussion would be required if $\eta(t)$ is singular and deterministic, for example a regular series of delta peaks $\eta(t) = \sum_n \delta(t - n)$.

c) Itô convention and Itô calculus

The simpler choice which first comes in mind is to consider that the process and the increment at equal time are independent. This is a natural choice if one discretizes the evolution, as explained

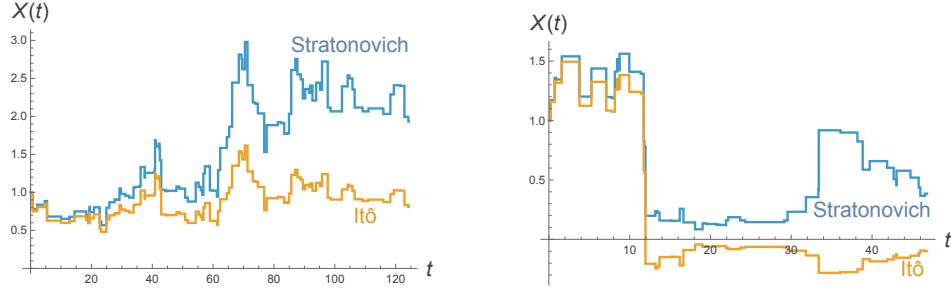


Figure 3.1: *Simulation of the solution of the SDE $\frac{d}{dt}X(t) = X(t)\eta(t)$ for $\eta(t) = \sum_n \alpha_n \delta(t - t_n)$ with Itô prescription (3.18) (orange) and Stratonovich prescription (3.19) (blue), for the same noise $\eta(t)$. Here, the coefficients α_n are i.i.d. with a symmetric distribution $p(\alpha) = (2a)^{-1}e^{-|\alpha|/a}$. The Itô evolution is always below the Stratonovich evolution. Left : $N_j = 100$ jumps with amplitude $a = 0.1$. Right : for larger jumps, $a = 0.25$, the Itô jumps may likely drag the process in the negative axis, whereas the Stratonovich solution always remains positive ($N_j = 50$ jumps).*

above, see Eq. (3.15). This is appropriate for numerical simulations. This is known as the *Itô convention*, corresponding to the “proposal 1” discussed above (and to discretization procedure, Eq. (3.15)). In order to specify in which sense the SDE is understood, we add a prescription

$$dX(t) = a(X(t))dt + b(X(t))dW(t) \quad (\text{Itô}). \quad (3.20)$$

where the label “Itô” indicates that

$$\boxed{\text{Itô : } X(t) \text{ and the noise } dW(t) \text{ are independent at coinciding times}}$$

There is a subtle point here : causality means that the process only depends on the noise in the past, i.e. $\langle \varphi(X(t))dW(t') \rangle \propto \theta_H(t - t')$ for any function $\varphi(x)$. Assuming here independence at *equal* time

$$\langle \varphi(X(t))dW(t) \rangle = 0 \quad (\text{Itô}) \quad (3.21)$$

means that Itô choice corresponds to $\theta_H(0) = 0$. Itô calculus also makes use of

$$\boxed{dW(t)^2 = dt} \quad (3.22)$$

(we can omit the averaging when considering the infinitesimal noise increment ; see the appendix of this chapter in order to understand why $\langle \cdots \rangle$ can be omitted for infinitesimal increments, while it should be kept for $\langle \delta W(t)^2 \rangle = \delta t$). Roughly speaking we have $dW(t) \sim \mathcal{O}(\sqrt{\delta t})$ and for this reason $dW(t)^{2+n} = 0$ for $n > 0$.

Itô formula.— An important formula concerns the change of variable $X(t) \rightarrow \varphi(X(t))$, where $X(t)$ is a stochastic process and $\varphi(x)$ is a regular function, differentiable a least twice. If $X(t)$ obeys (3.20), we deduce

$$\boxed{d\varphi(X(t)) = \left[\varphi'(X) a(X) + \frac{1}{2} \varphi''(X) b(X)^2 \right] dt + \varphi'(X) b(X) dW(t) \quad (\text{Itô})} \quad (3.23)$$

(in the r.h.s. X stands for $X(t)$). This is known as the “*Itô formula*” (or the “*Itô lemma*”).

Proof : using Itô calculus, we must keep a second order term in equalities between infinitesimal increments :

$$d\varphi(X) = \varphi'(X) dX + \overbrace{\frac{1}{2}\varphi''(X) dX^2}^{\text{novelty of Itô calculus}}. \quad (3.24)$$

This is due to (3.22) : a term $\mathcal{O}(dt)$ is produced by the $\mathcal{O}(dX^2)$ term : $dX^2 = [a(X)dt + b(X)dW(t)]^2 = b(X)^2 dW(t)^2 + 2a(X)b(X)dW(t)dt + a(X)^2 dt^2 = b(X)^2 dt$. The first term dropped was $\mathcal{O}(dW(t)dt) = \mathcal{O}(dt^{3/2})$ ($= 0$ in differential calculus).

 **Exercice 3.2 :** Write the Itô formula for the multiplicative noise $dX(t) = \sigma X(t) dW(t)$. Then apply the formula to $\varphi(x) = x^2$

Itô formula implies that **“Itô calculus” does not correspond with the “usual” differential calculus** when $W(t)$ and $X(t)$ are regular (differentiable) functions. Indeed, (3.23) shows that

$$\boxed{\text{!} \quad d\varphi(X(t)) \neq \varphi'(X(t)) dX(t) \quad (\text{Itô})} \quad (3.25)$$

Despite this drawback, the Itô calculus is widely used by probabilists (and arises in certain specific physical situations). It is also natural for financial mathematics, which is not a surprise as the time is discrete in finance : it corresponds to the discretization scheme mentioned above, Eq. (3.15), where δt is the elementary time step for reactualization of assets on markets.

Remark : Within Itô convention, the process $X(t)$ and the noise $dW(t)$ are independent *at equal time*. It follows that averaging (3.23) is straightforward (this is the interest of Itô’s convention!) and gives

$$\frac{d}{dt} \langle \varphi(X(t)) \rangle = \langle \varphi'(X(t)) a(X(t)) \rangle + \frac{1}{2} \langle \varphi''(X(t)) b(X(t))^2 \rangle. \quad (3.26)$$

Related FPE.— One can immediately deduce the FPE related to the Itô equation (3.20). Write (3.26) as

$$\frac{\partial}{\partial t} \int dx P_t(x) \varphi(x) = \int dx P_t(x) \left[\varphi'(x) a(x) + \frac{1}{2} \varphi''(x) b(x)^2 \right] \quad (3.27)$$

In the r.h.s, integrations by part allow to factorize $\varphi(x)$. Because the relation is true $\forall \varphi(x)$, we conclude that

$$\boxed{\frac{\partial P_t(x)}{\partial t} = -\frac{\partial}{\partial x} [a(x) P_t(x)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [b(x)^2 P_t(x)]} \quad (3.28)$$

How to deduce the FPE from the SDE in a simple manner ? Above, the relation between the Itô SDE and the FPE was demonstrated by introducing a test function. A simpler way is to use $\langle dW(t)^2 \rangle_{\text{noise}} = dt$ (physicist’s notation) and to remark that the drift and the “diffusion” are given by

$$\langle a(X) \rangle = \frac{\langle dX \rangle_{\text{noise}}}{dt} \quad \text{and} \quad \langle b(X)^2 \rangle = \frac{\langle dX^2 \rangle_{\text{noise}}}{dt} \quad (3.29)$$

As an application we consider the multidimensional case

$$dX_i(t) = a_i(\vec{X}) dt + b_{ij}(\vec{X}) dW_j(t) \quad (\text{Itô}). \quad (3.30)$$

with $\langle dW_i(t) dW_j(t) \rangle_{\text{noise}} = \delta_{ij} dt$. Only the diffusion term is more complicated

$$\frac{\langle dX_i dX_j \rangle_{\text{noise}}}{dt} = \langle b_{ik} b_{jk} \rangle \quad (3.31)$$

(with Einstein's convention for implicit summation over repeated indices). Then

$$\partial_t P_t(\vec{x}) = -\partial_i [a_i(\vec{x}) P_t(\vec{x})] + \frac{1}{2} \partial_i \partial_j [b_{ik}(\vec{x}) b_{jk}(\vec{x}) P_t(\vec{x})] . \quad (3.32)$$

An alternative simple way to prove this result is to generalize the Itô formula and consider $d\varphi(\vec{X}(t))$.

Application : Kramers and Smoluchowski equations.— Consider the two coupled SDE (equivalent to the Newton equation)

$$\begin{cases} dx = v dt \\ dv = \left(-\frac{1}{\tau} v + \frac{1}{m} F(x)\right) dt + \frac{1}{m} \sqrt{2k_B T \gamma} dW(t) \end{cases} \quad (3.33)$$

for the motion of a particle of position x and velocity v . The particle is submitted to a friction force $F_f = -(m/\tau) v$, a conservative force $F(x)$ and a Langevin force $\xi(t) = \sqrt{2k_B T \gamma} \eta(t)$. The drift terms in the two SDE are $a_x = \frac{\langle dx \rangle_{\text{noise}}}{dt} = v$ and $a_v = \frac{\langle dv \rangle_{\text{noise}}}{dt} = -\frac{v}{\tau} + \frac{F(x)}{m}$, respectively. The diffusive terms are $b_{xx}^2 = \langle dx^2 \rangle_{\text{noise}}/dt = v^2 dt \rightarrow 0$, $b_{vv}^2 = \langle dv^2 \rangle_{\text{noise}}/dt = 2k_B T \gamma / m^2 = 2D/\tau^2$ and $b_{xv}^2 = \langle dx dv \rangle_{\text{noise}}/dt \sim \langle v dW(t) \rangle_{\text{noise}} \rightarrow 0$. Finally, the FPE for the joint distribution is

$$\left(\partial_t + v \partial_x + \frac{F(x)}{m} \partial_v \right) P_t(x, v) = \frac{1}{\tau} \partial_v \left(v + \frac{D}{\tau} \partial_v \right) P_t(x, v) \quad (3.34)$$

This equation is called the *Kramers equation*. It is solved for $F(x) = 0$ in Exercise 2.13.

✎ **Exercice 3.3 Smoluchowski equation :** Using the overdamped limit introduced in chapter 1 get an equation for the marginal distribution $W_t(x) = \int dv P_t(x, v)$ in the limit of strong friction.

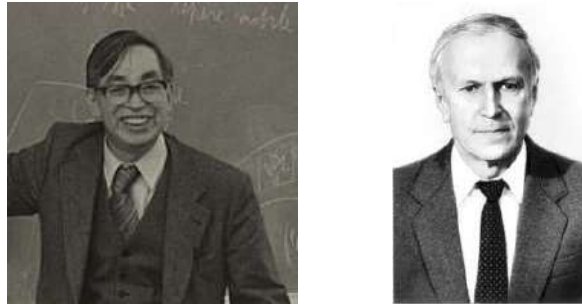


Figure 3.2: Kiyoshi Itô (1915-2008) and Ruslan Leont'evich Stratonovich (1930-1997).

d) Stratonovich convention

δ -function does not exist in real life : for physicists, it is only a convenient mathematical model for an extremely narrow function. Similarly, white noise is only a mathematical model and physicists should find more natural to think at the Gaussian white noise in the SDE (3.14) as the limit of a regular (non white) Gaussian noise with a small but finite correlation time. A concrete example of such a “real life” correlator is ¹⁷

$$\langle \eta^\epsilon(t) \eta^\epsilon(t') \rangle = \frac{1}{2\epsilon} e^{-|t-t'|/\epsilon} \quad (3.35)$$

¹⁷This choice is interesting because, from the theoretical point of view, we know how to generate such a non white noise from a white noise. We can simply solve the SDE $\frac{d}{dt} \eta^\epsilon(t) = -(1/\epsilon) \eta^\epsilon(t) + (1/\epsilon) \tilde{\eta}(t)$ where $\tilde{\eta}(t)$ is a white noise (this was done in chapter 1).

with ϵ “small” (much smaller than all characteristic time scales of the problem). It is convenient to introduce the notation $W^\epsilon(t) = \int_0^t du \eta^\epsilon(u)$, which is regular (not only continuous, but also *differentiable*). In general we have

$$\langle [\partial_t W^\epsilon(t)]^2 \rangle \sim 1/\epsilon \quad (3.36)$$

(for the correlator (3.35), we have $\langle [\partial_t W^\epsilon(t)]^2 \rangle = 1/(2\epsilon)$). Let us write the regularised SDE

$$dX^\epsilon(t) = \alpha(X^\epsilon) dt + \beta(X^\epsilon) dW^\epsilon(t). \quad (3.37)$$

The noise is differentiable once, hence the regularised process $X^\epsilon(t)$ is also differentiable (this differential equation is well defined mathematically). The standard rules of differential calculus for regular functions apply to $X^\epsilon(t)$: hence, the $\epsilon \rightarrow 0^+$ limit of this equation does not correspond to the Itô prescription, i.e. is NOT the SDE $dX(t) = \alpha(X) dt + \beta(X) dW(t)$ (Itô).

Stratonovich SDE.— The limit $\epsilon \rightarrow 0^+$ of the regularised SDE (3.37) is our definition of the “*Stratonovich SDE*” :

$$dX(t) = \alpha(X) dt + \beta(X) dW(t) \quad (\text{Stratonovich}) \quad (3.38)$$

Because it is defined as a limit of a regular process,

ordinary rules of differential calculus hold within the Stratonovich convention

(this is its interest !): what is true for $\epsilon > 0$ remains true at $\epsilon = 0$, which is not a singular point (this is the spirit of regularization in Physics). The drawback is that, as it differs from the Itô convention, the correlations at equal times are more difficult to characterize

Stratonovich : $X(t)$ and the noise $dW(t)$ are correlated at coinciding times

As stressed aboven $X(t)$ and $dW(t')$ are uncorrelated for $t' > t$, as the process depends only on the noise in the *past* : $\langle \varphi(X(t)) dW(t') \rangle \propto \theta_H(t - t')$ (causality). Within the Stratonovich convention, the process and the noise are correlated at *equal time* (these correlations are studied in Exercise 3.6 below). Note that this has to do with our definition of $\theta_H(0)$: in Itô convention $\theta_H(0) = 0$ while in Stratonovich convention $\theta_H(0) = 1/2$.¹⁸

Two questions now arise :

- Q1: what is the FPE related to the Stratonovich SDE (3.38) ?
- Q2: what is the relation between the Stratonovich SDE (3.38) and the Itô SDE (3.20) (if they both aim to describe the same process) ?

Stratonovich SDE and FPE.— In the Exercise 3.4 below, we demonstrate that, using the usual rules of differential calculus (for regular functions), the Stratonovich SDE (3.38) is related to the FPE

$$\left[\frac{\partial P_t(x)}{\partial t} = -\frac{\partial}{\partial x} [\alpha(x) P_t(x)] + \frac{1}{2} \frac{\partial}{\partial x} \left[\beta(x) \frac{\partial}{\partial x} [\beta(x) P_t(x)] \right] \right] \quad (3.39)$$

¹⁸This is consistent with the above discussion : the correlations of the noise are regularized as $\delta \rightarrow \delta^\epsilon$, which should correspond to a regularized Heaviside function $\theta_H^\epsilon(x) = \int_{-\infty}^x dt \delta^\epsilon(t)$. Clearly $\theta_H^\epsilon(0) = 1/2$ follows from $\delta^\epsilon(-t) = \delta^\epsilon(t)$.

✎ **Exercice 3.4 Proof of Eq. (3.39) :** We consider the Stratonovich SDE $\frac{dX}{dt} = \alpha(X) + \beta(X)\eta(t)$. For additive noise ($\beta(x) = \text{cste}$) the mapping onto the FPE is simple and has been discussed above. Difficulties arise for **multiplicative noise**. To circumvent this, we perform a transformation of the SDE which leads to additive noise. Interpreting the SDE in the Stratonovich sense, we perform the change of variable by using ordinary rules of differential calculus.

- a) Consider the map $Z(t) = \int^{X(t)} dx/\beta(x)$. Write the SDE for $Z(t)$.
- b) Give the FPE for $Q_t(z)$, the distribution of $Z(t)$.
- c) Deduce that the FPE for $P_t(x)$, the distribution of $X(t)$, is given by (3.39).

Itô SDE versus Stratonovich SDE.— With this result in hand, it is easy to answer to the second question : we can rewrite the FPE Eq. (3.39) as

$$\frac{\partial P_t(x)}{\partial t} = -\frac{\partial}{\partial x} \left[\left(\alpha(x) + \frac{1}{2}\beta(x)\beta'(x) \right) P_t(x) \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [\beta(x)^2 P_t(x)] \quad (3.40)$$

which corresponds to (3.20) if

$$a(x) = \alpha(x) + \frac{1}{2}\beta(x)\beta'(x) = \alpha(x) + \frac{1}{4}[\beta(x)^2]' \quad \text{and} \quad b(x) = \beta(x). \quad (3.41)$$

More physically, we can rewrite the relation in terms of the drift and the diffusion constant $D(x)$. Then, the Itô SDE $dX = F_{\text{Itô}}(X)dt + \sqrt{2D(X)}dW(t)$ is related to the Stratonovich SDE $dX = F_{\text{Strato}}(X)dt + \sqrt{2D(X)}dW(t)$ through

$$\boxed{F_{\text{Itô}}(x) = F_{\text{Strato}}(x) + \frac{1}{2}D'(x)} \quad (3.42)$$

The term $D'(x)/2$ is known as the “Wong-Zakai correction term”, from Ref. [?].

Remark 1 : Note that Mathematicians follow a different strategy to define the Stratonovich SDE : in 1961, Stratonovich introduced a “symmetrized” form of stochastic integrals and differential forms (at the end of this chapter, the appendix on stochastic integrals give a flavour of this strategy).

Remark 2 (regularization in practice) : The study of the Langevin equation (chapter 1) suggests another approach to the problem. A way to introduce a regular noise with $\langle \eta^\epsilon(t)\eta^\epsilon(t') \rangle = \frac{1}{2\epsilon}e^{-|t-t'|/\epsilon}$ is to add to (3.37) a differential equation for the noise. We recognize the correlations obtained in chapter 1 for the Ornstein-Uhlenbeck process. This provides a concrete way to analyze the limit $\epsilon \rightarrow 0^+$: we can study the couple of differential equations

$$\begin{cases} \partial_t X^\epsilon(t) = a(X^\epsilon) + b(X^\epsilon)\eta^\epsilon(t) \\ \epsilon \partial_t \eta^\epsilon(t) = -\eta^\epsilon(t) + \tilde{\eta}(t) \end{cases} \quad (3.43)$$

where $\tilde{\eta}(t)$ is a normalised Gaussian white noise (the first is Eq. (3.37)). In this perspective, $\eta^\epsilon(t)$ is called a “**colored noise**” for the process $X^\epsilon(t)$. The model is also known as the “*active Ornstein-Uhlenbeck process*” : it is one possible model for active particle, which are models for which the noise exhibits some persistence over time scale ϵ . Clearly, the limit $\epsilon \rightarrow 0^+$ is the white noise limit, $\eta^\epsilon(t) \rightarrow \tilde{\eta}(t)$ (cf. the discussion of the overdamped regime in chapter 1). The analysis of the two coupled SDE in this limit can be achieved thanks to a “projection method” (see [12, 53] for discussions).

To close the paragraph, I emphasize :

- when $b(x)$ is not constant (case of multiplicative noise), the two SDE $dX = a(X)dt + b(X)dW(t)$ (Itô) and $dX = a(X)dt + b(X)dW(t)$ (Stratonovich) describe two *different* processes (related to different FPEs).
- Conversely $dX = a(X)dt + b(X)dW(t)$ (Itô) and $dX = \alpha(X)dt + \beta(X)dW(t)$ (Stratonovich) describe the *same* process provided (3.41) hold (then, they are related to the same FPE).

✎ **Exercice 3.5 Stratonovich corresponds to standard differential calculus :** Let us check that the rules defined above for Stratonovich SDE are compatible with the usual rules of differential calculus for ordinary regular functions. Using the relation (3.41), transform the Itô formula (3.23) in the Stratonovich convention and check that

$$d\varphi(X(t)) = \varphi'(X(t)) dX(t) \quad (\text{Stratonovich}). \quad (3.44)$$

I.e. within the Stratonovich's prescription, standard rules of differential calculus for regular functions do apply. This is precisely the purpose of Stratonovich's interpretation.

✎ **Exercice 3.6 Correlation between the process and the noise (Stratonovich) :** Consider the Stratonovich equation (3.38).

a) Denoting $\eta(t) = dW(t)/dt$, show that $\langle \beta(X(t)) \eta(t) \rangle$ can be expressed as the average of a function of $X(t)$.

Hint : use the relation between Itô and Stratonovich SDE.

b) Let us now consider a more general problem and analyze $\langle \Phi(X(t)) \eta(t) \rangle$, where $\Phi(x)$ is an arbitrary function (in general $\Phi \neq \beta$) and where $X(t)$ solves (3.38). Noticing that the Gaussian white noise $\eta(t)$ is a Gaussian field and using the Furutsu-Novikov theorem (B.89), show that

$$\langle \Phi(X(t)) \eta(t) \rangle = \frac{1}{2} \langle \Phi'(X(t)) \beta(X(t)) \rangle \quad (\text{Stratonovich}) \quad (3.45)$$

Remark : obviously $\langle \Phi(X(t)) \eta(t) \rangle = 0$ in Itô convention.

e) **Take home message**

- Itô SDE appears as the limit of a discrete recurrence (3.15) in which the process and the noise are independent at equal times (by construction). Such a discrete time recurrence implies the existence of a finite time step, like for the evolution of assets on financial markets.
- Itô calculus is simple, however one must take care that the rules of differential calculus should be changed : expansion must be performed up to second order $d\varphi(X) = \varphi'(X) dX + \frac{1}{2} \varphi''(X) dX^2$, in order to account that dX^2 contains a contribution $\mathcal{O}(dt)$ due to the property $dW(t)^2 = dt$. We recover easily the Itô formula from this remark.
- The relation between Itô SDE with FPE is given by Eq. (3.28) (straightforward to deduce from the Itô formula).
- If a SDE appears in a physical continuous model, most frequently the white noise should be understood as a limit of a regular (non white/colored) noise with symmetric correlation function, hence the SDE must be interpreted in the Stratonovich sense. Within the Stratonovich convention, ordinary rules of differential calculus hold.

- The relation between the Stratonovich SDE (3.38) and the FPE (3.39) is important.
- Itô SDE (3.20) and Stratonovich SDE (3.38) are related through a transformation of the drift term :

$$\begin{aligned} \text{Itô} \rightarrow \text{Stratonovich} : \quad \alpha(x) &= a(x) - \frac{1}{2}b(x)b'(x) \quad \& \quad \beta(x) = b(x) \\ \text{Stratonovich} \rightarrow \text{Itô} : \quad a(x) &= \alpha(x) + \frac{1}{2}\beta(x)\beta'(x) \quad \& \quad b(x) = \beta(x) \end{aligned}$$

or more physically, Eq. (3.42).

Bibliography : More can be found in the book of Gardiner [12]. For a presentation for mathematicians, see the book [32].

✎ **Exercice 3.7 A student makes a computer simulation of a SDE :**

- a) A student from the M2 PCS is doing an internship in finance. His advisor asks him to simulate numerically the SDE controlling the evolution of an asset, $dS(t) = [r dt + \sigma dW(t)] S(t)$, where the actualization rate contains a fixed part, r , and a fluctuating part controlled by the volatility σ . What recurrence should the student implement on the computer to simulate this evolution ?
- b) Another PCS student is doing an internship in soft matter physics. His advisor asks him to simulate the equation for the position of a particle in a viscous fluid, $dX(t)/dt = F(X(t)) + \sqrt{2D(X(t))}\eta(t)$, where $\eta(t)$ is a Gaussian white noise. Same question.

f) A simple illustration : geometric Brownian motion (evolution of an asset)

In finance, the simplest model for the time evolution of an asset is to assume a growth rate with a fixed part and a random part, which corresponds to the simple SDE (“geometric Brownian motion”)

$$dX(t) = X(t) [r dt + \sigma dW(t)] \quad (\text{Itô}). \quad (3.46)$$

(the amplitude σ of the fluctuating part is the “volatility”).¹⁹ If $W(t)$ would be a regular (differentiable) function, one would simply divide the equation by X and use $d \ln X = dX/X$. This cannot be done here : within Itô calculus, the expression of $d \ln X$ should account for the second order term dX^2 . We can proceed in two equivalent manners.

Method 1 (use Itô formula) : Application of Itô formula is easy from

$$d(\ln X) = \frac{dX}{X} - \frac{dX^2}{2X^2} \quad (\text{Itô}). \quad (3.47)$$

Here we have $dX(t)^2 = \sigma^2 X(t)^2 dW(t)^2 = \sigma^2 X(t)^2 dt$. Thus

$$d \ln X(t) = \left(r - \frac{\sigma^2}{2} \right) dt + \sigma dW(t) \quad (\text{Itô}). \quad (3.48)$$

Because the noise in this new SDE is additive, the SDE can be as well be interpreted in the Stratonovich sense. One can now safely integrate the equation $\int_0^t d \ln X(t') = (r - \frac{\sigma^2}{2})t + \sigma \int_0^t dW(t')$, i.e.

$$X(t) = X(0) e^{(r - \frac{\sigma^2}{2})t + \sigma W(t)} \quad (3.49)$$

¹⁹Few typical values for the rate of return in 2025 : (i) the Livret A, which many ordinary people have, has a rate 1.7 % ($r = 1.017$), below inflation. (ii) The “assurance vie” (in €), which is known to be highly secured, has rate 2.5 % ($r = 1.025$). (iii) The 1800 super rich French people, with wealth larger than 100 M€, benefit from an average rate of return of their wealth equal to 6 % (according to Gabriel Zucman who proposed to tax them with a rate of 2 % of their wealth).

This analysis shows that one must be careful with integration of Itô differential equation.

Note that a naive integration of (3.46) with the usual rules of integration (i.e. interpreting (3.46) within the Stratonovich sense), would have missed an exponential factor $e^{-\sigma^2 t/2}$.

Method 2 (use the relation to the Stratonovich SDE) : Using the rule given above, we find that the Itô SDE is related to the Stratonovich SDE

$$dX(t) = \left(r - \frac{\sigma^2}{2}\right) X(t) dt + \sigma X(t) dW(t) \quad (\text{Stratonovich}). \quad (3.50)$$

Now, we can divide the Stratonovich SDE by $X(t)$, use $dX/X = d \ln X$ (usual rules of differential calculus) and integrate. We obtain the same result, Eq. (3.49), as it should.

Let us stress an advantage of the Itô SDE : Using independence of $X(t)$ and the noise $dW(t)$ at the same time, we can average the Itô equation, leading to $\frac{d}{dt} \langle X(t) \rangle = r \langle X(t) \rangle$, hence $\langle X(t) \rangle = \langle X(0) \rangle e^{rt}$. This is pretty straightforward.

We can recover this result as follows : we average the solution (3.49) for fixed $X(0)$ in order to check that it is consistent with this result :

$$\langle X(t) \rangle = X(0) e^{\left(r - \frac{\sigma^2}{2}\right)t} \left\langle e^{\sigma W(t)} \right\rangle = X(0) e^{\left(r - \frac{\sigma^2}{2}\right)t + \frac{1}{2}\sigma^2 \langle W(t)^2 \rangle} = X(0) e^{rt} \quad (3.51)$$

where we have used the Gaussianity of $W(t)$. It was more direct to average the Itô SDE.

Historical note on Doblin-Itô calculus

Until 2000, Itô was considered as the founder of what is usually denoted today the “Itô calculus”. However in 2000, a sealed envelope (“pli cacheté” number 11-668), received in 1940 from a young mathematician named Vincent Doblin (born Wolfgang Döblin), was opened at Académie des Sciences de Paris, which showed that Doblin’s contribution anticipated the work of Itô on stochastic calculus. Hence, we should rather name it “Doblin-Itô calculus”.



Figure 3.3: Vincent Doblin (1915-1940). A page of the pli cacheté (from [5]).

Wolfgang Döblin was the son of a well-known german writer, Alfred Döblin. Because he was jewish and opponent to the nazism, Alfred Döblin escaped Germany to Zürich at the begining of 1933 with part of his family, followed by his son Wolfgang. They arrived in Paris in the fall of 1933. Wolfgang obtained the french nationality in 1936, becoming “Vincent Doblin”. In 1938 he passed his PhD, under the supervision of the famous mathematician Maurice Fréchet (Fig. ??), however, at the end of 1938, he was incorporated in the French army. Refusing to serve as an officer, he was affected to the communications. During this period in the army, at the begining of the war, he was sent to the Ardennes and was able to produce important scientific results, which he chose to send to the Académie des Sciences under the form of a “pli cacheté”, entitled “sur l’équation de Kolmogoroff, par Vincent Doblin”. Just after the collapse

of the French army, as his company was surrounded by Germans in the Vosges region, Vincent Doblin tried unsuccessfully to cross the German lines and eventually preferred to commit suicide rather than being captured. It was only possible to open the “pli cacheté” 60 years after his death. Although Vincent Doblin was already known in the mathematics community despite his youth, the importance of his contribution was not anticipated before 2000.

To learn more : look at the article [5] (available on the internet) written by the two probabilists Bernard Bru and Marc Yor, who analyzed the pli cacheté and recognized its scientific importance. Or the book by Marc Petit [38].

APPENDIX : Stochastic integrals

If you feel unsatisfactory with the above presentation of Itô/Stratonovich convention, you can read this paragraph (borrowed from chapter 4 of [12]). Instead of considering the SDE, one considers integrals of the form $\int_0^t dW(t') G(t')$ which requires the same discussion as for SDE.

Itô integral.— One defines the Itô integral as

$$\text{Itô} \int_0^t dW(t') G(t') \stackrel{\text{def}}{=} \text{ms-lim}_{N \rightarrow \infty} \sum_{i=1}^N \delta W_i G(t_{i-1}) \quad (3.52)$$

where $\delta W_i = W(t_i) - W(t_{i-1})$. Here “ms-lim” stands for “mean-square limit” of a random variable, meaning that :

$$\text{ms-lim}_{N \rightarrow \infty} X_N = X_\infty \quad \text{if} \quad \lim_{N \rightarrow \infty} \langle [X_N - X_\infty]^2 \rangle = 0. \quad (3.53)$$

Let us study an example. Consider the integral $\text{Itô} \int_0^t dW(t') W(t')$. One has to analyze the sum

$$\begin{aligned} \sum_{i=1}^N \delta W_i W_{i-1} &= \frac{1}{2} \sum_{i=1}^N [(\delta W_i + W_{i-1})^2 - W_{i-1}^2 - \delta W_i^2] = \frac{1}{2} \sum_{i=1}^N W_i^2 - \frac{1}{2} \sum_{i=0}^{N-1} W_i^2 - \frac{1}{2} \sum_{i=1}^N \delta W_i^2 \\ &= \frac{1}{2} [W_N^2 - W_0^2] - \frac{1}{2} \sum_{i=1}^N \delta W_i^2 \end{aligned} \quad (3.54)$$

It is easy to show that $\text{ms-lim}_{N \rightarrow \infty} \sum_{i=1}^N \delta W_i^2 = t$ (this is the reason why one writes $dW(t)^2 = dt$ without the average). Thus

$$\text{Itô} \int_0^t dW(t') W(t') = \frac{1}{2} [W(t)^2 - W(0)^2 - t] \quad (3.55)$$

which differs (by $-t/2$) from the usual Riemann integral of a regular function.

Stratonovich integral.— Now introduce the definition of the Stratonovich integral

$$\int_0^t dW(t') G(t') \stackrel{\text{def}}{=} \text{ms-lim}_{N \rightarrow \infty} \sum_{i=1}^N \delta W_i \frac{G(t_i) + G(t_{i-1})}{2} \quad (3.56)$$

(I use the standard notation for integration, anticipating that it will coincide with usual Riemann integrals).

Consider now the same integral as before $\int_0^t dW(t') W(t')$ with the new convention. This time, one deals with

$$\sum_{i=1}^N \delta W_i \frac{W_i + W_{i-1}}{2} = \frac{1}{2} \sum_{i=1}^N W_i^2 - \frac{1}{2} \sum_{i=1}^N W_{i-1}^2 = \frac{1}{2} W_N^2 - \frac{1}{2} W_0^2 \quad (3.57)$$

so that we have recovered

$$\int_0^t dW(t') W(t') = \frac{1}{2} [W(t)^2 - W(0)^2] \quad (3.58)$$

as for the integration of regular functions.

From stochastic integral to SDE : in the books [32, 12], stochastic integrals are first discussed along these lines, then SDE are introduced as derivatives of stochastic integrals.

☺ Important points

- Understand the difference between Itô and Stratonovich conventions (for multiplicative noise).
- Remember $dW(t)^2 = dt$ and $dW(t)^{n+2} = 0$.
- For Itô calculus : be careful with differential calculus ! Be able to recover the Itô formula.
- Be familiar with the relations between SDE (Itô or Stratonovich) and the FPE.

Additional exercises

✎ **Exercise 3.8 Vanishing of the mean velocity :** Consider the SDE for the drift $F(x)$ and the x -dependent diffusion constant $D(x)$:

$$dX(t) = F(X) dt + \sqrt{2D(X)} dW(t) \quad (\text{Stratonovich}). \quad (3.59)$$

What is the condition ensuring that $\frac{d}{dt} \langle X(t) \rangle = 0$?

✎ **Exercise 3.9 Correlation between $X(t)$ and $W(t)$ at coinciding times :** Consider a process $X(t)$ which obeys the Itô SDE (3.20). We have $\langle dW(t) X(t) \rangle = 0$. The aim of the exercise is to study the correlation $C_{XW}(t) = \langle X(t) W(t) \rangle = \int^t \langle dW(t') X(t) \rangle$.

a) Show that the correlator obeys

$$\frac{d}{dt} \langle X(t) W(t) \rangle = \langle a(X(t)) W(t) \rangle + \langle b(X(t)) \rangle. \quad (3.60)$$

Hint : Consider $X(t)$ and $Y(t)$ two stochastic processes defined by two Itô SDEs. Write the Itô formula for $d\varphi(X, Y)$. Apply this for $Y = W$.

b) *Additive noise :* Apply this to the Ornstein-Uhlenbeck process $dX(t) = -k X(t) dt + \sigma dW(t)$.

c) Discuss also the case where $k \rightarrow -k$. Analyze $\langle X(t) W(t) \rangle / \sqrt{\langle X(t)^2 \rangle_c \langle W(t)^2 \rangle}$.

d) *Multiplicative noise :* Consider now the Itô SDE $dX(t) = X(t) [r dt + \sigma dW(t)]$ (studied below). Analyze $\langle X(t) W(t) \rangle / \sqrt{\langle X(t)^2 \rangle_c \langle W(t)^2 \rangle}$.

✎ **Exercise 3.10 Ornstein-Uhlenbeck process for time-dependent stiffness and the Brownian bridge :** We consider the Ornstein-Uhlenbeck process (harmonic oscillator in the overdamped regime) for a time-dependent stiffness

$$\frac{dX(t)}{dt} = -k(t) X(t) + \eta(t) \quad (3.61)$$

where the stiffness $k(t)$ is a known function and $\eta(t)$ a normalised Gaussian white noise.

a) Solve the SDE for $X(0) = x_0$; it is convenient to introduce $K(t) = \int_0^t du k(u)$.

b) Deduce $\langle X(t) \rangle$ and $\text{Var}[X(t)]$ (for x_0 fixed). Give $P(x, t|x_0, 0)$.

c) **Application for the Brownian bridge :** We now consider the SDE

$$\frac{dX(t)}{dt} = -\frac{X(t)}{T-t} + \eta(t) \quad \text{for } t \in [0, T] \quad (3.62)$$

Give the integral representation of the solution. Discuss $X(t)$ for $t \rightarrow T^-$.

d) Deduce $\langle X(t) \rangle$ and the correlator $C(t, t') = \langle X(t)X(t') \rangle$. What is the nature of the distribution ?

e) We now introduce the process $B(t) = W(t) - W(T)(t/T)$, where $W(t)$ is the Wiener process, with $W(0) = 0$. Plot a typical realization of $B(t)$. Compute $\langle B(t) \rangle$ and $C_B(t, t') = \langle B(t)B(t') \rangle$. Comment.

f) Propose the SDE describing a Bridge with $X(0) = x_0$ and $X(T) = x_f$. Give the integral representation of the solution.

✎ **Exercice 3.11 Hyperbolic Brownian motion :** The hyperbolic Poincaré half plane $\mathbb{H}_2$ is a surface of constant negative curvature with metric $(dx^2 + dy^2)/y^2$, for $x \in \mathbb{R}$ and $y \in \mathbb{R}^+$. It is possible to analyze the free diffusion on $\mathbb{H}_2$, which can be described by two coupled SDE

$$\begin{cases} dx = \sqrt{2D} y dW_x(t) \\ dy = \sqrt{2D} y dW_y(t) \end{cases} \quad (\text{It}\hat{o}) \quad (3.63)$$

where $W_x(t)$ and $W_y(t)$ are two independent Wiener processes, i.e. $dW_x(t)^2 = dW_y(t)^2 = dt$ and $dW_x(t)dW_y(t) = 0$.

a) **Preliminary (Stratonovich/Itô conversion trick) :** Starting from the Stratonovich SDE $dx = \alpha(x) dt + b(x) dW(t)$, check that we can write the corresponding Itô SDE as

$$dx = \alpha(x) dt + b\left(x + \frac{1}{2}dx_{\text{noise}}\right) dW(t) \quad (\text{It}\hat{o}) \quad (3.64)$$

where $dx_{\text{noise}} = b(x) dW(t)$. Give the corresponding trick allowing to get rapidly the Stratonovich SDE corresponding to the Itô SDE $dx = a(x) dt + b(x) dW(t)$.

b) Deduce the Stratonovich SDEs describing hyperbolic Brownian motion.

c) Which set of SDEs is the most convenient for integration ? Deduce two integral representations for $x(t)$ and $y(t)$ for initial conditions $x(0) = 0$ and $y(0) = 1$.

d) Deduce $\langle x(t) \rangle$ and $\langle y(t) \rangle$. Compute also $\text{Var}[x(t)]$ and $\text{Var}[y(t)]$. Comment.

✎ **Exercice 3.12 Electromagnetic noise :** We consider a model of electromagnetic noise : the two components of the electric field $E_x + iE_y$ obey the two uncoupled SDE

$$\begin{cases} dE_x(t) = -\gamma E_x(t) dt + \sqrt{D} dW_x(t) \\ dE_y(t) = -\gamma E_y(t) dt + \sqrt{D} dW_y(t) \end{cases} \quad (3.65)$$

where W_x and W_y are two independent Wiener processes, hence we can write

$$dW_x^2 = dW_y^2 = dt \text{ and } dW_x dW_y = 0 \quad (3.66)$$

(remember that averages can be omitted for elementary differential increments).