# Chapter 3

# Introduction to stochastic processes

In this chapter we review the basic concepts of what a stochastic process is. Our aim is not to be rigorous on the mathematical side but rather to focus on the physical insights behind the concepts. The name "stochastic process" is usually associated to a trajectory in phase space which is random enough to demand a probabilistic description. A paradigmatic example is that of the Brownian motion.

## 3.1 Brownian motion

The botanist Robert Brown discovered in 1827 that particles in a pollen in suspension execute random movements which he even interpreted initially as some sort of life. It is not so well known that L. Boltzmann knew as early as 1896 the reason for this erratic movement when he wrote "... very small particles in a gas execute motions which result from the fact that the pressure on the surface of the particles may fluctuate". However, it was A. Einstein in 1905 who successfully introduced the first mathematical treatment of the erratic movement of the Brownian particles. Rather than focusing on the (complicated) trajectory of a single particle, Einstein introduced a probabilistic description valid for an ensemble of Brownian particles. First Einstein introduced the concept of a coarse–grained description defined by a time $\tau$ such that different parts of the trajectory separated by a time $\tau$ or larger can be considered independent. No attempt is made to characterize the dynamics at a time scale smaller than this coarse–grain time $\tau$. Instead we consider snapshots of the system taken at time intervals $\tau$ (see Figure 3.1).

The second concept, probabilistic in nature, introduced by Einstein is that of the probability density function, $f(\vec{\Delta})$, for the three-dimensional distance $\vec{\Delta} = (\Delta_x, \Delta_y, \Delta_z)$ traveled by the Brownian particle in a time interval $\tau$. One could assume, for instance, a Gaussian form for $f(\vec{\Delta})$, but this it not necessary. In fact, the only assumption one needs about the function $f(\vec{\Delta})$ (besides the general condition of non-negativity and normalization) comes from the fact that the collisions of the fluid molecules and the Brownian particle occur with the same probability in any direction[1]. The absence of

[1]We also disregard the effect of gravity in the Brownian particle which would lead to a preferred

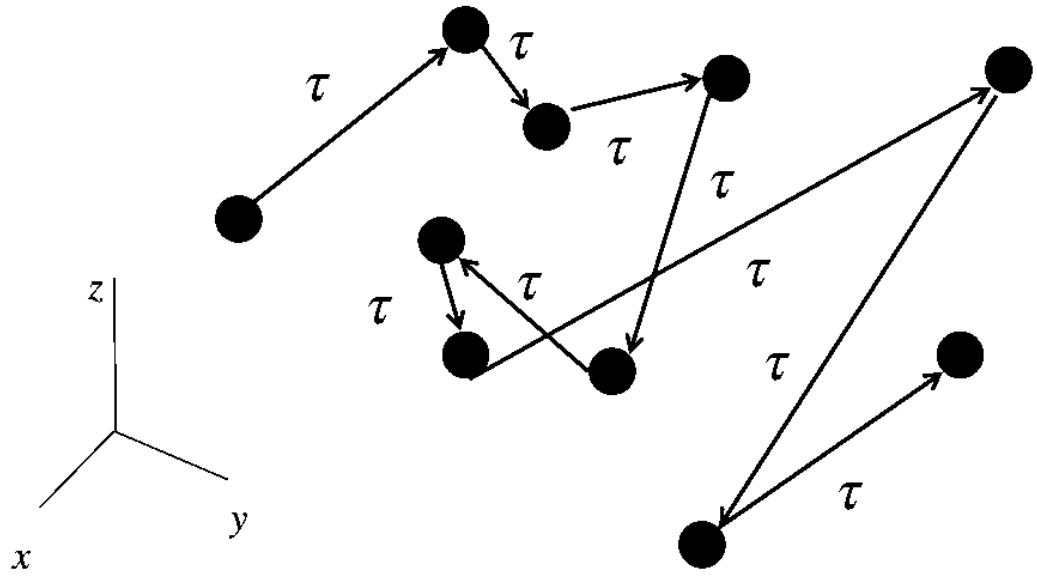


Figure 3.1: Schematic representation of the movement of a Brownian particle

preferred directions translates to a symmetry condition for $f(\vec{\Delta})$:

$$f(-\vec{\Delta}) = -f(\vec{\Delta}). \tag{3.1}$$

The third step in this description is to consider an ensemble of $N$ Brownian particles in a large enough system. Also we focus on spatial scales which are much larger than the size of a Brownian particle so that we can define a density of particles $n(\vec{x}, t)$ such that $n(\vec{x}, t)d\vec{x}$ is the number of particles in the interval $(\vec{x}, \vec{x} + d\vec{x})$ at time $t$. From the assumption that the parts of the trajectories separated a time interval $\tau$ are statistically independent, it follows that the number of particles at location $\vec{x}$ at time $t + \tau$ will be given by the number of particles at location $\vec{x} - \vec{\Delta}$ at time $t$ multiplied by the probability that the particle jumps from $\vec{x} - \vec{\Delta}$ to $\vec{x}$ which is $f(\vec{\Delta})$ and integrated for all the possible $\vec{\Delta}$ values:

$$n(\vec{x}, t + \tau) = \int_{\mathbb{R}^3} n(\vec{x} - \vec{\Delta}, t) f(\vec{\Delta})\, d\vec{\Delta}. \tag{3.2}$$

This is the basic evolution equation for the number density $n(\vec{x}, t)$. From the physical point of view it is a continuity equation whose interpretation is that Brownian particles can not be created nor can they disappear due to the collisions with the fluid molecules. By Taylor expanding the above expression and making use of the symmetry relation eq.(3.1) one gets the diffusion equation:

$$\frac{\partial n}{\partial t} = D\nabla^2 n, \tag{3.3}$$

direction in the movement.

where the *diffusion constant* $D$ is given in terms of the pdf $f(\vec{\Delta})$ by:

$$D = \frac{1}{2\tau}\int_{-\infty}^{\infty} \vec{\Delta}^2 f(\vec{\Delta}) d\vec{\Delta} = \frac{\langle \vec{\Delta}^2 \rangle}{2\tau} \tag{3.4}$$

If the initial condition is that all $N$ particles are located at the origin, $n(\vec{x}, t=0) = N\delta(\vec{x})$ the solution of the diffusion equation is:

$$n(\vec{x}, t) = \frac{N}{(4\pi D t)^{3/2}} \mathrm{e}^{-\vec{x}^2/4Dt}, \tag{3.5}$$

from where it follows that the average position of the Brownian particle is $\langle x(t) \rangle = 0$ and that the average square position increases linearly with time, namely:

$$\langle \vec{x}(t)^2 \rangle = 6Dt. \tag{3.6}$$

This prediction has been successfully confirmed in experiments and contributed to the acceptance of the atomic theory.

However successful Einstein's approach was, it is very phenomenological and it can not yield, for instance, an explicit expression that allows the calculation of the diffusion constant in terms of microscopic quantities. Langevin (1908) initiated a different approach which, in some way, can be considered complementary of the previous one. In his approach, Langevin focused on the trajectory of a single Brownian particle and wrote down Newton's equation Force = mass $\times$ acceleration. The trajectory of the Brownian particle is highly erratic and therefore its description would demand a peculiar kind of force. Langevin considered two types of forces acting on the Brownian particle: usual friction forces that, according to Stokes law, would be proportional to the velocity, and a sort of "fluctuating" force $\vec{\xi}(t)$ which represents the "erratic" force which comes from the action of the fluid molecules on the Brownian particle. The equation of motion becomes then:

$$m\frac{d\vec{v}}{dt} = -6\pi\eta a\vec{v} + \vec{\xi}, \tag{3.7}$$

$\eta$ is the viscosity coefficient and $a$ is the radius of the Brownian particle (which is assumed to be spherical). Multiplying both sides of eq.(3.7) by $\vec{x}$ one gets

$$\frac{m}{2}\frac{d^2\vec{x}^2}{dt^2} - m\left(\frac{d\vec{x}}{dt}\right)^2 = -3\pi a\eta\frac{d\vec{x}^2}{dt} + \vec{x}\cdot\vec{\xi}. \tag{3.8}$$

Langevin made two assumptions about the fluctuating force $\vec{\xi}(t)$: that is has mean $0$ (collisions do not push the Brownian particle in any preferred direction) and that it is uncorrelated to the actual position of the Brownian particle (the action of the molecules of fluid on the Brownian particle is the same no matter the location of the Brownian particle):

$$\begin{aligned} &\langle \vec{\xi}(t) \rangle = 0, \\ &\langle \vec{x}\cdot\vec{\xi} \rangle = \langle \vec{x} \rangle \cdot \langle \vec{\xi} \rangle = 0. \end{aligned} \tag{3.9}$$

Taking averages in Eq. (3.8) with respect to all realizations of the random force $\vec{\xi}(t)$ and using the previous conditions on $\vec{\xi}(t)$ one gets:

$$\frac{m}{2}\frac{d^2\langle\vec{x}^2\rangle}{dt^2} = m\langle\vec{v}\rangle - 3\pi a\eta\frac{d\langle\vec{x}^2\rangle}{dt}, \tag{3.10}$$

which is an equation for the average square position of the Brownian particle. Langevin assumed that we are now in a regime in which thermal equilibrium between the Brownian particle and the surrounding fluid has been reached. In particular, this implies that, according to the equipartition theorem, the average kinetic energy of the Brownian particle is $\langle m\vec{v}^2/2\rangle = 3kT/2$ ($k$ is Boltzmann's constant and $T$ is the fluid temperature). One can now solve Eq. (3.10) and find that, after some transient time, the asymptotic mean square displacement is given by:

$$\langle\vec{x}^2\rangle = \frac{kT}{\pi\eta a}t \tag{3.11}$$

This is nothing but Einstein's diffusion law, but now we have an explicit expression for the diffusion coefficient in terms of other macroscopic variables:

$$D = \frac{kT}{6\pi\eta a} \tag{3.12}$$

Langevin's random force $\vec{\xi}(t)$ is an example of a stochastic process. It is time we proceed to a more precise definition of what a stochastic process is. The natural machinery is that of probability theory.

## 3.2 Stochastic Processes

In Chapter 1 we have introduced the concept of a random variable $\hat{x}$ resulting from a probabilistic experiment. We now define a stochastic process as a family $\hat{x}(t)$ of random variables depending on some continuous real parameter $t$. In most applications, $t$ is a physical time and the stochastic process can be thought as performing multiple probabilistic experiments one at each time instant. The trajectory followed by the system depends on the outcome of each probabilistic experiment. As a consequence the knowledge of the initial condition $x_0$ at time $t_0$ is not enough to determine the position in phase space of the system at a later time $t_1$. Instead, the trajectory, and therefore the final state of the system, acquires a probabilistic nature. To fully determine the final state, it is also necessary to know the outcome of all the successive probabilistic experiments between the initial time $t_0$ and the final time $t_1$. In fact, each possible set of successive outcomes determines a possible trajectory for the system, all starting at $x_0$ at time $t_0$ but ending at different locations at time $t_1$. The stochastic process can be seen as the collection of all these possible trajectories.

Probably the most well–known example of a stochastic process is that of the random walk. The probabilistic experiment is now a series of binary results representing, for instance, the outcome of repeatedly tossing a coin:

$$(0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, \dots)$$

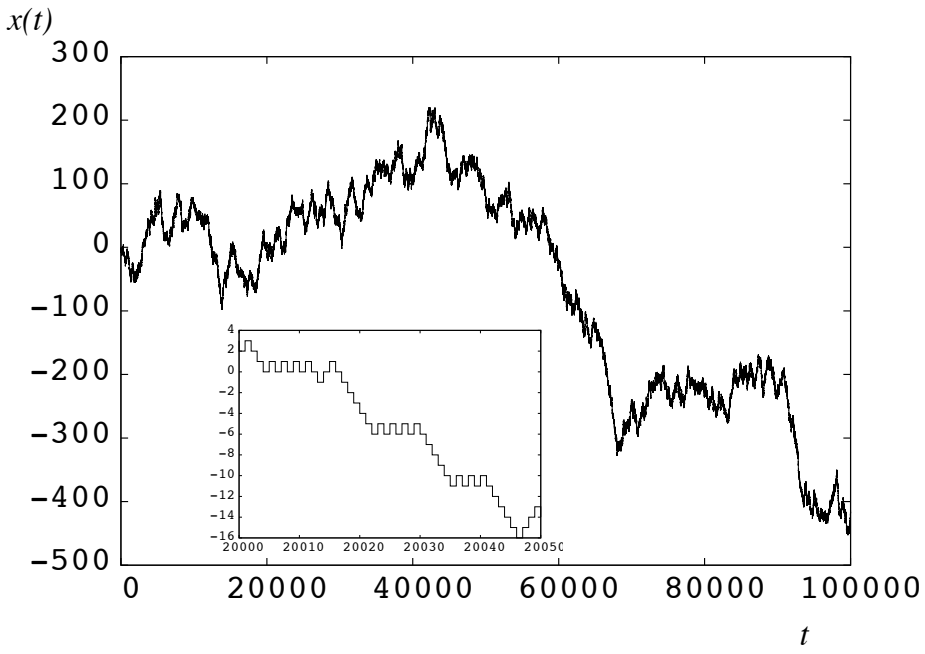


Figure 3.2: Example of a random walk trajectory. We have taken $\tau = a = 1$ and plotted the resulting trajectory after a large number, $10^5$, of steps. In the insert, we see the fine detail with the discrete jumps occurring at times multiples of $\tau$.

where $1$ means "heads" and $0$ means "tails". Consider that the tossing takes place at given times $0, \tau, 2\tau, ....$ To the outcome of this set of probabilistic experiments we associate a one-dimensional function $x(t)$ which starts at $x(0) = 0$ and that moves to the left (right) at time $k\tau$ an amount $a$ $(-a)$ if the $k$–th result of the tossed coin was 0 (1). In the intermediate times between two consecutive tossings, namely in the times between $k\tau$ and $(k+1)\tau$ the system just remains in the same location. Figure 3.2 shows the trajectory for the above result $u$.

What does one mean by characterizing a stochastic process? Since it is nothing but a continuous family of random variables, a stochastic process will be completely characterized when we give the joint probability density function for the arbitrary set $\{\hat{x}(t_1), \hat{x}(t_2), \ldots, \hat{x}(t_m)\}$, i.e. when we give the function $f(x_1, \ldots, x_m; t_1, \ldots, t_m)$ for arbitrary $m$. This function is such that

$$f(x_1, \ldots, x_m; t_1, \ldots, t_m) dx_1 \ldots dx_m \tag{3.13}$$

represents the probability that the random variable $\hat{x}(t_1)$ takes values in the interval $(x_1, x_1 + dx_1)$, the random variable $\hat{x}(t_2)$ takes values in the interval $(x_2, x_2 + dx_2)$, etc.[2]

[2]In a different language, we can say that a complete characterization of the trajectory is obtained by giving the functional probability density function $f([x(t)])$.

We can see here a generalization of the successions of random variables presented in 1.4. As $t$ is a continuous variable, we have here formally a non-numerable infinite number of random variables $\hat{x}(t)$. However, when we extract from this a finite set $\hat{x}(t_1), \ldots, \hat{x}(t_m)$ we can use the same definitions and results than in the case of a succession of random variables. For example, we extend now the Markov property to a stochastic process.

A stochastic process is said to be a Markov process if the rather general conditional probability

$$f(x_m; t_m | x_1, \ldots, x_{m-1}; t_1, \ldots, t_{m-1}) \equiv \frac{f(x_1, \ldots, x_m; t_1, \ldots, t_m)}{f(x_1, \ldots, x_{m-1}; t_1, \ldots, t_{m-1})} \tag{3.14}$$

is equal to the two–times conditional probability

$$f(x_m; t_m | x_{m-1}; t_{m-1}) \equiv \frac{f(x_{m-1}, x_m; t_{m-1}, t_m)}{f(x_{m-1}; t_{m-1})} \tag{3.15}$$

for all times $t_m > t_{m-1} > \ldots t_1$. This Markov property that, loosely speaking, means that the probability of a future event depends only on the present state of the system and not on the way it reached its present situation, allows to compute the $m$–times pdf as:

$$\begin{aligned} &f(x_1, \ldots, x_m; t_1, \ldots, t_m) = \\ &f(x_m; t_m | x_{m-1}; t_{m-1}) f(x_{m-1}; t_{m-1} | x_{m-2}; t_{m-2}) \ldots f(x_2; t_2 | x_1; t_1) f(x_1; t_1). \end{aligned} \tag{3.16}$$

The random walk constitutes an example of Markov process, since the probability of having a particular value of the position at time $(k+1)\tau$ depends only on the particle location at time $k\tau$ and not on the way it got to this location.

A particular case of Markov process is a complete independent process in which an arbitrary set of random variables at different times are independent and then we are able to write:

$$f(x_1, \ldots, x_m; t_1, \ldots, t_m) = f(x_1; t_1) f(x_2; t_2) \ldots f(x_m; t_m). \tag{3.17}$$

Another important example of random process is that of Gaussian process in which the $m$–times pdf admits an explicit form generalizing 1.5, namely:

$$f(x_1, \ldots, x_m; t_1, \ldots, t_m) = \sqrt{\frac{|S|}{(2\pi)^m}} \exp\left[ -\frac{1}{2} \sum_{i,j=1}^{m} (x_i - b_i) S_{ij} (x_j - b_j) \right], \tag{3.18}$$

where

$$\begin{aligned} b_i &= \langle x(t_i) \rangle, \\ (S^{-1})_{ij} &= \langle x(t_i) x(t_j) \rangle - \langle x(t_i) \rangle \langle x(t_j) \rangle. \end{aligned} \tag{3.19}$$

For Gaussian processes, the explicit form of the pdf (3.18) is rarely written, rather the process is characterized giving the mean value $\langle \hat{x}(t) \rangle$ and the correlation function $\langle \hat{x}(t) \hat{x}(t') \rangle$.

## 3.3 Stochastic Differential Equations

A stochastic differential equation is a differential equation which contains a stochastic process $\hat{\eta}(t)$, that is an equation of the form

$$\frac{d\hat{x}(t)}{dt} = G(\hat{x}, t, \hat{\eta}(t)), \tag{3.20}$$

where $G$ is a given function that depends, in general, on the variable $x(t)$, on the time $t$ and on the stochastic process $\hat{\eta}(t)$. A stochastic differential equation can be seen as a family of ordinary differential equations, one for each outcome of all the successive probabilistic experiments associated to the stochastic process $\hat{\eta}(t)$. As a consequence for any given initial condition $x_0$ at time $t_0$ one has a family of possible trajectories. Therefore $\hat{x}(t)$, which is the collection of all these possible trajectories, has to be viewed also as a stochastic process and this is why we label it with the "hat" symbol. However, $\hat{x}(t)$ is not an arbitrary stochastic process, rather it depends on $\hat{\eta}(t)$ in a specific manner determined by the stochastic differential equation and, as a consequence, the statistical properties of $\hat{x}(t)$ depend on the statistical properties of $\hat{\eta}(t)$.

Strictly speaking, "solving the stochastic differential equation" means to provide the complete characterization of the stochastic process $\hat{x}(t)$, namely to give all the $m$–times pdfs $f(x_1, \ldots, x_m; t_1, \ldots, t_m)$, in terms of the statistical properties of $\hat{\eta}(t)$. However, one has to understand that a complete characterization of a general stochastic process implies the knowledge of a function of an arbitrary number of parameters and is very difficult to carry out in practice. In many occasions one is happy if can give just the one–time pdf $f(x;t)$ and the two–times pdf $f(x_1, x_2; t_1, t_2)$. In terms of those, it is possible to compute trajectory averages:

$$\langle \hat{x}(t)^n \rangle = \int_{-\infty}^{\infty} dx\, x^n f(x;t), \tag{3.21}$$

and time correlations:

$$\langle \hat{x}(t_1)\hat{x}(t_2) \rangle = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, x_1 x_2 f(x_1, x_2; t_1, t_2). \tag{3.22}$$

In general the function $G$ can depend on the stochastic process $\hat{\eta}(t)$ in an arbitrary way. However many physical, chemical or biological systems can be described by stochastic differential equations in which $\hat{\eta}(t)$ appears linearly, namely

$$\frac{d\hat{x}}{dt} = q(\hat{x}) + g(\hat{x})\hat{\eta}(t). \tag{3.23}$$

This kind of stochastic differential equations, which are the only ones to be considered in this book, are called Langevin equations. In this case the independent stochastic process $\hat{\eta}(t)$ is usually refereed as "noise", notation that comes from the early days of radio broadcasting when the random fluctuations in the electrical signals taking place in the emitter, during the propagation in the atmosphere or at the receiver device lead to noises that were actually heard on top of the radio emission. Following this notation,

in Eq. (3.23) the term $g(\hat{x})\hat{\eta}(t)$ is refereed as the "noise term" while $q(\hat{x})$ is the "deterministic term" or "drift term". One distinguishes the case in which the function $g(\hat{x})$ is a constant, in which the noise is said to be *additive*. Otherwise, the noise is said to be *multiplicative*.

For the sake of simplicity in the notation, from now on we will drop the "hats" from the stochastic process and therefore we write the Langevin differential equation as

$$\frac{dx}{dt} = q(x) + g(x)\eta(t). \tag{3.24}$$

We have already encountered an example of Langevin differential equation in Section 3.1, Eq. (3.7), the equation introduced by Langevin himself to describe the movement of a Brownian particle. In this case the independent stochastic process is the random force that acts on the Brownian particle and models the collisions of the water molecules and it appears in the equation as an additive noise and the deterministic term is the drag induced by the water viscosity. And, somehow, we have already "solved" Langevin when we have determined some statistical properties of the movement of the Brownian particle, such as the mean square displacement.

## 3.4 White noise

We proceed now to characterize the stochastic process $\xi(t)$ that appears in the Langevin equation for the Brownian motion. To do this we first start with the characterization of another process we have already encountered, the one-dimensional random walk. Starting at $x = 0$ at time $t = 0$, the location of the random walker after tossing the coin $n$ times is given by the number of steps taken in the sense that $x$ increases $n_1$ (number of "heads") minus the number of steps taken in the opposite sense $n_0$ (number of "tails"), $x(n\tau) = (n_1 - n_0)a = (2n_1 - n)a$. The probability of having $n_1$ "heads" after $n$ throws is given by the binomial expression

$$P(n_1) = \binom{n}{n_1} 2^{-n}. \tag{3.25}$$

Therefore, the probability that the walker is at a location $x = ra$ after a time $t = n\tau$ is given by

$$P(x(n\tau) = ra) = \binom{n}{\frac{n+r}{2}} 2^{-n}. \tag{3.26}$$

From which it follows using 1.31-1.32:

$$\begin{aligned} \langle x(n\tau)\rangle &= 0, \\ \langle x(n\tau)^2\rangle &= na^2. \end{aligned} \tag{3.27}$$

As explained in section 1.3 for $n \gg 1$ the binomial distribution can be approximated by a Gaussian distribution (de Moivre–Laplace theorem):

$$P(x(n\tau) \leq ra) = \frac{1}{2} + \operatorname{erf}\left(\frac{r}{\sqrt{n}}\right). \tag{3.28}$$

We now take the continuum limit defined by:

$$n \to \infty,\ \tau \to 0,\ r \to \infty,\ a \to 0, \tag{3.29}$$

while preserving a finite value for

$$t = n\tau,\ x = ra,\ D = a^2/\tau. \tag{3.30}$$

In this limit the random walk process is called the Wiener process $W(t)$ and equation (3.28) can be written as:

$$P(W(t) \leq x) = \frac{1}{2} + \mathrm{erf}\left(\frac{x}{\sqrt{Dt}}\right), \tag{3.31}$$

which is the probability distribution function of a Gaussian variable with zero mean and variance $Dt$. The corresponding probability density function is:

$$f(x;t) = \frac{1}{\sqrt{2\pi Dt}} \exp\left(-\frac{x^2}{2Dt}\right). \tag{3.32}$$

The Wiener process inherits the Markovian and the Gaussian character of the random walk process. Therefore it can be characterized by giving its mean value and the two times correlation function, which can be computed as:

$$\langle W(t) \rangle = 0, \tag{3.33}$$

$$\langle W(t_1)W(t_2) \rangle = D \min(t_1, t_2). \tag{3.34}$$

A plot of a typical realization of the Wiener process is shown in figure 3.3.

The random walk process was a sequence of step functions. As a consequence the Wiener process is continuous but it does not have a well defined first derivative[3]. Still we will define now the *white–noise* random process $\xi(t)$ as the derivative of the Wiener process. Since we just said that the Wiener process does not have a well-defined derivative, it is not surprising that the result depends on the way the derivative is performed. We first go back to the discrete random walk process $x(t)$ and define a new stochastic process $w_\epsilon$ as:

$$w_\epsilon(t) = \frac{x(t+\epsilon) - x(t)}{\epsilon}. \tag{3.35}$$

A sketch of the process $w_\epsilon$ is given in figure 3.4. $w_\epsilon(t)$ is a Gaussian process since it is a linear combination of Gaussian processes. Therefore, it is sufficiently defined by its mean and correlations:

$$\langle w_\epsilon(t) \rangle = 0, \tag{3.36}$$

$$\langle w_\epsilon(t_1)w_\epsilon(t_2) \rangle = \begin{cases} 0, & t_1 - t_2 < -\epsilon, \\ a^2/(\tau\epsilon)(1 + (t_1 - t_2)/\epsilon), & -\epsilon \leq t_1 - t_2 \leq 0, \\ a^2/(\tau\epsilon)(1 - (t_1 - t_2)/\epsilon), & 0 \leq t_1 - t_2 \leq \epsilon, \\ 0, & t_1 - t_2 > \epsilon. \end{cases} \tag{3.37}$$

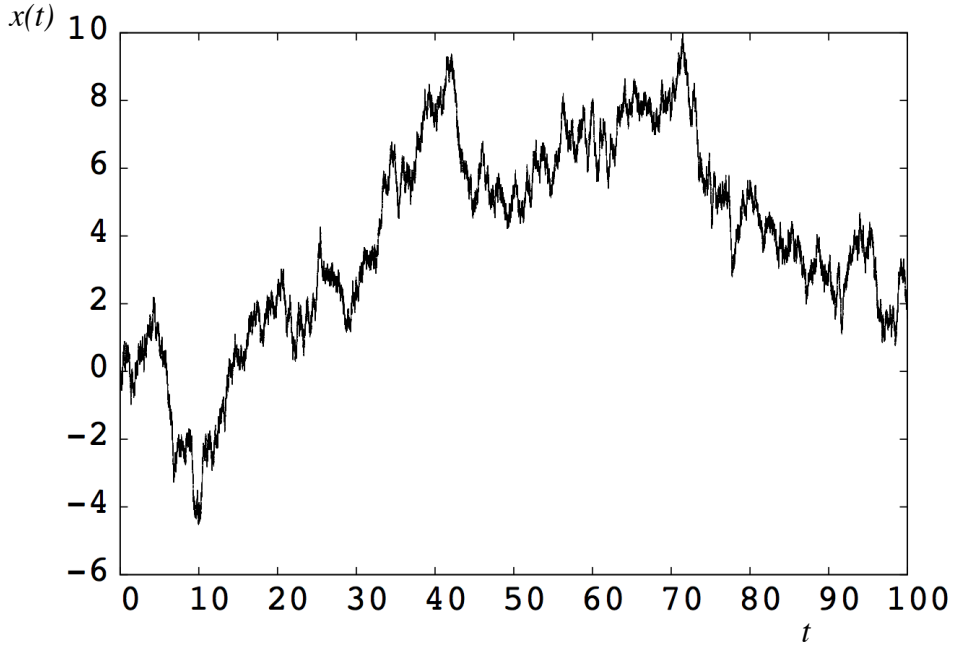


Figure 3.3: A typical realization of the Wiener process generated using the random walk with $\tau = 10^{-4}$ and $a = 10^{-2}$.

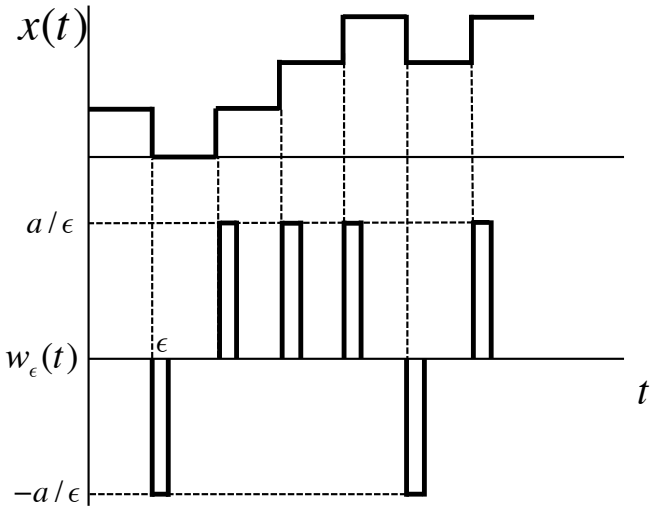


Figure 3.4: Random walk and its derivative

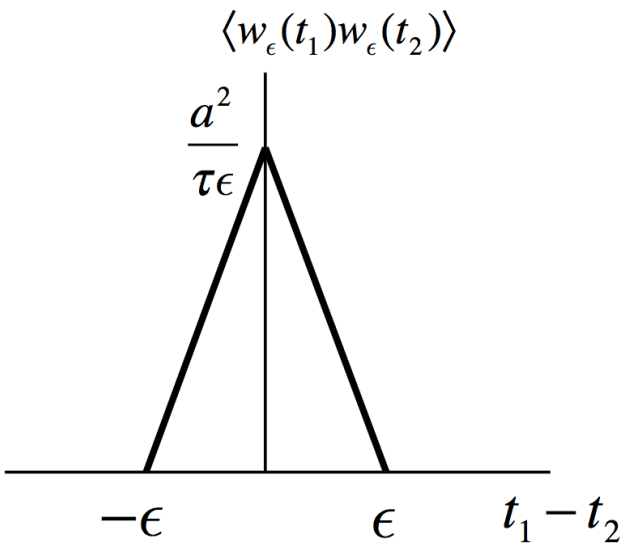


Figure 3.5: Correlation function for the derivative of the random walk process, as given by Eq. (3.37).

The shape of the correlation function is shown in figure 3.5.

In the limit $\epsilon \to 0$ the process $w(t) = \lim_{\epsilon \to 0} w_\epsilon(t)$ becomes the derivative of the random walk process and the correlation function (3.37) becomes a delta function

$$\langle w(t_1) w(t_2) \rangle = (a^2/\tau)\delta(t_1 - t_2). \tag{3.38}$$

If we go now to the limit defined by (3.29) and (3.30), the random walk tends to the Wiener process and its derivative can be written as $w(t) = D^{1/2}\xi(t)$. Where $\xi(t)$ is a Markovian, Gaussian process of zero mean and correlations:

$$\langle \xi(t) \rangle = 0, \tag{3.39}$$

$$\langle \xi(t_1)\xi(t_2) \rangle = \delta(t_1 - t_2), \tag{3.40}$$

known as white–noise. It is the derivative of the Wiener process (for $D = 1$):

$$\xi(t) = \frac{dW(t)}{dt}. \tag{3.41}$$

Physically, the white noise can be understood as a series of pulses each of them being very short but very intense, in a way that their effect is finite. The pulses are independent among them, which represents perturbations acting on the system in random directions so that the average of all the perturbations is zero (see (3.39). The name comes from the fact that its power spectral density (the Fourier transform of the correlation function) is flat, namely it is the same at all frequencies.

[3] In fact it is a fractal of dimension $1/2$.

## 3.5 Stochastic integrals. Itô and Stratonovich interpretations.

In the previous section the white noise we have introduced the white noise stochastic process as the derivative of the Wiener process. Since the Wiener process does not have a well-defined derivative, the precise interpretation of the white noise process depends on the way the derivative is performed. The most widely used interpretations are the those of Itô and Stratonovich. To illustrate this let us consider the integral

$$\int_t^{t'} f(x(s))\xi(s)ds, \tag{3.42}$$

where $f(x(t))$ is an arbitrary function of the stochastic process $x(t)$ whose dynamics is given by a Langevin equation of the form

$$\frac{dx}{dt} = q(x) + g(x)\xi(t). \tag{3.43}$$

The integral (3.42) is sometimes also written as

$$\int_t^{t'} f(x(s))dW\,, \tag{3.44}$$

although here we will mainly use the notation $\xi(s)ds$ rather than $dW$. Since the integral depends on a stochastic process, each realization of the stochastic process will lead to a different value for the integral, thus integrals of this form are called stochastic integrals.

Now we first proceed to compute the integral (3.42) for $t' = t+h$ in the limit $h \to 0$. In the so-called Itô interpretation the result of the integral is

$$\int_t^{t+h} f(x(s))\xi(s)ds = f(x(t))[W(t+h) - W(t)], \tag{3.45}$$

namely the function is evaluated at the initial time. In the so-called Stratonovich interpretation the result of the integral is

$$\int_t^{t+h} f(x(s))\xi(s)ds = f\left(\frac{x(t) + x(t+h)}{2}\right)[W(t+h) - W(t)], \tag{3.46}$$

that is, the function is evaluated at a point which is the average of the value of $x(t)$ at the initial and final times. In some sense, the existence of the different definitions is related to the definition of the following expression:

$$\int_0^{\infty} dt\ \delta(t) \tag{3.47}$$

which is equal to $1$ (Itô) or to $1/2$ (Stratonovich).

For a finite integration time we divide the integration interval $[t, t']$ in $N$ subintervals take the limit $N \to \infty$. In Itô calculus, this leads to

$$\int_t^{t'} f(x(s))\xi(s)ds = \lim_{N\to\infty} \sum_{i=1}^{N} f(x(t_{i-1}))[W(t_i) - W(t_{i-1})]\,, \tag{3.48}$$

where $t_i = t + ih$ with $h = (t' - t)/N$. In Stratonovich calculus the integral is given by

$$\int_t^{t'} f(x(s))\xi(s)ds = \lim_{N\to\infty} \sum_{i=1}^{N} f\left(\frac{x(t_{i-1}) + x(t_i)}{2}\right)[W(t_i) - W(t_{i-1})]\,. \tag{3.49}$$

In general one could have evaluated the stochastic integral as

$$\begin{aligned}\int_t^{t'} f(x(s))\xi(s)ds &= \\ &= \lim_{N\to\infty} \sum_{i=1}^{N} f\left(\alpha x(t_{i-1}) + (1-\alpha)x(t_i)\right)[W(t_i) - W(t_{i-1})]\,,\end{aligned} \tag{3.50}$$

with any arbitrary $\alpha$ such that $0 \leq \alpha \leq 1$. Ito interpretation corresponds to take $\alpha = 1$ while Stratonovich corresponds to $\alpha = 1/2$

In the past there was much arguing about to which is the "correct" interpretation. There is no such a thing. It is just a matter of convention. This means that a Langevin stochastic differential equation with a white noise eq. (3.43) is not completely defined unless we say in which interpretation should be considered. Once the interpretation have been fixed then any expression such as equation (3.42) has a unique meaning and its result can be uniquely determined.

From a physical point of view the Stratonovich interpretation turns out to be the more "natural" since it commutes with the limit given by (3.29) and (3.30). In this text, unless otherwise stated we will always use the Stratonovich interpretation. Besides, the Stratonovich interpretation allows for the use of the familiar rules of calculus, such as the change of variables in an integration. Instead in the Itô interpretation, in some instances, one can not use the ordinary to rules of calculus and instead use specific ones, which tend to be more cumbersome. One advantage of the Itô interpretation is that the stochastic process at time $t$, $x(t)$, is not correlated with the white noise acting on the system at the same time $\xi(t)$, that is $\langle x(t)\xi(t)\rangle = 0$. This is not the case of the Stratonovich interpretation.

To illustrate the differences between Itô and Stratonovich interpretations, let us consider the simple integral

$$\int_{t_a}^{t_b} W(s)\xi(s)ds\,. \tag{3.51}$$

In Itô calculus, this leads to

$$\int_{t_a}^{t_b} W(s)\xi(s)ds = \lim_{N\to\infty}\sum_{i=1}^{N} W(t_{i-1})\left[W(t_i)-W(t_{i-1})\right] =$$
$$= \frac{1}{2}\left[W(t_b)^2 - W(t_a)^2\right] - \frac{1}{2}\lim_{N\to\infty}\sum_{i=1}^{N}\left[W(t_i)-W(t_{i-1})\right]^2 \tag{3.52}$$

In Stratonovich calculus this leads to

$$\int_{t_a}^{t_b} W(s)\xi(s)ds = \lim_{N\to\infty}\sum_{i=1}^{N} \frac{1}{2}\left[W(t_{i-1})+W(t_i)\right]\left[W(t_i)-W(t_{i-1})\right] =$$
$$= \lim_{N\to\infty}\frac{1}{2}\sum_{i=1}^{N}\left[W(t_i)^2 - W(t_{i-1})^2\right] = \frac{1}{2}\left[W(t_b)^2 - W(t_a)^2\right] . \tag{3.53}$$

Thus while the Stratonovich calculus leads to a result for the integral which is the one that one would expect from ordinary calculus, in the Itô interpretation there is a non intuitive additional term. Technically the limit is to be understood as mean square limit. A series $F_n$ converge $F$ in mean square if

$$\lim_{n\to\infty}\langle (F_n - F)^2\rangle = 0. \tag{3.54}$$

It can be shown, (see exercise 3) that the mean square limit of $\sum_{i=1}^{N}\left[W(t_i)-W(t_{i-1})\right]^2$ is $t_b - t_a$. Therefore in Itô calculus

$$\int_{t_a}^{t_b} W(s)\xi(s)ds = \frac{1}{2}\left[W(t_b)^2 - W(t_a)^2 - (t_b - t_a)\right] . \tag{3.55}$$

We should also say that there is a simple relation between the Langevin stochastic differential equations written in both interpretations. The rule is that the Langevin equation

$$\frac{dx}{dt} = q_{\rm I}(x) + g_{\rm I}(x)\xi(t) \tag{3.56}$$

in the Itô sense is equivalent to the Langevin equation

$$\frac{dx}{dt} = q_{\rm I}(x) - \frac{1}{2}g_{\rm I}(x)g_{\rm I}'(x) + g_{\rm I}(x)\xi(t) \tag{3.57}$$

in the Stratonovich sense. Therefore the equivalent Stratonovich equation to a given Itô equation has an additional drift term. Notice that when the noise is additive, $g(x)$ is a constant, then the Langevin equation is the same in both interpretations.

## 3.6 The Ornstein-Uhlenbeck process

The white noise we have introduced in Section 3.4 is in fact a mathematical idealization that allows for some simplifications and also to obtain some analytical results. For instance, it can be shown that the solution $x(t)$ of a Langevin stochastic differential equation (3.24) is a Markov process if the noise is white. However random fluctuations affecting physical, chemical or biological systems have typically a non zero correlation time. For example in the Brownian process, there is timescale given by the collision time of the water molecules with the Brownian particle beyond which fluctuations can not be considered uncorrelated. It just happens that this time scale is much smaller than the typical time scales in which the Brownian particle moves, and therefore the white noise approximation is justified when one is interested in describing the "slow" movement of the "big" Brownian particle. However there are instances in which the separation between the time scales is not so big and then one must consider a Langevin stochastic differential equation in which the noise term does not have a zero correlation time. An example is the emission in a dye laser, in which the hydrodynamic fluctuations in the flow of the dye are at a quite slow time scale as compared with the characteristic scales for stimulated emission.

A stochastic process widely used to model fluctuations with a finite correlation time is the Ornstein-Uhlenbeck noise $\xi^{ou}(t)$. It is formally defined as a Gaussian Markov process characterized by

$$\langle \xi^{ou}(t) \rangle = 0 \tag{3.58}$$

$$\langle \xi^{ou}(t)\xi^{ou}(s) \rangle = \frac{1}{2\tau} e^{-|t-s|/\tau}, \tag{3.59}$$

namely, the correlations of the noise decay exponentially in time with a characteristic time scale given by the parameter $\tau$. In the limit $\tau \to 0$ the exponential becomes a Dirac delta function and the Ornstein-Uhlenbeck process becomes a white noise process. We note that when performing this limit the resulting stochastic differential equation should be interpreted in the Stratonovich sense. It can also be shown that, up to a change of variables, the Ornstein Uhlenbeck process is the only Gaussian, Markov, stationary process.

The Ornstein Uhlenbeck process can be seen as the solution of the Langevin stochastic differential equation

$$\frac{d\xi^{ou}(t)}{dt} = -\frac{1}{\tau}\xi^{ou}(t) + \frac{1}{\tau}\xi(t), \tag{3.60}$$

where $\xi(t)$ is a white noise defined by eqs. (3.39) and (3.40), and taking as initial condition that $\xi^{ou}(0)$ is a Gaussian random variable of mean $0$ and variance $(2\tau)^{-1}$ (see exercise 1). Furthermore, for an arbitrary initial condition the stationary solution of (3.60) is also an Ornstein-Uhlenbeck process (see exercise 1).

**3.6.0.0.10 Colored noise** The Ornstein-Uhlenbeck process is an example of "colored noise". By this notation one usually refers to stochastic processes acting on

Langevin equations for which the correlation time is not a Dirac delta function. The name comes as opposition to white noise which has a flat spectrum. Instead, noises that have a finite correlation time have a non flat spectrum. We also mention that although in principle "colored noise" could refer to any noise with finite correlation time, in many instances it is used as synonym of Ornstein-Uhlenbeck process.

## 3.7 The Fokker–Planck equation.

So far we have focused on trajectories to describe stochastic processes following the initial Langevin approach for Brownian motion. The alternative approach introduced by Einstein focuses on probabilities rather than in trajectories and, as it happens with the Langevin approach, it can be extended way beyond the Brownian motion. In what follows we are going to determine an equation for one time probability distribution for a stochastic process described by a Langevin equation with white noise. This is called the Fokker-Planck equation and it is a generalization of the diffusion equation obtained by Einstein to describe the Brownian process.

To be more precise, we want to find an equation for the one-time probability distribution function $f(x,t)$ for a stochastic process $x(t)$ which arises as a solution of a stochastic differential equation with a Gaussian white noise $\xi(t)$ defined in (3.39) and (3.40,

$$\frac{dx}{dt} = q(x) + g(x)\xi(t). \tag{3.61}$$

This is to be understood in the Stratonovich interpretation. Let us consider first the corresponding deterministic equation:

$$\frac{dx}{dt} = q(x), \qquad x(t=0) = x_0. \tag{3.62}$$

The solution of this equation is a (deterministic) function $x(t) = F(t, x_0)$. Nevertheless for our purposes, we can see $x(t)$ as a random variable whose probability density function $\rho(x;t)$ gives the probability to find the system at a given location $x$ in phase-space at a time $t$. Since the trajectory followed by the system is uniquely determined once the initial condition is given, this probability is zero everywhere except at the location $x = F(t, x_0)$. Therefore the probability density function is a delta–function:

$$\rho(x;t) = \delta(x - F(t, x_0)). \tag{3.63}$$

We can now transform $x(t)$ into a stochastic process by simply letting the initial condition $x_0$ become a random variable. We have now an ensemble of trajectories, each trajectory starting from a different initial condition $x_0$. In this case the probability density function $\rho(x,t)$ is obtained by averaging the above pdf over the distribution of initial conditions:

$$\rho(x;t) = \langle \delta(x - F(t, x_0)) \rangle_{x_0}. \tag{3.64}$$

In fact $\rho(x;t)$ can be seen as a density function, to visualize it assume that at time $t=0$ we have unleashed a large number of particles $N$ with one particle at each of the initial conditions we are considering in our ensemble. We let the particles evolve according to the dynamics and after a time $t$, $\rho(x;t)dx$ measures the fraction of particles located in the interval $(x, x+dx)$:

$$\rho(x;t)dx = \frac{\#\text{particles in } (x, x+dx) \text{ at time t}}{\text{total } \#\text{ of particles}} = \frac{n(x,t)}{N}dx. \tag{3.65}$$

Assuming there are no sinks not sources of particles, then the number of particles at a given location changes due to the number of particles crossing the borders. Let $J(x,t)$ be the flux of particles

$$J(x,t)dt = \frac{\#\text{particles that cross the point } x \text{ in the interval } (t, t+dt)}{N}. \tag{3.66}$$

$J$ has a direction. For convenience we consider $J>0$ if the particle moves to the right. Then

$$\rho(x;t+dt)dx - \rho(x;t)dx = J(x,t)dt - J(x+dx,t)dt. \tag{3.67}$$

The first term on the LHS is the final number of particles at $x$ while the second is the initial one. Thus the LHS corresponds to the variation of the number of particles at $x$ from time $t$ to time $t+dt$. In the right-hand-side, the first term is the number of particles that have entered or leaved the interval $(x, x+dx)$ through the left boundary while the second measures the number of particles that cross the right boundary. Therefore one has

$$\frac{\rho(x;t+dt)-\rho(x;t)}{dt} = \frac{J(x,t)-J(x+dx,t)}{dx}, \tag{3.68}$$

which in the continuous limit corresponds to

$$\frac{\partial \rho(x;t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x}, \tag{3.69}$$

which is the continuity equation.

Since particles move following a deterministic dynamics given by eq. (3.62) trajectories can not cross, and therefore in a one-dimensional system they can not advance one to the other. The particles that will cross the point $x$ in the interval $(t, t+dt)$ are all those located in the interval $dx = q(x,t)dt$. Therefore

$$J(x,t) = \frac{n(x,t)}{N}\frac{dx}{dt} = \rho(x,t)q(x,t). \tag{3.70}$$

Replacing this in eq. (3.69) one gets the Liouville equation

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial x}[q(x)\rho]. \tag{3.71}$$

We consider now the full stochastic differential equation (3.61). We can repeat the above argument for a given realization of the noise term. The probability density function $f(x;t)$ will be the average of $\rho(x;t)$ with respect to the noise distribution:

$$f(x;t) = \langle \rho(x;t) \rangle_\xi = \langle \delta(x - x(t,x_0)) \rangle_{x_0,\xi}. \tag{3.72}$$

$\rho(x;t)$ satisfies the Liouville equation (3.69). Now the current is given by

$$J(x,t) = \rho(x;t)\frac{dx}{dt} = \rho(x;t)\left[q(x) + g(x)\xi(t)\right]. \tag{3.73}$$

Therefore

$$\frac{\partial \rho(x;t)}{\partial t} = -\frac{\partial}{\partial x}[(q(x) + g(x)\xi(t))\rho]. \tag{3.74}$$

By taking averages over the noise term we get:

$$\begin{aligned}\frac{\partial f(x;t)}{\partial t} &= \langle \frac{\partial \rho(x;t)}{\partial t} \rangle_\xi = -\frac{\partial}{\partial x}\langle \left[q(x) + g(x)\xi(t)\right] \rho(x;t) \rangle_\xi \\ &= -\frac{\partial}{\partial x}\left[q(x)\langle \rho(x;t)\rangle_\xi\right] - \frac{\partial}{\partial x}\left[g(x)\langle \xi(t)\rho(x;t)\rangle_\xi\right].\end{aligned} \tag{3.75}$$

The averages on the first term of the right-hand-side can be easily performed $\langle \rho(x;t) \rangle_\xi = f(x;t)$. However averages on the second term of the right-hand-side are more cumbersome, since one has to keep in mind that the density distribution $\rho$ depends on $x(t)$ which in itself depends functionally on the noise through the dynamics (3.61). In fact, to be precise the average on the second term of the right-hand-side should be written as $\langle \xi(t)\rho(x[\xi(t)];t)\rangle_\xi$. This average can be done by using Novikov's theorem[4], which establishes that for any Gaussian stochastic process $\xi_G(t)$ with zero mean, $\langle \xi_G(t) \rangle = 0$, and for any functional of the noise $\mathcal{F}[\xi_G(t)]$ one has

$$\langle \xi_G(t)\mathcal{F}[\xi_G(t)] \rangle_{\xi_G} = \int_0^t ds \langle \xi_G(t)\xi_G(s)\rangle_{\xi_G} \langle \frac{\delta \mathcal{F}}{\delta \xi_G(s)} \rangle_{\xi_G}, \tag{3.76}$$

where the last term is the functional derivative of $\mathcal{F}[\xi_G(t)]$ with respect to $\xi_G(s)$. We note that Novikov theorem is quite general, it only requires the noise to be Gaussian but it does not requires to be white, namely, the noise correlations do not need to be a delta function.

In our case we have a Gaussian white noise, for which the correlation is a delta function. This allows to evaluate the integral easily:

$$\begin{aligned}\langle \xi(t)\rho(x[\xi(t)];t)\rangle_\xi &= \int_0^t ds\ \delta(t-s) \left\langle \frac{\delta \rho(x[\xi(t)];t)}{\delta \xi(s)} \right\rangle_\xi \\ &= \frac{1}{2}\left\langle \frac{\delta \rho(x[\xi(t)];t)}{\delta \xi(s)}\bigg|_{s=t} \right\rangle_\xi .\end{aligned} \tag{3.77}$$

[4]We use the extension to a Gaussian process of the theorem whose proof was proposed in exercise 1.**??** for a finite number of Gaussian variables.

By using functional calculus, this can be computed as:

$$\left\langle \left.\frac{\delta\rho(x[\xi(t)];t)}{\delta\xi(s)}\right|_{s=t}\right\rangle_\xi = \left\langle \left.\frac{\delta x(t)}{\delta\xi(s)}\right|_{s=t}\frac{\partial\rho(x;t)}{\partial x(t)}\right\rangle_\xi$$
$$= -\frac{\partial}{\partial x}\left\langle \left.\frac{\delta x(t)}{\delta\xi(s)}\right|_{s=t}\rho(x;t)\right\rangle_\xi . \tag{3.78}$$

To evaluate the functional derivative of the stochastic process $x(t)$ with respect to the noise we use a formal solution of eq.(3.61):

$$x(t) = x_0 + \int_0^t ds\; q(x(s)) + \int_0^t ds\; g(x(s))\xi(s). \tag{3.79}$$

Then

$$\left.\frac{\delta x(t)}{\delta\xi(s)}\right|_{s=t} = g(x(t)), \tag{3.80}$$

and

$$\left\langle \left.\frac{\delta\rho(x[\xi(t)];t)}{\delta\xi(s)}\right|_{s=t}\right\rangle_\xi = -\frac{\partial}{\partial x}[g(x(t))\langle\rho(x;t)\rangle_\xi] = -\frac{\partial}{\partial x}[g(x)f(x;t)]. \tag{3.81}$$

Using this result in Eq. (3.75) we get, finally, the Fokker–Planck equation for the probability density function:

$$\frac{\partial f(x;t)}{\partial t} = -\frac{\partial}{\partial x}\left[q(x)f(x;t)\right] + \frac{1}{2}\frac{\partial}{\partial x}\left[g(x)\frac{\partial}{\partial x}\left[g(x)f(x;t)\right]\right]. \tag{3.82}$$

The procedure we have used to derive the Fokker-Planck equation for a single variable can be extended to the case of several variables. Consider a set of stochastic variables $\vec{x} = (x_1, x_2, ..., x_N)$ whose dynamics is given by the set of Langevin equations to be considered in the Stratonovich interpretation

$$\frac{dx_i}{dt} = q_i(\vec{x},t) + \sum_{j=1}^{N} g_{ij}(\vec{x},t)\xi_j(t), \quad i = 1...N, \tag{3.83}$$

where we allow for the drift terms $q_i(\vec{x},t)$ and diffusion terms $g_{ij}(\vec{x},t)$ to explicitly depend on the time. $\xi_j(t)$ are uncorrelated Gaussian white noises with zero mean, that is

$$\langle\xi_i(t)\rangle = 0 \quad \langle\xi_i(t)\xi_j(s)\rangle = \delta_{i,j}\delta(t-s). \tag{3.84}$$

By using a straightforward extension of the method used for one variable, one can prove that the one-time probability density function $f(\vec{x};t)$ satisfies the following multivariate Fokker-Planck equation

$$\frac{\partial f(\vec{x};t)}{\partial t} = -\sum_{i=1}^{N}\frac{\partial}{\partial x_i}\left[q_i(\vec{x},t)f(\vec{x};t)\right]$$
$$+\frac{1}{2}\sum_{i,j,k=1}^{N}\frac{\partial}{\partial x_i}\left[g_{ik}(\vec{x},t)\frac{\partial}{\partial x_j}\left[g_{jk}(\vec{x},t)f(\vec{x};t)\right]\right]. \tag{3.85}$$

It can also be written in the form of a continuity equation

$$\frac{\partial f(\vec{x};t)}{\partial t} + \sum_{i=1}^{N} \frac{\partial}{\partial x_i} J_i(\vec{x},t) = 0, \tag{3.86}$$

where the probability currents $J_i(\vec{x},t)$ are given by

$$J_i(\vec{x},t) = q_i(\vec{x},t) f(\vec{x};t) - \frac{1}{2} \sum_{j,k=1}^{N} g_{ik}(\vec{x},t) \frac{\partial}{\partial x_j} \left[ g_{jk}(\vec{x},t) f(\vec{x};t) \right]. \tag{3.87}$$

If $q_i(\vec{x},t)$ and $g_{ij}(\vec{x},t)$ do not depend explicitly on time, namely $q_i(\vec{x},t) = q_i(\vec{x})$ and $g_{ij}(\vec{x})$ then the Fokker-Planck is called homogeneous.

**3.7.0.0.11 Stationary solution** In general, solutions of the Fokker-Planck equation are difficult to be found analytically. An exception is the stationary solution in the case of a single variable. In this case imposing

$$\left. \frac{\partial f(x;t)}{\partial t} \right|_{f(x,t)=f^{\mathrm{st}}(x)} = 0, \tag{3.88}$$

one has that in the stationary state the current must fulfill

$$\frac{\partial J^{\mathrm{st}}(x)}{\partial x} = 0, \tag{3.89}$$

namely $J^{\mathrm{st}}(x)$ must be a constant. The simplest situation is when this constant is zero which means that in the stationary state there is no flux of probability, then one has

$$q(x) f^{\mathrm{st}}(x) - \frac{g(x)}{2} \frac{d}{dx} \left[ g(x) f^{\mathrm{st}}(x) \right] = 0. \tag{3.90}$$

This can be rewritten as

$$\frac{2q(x) - g(x)g'(x)}{g^2(x)} dx = \frac{df^{\mathrm{st}}(x)}{f^{\mathrm{st}}(x)}, \tag{3.91}$$

Integrating at both sides, and considering that $a$ and $b$ are the lower and upper boundaries for the variable $x$ one has

$$\int_a^x dy \frac{2q(y) - g(y)g'(y)}{g^2(y)} = \ln[f^{\mathrm{st}}(y)]|_a^x \tag{3.92}$$

Then

$$f^{\mathrm{st}}(x) = \mathcal{Z}\Psi(x), \tag{3.93}$$

where

$$\Psi(x) = \exp \int_a^x dy \frac{2q(y) - g(y)g'(y)}{g^2(y)} = \frac{|g(a)|}{|g(x)|} \exp \int_a^x dy \frac{2q(y)}{g^2(y)}, \tag{3.94}$$

and $\mathcal{Z}$ is a normalization constant such that

$$\int_a^b f^{\text{st}}(x) = 1. \tag{3.95}$$

We now consider the case in which the stationary current is a nonzero constant $J^{\text{st}}(x) = J$. This situation, in which there is a stationary flux of probability, is usually encountered in systems with periodic boundary conditions, for which $f^{\text{st}}(a) = f^{\text{st}}(b)$, being $a$ and $b$ the left and right boundaries of the system respectively. $J$ is not arbitrary, its value is determined by the normalization of $f^{\text{st}}(x)$ and by the boundary value of the probability density function. In this case one has

$$q(x)f^{\text{st}}(x) - \frac{g(x)}{2}\frac{d}{dx}\left[g(x)f^{\text{st}}(x)\right] = J, \tag{3.96}$$

which is an equation similar to (3.89) but with an additional inhomogeneous term. It can be shown (see exercise 4) that the current is given by

$$J = \left[\frac{g(b)^2}{\Psi(b)} - \frac{g(a)^2}{\Psi(a)}\right] \frac{f^{\text{st}}(a)}{\int_a^b \frac{dy}{\Psi(y)}} \tag{3.97}$$

and the stationary solution is

$$f^{\text{st}}(x) = \mathcal{C}\frac{\frac{g^2(b)}{\Psi(b)}\int_a^x \frac{dy}{\Psi(y)} + \frac{g^2(a)}{\Psi(a)}\int_x^b \frac{dy}{\Psi(y)}}{\frac{g^2(x)}{\Psi(x)}\int_a^b \frac{dy}{\Psi(y)}}, \tag{3.98}$$

where $\mathcal{C}$ is a normalization constant.