Appendix A
Calculus of Variations

The calculus of variations appears in several chapters of this volume as a means to formally derive the fundamental equations of motion in classical mechanics as well as in quantum mechanics. Here, the essential elements involved in the calculus of variations are briefly summarized.[1]

Consider a functional $\mathcal{F}$ depending on a function y of a single variable x (i.e., $y = y(x)$) and its first derivative $y' = dy/dx$. Moreover, this functional is defined in terms of the line or path integral

$$\mathcal{F}[I] = \int_{x_a}^{x_b} I(y, y', x)dx. \tag{A.1}$$

Accordingly, the value of $\mathcal{F}$ will depend on the path chosen to go from x_a to x_b. The central problem of the *calculus of variations* [1–3] consists of determining the path $y(x)$ that makes $\mathcal{F}$ an extremum (a maximum, a minimum or a saddle point). In other words, this is equivalent to determining the conditions for which (A.1) acquires a stationary value or, equivalently, is invariant under first-order variations (or perturbations) of the path $y(x)$, i.e.,

$$\delta\mathcal{F} = \delta \int_{x_a}^{x_b} I\, dx = \int_{x_a}^{x_b} \delta I\, dx = 0. \tag{A.2}$$

Let us thus define the quantities $\delta y = Y(x) - y(x)$ and $\delta I = I(Y, Y', x) - I(y, y', x)$, where $Y(x)$ denotes the perturbed path. Variations are taken with respect to the same x value, so $\delta x = 0$. It is straightforward to show that $\delta y' = d(\delta y)/dx$ and therefore

[1] The brief description of the essential elements involved in the calculus of variations presented here can be complemented with more detailed treatments, which can be found in well-known textbooks on mathematical physics, e.g., [1–3].

A. S. Sanz and S. Miret-Artés, *A Trajectory Description of Quantum Processes. I. Fundamentals*, Lecture Notes in Physics 850, DOI: 10.1007/978-3-642-18092-7,

$$\delta I = \left(\frac{\partial I}{\partial y} + \frac{\partial I}{\partial y'} \frac{d}{dx} \right) \delta y. \tag{A.3}$$

Substituting this expression into (A.2) and then integrating by parts yields

$$\int_{x_a}^{x_b} \left(\frac{\partial I}{\partial y} - \frac{d}{dx} \frac{\partial I}{\partial y'} \right) \delta y \, dx = 0, \tag{A.4}$$

since, at the boundaries, $\delta y(x_a) = \delta y(x_b) = 0$. Because δy is an arbitrary, infinitesimal increment, it can be chosen so that the integrand in (A.4) vanishes. This leads to the well-known *Euler–Lagrange equation*,

$$\frac{\partial I}{\partial y} - \frac{d}{dx} \frac{\partial I}{\partial y'} = 0. \tag{A.5}$$

The function y satisfying this equation, if it exists, is said to be an *extremal curve* or *extremal*.

Equation (A.5) can also be recast as

$$\frac{\partial I}{\partial x} - \frac{d}{dx} \left(I - y' \frac{\partial I}{\partial y'} \right) = 0, \tag{A.6}$$

which arises after taking into account the dependence of I on x, y and y' as well as the fact that

$$\frac{d}{dx} = \frac{\partial}{\partial x} + y' \frac{\partial}{\partial y} + y'' \frac{\partial}{\partial y'}. \tag{A.7}$$

Equation (A.6) is useful whenever I does not depend explicitly on x, for it becomes

$$I - y' \frac{\partial I}{\partial y'} = \text{constant}, \tag{A.8}$$

which is also an extremal.

Consider now that I depends on several functions $y_1, y_2, \ldots, y_N$ of x and their respective derivatives, $y_1', y_2', \ldots, y_N'$. Then, proceeding in a similar way, a functional

$$\mathcal{F}[I] = \int_{x_a}^{x_b} I(y_1, y_2, \ldots, y_N, y_1', y_2', \ldots, y_N', x) dx \tag{A.9}$$

can be defined, which becomes an extremum or stationary when the set of Euler–Lagrange equations

$$\frac{\partial I}{\partial y_i} - \frac{d}{dx} \frac{\partial I}{\partial y_i'} = 0, \qquad i = 1, 2, \ldots, N, \tag{A.10}$$

is satisfied. However, it could happen that the search for an extremum condition is subject to a constraint, as in the so-called *isoperimetric problems* (e.g., determining the closed plane curve of maximum area and fixed perimeter). In such cases, given a set $J_1, J_2, \ldots, J_M$ of constraining conditions that depend on x and the $y_i (i = 1, 2, \ldots, N)$, the set of N equations (A.10) is replaced by the set of $N + M$ equations

$$\begin{cases} \dfrac{\partial I}{\partial y_i} - \dfrac{d}{dx}\dfrac{\partial I}{\partial y_i'} + \displaystyle\sum_{j=1}^{M} \lambda_j(x)\dfrac{\partial J_j}{\partial y_i} = 0 \\ J_j(x, y_1, y_2, \ldots, y_N) = 0 \end{cases} . \tag{A.11}$$

The λ_j functions are the so-called *Lagrange undetermined multipliers*, M unknown functions of x (or constants) which have to be determined in order to obtain a full (complete) solution to the problem.

If the constraints in (A.11) are specified by a set of M functional integral constraints,

$$\mathcal{F}_j = \int_{x_a}^{x_b} J_j(y_1, y_2, \ldots, y_N, y_1', y_2', \ldots, y_N', x)dx = c_j, \tag{A.12}$$

where all c_j are constant and the $\mathcal{F}_j$ are extrema for the y_i, a function

$$K = I + \sum_{j=1}^{M} \lambda_j J_j \tag{A.13}$$

can be defined. Proceeding as before, one finds that these functions have to satisfy the Euler–Lagrange equation

$$\frac{\partial K}{\partial y_i} - \frac{d}{dx}\frac{\partial K}{\partial y_i'} = 0, \qquad i = 1, 2, \ldots, N, \tag{A.14}$$

as well as the integral constraints (A.12).

In the particular case of mechanical systems, when the variational principle is applied, power series expansions up to the third order in the displacement are often considered. In these series expansions, the zeroth-order term gives us the action integral along the reference trajectory; the second-order is called the first variation, which vanishes for any path due to the stationarity condition; the third-order or second variation provides us with information about the nature of the stationary value (maximum, minimum or saddle point) by analyzing the eigenvalues of the matrix associated with the corresponding quadratic form in the displacements.

The formalism described above is rather general. As seen in Chap. 3, for example, it is closely related to the formal derivation of Schrödinger's wave equation. In this case, instead of several functions y_i of a single variable x, one considers a function ψ of several variables x_i. These functions are usually called *field functions* or *fields*. Furthermore, a subtle conceptual difference can be found in the application of the

calculus of variations in classical and in quantum mechanics. In classical mechanics, it is tightly connected to the concept of energy (Hamiltonian); different solutions are then obtained from its application, namely the classical trajectories. In quantum mechanics, though, this idea is extended to functionals of a *single* dependent variable (the *wave function* field) and several independent variables, thus generalizing the classical case. Thus, rather than keeping constant the energy along a given path, energy conservation appears in the calculation of the average or expectation value of such an *observable*.

References

1. Morse, P.M., Feshbach, H.: Methods of Theoretical Physics. McGraw-Hill, New York (1953)
2. Margenau, H., Murphy, G.M.: The Mathematics of Physics and Chemistry, 2nd edn. Van Nostrand, New York (1956)
3. Arnold, V.I.: Mathematical Methods of Classical Mechanics. Springer, New York (1989)

Appendix B
Stochastic Processes

The theory of stochastic processes is of fundamental interest to understand the theory of open classical and quantum systems, as seen in Chaps. 2, 5 and 6. In this regard, in order to make this volume self-contained as far as possible and to better understand the dynamics of open systems, some elementary concepts on probability theory and stochastic processes are introduced here.[1]

B.1 Random or Stochastic Variables

A probability or measure space is defined by three mathematical objects $(\Omega, \mathcal{A}, P)$. In this triad, Ω represents the set of elementary outcomes or *sample space*, with such elements usually denoted by ω, with $\omega \in \Omega$; $\mathcal{A}$ is a *collection* or *field of events*; and P is a *probability measure*.

The field of events $\mathcal{A}$ is also called a σ-algebra of events. It consists of a family or set of subsets of Ω, such that:

1. $\Omega \in \mathcal{A}$ and $\emptyset \in \mathcal{A}$.
2. If $A \in \mathcal{A}$, then $\bar{A} = \Omega - A \in \mathcal{A}$.
3. If $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}, A \cap B \in \mathcal{A}$ and $A - B \in \mathcal{A}$.

In brief, the σ-algebra forms a closed system under the union ($\cup$), intersection ($\cap$) and complementary set operations.

A probability measure is simply a map $P : \mathcal{A} \to \mathbb{R}$, which assigns a real number to each event A of the σ-algebra. The probability of an event A, denoted by $P(A)$, satisfies the Kolmogorov axioms of probability, from which a series of consequences arise. For example, the probability $P(A)$ has a numeric bound for all A, namely $0 \leq P(A) \leq 1$. Thus, $P(\Omega) = 1$ characterizes a *certain event* and $P(\emptyset) = 0$ an *impossible event*; $P(\bar{A}) = 1 - P(A)$ represents the probability of the *complementary event* of A. If A_i is a countable collection of non-overlapping (or *mutually exclusive*)

[1] For further reading on these issues, the interested reader may consult more specialized works, e.g., [1–10].

A. S. Sanz and S. Miret-Artés, *A Trajectory Description of Quantum Processes. I. Fundamentals*, Lecture Notes in Physics 850, DOI: 10.1007/978-3-642-18092-7,

sets, i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$, then $P(A_1 \cup A_2 \cup \cdots) = \sum_i P(A_i)$. On the contrary, given A and B, if $A \cap B \neq \emptyset$, the non-vanishing probability $P(A \cap B)$ represents the *joint probability* that any element of the intersection occurs; two events are called independent when $P(A \cap B) = P(A)P(B)$. Similarly, the *conditional probability* that an event A occurs, given another event B also occurs (i.e., $P(B) \neq 0$), is defined as $P(A|B) = P(A \cap B)/P(B)$.

A *random* or *stochastic variable* X is a function from the sample space Ω into the set of real (or complex) numbers with the property $A = \{\omega / X(\omega) \leq x\} \in \mathcal{A}$, with $x \in \mathbb{R}$ being a *realization* of X. This random variable or function is said to be measurable with respect to the σ-algebra of $\mathcal{A}$, this definition often being expressed in terms of the inverse image, as $A = X^{-1}((-\infty, x]) \in \mathcal{A}$, with $x \in \mathbb{R}$. Accordingly, the probability of an observable event of $\mathcal{A}$ is defined as $P_X(A) = P(X^{-1}(A)) = P(\{\omega / X(\omega) \in \mathcal{A}\})$. In order to formalize this statement, the σ-algebra of *Borel sets* of $\mathbb{R}$, denoted by $\mathcal{B}$, is introduced, which is defined as the smallest σ-algebra containing all subsets of the form $(-\infty, x)$, with $x \in \mathbb{R}$—in particular, it contains all open and closed intervals of the real axis. A function is then measurable when, for any Borel set $B \in \mathcal{B}$, the pre-image $A = X^{-1}(B)$ belongs to the σ-algebra $\mathcal{A}$ of events. The *distribution function* or *probability distribution* of the random variable is defined as $F_X(x) = P(\{\omega / X(\omega) \leq x\}) = P(X \leq x)$, with $x \in \mathbb{R}$, which satisfies

1. $F_X(-\infty) = 0$.
2. $F_X(+\infty) = 1$.
3. $F_X(x)$ is an increasing monotonically right continuous function.

In many applications, continuous random variables are often found, with their *probability density* being defined as $p_X(x) = dF_X(x)/dx$ or, equivalently, as

$$F_X(x) = \int_{-\infty}^{x} p_X(x')dx', \tag{B.1}$$

so that

$$\begin{aligned} dF_X(x) &= F_X(x + dx) - F_X(x) = p_X(x)dx \\ &= P(x \leq X < x + dx) = P(d\omega). \end{aligned} \tag{B.2}$$

For random vectors or an arbitrary collection of d random variables, similar functions and densities can be defined in $\mathbb{R}^d$—if $d = 2$, they are called bivariate distributions, while for any general d, they are multivariate ones. This information provides a complete characterization of the random variable X. Thus, if the density of a random variable exists, the probability that x will be contained in the interval $(x, x + dx)$ goes to zero with dx. Therefore, the probability that X has exactly an x value is zero. Sets containing one single point as well as any set only containing a countable number of points have zero probability. In probability theory, all equalities are at best only *almost certainly true*, *almost surely* or *with probability one*. Very often X drops in F_X and p_X.

Functions of random variables, $Y = g(X)$, can also be defined. Thus, if F_X is the probability distribution of X, then $P_Y(B) = P_X(g^{-1}(B))$ and

$$p_Y(y) = \int \delta(y - g(x)) p_X(x) dx, \tag{B.3}$$

where δ denotes the Dirac δ-function; if $Y = X_1 + X_2$, then $g(x_1, x_2) = x_1 + x_2$; and, if X_1 and X_2 are independent, then

$$p_Y(y) = \int p_{X_1}(x_1) p_{X_2}(y - x_1) dx_1, \tag{B.4}$$

which is the convolution of the densities associated with X_1 and X_2.

Formally, the average, expectation or mean value of a discrete random variable X is defined as

$$E\{X\} = \sum_i x_i P(A_i), \tag{B.5}$$

which is also often denoted as $\bar{X}$ or $\langle X \rangle$. On the other hand, if X is a continuous random variable,

$$E\{X\} = \int_\Omega X(\omega) dP(\omega) = \int_\Omega X(\omega) P(d\omega), \tag{B.6}$$

which can also be expressed in a more familiar form as

$$E\{X\} = \int_{\mathbb{R}} x dF_X(x) = \int_{\mathbb{R}} x p_X(x) dx. \tag{B.7}$$

The mean value (B.7) is also known as the *first moment* of the distribution function; higher nth-order moments $E\{X^n\}$ can be defined in a similar fashion provided $p_X(x)$ exists. In this regard, it is worth mentioning that the knowledge of all moments is not a sufficient condition to uniquely determine $p_X(x)$. From (B.7), the expectation value of functions of a random variable can also be defined. For example, if the mean value is known, the central moments of a random variable are defined as

$$E\{(\delta X)^r\} = E\{(X - E\{X\})^r\} = \int_{\mathbb{R}} (x - E\{X\})^r p_X(x) dx. \tag{B.8}$$

The second central moment, σ^2, defined as

$$\mathrm{Var}(X) = \Delta X = E\{(X - E\{X\})^2\} = E(X^2) - E(X)^2, \tag{B.9}$$

also denoted as ΔX, is known as the *variance*, *mean-square deviation* or *fluctuation*. This moment is a measure of the width of the probability density or, in other words, of the fluctuations of X with respect to its mean value, δX. The number and location

of the extrema of the probability distribution are important, because their maxima are the so-called most probable states (which is a local property).

When there are two random variables, it is very often interesting to know if they are correlated or not. A measure of their degree of correlation is given by the so-called *covariance*, defined as

$$\begin{aligned}\sigma_{XY} &= E\{(X - E\{X\})(Y - E\{Y\})\} \\ &= \int_{\mathbb{R}} (x - E\{X\})(y - E\{Y\}) p_{XY}(x, y) dxdy,\end{aligned} \tag{B.10}$$

where $p_{XY}(x, y)$ is the *joint probability density*. If X and Y are independent (or uncorrelated), then $\sigma_{XY} = 0$ (the opposite is not true in general). For several random variables, a *covariance matrix* can be defined, which is symmetric and positive semi-definite. The off-diagonal elements are called covariances and are a measure of the linear dependence of two random variables. Hence, if the variables are pairwise independent, the covariance matrix will be diagonal. Correlation coefficients between two random variables are then defined by their covariance divided by the square root of the product of their respective variances, i.e., $C(X, Y) = \sigma_{XY}/\sqrt{\Delta X \Delta Y}$. It is clear that $0 \leq |C(X, Y)| \leq 1$.

The *characteristic function* of a random variable X is defined as the Fourier transform of its probability density,

$$G(\xi) = E\{e^{i\xi X}\} = \int p_X(x) e^{i\xi x} dx, \tag{B.11}$$

for a real number ξ—note the close similarity between this form and the transformations that allow to pass in quantum mechanics from the configuration to the momentum space and vice versa. This function is also called the generating function of all moments of the random variable, since the nth derivative of $G(\xi)$ evaluated at $\xi = 0$ gives the nth moment, $E\{X^n\}$—of course, this relies on the implicit assumption that $p_X(x)$ is sufficiently regular and therefore the exponential admits a Taylor series expansion. From the Fourier inversion formula, $p_X(x)$ can be determined with probability one. As it can be shown, a sequence of probability densities converges to a limiting probability density if the corresponding characteristic functions converge to the characteristic function associated with the limiting probability density. A straightforward generalization to more variables can easily be carried out. In such a case, if all of them are independent, the corresponding characteristic function will be the product of the individual characteristic functions. Similarly, the logarithm of the characteristic function generates all the *cumulants* of X. The first cumulant is the mean value, while the second cumulant is the covariance of two random variables. Finally, X can also be given by the integral of a stochastic process.

When two random variables are not statistically independent, there is some information about one of them with respect to the other. In such a case, one can define

the so-called *marginal distributions*,

$$p_X(x) = \int_{\mathbb{R}} p_{XY}(x, y)dy, \tag{B.12a}$$

$$p_Y(y) = \int_{\mathbb{R}} p_{XY}(x, y)dx. \tag{B.12b}$$

The conditional expectation value of the random variable X is defined as

$$E\{X|B\} = \int_{\Omega} X(\omega)P(d\omega|B), \tag{B.13}$$

although it can also be written as

$$E\{X|B\}P(B) = \int_{B} X(\omega)P(d\omega). \tag{B.14}$$

Many times it is necessary to establish a set of conditions with respect to a collection of events or to the history of events. This is done, for example, to make the best possible prediction of the actual random variable X knowing the available or previous information. The sub-σ-algebra $\mathcal{C}$ generated by $\{A_i\}$, with $\mathcal{C} \subset \mathcal{A}$, is precisely the available information. The best possible prediction or estimate of X is another random variable Y which has to be $\mathcal{C}$-measurable and fulfills the condition

$$\int_{A_i} Y(\omega)P(d\omega) = \int_{A_i} X(\omega)P(d\omega), \tag{B.15}$$

with $A_i \subset \mathcal{C}$. According to the *Radon–Nikodym theorem* [11, 12], there is a random variable Y with the above properties, such that it is almost surely unique and can be expressed as

$$Y = E\{X|\mathcal{C}\} = \int_{\Omega} X(\omega)P(d\omega|\mathcal{C}). \tag{B.16}$$

The most important properties of conditional expectations are:

- $E\{E\{X|\mathcal{C}\}\} = E\{X\}$.
- If $X \geq 0$, then $E\{X|\mathcal{C}\} \geq 0$.
- If X is measurable with respect to $\mathcal{C}$, then $E\{X|\mathcal{C}\} = X$.
- If $E\{X\} < \infty$ and $E\{Y\} < \infty$, then $E\{aX + bY|\mathcal{C}\} = aE\{X|\mathcal{C}\} + bE\{Y|\mathcal{C}\}$, with a and b being constant.
- If X and $\mathcal{C}$ are independent, then $E\{X|\mathcal{C}\} = E\{X\}$.
- If $\mathcal{C}_1$ and $\mathcal{C}_2$ are sub-σ-algebras of $\mathcal{A}$, such that $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{A}$, then $E\{E\{X|\mathcal{C}_2\}|\mathcal{C}_1\} = E\{E\{X|\mathcal{C}_1\}|\mathcal{C}_2\} = E\{X|\mathcal{C}_1\}$.

A particular class of conditional expectation values is obtained when the conditioning is considered with respect to another random variable. In this case, $E\{X|\mathcal{C}\} =$

$E\{X|Z\}$, when the σ-algebra $\mathcal{C}$ is generated by the random variable Z. Since $E\{X|Z\}$ is a random variable, it can be shown by means of a theorem that $E\{X|Z\} = h(Z)$, i.e., the corresponding random variable can be written as a measurable function h of Z. This function is real-valued and almost surely uniquely defined. One can therefore write $E\{X|Z = z\} = h(z)$ and, if the marginal probability density

$$p_Z(z) = \int_{\mathbb{R}} p_{XZ}(x, z)dx \tag{B.17}$$

is positive, then

$$p(x|z) = \frac{p_{XZ}(x, z)}{p_Z(z)}. \tag{B.18}$$

From these expressions, one finds

$$E\{x|z\} = \int_{\mathbb{R}} xp(x|z)dx \tag{B.19a}$$

$$E\{E\{x|z\}\} = E\{x\} = \int_{\mathbb{R}} xp_{XZ}(x, z)dxdz. \tag{B.19b}$$

One of the most widely used probability distributions is the *Gaussian* or *normal distribution*. According to the *central limit theorem*, any random variable X given by the sum of N statistically independent and identically distributed random variables becomes Gaussian or normally distributed. That is, in the limit $N \to \infty$ and provided the first and second moments do not diverge (in $p_X(x)$ very often X drops in most of textbooks),

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}. \tag{B.20}$$

Within this context, the concept of limiting distribution readily appears, not related to the regular behavior of the moments of the distribution itself, but to stability. It can be shown that a probability distribution can only be a limiting distribution if it is a *stable* or *Lévy distribution*. In this class of distributions, the logarithm of their characteristic functions must satisfy a certain mathematical expression [7] and display long-range, inverse power-law tails which may lead to a divergence of even the lowest order moments. For example, the second moment of the *Cauchy* or *Lorentzian distribution* is infinite, going as $|x^{-(1+\alpha)}|$ when $0 < \alpha < 2$ and $|x| \to \infty$.

Another important distribution is the *binomial distribution*, which is a discrete probability distribution accounting for the number of successful events with probability p in a sequence of n independent experiments with two possible outcomes (e.g., yes/no, 0/1). This type of distribution describes, for example, a random walker on a line, stepping forward and backwards randomly. The binomial distribution converges to a Gaussian one in the limit of a large number of jumps, as can be easily shown using

Stirling's formula. An "intermediate" case between both distributions is the *Poisson distribution*, which is continuous and appears in the joint limit $p \to 0$ and $n \to \infty$, but keeping constant the quantity $\lambda = pN$.

The concept of limit of a sequence of random variables arises naturally in many different physical situations, though there is not a unique way to define it. Some of these definitions are:

1. The *almost certainly* or *surely limit,*

$$X = \lim_{n\to\infty} X_n, \tag{B.21}$$

   i.e., X_n converges almost surely to X, is the simplest definition. In brief, this limit is more explicitly expressed as

$$X = \text{as}\text{–}\lim_{n\to\infty} X_n. \tag{B.22}$$

2. The *mean square limit* or limit in the mean,

$$X = \text{ms}\text{–}\lim_{n\to\infty} X_n, \tag{B.23}$$

   which implies that

$$\text{ms}\text{–}\lim_{n\to\infty} E\{(X_n - X)^2\} = 0. \tag{B.24}$$

3. The *stochastic limit* or *limit in probability*,

$$X = \text{st}\text{–}\lim_{n\to\infty} X_n. \tag{B.25}$$

   According to this definition

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 0 \tag{B.26}$$

   for positive ϵ.
4. The *limit in distribution*,

$$E\{f(X)\} = \lim_{n\to\infty} E\{f(X_n)\}, \tag{B.27}$$

   for any continuous bound function $f(x)$.

B.2 Stochastic Processes

A family of random variables indexed by the parameter time is known as a *random* or *stochastic process*, denoted by a collection of real (or complex) numbers

$\{X_t(\omega)\}$ or $\{X(\omega, t)\}$. If time is fixed and ω varies over the sample space, the random variable is measurable in the sense that the pre-images of any Borel set in $\mathbb{R}$ must belong to the σ-algebra of events of the probability space considered. Conversely, if ω is fixed and t varies on a given interval T, the function is real-valued on the time axis. This is called a *realization*, *stochastic trajectory* or *sample path of the stochastic process* X_t. A stochastic process can then be regarded as a map $X : \Omega \times T \to \mathbb{R}$. Multivariate stochastic processes X_t are defined as vector stochastic processes with a given number of components, each one being a real-valued stochastic process.

A stochastic process is characterized by a hierarchy of joint distribution functions $F(x, t) = P(X_t \leq x)$, $F(x_1, t_1; x_2, t_2) = P(X_{t_1} \leq x_1; X_{t_2} \leq x_2)$, etc., which satisfies two properties:

1. *Symmetry*: two distribution functions differing by a permutation of n time values are equal.
2. *Compatibility*: the lower members of the hierarchy can be obtained from the higher ones.

In terms of the probability density, if

$$F(x_1, t_1; \cdots, x_n, t_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} dx_1' \cdots dx_n' p(x_1', t_1; \cdots; x_n', t_n), \tag{B.28}$$

the compatibility property then reads as

$$p(x_1, t_1; \cdots; x_m, t_m) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} dx_{m+1}' \cdots dx_n' p(x_1, t_1; \cdots; x_n, t_n), \tag{B.29}$$

for $m < n$. According to Kolmogorov's fundamental theorem, for every hierarchy of joint distribution functions satisfying both properties, there exist a probability space $(\Omega, \mathcal{A}, P)$ and a stochastic process X_t defined on it, which possess the given distribution functions.

Two stochastic processes are said to be equivalent if they have an identical hierarchy of joint distribution functions—which does not mean that the realization of the two processes are identical. A stochastic process is said to have almost surely continuous sample paths if

$$P(\{\omega / X_t(\omega) \text{ is a continuous function on time}\}) = 1. \tag{B.30}$$

Taking into account the several definitions given above for the limit of a random variable sequence, a series of definitions of continuity for a stochastic process can also be established:

1. X_t is *continuous in probability* if for every t and positive ϵ

$$\lim_{s \to t} P(|X_s - X_t| > \epsilon) = 0. \tag{B.31}$$

2. X_t is *continuous in mean square* if for every t

$$\lim_{s \to t} E\{(X_s - X_t)^2\} = 0. \tag{B.32}$$

3. X_t is *continuous almost surely* if for every t

$$P(\{\omega / \lim_{s\to t} X_s(\omega) = X_t(\omega)\}) = 1. \tag{B.33}$$

In any of the three definitions, the continuity of the sample path is guaranteed only if the probability that a discontinuity in such paths occurs at a given time is zero.

A stochastic process is called *stationary* if all its finite-dimensional probability densities are invariant with respect to time shifts. Thus, the expectation value of a stationary stochastic process will be constant with time and the two-event (or two-dimensional) probability density will only depend on the time difference,

$$p(x_1, t_1; x_2, t_2) = p(x_1, x_2; t_2 - t_1). \tag{B.34}$$

Taking this into account, a (two-event) correlation function can be defined as

$$C(|t_2 - t_1|) = E\{(X_{t_1} - E\{X_{t_1}\})(X_{t_2} - E\{X_{t_2}\})\}, \tag{B.35}$$

from which the memory or correlation time of a stationary stochastic process X_t is given by

$$\tau_{\text{corr}} = \frac{1}{C(0)} \int_0^\infty C(\tau) d\tau, \tag{B.36}$$

with $C(\tau) = E\{(X_t - E\{X_t\})(X_{t+\tau} - E\{X_{t+\tau}\})\}$. This timescale can be considered as a measure for the rapidity of the stochastic process fluctuations. For example, short-memory or short-correlation times imply faster decreasing correlation functions. When dealing with realizations or stochastic trajectories, time averages can also be computed. For random variables, their mean value is defined as

$$\langle x \rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t) dt \tag{B.37}$$

and the corresponding autocorrelation function as

$$C(\tau) = \langle x(t)x(t+\tau) \rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau) dt. \tag{B.38}$$

If time averages and ensemble averages are equal, the stochastic process is said to be ergodic.

According to the spectral decomposition theorem, X_t can be written as a Fourier integral with random coefficients

$$X_t = \int_{\mathbb{R}} e^{i\nu t} dZ_\nu, \tag{B.39}$$

where Z_ν is a stochastic process in the complex space. It has uncorrelated increments of zero mean value. As it can be shown,

$$C(\tau) = \int_{\mathbb{R}} e^{i\nu\tau} S(\nu) d\nu, \tag{B.40}$$

where $S(\nu)$ is known as the *power* or *frequency spectrum*, which satisfies the property $E\{dZ_\nu dZ_\nu^*\} = \int S(\nu)d\nu$. This quantity is a measure of the mean square power with which an oscillation of frequency ν contributes to the process X_t. Accordingly, an effective band width can be defined,

$$\nu_{\text{eff}} = \frac{1}{S(0)} \int_0^\infty S(\nu) d\nu, \tag{B.41}$$

which becomes very broad for small correlation times.

In stochastic dynamics, there are several stochastic processes which play a very important role [7, 13]: Wiener processes, Ornstein–Uhlenbeck processes and Poisson processes. They will be briefly revised below.

B.2.1 Wiener Processes and Brownian Motion

A *Wiener* (W) *process* [14] is a mathematical model that describes the Brownian motion undergone by small particles. This type of process presents the following characteristics:

1. It undergoes very rapid motions as, for example, it happens in low viscosity fluids and at high temperatures.
2. Ceaseless motion with very irregular trajectories (the velocity of Brownian particles is undefined).

If W_t is used to denote the position of the Brownian particle from some arbitrary point at $t_0 = 0$, then $W_0 = 0$. W-processes are Gaussian, i.e., all the finite dimensional probability distributions are Gaussian and have stationary independent increments. This means that the increments $W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent for $t_1 < \ldots < t_n$. The corresponding hierarchy of probability densities is given by

$$p(x, t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} \tag{B.42}$$

and

$$p(x_1, t_1; \ldots; x_n, t_n) = p(x_1, t_1) p(x_2 - x_1, t_2 - t_1) \cdots p(x_n - x_{n-1}, t_n - t_{n-1}). \tag{B.43}$$

Although the probability density of a given increment is Gaussian, this stochastic process is not stationary itself. Furthermore, expectation values satisfy the following properties:

1. $E\{W_t\} = 0$.
2. $E\{W_t W_s\} = \min(t, s)$.
3. $E\{W_t^2\} = t$.

Wiener sample paths are continuous functions with probability one, but nowhere differentiable. In other words, $(W_{t+h} - W_t)/h$ is Gaussian distributed, but when h goes to zero the Gaussian distribution diverges. Moreover, the limit in the mean square,

$$\lim_{n\to\infty} \sum_{k=1}^{n} (W_{t_k} - W_{t_{k-1}})^2 = t - s, \tag{B.44}$$

is almost surely, where $s = t_0^{(n)} < \ldots < t_n^{(n)} = t$ is a sequence of partitions of the interval $[s, t]$, such that the size of each partition goes to zero for $n \to \infty$.

B.2.2 Ornstein–Uhlenbeck Processes and Brownian Motion

In *Ornstein–Uhlenbeck* (OU) *processes* [15], the role of the stochastic process is assigned to the velocity of the Brownian particle. Thus, the particle position (which is no longer a W-process) can then be obtained by integration. OU-processes are stationary, with $E\{X_t\} = 0$, and their correlation functions are given by decreasing exponentials,

$$E\{X_t X_s\} = \frac{\sigma^2}{2\gamma} e^{-\gamma|t-s|}, \tag{B.45}$$

where γ is the damping rate. Like W-processes, OU-processes are also Gaussian and do not have independent but correlated increments. The hierarchy of probability densities is given by

$$p(x) = \sqrt{\frac{\gamma}{\pi\sigma^2}} e^{-\gamma x^2/\sigma^2} \tag{B.46}$$

and

$$p(x_1, t_1; \ldots; x_n, t_n) = p(x_1) p(x_2, x_1; t_2 - t_1) \cdots p(x_n, x_{n-1}; t_n - t_{n-1}), \tag{B.47}$$

with

$$p(x, y; \Delta t) = \sqrt{\frac{\gamma}{\pi\sigma^2(1 - e^{-2\gamma\Delta t})}} e^{-\gamma(y - xe^{-\gamma\Delta t})^2/\sigma^2(1-e^{-2\gamma\Delta t})}. \tag{B.48}$$

As mentioned above, the integral of an OU-process,

$$Y_t = \int_0^t X_s ds, \tag{B.49}$$

with $Y_0 = 0$, renders the Brownian particle position. This integration has to be understood as *realization-wise*, i.e., as almost surely. Moreover, since X_t is a continuous function of x with probability one, the integral is well-defined—the integral

over a Gaussian process is another Gaussian process. Thus, $E\{Y_t\} = 0$, while the covariance $E\{Y_t Y_s\}$ is different from zero. The W-process is recovered in the limit $\sigma \to \infty$ and $\gamma \to \infty$, with σ^2/γ^2 constant, usually equal to 1.

B.2.3 Poisson Processes

In order to describe discrete steps, *Poisson* (P) *processes* [16] in terms of discrete variables with independent increments are considered, rather than continuous W-processes. Considering all sample paths start from zero at time zero ($X_0 = 0$), the hierarchy of probability densities are Poisson distributions,

$$p(j, t) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}, \tag{B.50}$$

with $j = 0, 1, \ldots$ and zero values for $j < 0$, and

$$p(j_1, t_1; \ldots; j_n, t_n) = p(j_1, t_1) p(j_2 - j_1, t_2 - t_1) \cdots p(j_n - j_{n-1}, t_n - t_{n-1}). \tag{B.51}$$

The increments of a P-process are also stationary, although the P-process itself is not stationary in a strict sense, since

1. $E\{X_t\} = \lambda t$
2. $E\{X_t X_s\} = \lambda \min(t, s)$

P-processes do not have almost surely continuous sample paths and are not almost surely differentiable functions.

B.3 Markov Processes and Noise

In general, if the power spectrum has a finite effective frequency band, the process is said to keep *memory* of its past. This is often described by means of a *colored noise*. However, there are many physical situations of interest where the fluctuations of a surrounding or *environment* are very fast. This gives rise to very broad effective frequency bands, which may cover frequencies even higher than those characterizing the system, described by some stochastic process X_t. In the limit $\tau_{\text{corr}} = 0$ or *zero memory*, the subsequent values of the stochastic variables describing the process at each time are independent, i.e., they are completely random. In these cases, the changes induced in the system by the environment will essentially depend on the strength of the latter fluctuations, increasing as such fluctuations become larger. For example, consider an OU-process. If $\tau_{\text{corr}} \to 0$ and $\sigma \to \infty$, the power spectrum is constant (flat) and therefore its corresponding correlation function becomes a Dirac δ-function. Actually, this δ-function can also be obtained from a Gaussian function whose variance goes to zero, i.e.,

$$\delta(t-s) = \lim_{\sigma^2 \to 0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-s)^2/2\sigma^2}. \tag{B.52}$$

A δ-correlated process, with a flat spectrum, is called a *white noise*, since all frequencies are present and contribute equally, as in the case of white light. Since the OU-process is a Gaussian process, in the no-memory limit this process is known as *Gaussian white noise*. White noise can also be interpreted as the time-derivative of a process with stationary independent increments. In this sense, Gaussian white noise would be the time-derivative of a W-process; similarly, the same can be said for a P-process. However, neither the W-process nor the P-process are differentiable in the mean square sense. Thus, one can write

$$\xi_t = \dot{W}_t, \tag{B.53}$$

with $E\{\dot{W}_t\} = 0$ and $E\{\dot{W}_t \dot{W}_s\} = \delta(t-s)$.

Markov processes are stochastic processes with their time-evolution only depending on the present time, thus displaying a very short memory—i.e., the past history is rapidly forgotten. The hierarchy of joint probabilities can be reconstructed from just two distribution functions. It can be shown that a system is Markovian if the fluctuations are white, whereas non-Markovian systems are ruled by colored noise. Accordingly, the mathematical condition for a process to be Markovian is

$$P(X_t \in B / X_{t_m} = x_m, \ldots, X_{t_1} = x_1) = P(X_t \in B / X_{t_m} = x_m). \tag{B.54}$$

This condition holds for all ordered set of times $t_1 < t_2 < \ldots < t_m < t$, for all Borel sets B and all $x_1, x_2, \ldots, x_m \in \mathbb{R}^d$. Thus, the probability of the event $X_t \in B$, conditioned on m previous events, only depends on the latest event $X_{t_m} = x_m$,

$$p(x, t|x_m, t_m; \ldots; x_1, t_1) = p(x, t|x_m, t_m). \tag{B.55}$$

This conditional probability is also called *transition probability* or *propagator*,

$$T(x, t|x', t') = p(x, t|x', t'), \tag{B.56}$$

which satisfies the following properties:

1. $\int T(x, t|x', t')dx = 1$ (normalization).
2. $\lim_{t \to t'} T(x, t|x', t') = \delta(x - x')$.

When the propagator only depends on the time-difference $t' - t$, it is called *homogeneous*. In this sense, for example, the W-process is a homogeneous process in time, although it is not stationary.

The so-called *Chapman–Kolmogorov* (CK) *equation* for the propagator is

$$T(x_3, t_3|x_1, t_1) = \int T(x_3, t_3|x_2, t_2) T(x_2, t_2|x_1, t_1) dx_2, \tag{B.57}$$

which in its differential version reads as

$$\frac{\partial T(x, t|x', t')}{\partial t} = \mathcal{L}(t) T(x, t|x', t'), \tag{B.58}$$

where $\mathcal{L}(t)$ is a linear operator generating infinitesimal time translations,

$$\mathcal{L}(t) \equiv \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int T(x, t + \Delta t | x', t) dx' - \delta(x - x'). \tag{B.59}$$

For a homogeneous Markov process, $\mathcal{L}$ is time-independent. In this case, (B.59) can be formally solved to give

$$T_\tau(x|x') = e^{\mathcal{L}\tau} \delta(x - x'), \tag{B.60}$$

for $\tau \geq 0$. The one-parameter family $\{T_\tau / \tau \geq 0\}$ represents a dynamical semigroup [17], since τ is restricted to nonnegative values. This fact is related to irreversibility, which is the characteristic feature of any stochastic process. In other words, the exponential operator is not invertible in the total space of all probability distributions. The simplest Markov process is given by a deterministic process: $\dot{x} = f(x)$, with $x(t) \in \mathbb{R}^d$ and $f(x) \in \mathbb{R}^d$. Consider the phase flow associated with such a differential equation is denoted by $\Phi_t(x)$ (see Chap. 1) and the phase curve with time is obtained for a fixed x and initial condition $\Phi_0(x) = x$. The corresponding propagator is [17]

$$T(x, t|x', t') = \delta(x - \Phi_{t-t'}(x')), \tag{B.61}$$

By appealing to the semigroup property $\Phi_t(\Phi_s(x)) = \Phi_{t+s}(x)$, it can be shown that the CK equation is satisfied by the propagator (B.61). The differential CK equation for a deterministic process is the Liouville equation,

$$\frac{\partial T(x, t|x', t')}{\partial t} = -\sum_i \frac{\partial \left[f_i(x) T(x, t|x', t') \right]}{\partial x_i}, \tag{B.62}$$

where only the initial conditions are assumed to be random. The time-evolution of this equation describes the deterministic drift.

Sometimes it is necessary to describe instantaneous jump processes. The corresponding differential CK equation is then given by the master equation

$$\begin{aligned} \partial_t T(x, t|x', t') &= \mathcal{L}(t) T(x, t|x', t') \\ &= \int \left[W(x|x'', t) T(x'', t|x', t') - W(x''|x, t) T(x, t|x', t') \right] dx'', \end{aligned} \tag{B.63}$$

where $W(x|x', t)$ is the *transition rate* accounting for the instantaneous jump from the state x' at time t to the state x. The total rate for a jump at time t, $\Gamma(x', t)$, is obtained by integrating $W(x|x', t)$ over x,

$$\Gamma(x', t) = \int W(x|x', t) dx. \tag{B.64}$$

A more standard form for (B.63) is

$$\frac{\partial p(x,t)}{\partial t} = \int \left[W(x|x',t)p(x',t) - W(x'|x,t)p(x,t)\right]dx', \tag{B.65}$$

where $p(x,t)$ is the probability density. The so-called Kramers–Moyal expansion of the master equation can be obtained from a different rewriting of (B.63) after performing an appropriate Taylor series expansion. This leads to a partial differential equation of infinite order.

The continuity condition for Markov processes arises when the probability for a transition, during an increment of time with size larger than a given small amount, decreases more rapidly than the increment of time as it goes to zero. Diffusion processes satisfy this condition, but they are not deterministic. The corresponding differential CK equation is the Fokker–Planck equation expressed in terms of the probability density,

$$\frac{\partial p(x,t)}{\partial t} = -\sum_i \frac{\partial\left[g_i(x,t)p(x,t)\right]}{\partial x_i} + \frac{1}{2}\sum_{i,j} \frac{\partial^2\left[D_{ij}p(x,t)\right]}{\partial x_i \partial x_j}, \tag{B.66}$$

where g_i and D_{ij} are the first and second moments of the jump distribution (the drift and diffusion coefficients, respectively). The D matrix is known as the *diffusion matrix*, which is symmetric and positive semidefinite. The Fokker–Planck equation (B.66) can be seen as a truncation of the Kramers–Moyal expansion to second order. Moreover, it can also be recast as a continuity equation,

$$\frac{\partial p(x,t)}{\partial t} + \sum_i \frac{\partial J_i(x,t)}{\partial x_i} = 0, \tag{B.67}$$

after defining the probability current density

$$J_i(x,t) = g_i(x,t)p(x,t) - \frac{1}{2}\sum_j \frac{\partial\left[D_{ij}p(x,t)\right]}{\partial x_j}. \tag{B.68}$$

Similar equations can also be written for the propagator.

Piecewise deterministic processes, arising from the combination of deterministic and jump processes, are also very important in many applications of the theory of open systems. The corresponding differential CK equation is known as the *Liouville master equation*, which in terms of the propagator, reads as

$$\begin{aligned}\frac{\partial T(x,t|x',t')}{\partial t} = &-\sum_i \frac{\partial[g_i(x,t)T(x,t|x',t')]}{\partial x_i} \\ &+ \int [W(x|x'')T(x'',t|x',t') - W(x''|x)T(x,t|x',t')]dx''.\end{aligned} \tag{B.69}$$

The so-called *waiting time distribution*, $F(\tau|x', t')$, is the probability for the next jump to occur during the time interval $[t', t'+\tau]$ starting from x' at time t'. Its general expression is given by

$$F(\tau|x', t') = 1 - e^{-\int_0^\tau ds\Gamma(\Phi_s(x'))}. \tag{B.70}$$

If one has a pure jump process, the drift is zero and the waiting time distribution is an exponential function.

If the propagator $T(x, t|x', t') = T_{t-t'}(x - x')$ is invariant with respect to space-time translations, the stochastic process is a *Lévy process* [18, 19] (homogeneous both in space and time). These two requirements allow to directly work with the integral CK equation. Lévy processes are also stable if by performing a time scaling the new process can be expressed as the original one multiplied by a scaling factor. This property is called *self-similarity* or *fractality*.

B.4 Stochastic Differential Equations

In order to describe piecewise deterministic processes in terms of random variables instead of probability densities or propagators, a stochastic differential equation (SDE) is necessary to account for their time-evolution. For example, in the case of a Gaussian white noise, the time-evolution of a one-dimensional random variable X_t is governed by the SDE

$$dX_t = f(X_t)dt + bg(X_t)dW_t. \tag{B.71}$$

In integral form, this equation reads as

$$X_t = X_0 + \int_0^t f(X_s)ds + b\int_0^t g(X_s)dW_s, \tag{B.72}$$

where the first integral can be understood as an ordinary Riemann integral, while the b coefficient of the second one is related to the diffusion coefficient. This second integral, on the contrary, is problematic due to the intrinsic properties of the W-process. While the Riemann integral is independent of the different evaluation points along the interval $[0, t]$, the same does not hold for the stochastic integral. To obtain an unambiguous definition of the stochastic integral, a choice of the evaluation points has to be previously decided. Thus, if

$$\tau_i^{(n)} = (1-\alpha)t_{i-1}^{(n)} + \alpha t_i^{(n)}, \tag{B.73}$$

with $0 \leq \alpha \leq 1$ and $t_0 = t_1^{(n)} < \cdots < t_n^{(n)} = t$, there are two reasonable choices (which lead to desirable features):

1. The *Itô integral*, if $\alpha = 0$.
2. The *Stratonovich integral*, if $\alpha = 1/2$.

Although the latter conserves the ordinary rules of integration, the Itô integral is much more appealing from a mathematical viewpoint, for it retains the important properties of the W-process—indeed, for a white noise it is the only reasonable choice. The integral of the quantity $W_s dW_s$ within the interval $[0, t]$ can be used to illustrate how each one of these integration schemes works. In the case of Itô integration,

$$\int_{t_0}^{t} W_s dW_s = \frac{1}{2}\left[(W_t^2 - W_{t_0}^2) - (t - t_0)\right], \tag{B.74}$$

while in the case of Stratonovich integration, the integral becomes

$$\int_{t_0}^{t} W_s dW_s = \frac{1}{2}\left(W_t^2 - W_{t_0}^2\right). \tag{B.75}$$

An important mathematical result establishes that any stochastic integral based on a different choice of α can be expressed as the sum of the corresponding Itô integral plus an ordinary Riemann integral.

Only a certain class of stochastic processes can be used in the Itô integral, the so-called *non-anticipating stochastic processes*, G_t. These processes have the property that G_t is only known from the past history of the W-process up to time t. That is, G_t is independent of all increments of the W-process for times greater than t. Thus, the stochastic process $Y_t = \int_{t_0}^{t} G_s(\omega) dW_s(\omega)$ is an Itô integral if:

1. Y_t is a non-anticipating process.
2. Y_t has almost surely continuous sample paths.
3. If $\int_{t_0}^{t} E\{G_s^2\} ds < \infty$, then $E\{Y_t\} = 0$ and $E\{Y_t Y_s\} = \int_{t_0}^{min(t,s)} E\{G_u^2\} du$.
4. $E\{Y_t | \mathcal{F}_s\} = Y_s$ (the martingale property) with $s \leq t$; $\mathcal{F}_s$ with $s \geq 0$ is a filtration or family of sub-σ algebras of $\mathcal{F}$ if $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$, with $s \leq t$.

The Stratonovich SDE is similar to the corresponding Itô one, but with the difference that integration follows the usual rules of the Riemann integral, as said above. Thus, given a SDE, it is always possible to change of integration scheme or interpretation, Itô or Stratonovich. For additive noise ($g(X_t) =$ constant), there is no difference between both schemes. However, for multiplicative noise ($g(X_s) \neq$ constant), there are important differences. Both integrals lead to mathematically consistent calculus. However, the question on which form is the correct one to describe physical systems has led to a very long controversy [20] and, in the end, the last word always comes from the comparison with the experiment. In any case, the standard Langevin equation can always be expressed in terms of a SDE given by (B.71).

The Itô stochastic calculus is based on the fact that second order terms have to be retained. Thus, whereas in ordinary calculus $(dt)^2 \to 0$ and $dt\, dW_t \to 0$, in

stochastic calculus $(dW_t)^2 = dt$. For example, given $f = f(W_t, t)$, the Taylor series expansion of f leads to

$$df(W_t, t) = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W_t^2}\right)dt + \frac{\partial f}{\partial W_t}dW_t. \tag{B.76}$$

A stochastic process X_t obeys an Itô SDE,

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t, \tag{B.77}$$

whenever

$$X_t = X_{t_0} + \int_{t_0}^{t} a(X_{t'}, t')dt' + \int_{t_0}^{t} b(X_{t'}, t')dW_{t'} \tag{B.78}$$

for all t_0 and t. If the solution of the Itô SDE is unique, then it is Markovian. If X_t fulfills the Itô SDE and f is a functional of X_t and t, then

$$\begin{aligned} df(X_t, t) = &\left[\frac{\partial f(X_t, t)}{\partial t} + a(X_t, t)\frac{\partial f(X_t, t)}{\partial X_t} + \frac{1}{2}b^2(X_t, t)\frac{\partial^2 f(X_t, t)}{\partial X_t^2}\right]dt \\ &+ b(X_t, t)\frac{\partial f(X_t, t)}{\partial X_t}dW_t, \end{aligned} \tag{B.79}$$

which is known as *Itô formula*. It can be proven that if a and b depend explicitly on time, the SDE defines a diffusion process. The stochastic equation is always linear in dW_t. A generalization to more than one dimension is quite straightforward.

Following the similar prescription, in general, given a function $f(\mathbf{x}(t), t)$ of a vector stochastic variable $\mathbf{x}(t)$, a second-order series expansion in increments leads to

$$df(\mathbf{x}, t) = \frac{\partial f(\mathbf{x}, t)}{\partial t}dt + \nabla f(\mathbf{x}, t) \cdot d\mathbf{x} + \frac{1}{2}\nabla^2 f(\mathbf{x}, t)(d\mathbf{x})^2. \tag{B.80}$$

If $\mathbf{x}(t)$ satisfies the Itô SDE

$$d\mathbf{x}(t) = \mathbf{a}(\mathbf{x}, t)dt + bd\mathbf{W}(t), \tag{B.81}$$

then

$$df(\mathbf{x}, t) = \left[\frac{\partial f(\mathbf{x}, t)}{\partial t} + \mathbf{a}(\mathbf{x}, t)\nabla f(\mathbf{x}, t) + \frac{1}{2}b^2\nabla^2 f(\mathbf{x}, t)\right]dt + b\nabla f(\mathbf{x}, t)d\mathbf{W}(t). \tag{B.82}$$

In this diffusion process, the probability distribution is described by the Fokker–Planck equation

$$\frac{\partial F(\mathbf{x}, t)}{\partial t} = -\nabla \cdot [\mathbf{a}(\mathbf{x}, t)F(\mathbf{x}, t)] + \frac{b^2}{2}\nabla^2 F(\mathbf{x}, t). \tag{B.83}$$

In some applications it is also very useful to define the so-called *mean forward* and *backward derivatives* of a given function $f(\mathbf{x}(t), t)$ [21],

$$\begin{aligned} D_+ f(\mathbf{x}(t), t) &= \frac{\partial f(\mathbf{x}, t)}{\partial t} + \mathbf{a}_+(\mathbf{x}, t)\dot{\nabla} f(\mathbf{x}, t) + \frac{1}{2} b^2 \nabla^2 f(\mathbf{x}, t), \\ D_- f(\mathbf{x}(t), t) &= \frac{\partial f(\mathbf{x}, t)}{\partial t} + \mathbf{a}_-(\mathbf{x}, t)\dot{\nabla} f(\mathbf{x}, t) - \frac{1}{2} b^2 \nabla^2 f(\mathbf{x}, t), \end{aligned} \tag{B.84}$$

respectively, where $D_+\mathbf{x}(t) = \mathbf{a}_+(\mathbf{x}, t)$ and $D_-\mathbf{x}(t) = \mathbf{a}_-(\mathbf{x}, t)$. If $\mathbf{x}(t)$ defines the position of a given particle and is symmetric under time reversal, i.e., $\mathbf{x}(t) = \mathbf{x}(-t)$, the corresponding Itô SDEs are

$$\begin{aligned} d\mathbf{x}(t) &= \mathbf{a}_+(\mathbf{x}, t)dt + b d\mathbf{W}(t), \\ d\mathbf{x}(t) &= \mathbf{a}_-(\mathbf{x}, t)dt + b d\mathbf{W}_-(t), \end{aligned} \tag{B.85}$$

where $E[d\mathbf{W}_-(t) \cdot d\mathbf{W}_-(t)] = |dt| = -dt$ for the reverse time flow. The Newtonian velocity is then defined as the mean velocity (or flow velocity) of a Brownian particle,

$$\mathbf{v}(\mathbf{x}, t) = \frac{1}{2}\left[\mathbf{a}_+(\mathbf{x}, t) + \mathbf{a}_-(\mathbf{x}, t)\right]. \tag{B.86}$$

If both derivatives are not equal, their difference defines the vector field

$$\mathbf{u}(\mathbf{x}, t) = \frac{1}{2}\left[\mathbf{a}_+(\mathbf{x}, t) - \mathbf{a}_-(\mathbf{x}, t)\right], \tag{B.87}$$

which is called the *osmotic velocity* [21]. Similarly, a mean acceleration can be defined as

$$\mathbf{A}(\mathbf{x}, t) = \frac{1}{2}\left[D_+\mathbf{a}_-(\mathbf{x}, t) + D_-\mathbf{a}_+(\mathbf{x}, t)\right]. \tag{B.88}$$

One of the central aspects of stochastic dynamics is the concept of *noise-induced transition*. Under some conditions, the influence of environmental fluctuations can be far from being negligible. Indeed, the effect of external noise may depend on the system state. A transition takes place at points of the parameter space (mean value of the external noise, its variance, its correlation time, etc.) where the functional form of the mapping from the sample space into the state space changes qualitatively—e.g., the number and location of extrema of the stationary probability density [5]. This probability density is usually written in terms of a "probability" potential which is subject to topological analysis. It can be shown that only when the external noise is multiplicative, new potential wells or states can be created and some of the existing ones will be destroyed. In this regard, it is of particular interest the so-called *stochastic resonance mechanism* [22].

B.5 Stochastic Processes in Quantum Mechanics

In quantum mechanics, the general probability theory has to be applied in a rather different way, within the context of Hilbert spaces $\mathcal{H}$ and noncommuting algebras of operators, variables and observables or measurable quantities. Details on the mathematical notions involved in the formulation of this theory and its statistical interpretation [23, 24], and more particularly in the theory of measurement [17, 25], will not be accounted for here. Only those elements which are more closely related to the issues considered throughout this monograph are briefly discussed. To simplify the notation, both the random variable and its Hermitian operator will be denoted by the same symbol A, being easy to discern when the latter plays the role of a variable or an operator, respectively.

The basic tool when dealing with statistical mixtures is the statistical or density operator ρ. This operator has to be understood as describing a statistical ensemble consisting of a large number, N, of identical quantum systems prepared in the following way: M sub-ensembles each one described by a normalized state vector $\psi_n (n = 1, \cdots, M)$ formed by N_n elements. Accordingly, the density operator reads as

$$\rho = \sum_n w_n |\psi_n\rangle\langle\psi_n|, \tag{B.89}$$

where the statistical weight of each state is given by $w_n = N_n/N$, with $N = \sum_{n=1}^{M} N_n$. The density operator is Hermitian or self-adjoint, positive and with trace equal to one. A further discussion on the density operator and its properties can be found, for example, in [24, 26].

An observable via its spectral family leads to a real random variable, which describes the probabilities for all possible measurement outcomes. The sample space is the real axis and the algebra of events $\mathcal{B}$ is given by the Borel sets of the real axis. Thus,

$$E\{A\} = \mathrm{Tr}\,\{A\rho\} \equiv \langle A\rangle, \tag{B.90}$$

where Tr stands for the trace operation. Within this context, the associated variance, which renders information about the dispersion of A, reads as

$$\sigma(A) \equiv (\Delta A)^2 = \mathrm{Tr}\,\{A^2\rho\} - (\,\mathrm{Tr}\,\{A\rho\})^2 \equiv \langle A^2\rangle - \langle A\rangle^2. \tag{B.91}$$

That is, according to this expression, there are no dispersion-free measurements.

Observables are the result of measurements and therefore display fluctuations around some average value. These average values can be obtained from pure or mixed state.[2] In the case of a pure state, if it corresponds with an eigenvector of A, the fluctuation is zero. However, those values have nothing to do with the typical

[2] The distinction between pure and mixed states is usually characterized by the definition of the convex linear combination in the convex set of density operators in the Hilbert space. Thus, ρ is a

errors associated with the measuring devices. In this regard, Heisenberg's uncertainty principle relates very precisely the fluctuations of two non-commuting operators, stating that the corresponding observables cannot be simultaneously measured with an infinite precision. Usually, after a measurement the system is projected in an eigenstate of the measuring device (ideal measurement). However, not always this is the case. Indeed, within the context of the quantum theory of measurement [25], Aharonov et al. [27] introduced the definition of *weak measurement*. This concept is based on assuming that the coupling between measuring device and observed system can be set so weakly that the uncertainty associated with a single measurement is still very large compared with the separation between the observable eigenvalues. Hence, after a weak measurement the system is not left in an eigenstate of the observable, but rather in a superposition of the unresolved eigenstates. This procedure is useful for the amplification and detection of weak effects [28–30], e.g., a direct detection of the photon wave function [9] or the associated photon trajectories [31].

In the case of dissipative and/or stochastic dynamics, the operator $\delta A = A - \langle A \rangle$ is also called a *fluctuation*, and it can be shown that $(\Delta(\delta A))^2 = (\Delta A)^2$ (see (B.9)). When one is interested in the average moments $\langle A^n \rangle$ (n being an integer) of a given operator A, sometimes it is more convenient to evaluate the so-called characteristic function of A, $\langle e^{i\xi A} \rangle$, where ξ is a real parameter—note that the nth moment arises from the partial differentiation with respect to $i\xi$ of nth order and evaluated at $\xi = 0$. Moreover, the Fourier transform of the characteristic function of A gives the probability distribution function or diagonal matrix elements of the density operator in the representation in which A is diagonal. In this regard, for example, the Wigner, Poisson and exponential distribution functions can be expressed as the Fourier transform of a given characteristic function [26]. When a set of observables commute, a characterization by joint probability distributions is also possible. For the particular case of the Wigner distribution, which is given in terms of the phase space coordinates (e.g., the position and momentum of a particle), see for example [26].

The noise in quantum mechanics has to be treated with some special care. For example, in the quantum Langevin equation the noise is a quantum operator and the noise (symmetric) autocorrelation function is a complex quantity because, in general, it does not commute at different times. At very low or zero temperatures, the noise is still correlated at very long times [32, 33]. Furthermore, it is not proportional to a Dirac delta function in time. Thus, we have the situation that, although there is no memory in the standard Langevin equation, the quantum process is not Markovian [34].

When dealing with open quantum systems, the total Hamiltonian is often expressed as the sum of the Hamiltonian accounting for the quantum system of interest (S), the Hamiltonian accounting for the environment (B) and their coupling. A measuring device can also be seen as an environment [17]. Many times, though, only the

---

convex linear combination of ρ_1 and ρ_2 when

$$\rho = \alpha\rho_1 + (1 - \alpha)\rho_2,$$

with $\alpha \in [0, 1]$. If ρ describes a pure state, then $\rho = \rho_1 = \rho_2$.

dynamics of the system of interest is relevant, which is obtained after tracing out over the environment (or bath) degrees of freedom. This gives rise to a *reduced* system dynamics, which is no longer unitary in time. Quantum Markov processes represent the simplest description of open (quantum) system dynamics. However, the extension of Markov processes to quantum mechanics requires the definition of a quantum dynamical semigroup [17]. Thus, consider a general dynamical map $V(t)$ which describes the transformation over time

$$\rho_S(t) = V(t)\rho_S(0),$$

where $\rho_S \equiv \mathrm{Tr}_B(\rho)$ is the *reduced density operator* or *reduced density matrix*. This map represents a convex-linear, completely positive and trace-preserving quantum operation. A quantum dynamical semigroup is defined as follows:

1. If $V(t)$ defines a continuous, one-parameter family of dynamical maps by varying t (with $t \geq 0$).
2. The Markov approximation for the homogeneous case is assumed.
3. The semigroup property $V(t)V(s) = V(t+s)$, with $t, s \geq 0$, holds.

This gives rise to a first-order linear differential equation for the reduced density operator, the so-called *Linblad equation*,

$$\dot{\rho}_S(t) = \mathcal{L}\rho_S(t), \tag{B.93}$$

where the generator $\mathcal{L}$ of the semigroup is a super-operator, since its action over operators yields another operator.

Coherence is a key issue in different branches of quantum mechanics. The opposite effect, decoherence, it appears when a quantum interference pattern is destroyed or suppressed. In a certain sense, it could be said that decoherence leads to the appearance of a classical world in quantum mechanics. Environment induced decoherence [25] is omnipresent and, in general, it is a short time phenomenon. When measuring, the dynamics of the system tends to be decoherent. In the weak coupling approximation between the system and environment, the trace operation carried out to obtain the reduced density matrix is key to the existence of any master equation. However, this operation is questionable when the system and environment are entangled at all times, including the initial state. The role of the initial conditions has also be widely discussed in the corresponding literature.

In order to define the state vector as a random variable in Hilbert space, it is necessary a new type of quantum-mechanical ensembles, which are not fully characterized by a density matrix [17]. This new ensemble is formed by M statistical ensembles of the type described above to define the statistical operator. Each one of the M ensembles consists of N_n elements prepared in a normalized state $|\psi_n\rangle$, in such a way that $N = \sum_{n=1}^{M} N_n$. This is the new sample space, where w_n is the measure corresponding to the ensemble characterized by the state vector $|\psi_n\rangle$.

Different types of measurements can be defined from different probability density functionals, $P[\psi]$. For example, for a subset A, such that $A \in \mathcal{A}$,

$$\mu(A) = \int_A D\psi D\psi^* P[\psi]. \tag{B.94}$$

If $P[\psi] = \delta(\psi - \psi_0)$, one obtains the Dirac measure. The volume element in Hilbert space can be chosen to be the Euclidean volume element [17]. The expectation values are then obtained from

$$E\{F[\psi]\} = \int D\psi D\psi^* P[\psi] F[\psi] \tag{B.95}$$

for a functional $F[\psi]$—e.g., $F[\psi] = \text{Tr}\,\{A\rho\} = \langle\psi|A|\psi\rangle$. Taking into account this new ensemble, the density operator can then be represented as

$$\rho = E\{|\psi\rangle\langle\psi|\} = \int D\psi D\psi^* P[\psi]|\psi\rangle\langle\psi|. \tag{B.96}$$

The time-dependence of the state vector leads to stochastic processes $|\psi(t)\rangle$ in Hilbert space, where the probability density functional is $P[\psi, t]$. The dynamics of the open quantum system can then be described by a differential stochastic equation for $|\psi(t)\rangle$ instead of a master equation. This is the idea behind what it is termed as *unravelling* of the master equation. The evolution in time of $P[\psi, t]$ can be expressed as

$$P[\psi, t] = \int D\phi D\phi^* T[\psi, t|\phi, t_0] P[\phi, t_0], \tag{B.97}$$

where T gives the conditional transition probability. Nevertheless, a similar prescription as before can be followed, which allows to describe the time-evolution of the wave function as a diffusion process. This process is governed by a SDE,

$$d|\psi(t)\rangle = -iK\{|\psi(t)\rangle\}dt + bM\{|\psi(t)\rangle\}dW(t), \tag{B.98}$$

where K is a drift operator and M is related to a certain diffusion operator. As can be noticed, here the noise is multiplicative, since the coefficient accompanying $dW(t)$ is not a constant (it depends on the wave function itself). This noise may lead to transitions, namely *noise-induced transitions*, which is an important aspect of the stochastic dynamics not fully developed yet.

References

1. Chandrasekhar, S.: Stochastic problems in physics and astronomy. Rev. Mod. Phys. **15**, 1–89 (1943)
2. Papoulis, A.: Probability, Random Variables and Stochastic Processes. McGraw-Hill, New York (1965)

3. van Kampen, N.G.: Stochastic Processes in Physics and Chemitry. North-Holland, Amsterdam (1981)
4. Risken, R.: The Fokker–Planck Equation. Springer, Berlin (1984)
5. Horsthemke, W., Lefever, R.: Noise-Induced Transition.. Springer, Berlin (1984)
6. Shiryaev, A.N.: Probability. Springer, Berlin (1995)
7. Wolfgang, P., Baschnagel, J.: Stochastic Processes. From Physics to Finance. Springer, Berlin (1999)
8. Margalef-Roig, J.: Curso de ocesos Estocásticos. CSIC, Madrid (2000)
9. Lundeen, J.S., Sutherland, B., Patel, A., Stewart, C., Bamber, C.: Direct measurement of the quantum wavefunction. Nature **474**, 188–191 (2011)
10. Gardiner, C.W.: Handbook of Stochastic Methods. Springer, Berlin (1983)
11. Radon, J.: Theorie und Anwendungen der Absolut Additiven Mengenfunktionen. Sitzungsber. Akad. Wissen. Wien **112**, 1295–1438 (1913)
12. Nikodym, O.: Sur une Généralisation des Intégrales de M. J. Radon. Fundamenta Mathematicæ **15**, 131–179 (1930)
13. Ross, S.M.: Stochastic Processes. Wiley, New York (1996)
14. Wiener, N.: The average of an analytical functional and the Brownian movement. Proc. Natl Acad. Sci. USA **7**, 294–298 (1921)
15. Uhlenbeck, G.E., Ornstein, L.S.: On the theory of Brownian Motion. Phys. Rev. **36**, 823–841 (1930)
16. Cox, D.R., Isham, V.I.: Point Process. Chapman & Hall, London (1980)
17. Breuer, H.-P., Petruccione, F.: The Theory of Open Quantum Systems. Clarendon Press, Oxford (2006)
18. Lévy, P.: Processus stochastiques et mouvement brownien. Gauthier-Villars, Paris (1965)
19. Applebaum, D.: Lévy processes—from probability to finance and quantum groups. Not. Am. Math. Soc. **51**, 1336–1347 (2004)
20. van Kampen, N.G.: Itô versus Stratonovich. J. Stat. Phys. **24**, 175–187 (1981)
21. Nelson, E.: Derivation of the Schrödinger equation from Newtonian mechanics. Phys. Rev. **150**, 1079–1085 (1966)
22. Gammaitoni, L., Hänggi, P., Jung, P., Marchesoni, F.: Stochastic resonance. Rev. Mod. Phys. **70**, 223–287 (1998)
23. Ballentine, L.E.: The statistical interpretation of Quantum Mechanics. Rev. Mod. Phys. **42**, 358–381 (1970)
24. Ballentine, L.E.: Quantum Mechanics. A Modern Development. World Scientific, Singapore (1998)
25. Wheeler, J.A., Zurek, W.H.: Quantum Theory and Measurement. Princeton University Press, Princeton (1983)
26. Louisell, W.H.: Quantum Statistical Properties of Radiation. Wiley Classics Library, New York (1973)
27. Aharonov, Y., Albert, D.Z., Vaidman, L.: How the resuslt of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. Phys. Rev. Lett. **60**, 1351–1354 (1988)
28. Ritchie, N.W.M., Story, J.G., Hulet, R.G.: Realization of a measurement of a "weak value". Phys. Rev. Lett. **66**, 1107–1110 (1991)
29. Wiseman, H.M.: Grounding Bohmian mechanics in weak values and bayesianism. New J. Phys. 9:165(1–12) (2007)
30. Mir, R., Lundeen, J.S., Mitchell, M.W., Steinberg, A.M., Garretson, J.L., Wiseman, H.M.: A double-slit 'which-way' experiment on the complementarity-uncertainty debate. New J. Phys. 9:287(1–11) (2007)
31. Kocsis, S., Braverman, B., Ravets, S., Stevens, M.J., Mirin, R.P., Shalm, L.K., Steinberg, A.M.: Observing the average trajectories of single photons in a two-slit interferometer. Science **332**, 1170–1173 (2011)

32. Ingold, G.-L.: Path Integrals and their Application to Dissipative Quantum Systems. Lecture Notes Physics, vol. 611, pp. 1–53. Springer, Berlin (2002)
33. Gardiner, C.W., Zoller, P.: Quantum Noise.. Springer Complexity, Berlin (2004)
34. Ford, G.W., Lewis, J.T., O'Connell, R.F.: Quantum Langevin equation. Phys. Rev. A **37**, 4419 (1988)

Index

A. S. Sanz and S. Miret-Artés, *A Trajectory Description of Quantum Processes. I. Fundamentals*, Lecture Notes in Physics 850, DOI: 10.1007/978-3-642-18092-7,