1
Basic Concepts of the Probability Theory

In order to formulate the theoretical concepts that will be crucial for the subsequent chapters, we must first mention some basic notions of the probability theory and their further ramifications. This chapter provides a brief outline. The interested reader will find a more detailed discussion of the relevant topics in textbooks and monographs on the probability theory and statistical physics (see the list of references at the end of the chapter).

1.1
Events, Set of Events, and Probability

Some of the typical situations considered in this chapter are as follows: a particle is in small volume (elementary volume) that includes a point X; N particles are in a small region of space; N_1 particles of type 1 and N_2 particles of type 2 are in a certain region. Speaking more generally, we shall be dealing with situations having to do with particle positions in space at different instants of time. Every such case can be thought of as a specific realization of some event. We shall define an event as an element of a certain space of events. In what follows, we will most often be using vector spaces defined by vectors $\boldsymbol{X}$ $(X_1, X_2, \ldots, X_n)$, where X_i are real numbers. The probability theory introduces the notion of a set of events. Then the condition that an event ω belongs to a set of events A is written as $\omega \in A$.

Consider an event of finding a particle in some volume element ΔV centered at a point $\boldsymbol{X}$. Let all such events form an ensemble of events denoted as $A(\Delta V, \boldsymbol{X})$. It turns out that it is possible to determine whether the particle is in the vicinity of X, but it makes no sense to determine whether the particle is exactly at the point X. Physically, the probability of finding a particle exactly at some given point is zero. As an idealization one can consider a probability distribution given by a delta function, in which case the probability can be a finite number, and it is associated with an infininesimal interval around this point as the interval tends to zero. Let $\omega(\boldsymbol{X})$ denote the event that a particle is found in infinitesimal volume element centered at $\boldsymbol{X}$. We can then ascribe a certain probability to the condition $\omega(\boldsymbol{X}) \in A(\Delta V, \boldsymbol{X})$. This is just the probability for the particle to be in the volume element ΔV centered at $\boldsymbol{X}$.

Statistical Microhydrodynamics. Emmanuil G. Sinaiski and Leonid Zaichik

ISBN: 978-3-527-40656-2

The probability $P(A)$ of a set of events is defined as a function of A which satisfies the following probability axioms:

1. $P(A) \geq 0$ for all A;
2. $P(\Omega) = 1$;
3. if $\{A_i\}$ is a finite or countable sequence of non-overlapping sets, that is, $(A_i \cap A_j = \phi)$, then $P(\bigcup_i A_i) = \sum_i P(A_i)$.

The two consequences of these axioms are:

1. $P(\bar{A}) = 1 - P(A)$;
2. $P(\phi) = 0$.

Here $\cap$ and $\cup$ denote, respectively, the intersection and union operations on sets; Ω is the set of all events; ϕ – the empty set; $\bar{A}$ – the complement of a set A, that is set of all events not belonging to A.

Coexisting with the notion of probability is the notion of frequency of an event. To understand the difference between probability and frequency of events, consider an event ω selected at random from the full set Ω. The number of occurrences of $\omega \in A$ in N trials gives us the relative frequency of realization of the event $\omega \in A$. When N is increased, the relative frequency goes to the limit $P(A)$, which is defined as the probability of the event A. At $N \gg 1$, it is safe to assume the relative frequency of an event to be equal to the probability of occurrence of this event presuming that relative frequency has been normalized.

The axiom 3 is given for a finite or countable number of sets. But often it is necessary to deal with uncountable infinite number of sets. For instance, when studying the motion of particles under the action of external forces, one has to deal with sets of particle positions in spacetime. Let $\boldsymbol{X}$ is the position of a particle in space. The probability for the particle to be exactly at point $\boldsymbol{X}$ (it would then belong to a set consisting of only one element) is equal to zero, while the probability to find the particle in the vicinity of that point (that is, in the finite volume element ΔV centered at $\boldsymbol{X}$) is nonzero. The region ΔV can be visualized as a union of an infinite number of one-element sets of the type $\boldsymbol{X}$. A direct application of axiom 3 to this case would produce an uncertainty of the type $0 \cdot \infty$. Therefore the axiom 3 is unsuitable for infinitive sequences of sets, and the probability for the event to belong to the set ΔV cannot be obtained as the sum of such probabilities for the sets $\boldsymbol{X} \subset \Delta V$.

Axiom 3 is applicable only to incompatible events, that is, mutually exclusive events that belong to non-overlapping sets. Consider now the case of intersecting sets and overlapping events, that is, events belonging to two or more sets at the same time. Such events are called joint. Consider two sets A and B, whose intersection $A \cap B$ is not empty. We say that ω belongs to the intersection ($\omega \in A \cap B$) if $\omega \in A$ and $\omega \in B$. Then the probability of the joint event ω can be written as

$$P(A \cap B) = P(\omega \in A \quad \text{and} \quad \omega \in B). \tag{1.1}$$

As examples of joint events, consider two situations, which will prove to be of interest further on:

1. At a given time, the volume element ΔV centered at the space point $\boldsymbol{X}$ contains N_1 particles of type 1 (first event) and N_2 particles of type 2 (second event). The probability of this happening is given by the joint probability of both events.
2. A volume element ΔV centered at a space point $\boldsymbol{X}$ contains N_1 particles of type 1 and N_2 particles of type 2 at the time t_1 (first event) and n_1 particles of type 1 and n_2 particles type 2 at the time t_2 (second event). The probability of the joint event is the joint probability of both events at times t_1 and t_2.

Sometimes one is interested in the probability of an event given the occurrence of some other event. For example, we may want to know the probability of finding a particle in a volume element ΔV centered at the point $\boldsymbol{X}$ at the time t given that at the time $t_0 < t$, it was located in a volume element ΔV_0 centered at the point $\boldsymbol{X}_0 \neq \boldsymbol{X}$. Actually, we consider the set of all events C, where C denotes an event of finding the particle in the volume element ΔV at the time t. The particle could get into this element from any initial spatial position (with different probabilities), but we are interested only in some of those positions, that is, in a subset B of the set A. The probability of such an event is called a conditional probability. Conditional probability is defined as the probability of realization of an event $\omega \in A$ under the condition that $\omega \in B$ and is equal to

$$P(A|B) = P(A \cap B)/P(B). \qquad (1.2)$$

The theory of stochastic processes is based (to a considerable degree) on the notion of joint probability. In this context, let us mention an important property of the joint probability. Suppose the full set Ω is divided into non-overlapping subsets B_i, that is,

$$B_i \cap B_j = \phi \quad \text{and} \quad \bigcup_i B_i = \Omega.$$

As far as

$$\bigcup_i (A \cap B_i) = A\mathbf{I}(\bigcup_i B_i) = A \cap W = A$$

and (see axiom 3)

$$\sum_i P(A \cap B_i) = P(\bigcup_i (A \bigcup_i B_i)) = P(A)$$

we find from (1.2):

$$\sum_i P(A|B_i)P(B_i) = P(A). \qquad (1.3)$$

Thinking of the subset B_i as a variable, one can see from the last relation that summation over all mutually exclusive possibilities (i.e. over all sets B_i) eliminates this variable from the outcome.

Yet another important notion is the notion of independent events. Two sets of events A and B are called independent if the probability for an event to belong to set A and the probability to belong to set B are not correlated. Then

$$P(A \cap B) = P(A)P(B). \tag{1.4}$$

1.2
Random Variables, Probability Distribution Function, Average Value, and Variance

The concept of a random variable is of primary importance in stochastic processes. A random variable $F(X)$ is defined as a function of the element X of the space of probabilistic events $\boldsymbol{X}$. An event is specified by X, so X now stands for the event previously denoted by ω. The examples of random variables include position, momentum, and spatial orientation of a particle driven by random external forces (Brownian motion, motion in a turbulent flow). The introduction of a random variable notation simplifies operations with functions of random variables, calculations of random variable distributions, of averages and other statistical characteristics of distributions. Furthermore, the introduction of continuous random variables enables us to operate with stochastic differential equations and study the change of random variables in space and in time in the same way as we study deterministic systems by using differential equations.

The frequency (or probability) of realization of a definite event is equal to some value between zero and one. If the events are mutually exclusive, the sum of probabilities must be equal to one. This means that one of the events will realize.

Statistical mechanics is usually concerned with continuous random variables, that is, variables that can assume a continuous range of values. As far as the probability to get any given value from a continuum of possible values is zero, and the sum of all probabilities is one, it is necessary to look at the probability of realization of an event that is associated with an infinitesimal interval (set) of values rather than a single value. This probability is also an infinitely small quantity having the same order as the length of the interval (measure of event) and so is proportional to the measure of events, that is, to $d\boldsymbol{X}$. Thus the probability that a random variable is contained in the interval $(\boldsymbol{X}, \boldsymbol{X}+d\boldsymbol{X})$ can be represented as

$$P(X \in (\boldsymbol{X}, \boldsymbol{X} + d\boldsymbol{X})) = p(X)d\boldsymbol{X}. \tag{1.5}$$

The function $p(\boldsymbol{X})$ is called the probability density function (PDF) or simply the probability density. The condition that the sum of probabilities for a continuous random variable is equal to one can be written in the integral form:

$$\int_X p(\boldsymbol{X})d\boldsymbol{X} = 1, \tag{1.6}$$

where X is the domain of the n-dimensional space in which $\boldsymbol{X}$ varies. The relation (1.6) can be interpreted as the normalization condition for the PDF.

The introduction of a PDF enables us to find statistical characteristics of the distribution of a random variable $\boldsymbol{X}$. The most important of them is the average value (aka mean value, or expectation value) of a random variable or random function:

$$\langle f \rangle = \int_X f(\boldsymbol{X})\, p(\mathbf{X}) d\boldsymbol{X}. \tag{1.7}$$

If $\boldsymbol{X}$ is a vector in an n-dimensional coordinate space, then (1.6) and (1.7) can be written in the coordinate form:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(X_1, X_2, \ldots, X_n)\, p(X_1, X_2, \ldots, X_n) dX_1, dX_2, \ldots, dX_n = 1.$$

$$\langle f \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(X_1, X_2, \ldots, X_n)\, p(X_1, X_2, \ldots, X_n) dX_1, dX_2, \ldots, dX_n.$$

Another statistical characteristic is the variance σ^2. For a one-dimensional space, the variance is defined as

$$\sigma^2 = \int_X (f - \langle f \rangle)^2\, p(\mathbf{X}) d\mathbf{X}. \tag{1.8}$$

The square root σ of the variance is called the standard deviation. Sometimes the PDF has the form of a function with a sharp peak at the point $\boldsymbol{X}=\boldsymbol{X}_0$. In limiting case it is infinite at $\boldsymbol{X}=\boldsymbol{X}_0$ and zero at $\boldsymbol{X}\neq\boldsymbol{X}_0$. Such a case arises when we idealize a process. For example, we can choose to regard a mass that is continuously distributed in a small volume element centered at $\boldsymbol{X}_0$ as localized at one space point $\boldsymbol{X}_0$ (i.e. as "point mass"). Than the density of the substance differs from zero only at this point and the integral (1.7) has the meaning of the total mass. A similar reasoning leads to the concept of a point force – the net force with which we replace a force that is continuously distributed over a small volume element. To ensure the existence of integrals of such functions, we have to extend the notion of a function, what is achieved by the introduction of generalized functions.

1.3 Generalized Functions

The simplest and most extensively used generalized function is Diracs delta function $\delta(\boldsymbol{X}-\boldsymbol{X}_0)$, which can be defined as the limit of following sequence (sometimes referred to as "delta sequence"):

$$\delta(\boldsymbol{X}-\boldsymbol{X}_0) = \lim_{m\to\infty}\left(\frac{m}{\sqrt{\pi}}\right)^n \exp(-m^2(\boldsymbol{X}-\boldsymbol{X}_0)^2). \tag{1.9}$$

Here n is the number of dimensions and, accordingly, $\boldsymbol{X}$ is an n-dimensional vector with components $X_1, X_2, \ldots, X_n$. Eq. (1.9) can also be written for one-dimensional sequences of $X_i - X_i^0$. Then the following identity will hold:

$$\delta(X_1-X_1^0)\delta(X_2-X_2^0)\ldots\delta(X_n-X_n^0) = \delta(\boldsymbol{X}-\boldsymbol{X}_0). \tag{1.10}$$

The limit on the right-hand side of Eq. (1.9) is 0 at $\boldsymbol{X} \neq \boldsymbol{X}_0$ and $+\infty$ at $\boldsymbol{X} = \boldsymbol{X}_0$. Therefore Diracs delta function is not a function in the usual sense and should not be interpreted as giving the value of the dependent variable at each point. What is important, however, is that this function is still integrable, and behaves similarly to ordinary functions in its capacity as an integrand. In particular, the integral of the scalar product of Diracs delta function $\delta(\boldsymbol{X}-\boldsymbol{X}_0)$ and an ordinary function $\varphi(\boldsymbol{X})$ equals

$$\begin{aligned}(\delta(\boldsymbol{X}-\boldsymbol{X}_0), \varphi(\boldsymbol{X})) &= \int_X \delta(\boldsymbol{X}-\boldsymbol{X}_0)\varphi(\boldsymbol{X})d\boldsymbol{X} \\ &= \lim_{m\to\infty}\int_X \left(\frac{m}{\sqrt{\pi}}\right)^n \exp(-m^2(\boldsymbol{X}-\boldsymbol{X}_0)^2)\varphi(\boldsymbol{X})d\boldsymbol{X} = \varphi(\boldsymbol{X}_0)\end{aligned}$$

provided the domain contains the point X_0.

Thus, by its definition, the Delta function has two basic properties:

$$\delta(\boldsymbol{X}) = \begin{cases} 0, & \text{for} \quad \boldsymbol{X} \neq 0, \\ 1, & \text{for} \quad \boldsymbol{X} = 0, \end{cases} \tag{1.11a}$$

$$\int_X \delta(\boldsymbol{X}-\boldsymbol{X}_0)\varphi(\boldsymbol{X})d\boldsymbol{X} = \varphi(\boldsymbol{X}_0). \tag{1.11b}$$

In the particular cases $\varphi(\boldsymbol{X}) = 1$ and $\varphi(\boldsymbol{X}) = \boldsymbol{X}$ one gets:

$$\int_X \delta(\boldsymbol{X}-\boldsymbol{X}_0)d\boldsymbol{X} = 1. \tag{1.12}$$

and

$$\int_X \delta(\boldsymbol{X}-\boldsymbol{X}_0)\boldsymbol{X}d\boldsymbol{X} = \boldsymbol{X}_0. \tag{1.13}$$

Hence, according to Eq. (1.7), $\delta(\boldsymbol{X}-\boldsymbol{X}_0)$ can be taken as a PDF such that the random variable $\boldsymbol{X}$ has the average value $\boldsymbol{X}_0$. For the one-dimensional case, the following equality can be written:

$$(X-X_0)\delta(X-X_0) = 0$$

or

$$X\delta(X-X_0) = X_0\delta(X-X_0). \tag{1.14}$$

Taking $\varphi(\boldsymbol{X}) = (\boldsymbol{X} - \boldsymbol{X}_0)^2$, we can write

$$\int_X \delta(\boldsymbol{X}-\boldsymbol{X}_0)(\boldsymbol{X}-\boldsymbol{X}_0)^2 d\boldsymbol{X} = (\boldsymbol{X}_0-\boldsymbol{X}_0)^2 = 0. \tag{1.15}$$

The left-hand side of (1.23) coincides with the definition of the variance for the PDF $\delta(\boldsymbol{X} - \boldsymbol{X}_0)$. Thus, its variance is zero, and the delta function describes the case when one knows for sure that $\langle \boldsymbol{X} \rangle = \boldsymbol{X}_0$.

The (one-dimensional) Cauchy sequence is not the only sequence converging to the delta function. For example, the sequence

$$\delta(X-X_0) = \lim_{\varepsilon \to 0} \frac{\varepsilon}{\pi(X-X_0)^2 + \varepsilon^2} \tag{1.16}$$

can be used as an alternative representation of the delta function.

The delta function is an infinitely differentiable function. Its derivative can be defined by differentiating the integral

$$\int \varphi(X) \frac{d}{dX} \delta(X-X_0) dX$$

by parts and using the property (1.11):

$$\int_X \delta'(X-X_0)\varphi(X) dX = -\int_X \delta(X-X_0)\varphi'(X) dX = -\varphi'(X_0), \tag{1.17}$$

from which there follows a useful symbolic equality

$$\delta'(X) = -\frac{\delta(X)}{X}, \quad (X \neq 0),$$

In the more general case, one can write

$$\delta^{(r)}(X) = (-1)^r \frac{\delta(X)}{X^r}, \quad (X \neq 0, \quad r = 0, 1, \ldots). \tag{1.18}$$

If a function $Y = f(X)$ is single-valued, that is, if it can be solved with respect to X in a unique way, then $X = f^{-1}(Y)$ and

$$\delta(Y - f(X)) = \frac{\delta(X - f^{-1}(Y))}{|df/dX|}. \tag{1.19}$$

A similar relation takes place for a vector function $\boldsymbol{Y}=\boldsymbol{f}(\boldsymbol{X})$:

$$\delta(\boldsymbol{Y}-\boldsymbol{f}(\boldsymbol{X})) = \frac{\delta(\boldsymbol{X}-\boldsymbol{f}^{-1}(\boldsymbol{Y}))}{\Delta}. \tag{1.19$'$}$$

where Δ is the determinant of the Jacobian $||\partial f_i/\partial X_j||$.
The delta function can be connected with the unit step function (Heaviside function) defined as

$$H(X) = \begin{cases} 0, & \text{for} \quad X<0, \\ 1, & \text{for} \quad X>0 \end{cases}$$

through the symbolic relation

$$\delta(X) = \frac{d\mathrm{H}}{dX}. \tag{1.20}$$

If there is more then one independent variable, one has to use partial derivatives of the delta function. For example, if we take the delta function as a generalized vector function $d(\boldsymbol{X} - \boldsymbol{X}_0)$, its gradient is defined as

$$\nabla\delta = \frac{\partial\delta(\boldsymbol{X}-\boldsymbol{X}_0)}{\partial\boldsymbol{X}} = \left(\frac{\partial\delta}{\partial X_1}, \frac{\partial\delta}{\partial X_2}, \ldots, \frac{\partial\delta}{\partial X_n}\right). \tag{1.21}$$

1.4
Methods of Averaging

When looking at the hydrodynamic characteristics of a turbulent flow or at the motion of particles under the action of random external forces, we notice one distinguishing feature shared by these two types of motion: the presence of random fluctuations. Because of fluctuations, the dependences of hydrodynamic field parameters on spacetime coordinates, and the configuration of particles in space at different moments look irregular and have a confusing pattern. If a process is repeated multiple times under the identical set of initial and boundary conditions, the observed values of field parameters and particle positions will be different. This necessitates the use of averaging methods in any study of random motions. Averaging allows us to make a transition from irregular characteristics to much more smooth and regular mean values. In practice the mean value is determined by averaging over the time interval,

$$\langle f(\mathbf{X},t)\rangle_T = \frac{1}{T}\int_{-T/2}^{T/2} \omega(\mathbf{X},\tau)\, f(\mathbf{X},t+\tau)\,d\tau, \tag{1.22}$$

or by averaging over the considered spatial region,

$$\langle f(\mathbf{X},t)\rangle_V = \frac{1}{V}\int_V \omega(\mathbf{x},t)\, f(\mathbf{X}+\mathbf{x},t)\,d\mathbf{x} \tag{1.23}$$

or, most generally, by spacetime averaging,

$$\langle f(\mathbf{X},t)\rangle_{TV} = \frac{1}{VT}\int_V \int_{-T/2}^{T/2} \omega(\mathbf{x},\tau)\, f(\mathbf{X}+\mathbf{x},t+\tau)\,d\mathbf{x}\,d\tau, \tag{1.24}$$

where $\omega(\mathbf{x}, t)$ is the weight function.

We can also introduce the autocorrelation function $\Psi(\tau)$, which is defined as follows: take a random function $f(t)$ at one and the same point of space but at instances of time t and $t+\tau$, form the product, and find its average value over the time interval $(0, T)$ for $T \to \infty$:

$$\Psi(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(t)\, f(t+\tau)\,dt. \tag{1.25}$$

This function plays an important role in many applications.

The three types of averaging mentioned above have one drawback, namely, they apply to only one instance of the process under consideration (turbulent velocity field, etc.). Another shortcoming is that one is faced with the problem of choosing the most convenient weight function.

If the process is repeated multiple times under the same initial conditions, we are dealing with many instances of the same process. In this case one can talk about a statistical set of identical processes (flows, particle motions etc.) taking place under fixed initial and external conditions. Let one and the same experiment be replicated N times under the same conditions, yielding different values u_i of one and the same parameter, for example, velocity u. By averaging the velocities u_i observed in a discrete set of similar tests, we obtain the mean value

$$\langle u(\mathbf{X},t)\rangle = \frac{1}{N}\sum_{i=1}^{N} u_i,$$

which is called the ensemble average. In many cases the ensemble average proves to be stable enough, in other words, the outcomes of a sufficiently large set of experiments show a very small variance.

Let a continuous random variable u $(-\infty < u < \infty)$ be characterized by the PDF $p(u)$. If we are interested in the value of u at one and the same space point M, then $p(u)du$ signifies the probability for u to be found in interval $(u,\ u+du)$. Then the ensemble average of u is

$$\langle u(\boldsymbol{X},t)\rangle = \int_{-\infty}^{\infty} u\, p(u)\, du. \tag{1.26}$$

By analogy, the ensemble average of any function F is equal to

$$\langle F(u(\boldsymbol{X},t))\rangle = \int_{-\infty}^{\infty} F(u)\, p(u) du. \tag{1.27}$$

Now, let u be measured at different spacetime points $M_1 = (X_1,\ t_1)$, $M_2 = (X_2,\ t_2)$, ..., $M_N = (X_N,\ t_N)$. The resulting values of u are denoted by u_1, u_2, ..., u_N. We then introduce the N-dimensional PDF $p(u_1, u_2, \ldots, u_N)$, where $p(u_1, u_2, \ldots, u_N) du_1 du_2 \ldots du_N$ means the probability of finding u_i in the interval $(u_i,\ u_i + du_i)$. We have used a common convention where the index stands for all N variations, that is $(u_i, u_i + du_i) = (u_1 + du_1, \ldots, u_n + du_n)$. The average of any function will be written as

$$\langle F\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} F(u_1, u_2, \ldots, u_N)\, p(u_1, u_2, \ldots, u_N) du_1 du_2 \ldots du_N. \tag{1.28}$$

By introducing an N-dimensional vector $\boldsymbol{u}(u_1,\ u_2,\ \ldots,\ u_N)$, we can rewrite the relation (1.28) in a more compact form:

$$\langle F\rangle = \int_{-\infty}^{\infty} F(\boldsymbol{u})\, p(\boldsymbol{u}) d\boldsymbol{u}. \tag{1.29}$$

Multidimensional PDFs are especially important for studying the behavior of an N-particle system in a random field of external forces. If u_i denotes the coordinate of the i-th particle, then the above-introduced PDF is called a multiparticle PDF. One-particle and two-particle PDFs are of particular interest in applications. Sometimes the two-particle PDF is also called "pair PDF" or "pair distribution". These PDFs can be derived from multidimensional PDFs by integrating them over all possible positions of the remaining particles. For instance, a single-particle PDF is obtained as

$$p(X_1) = \int_{-\infty}^{\infty}\int_{-8}^{\infty}\ldots\int_{-\infty}^{\infty} p(X_1, X_2, \ldots, X_N) dX_2 \ldots dX_N \tag{1.30}$$

Such PDFs are also called marginal PDFs.

If we consider spherical particles of different radii a_i, than the PDF $p(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N, a_1, \ldots, a_N)$ will be associated with the radius distribution in addition to the coordinate distribution, and $p(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N, a_1, \ldots, a_N)\, d\boldsymbol{X}_1 \ldots d\boldsymbol{X}_N\, da_1 \ldots da_N$ will have the meaning of probability to find the N-particle system in the volume element $(d\boldsymbol{X}_1 \ldots d\boldsymbol{X}_N)$ with particle radii lying in the interval $(a_1 + da_1), \ldots, (a_N + da_N)$. The corresponding single-particle PDF is

$$p(\boldsymbol{X}_1, a_1) = \int_{-\infty}^{\infty}\int_{0}^{\infty} p(\mathbf{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N, a_2, \ldots, a_N) d\boldsymbol{X}_2 \ldots d\boldsymbol{X}_N da_2 \ldots da_N. \quad (1.31)$$

If the radius a must be the same for all particles, then it is convenient to operate with particle distribution over the radius

$$n(\boldsymbol{X}, a) = N\, p(\boldsymbol{X}, a), \quad (1.32)$$

such that $n(\boldsymbol{X}, a)da$ is the probabilistic numerical concentration (aka number concentration) of particles with radius in the interval $(a + da)$ in the volume element $d\boldsymbol{X}$.

The multidimensional PDF should satisfy the following properties:

1. $p(\boldsymbol{u}) \geq 0$;
2. $\int_{-\infty}^{\infty} p(\boldsymbol{u}) d\boldsymbol{u} = 1$;
3. $p(u_1, u_2, \ldots, u_N) = p(u_{i_1}, u_{i_2}, \ldots, u_{i_N})$, where the set $\{i_1, i_2, \ldots, i_N\}$ is formed from the set $\{1, 2, \ldots, N\}$ by changing the order.
4. $p(u_1, u_2, \ldots, u_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(u_1, u_2, \ldots, u_n, u_{n+1}, \ldots, u_N) du_{n+1} \ldots du_N$ for $n < N$;
5. For independent random variables $u_1, u_2, \ldots, u_N$, there holds:

$$p(u_1, u_2, \ldots, u_N) = p(u_1)\, p(u_2), \ldots, p(u_N) \quad (1.33)$$

Property 3 is known as the symmetry property and property 4 – as the consistency property.

It is now time to discuss the connection between different types of averaging. In practice, we use time or space averaging rather than ensemble averaging, because the latter requires a large number of experiments. In Statistical Mechanics, ensemble averaging, that is, averaging over the set of all possible states, is often replaced by time or space averaging, with the implicit assumption that by increasing the averaging interval we can always make the average values converge to the corresponding ensemble averages. This assumption is called the ergodic hypothesis, or, in those special cases when it can be rigorously proved, the ergodic theorem.

When studying such problems as the flow of a disperse medium or the filtration of a fluid through a porous medium, one often uses the so-called Saffman step function:

$$H(\boldsymbol{X}) = \begin{cases} 0, & \text{for} \quad \boldsymbol{X} \quad \text{in rigid body}, \\ 1, & \text{for} \quad \boldsymbol{X} \quad \text{in fluid}. \end{cases} \tag{1.34}$$

This function depends on statistical parameters of the distribution of moving particles of the disperse phase or fixed particles of the porous medium. After averaging over the particle ensemble, we get

$$\langle H(\boldsymbol{X})\rangle = 1-\varphi, \tag{1.35}$$

where φ is volume concentration of particles.

One can use the Saffman function to perform space averaging of hydrodynamic parameters. For example, the velocity of the fluid $\boldsymbol{u}$ will be averaged as

$$\bar{\boldsymbol{u}} = \langle H\boldsymbol{u}\rangle/\langle H\rangle = \langle \boldsymbol{u}\rangle/(1-\varphi), \tag{1.36}$$

where $\bar{\boldsymbol{u}}$ is the mean-flow-rate velocity through the microcapillaries of the porous medium. It should not be confused with the ensemble average $\langle \boldsymbol{u}\rangle$, although for a highly permeable medium ($\varphi \ll 1$), the two velocities are equal: $\langle u\rangle = \bar{u}$.

1.5 Characteristic Functions

Instead of using the PDF $p(u_1, u_2, \ldots, u_N)$, it is often convenient to use its Fourier transform:

$$\varphi(\rho_1, \rho_2, \ldots, \rho_N) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp\left\{ i\sum_{k-1}^{N} \rho_k u_k \right\} p(u_1, u_2, \ldots, u_N)\, du_1 du_2 \ldots du_N$$

or, in the vector form,

$$\varphi(\rho) = \int_{-\infty}^{\infty} e^{i\rho\cdot\boldsymbol{u}}\, p(\boldsymbol{u})\, d\boldsymbol{u}. \tag{1.37}$$

Here ρ is an N-dimensional vector with components $(\rho_1, \rho_2, \ldots, \rho_N)$. The function $\varphi(\rho)$ is called the characteristic function or the moment-generating function. Because of Eq. (1.29), it can be represented as

$$\varphi(\rho) = \left\langle e^{i\rho\cdot\boldsymbol{u}} \right\rangle. \tag{1.38}$$

If the characteristic function is known, then the PDF is obtained as the inverse Fourier transform:

$$p(\boldsymbol{u}) = \frac{1}{(2\pi)^N} \int_{-\infty}^{\infty} e^{-i\rho\cdot\boldsymbol{u}} \varphi(\boldsymbol{\rho}) d\boldsymbol{\rho}. \tag{1.39}$$

So, the knowledge of the characteristic function is tantamount to the knowledge of the PDF. Hence the properties of the PDF are readily obtained from those of the characteristic function. The normalization condition for the PDF means that

$$\omega(\mathbf{0}) = \mathbf{1}. \tag{1.40}$$

For independent random variables we have, according to (1.33):

$$\varphi(\boldsymbol{\rho}) = \varphi(\rho_1)\varphi(\rho_2)\ldots\varphi(\rho_N). \tag{1.41}$$

The symmetry and consistency properties of characteristic function follow from properties 3 and 4 of the PDF:

$$\varphi(\rho_1, \rho_2, \ldots, \rho_N) = \varphi(\rho_{i_1}, \rho_{i_2}, \ldots, \rho_{i_N}), \tag{1.42}$$

$$\varphi(\rho_1, \rho_2, \ldots, \rho_n) = \varphi(\rho_1, \rho_2, \ldots, \rho_n, 0, 0, \ldots, 0), \tag{1.43}$$

where $i_1, i_2, \ldots, i_N$ is any combination of non-repeating numbers 1, 2, ..., N. In the last relation, $n < N$ and the number of zeros is equal to $N - n$. The property (1.57) allows us to obtain the characteristic function for a smaller number of dimensions (smaller number of particles) from the N-dimensional (N-particle) characteristic function, and then to get the corresponding marginal PDF by using the inverse Fourier transform (1.39). Therefore one can specify all PDFs, describing random variables at all possible points, through a single characteristic function known as the characteristic functional. In particular, for one-dimensional random function $u(X)$ defined on a finite interval $a \le X \le b$, the characteristic functional is

$$\Phi(\rho(X)) = \left\langle \exp\left\{ i\int_a^b \rho(X)u(X)dX \right\} \right\rangle, \tag{1.44}$$

where $\rho(X)$ is a function selected in such a way that the integral in the exponent converges. The left-hand side is a function of a function, which is why it is called a functional.

If

$$\rho(X) = \sum_{i=1}^{N} \rho_i \delta(X - X_i),$$

then Eq. (1.44) gives us:

$$\Phi(\rho(X)) = \left\langle \exp\left\{ i\sum_{k=1}^{N}\rho_k X_k \right\} \right\rangle = \varphi(\rho_1, \rho_2, \ldots, \rho_N), \tag{1.45}$$

Thus the characteristic functional turns into a characteristic function of the multidimensional PDF for $u(X_1)$, $u(X_2)$, ..., $u(X_N)$. Additional information about the characteristic functional can be found in Section 1.15.

1.6
Moments and Cumulants of Random Variables

To solve a specific problem in a rigourous way, one has to specify a multidimensional (multiparticle) PDF. However, one run into difficulties with this approach because the PDF cannot be determined with a sufficient accuracy. Furthermore, it is inconvenient to use because it results in cumbersome expressions. In practice, when solving applied problems, one usually considers only the more simple parameters that characterize specific statistical properties of the process. The most important of these parameters are moments.

Let us consider a set of N random variables with N-dimensional PDF $p(u_1, u_2, \ldots, u_N)$. The moments are defined as follows:

$$\begin{aligned} B_{k_1 k_2 \ldots k_N} &= \left\langle u_1^{k_1} u_2^{k_2} \ldots u_N^{k_N} \right\rangle \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} u_1^{k_1} u_2^{k_2} \ldots u_N^{k_N}\, p(u_1, u_2, \ldots, u_N)\, du_1 du_2 \ldots du_N, \end{aligned} \tag{1.46}$$

where k_1, k_2, ..., k_N are non-negative integers, whose sum $K = k_1 + k_2 + \ldots + k_N$ is called the moments order. It is evident that moments of the first order are simply the mean values of random variables u_1, u_2, ..., u_N.

In addition to ordinary moments (1.61) one often uses some combinations of moments. In particular, central moments are defined as moments of fluctuations (deviations of random variables u_1, u_2, ..., u_N from their mean values):

$$b_{k_1 k_2 \ldots k_N} = \left\langle (u_1 - \bar{u}_1)^{k_1} (u_2 - \bar{u}_2)^{k_2} \ldots (u_N - \bar{u}_N)^{k_N} \right\rangle. \tag{1.47}$$

For $N = 1$ and $k_1 = 2$, we get the second-order central moment $b_2 = \sigma^2$.

If u_i are the velocities of a turbulent flow at spatial points x_i, then the differences $u_i - \langle u_i \rangle$ have the meaning of velocity fluctuations at these points. Thus central moments characterize statistical properties of random variables – velocity fluctuations. In the case when u_i are the random positions of particles driven by a random external force, central moments characterize the statistical properties of the disperse phase in the suspension.

By removing brackets in Eq. (1.47) and making use of Eq. (1.46), we can obtain connections between central and ordinary moments. The case of $N = 1$ yields

$$b_1 = 0; \quad b_2 = B_2 - B_1^2; \quad b_3 = B_3 - 3B_1B_2 + 2B_1^3; \\ b_4 = B_4 - 4B_1B_3 + 6B_1^2B_2 - 3B_2^4 \quad \text{etc.} \tag{1.48}$$

When $\bar{u}_i = 0$, the central and ordinary moments coincide. The two combinations of central moments,

$$\sqrt{\beta_1} = \frac{b_3}{b_2^{3/2}}, \quad \beta_2 = \frac{b_4}{b_2^2} - 3 \tag{1.49}$$

serve as statistical characteristics of a random quantity and are called, repectively, the asymmetry and the excess.

The moments of random variables $u_1, u_2, \ldots, u_N$ can be expressed through a corresponding characteristic function $\varphi(\rho_1, \rho_2, \ldots, \rho_N)$ by comparing the relations (1.46) and (1.37):

$$B_{k_1k_2\ldots k_N} = (-i)^K \frac{\partial^K \varphi(\rho_1, \rho_2, \ldots, \rho_N)}{\partial\rho_1^{k_1}\partial\rho_2^{k_2}\ldots\partial\rho_N^{k_N}}\bigg|_{\rho_1=\rho_2=\ldots=\rho_N=0}, \tag{1.50}$$

from which one can see that moments can also be thought of as coefficients in the Taylor expansion of the characteristic function:

$$\varphi(\rho_1, \rho_2, \ldots, \rho_N) = \sum_{k_1,k_2,\ldots,k_N} i^K \frac{B_{k_1k_2\ldots k_N}}{k_1!k_2!\ldots k_N!}\rho_1^{k_1}\rho_2^{k_2}\ldots\rho_N^{k_N}. \tag{1.51}$$

Thus, if the moments are known, Eq. (1.51) gives us the characteristic function, and then the PDF follows from Eq. (1.39). It means that the PDF is uniquely defined by the moments of the distribution.

The other category of combinations of central moments are the so-called cumulants (aka semi-invariants) $S_{k_1k_2k_N}$. Let us introduce the logarithm of the characteristic function

$$\psi(\rho_1, \rho_2, \ldots, \rho_N) = \ln\varphi(\rho_1, \rho_2, \ldots, \rho_N), \tag{1.52}$$

which is called the generating function of cumulants. Cumulants can then be defined as the coefficients in the Taylor expansion of ψ in just as the moments were defined as the coefficients in the expansion (1.66):

$$\psi(\rho_1, \rho_2, \ldots, \rho_N) = \sum_{k_1,k_2,\ldots,k_N} i^K \frac{S_{k_1k_2\ldots k_N}}{k_1!k_2!\ldots k_N!}\rho_1^{k_1}\rho_2^{k_2}\ldots\rho_N^{k_N}, \tag{1.53}$$

$$S_{k_1k_2\ldots k_N} = (-i)^K \frac{\partial^K \varphi(\rho_1, \rho_2, \ldots, \rho_N)}{\partial\rho_1^{k_1}\partial\rho_2^{k_2}\ldots\partial\rho_N^{k_N}}\bigg|_{\rho_1=\rho_2=\ldots=\rho_N=0}, \tag{1.54}$$

Recalling that $\varphi(0, \ldots, 0) = 1$ and taking $N = 1$, we can express cumulants in terms of ordinary and central moments:

$$\begin{aligned} &S_1 = B_1; \quad S_2 = B_2 - B_1^2 = b_2; \quad S_3 = B_3 - 3B_1B_2 + 2B_1^3 = b_3; \\ &S_4 = B_4 - 4B_1B_3 - 3B_2^2 + 12B_1^2B_2 - 6B_1^4 = b_4 - 3b_2^2; \\ &S_5 = b_5 - 10b_2b_3 \quad \text{etc.} \end{aligned} \tag{1.55}$$

By the same token, moments could be expressed in terms of cumulants:

$$B_1 = S_1; \quad B_2 = S_2 + S_1^2; \quad B_3 = S_3 + 3S_2S_1 + S_1^2; \quad \text{etc.} \tag{1.56}$$

In the case of one-dimensional PDF $p(u)$ we have the following expressions for moments and cumulants:

$$B_n = \int_{-\infty}^{\infty} p(u)u^n du = \left(\frac{1}{i}\frac{d}{d\rho}\right)^n \varphi(\rho)\big|_{\rho=0}, \tag{1.57}$$

$$S_n = \left(\frac{1}{i}\frac{d}{d\rho}\right)^n \psi(\rho)\big|_{\rho=0}. \tag{1.58}$$

There is a recurrent relation between moments and cumulants. In one-dimensional case it has the following form:

$$B_0 = 1; \quad B_n = \sum_{k=0}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} S_k B_{n-k}; \quad (n = 1, 2, \ldots). \tag{1.59}$$

From this relation, one can derive Eq. (1.56).

1.7
Correlation Functions

In statistical mechanics of disperse media as well as in the turbulence theory, one often encounters random fields described by a random function $u(M)$ of a spacetime point M. Following the definition (1.46), let us call expressions

$$B_{uu\ldots u}(M_1, M_2, \ldots, M_k) = \langle u(M_1)u(M_2)\ldots u(M_k)\rangle \tag{1.60}$$

the k-th order moments of such a field. Generally speaking, some points may coincide. The number of different points is called the type of the moment. In this context, one can talk about single-point moments, two-point moments, and so on. The average values of products of several correleted random functions of different fields are called mixed moments.

Consider, for example, the field of velocities in a turbulent flow given by the velocity vector $\boldsymbol{u}(u_1, u_2, u_3)$. Components of this vector can be regarded as different mutually correlated random functions. Since the values of velocity components could be taken at the same point or at different points, there are moments of various types and orders. Of primary importance in statistical mechanics are two-point second-order moments known as correlation functions (or simple correlations):

$$B_{ij}(M_1, M_2) = \langle u_i(M_1)u_j(M_2)\rangle. \tag{1.61}$$

B_{ij} are the components of a second rank tensor $\boldsymbol{B}$. It is obvious that the relation (1.61) could be written in a matrix form:

$$\boldsymbol{B}(M_1, M_2) = \langle \boldsymbol{u}(M_1)\boldsymbol{u}^T(M_2)\rangle,$$

where the superscript T stands for transpose.

When $M_1, M_2, \ldots, M_K$ are points in spacetime, the corresponding moments and correlations are called spacetime moments/correlations. In statistical mechanics one usually has to deal with correlations of random functions at different points at one and the same instant of time or with correlations at one and the same point but at different instants of time. The former correlations are called spatial correlations and the latter – time correlations.

Let us metion some important properties of correlation functions. A correlation function $B_{uu}(M_1, M_2) = \langle u(M_1)u(M_2)\rangle$ is symmetric with respect to the pair of points M_1, M_2:

$$B_{uu}(M_1, M_2) = B_{uu}(M_2, M_1). \tag{1.62}$$

A quadratic form with coefficients B_{uu} (M_i, M_j) is always non-negative, that is,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} B_{uu}(M_i, M_j)X_iX_j \geqslant 0, \tag{1.63}$$

for all real X_i, non-negative integer n and any selection of points $M_1, M_2, \ldots, M_n$. In particular, at $n = 2$ the expression (1.63) becomes

$$|B_{uu}(M_1, M_2)| \leq |B_{uu}(M_1, M_1)|^{1/2}|B_{uu}(M_2, M_2)|^{1/2}. \tag{1.64}$$

In addition to the above-mentioned two-point moments of one random function at different points, $u(M_1)$ and $u(M_2)$, one can consider two-point moments of different random functions, $u(M_1)$ and $v(M_2)$. A mixed two-point moment B_{uv} (M_1, M_2) is called the mutual correlation function. Its properties are similar to those of "ordinary" moments. For example, the symmetry property still holds:

$$B_{uv}(M_1, M_2) = B_{vu}(M_2, M_1). \tag{1.65}$$

Two-point moments of orders higher than two are referred to as correlation functions of higher order.

By analogy, one can define two-point central moments of the second order:

$$\begin{aligned} b_{uu} &= \langle (u(M_1) - \langle u(M_1)\rangle)(u(M_2) - \langle u(M_2)\rangle)\rangle \\ &= B_{uu}(M_1, M_2) - \langle u(M_1)\rangle\langle u(M_2)\rangle, \end{aligned} \tag{1.66}$$

$$\begin{aligned} b_{uv} &= \langle (u(M_1) - \langle u(M_1)\rangle)(v(M_2) - \langle v(M_2)\rangle)\rangle \\ &= B_{uv}(M_1, M_2) - \langle u(M_1)\rangle\langle v(M_2)\rangle \end{aligned} \tag{1.67}$$

The variances of distributions of random variables u and v can be expressed through central moments:

$$\sigma_u^2(M) = b_{uu}(M, M), \quad \sigma_v^2(M) = b_{vv}(M, M). \tag{1.68}$$

Two-point central moments of the second order relate the deviations of random functions from their mean values (i.e. fluctuations) at two different points. This is why they are also called correlation functions of fluctuations.

Another important statistical parameter is the correlation coefficient, defined as

$$\Psi_{uu}(M) = \frac{b_{uu}(M_1, M_2)}{\sigma_u(M_1)\sigma_u(M_2)}, \quad \Psi_{uv}(M) = \frac{b_{uv}(M_1, M_2)}{\sigma_u(M_1)\sigma_v(M_2)}. \tag{1.69}$$

As a consequence of the Schwartz inequality, these coefficients satisfy $|\psi_{uu}| \leq 1$ and $|\psi_{uv}| \leq 1$. If the correlation coefficient vanishes, the correlation between fluctuations at different spatial points is absent.

An important property that follows from physical considerations is the damping of correlation between random variables at different spacetime points as we increase the distance between the points. As the distance goes to infinity, the correlation function will tend to zero. Of course, the ''distance'' that goes to infinity can be the geometrical distance, $|\boldsymbol{X}_2 - \boldsymbol{X}_1| \to \infty$ at $t_2 = t_1$, the temporal distance (time interval), $|t_2 - t_1| \to \infty$ at $\boldsymbol{X}_2 = \boldsymbol{X}_1$, or both.

1.8
Bernoulli, Poisson, and Gaussian Distributions

Let us consider the three distributions that are most frequently used in physical applications – the Bernoulli, Poisson, and Gaussian (normal) distributions.

Bernoulli distribution To begin, let us formulate the random walk problem, which will be considered in more detail farther on, in the chapter dedicated to Brownian motion. A particle undergoes a sequence of random displacements along a straight line. Each displacement is a step of the same length 1, and each step can be directed either forward or backward with the same probability of 0.5. The origin of the reference system is coincident with the initial position of the particle. Then the particles coordinate can assume only integer values $\ldots -N, -N+1, \ldots, 0, 1, \ldots, N-1, N \ldots$. The probability of finding the particle at a point m after N steps is given by the Bernoulli distribution,

$$P(m, N) = C^N_{(N+m)/2}\left(\frac{1}{2}\right)^N, \tag{1.70}$$

where $C^N_{(N+m)/2} = \frac{N!}{\left(\frac{1}{2}(N+m)\right)!\left(\frac{1}{2}(N-m)\right)!}$ are binomial coefficients.

The mean and root-mean-square displacements of the particle are, respectively,

$$\langle m\rangle = 0, \quad \sqrt{\langle m^2\rangle} = \sqrt{N}.$$

In the limiting case where $N \gg 1$ and $m \ll N$ this reduces to an asymptotic formula

$$P(m, N) \approx \left(\frac{2}{\pi N}\right)^{1/2} \exp\left(-\frac{m^2}{2N}\right). \tag{1.71}$$

Poisson distribution Let N particles be randomly distributed in a volume V. Then the probability of finding n particles in a volume element v, where v is a small part of V, is given by the Bernoulli distribution

$$P_N(n) = \frac{N!}{n!(N-n)!}\left(\frac{v}{V}\right)^n \left(1-\frac{v}{V}\right)^{N-n}. \tag{1.72}$$

For a given N, V and v, the mean value of n equals

$$\langle n\rangle = N\left(\frac{v}{V}\right) \equiv \nu.$$

In the limiting case where $N \to \infty$ and $V \to \infty$ but v remains finite, the distribution (1.72) tends asymptotically to the Poisson distribution

$$P(n) = \frac{\nu^n e^{-\nu}}{n!}. \tag{1.73}$$

If v is large and v is of the same order as $\boldsymbol{n}$, the Poisson distribution is close to the distribution

$$P(n) = \left(\frac{1}{2\pi\nu}\right)^{1/2} \exp\left(-\frac{(n-\nu)^2}{2\nu}\right). \tag{1.74}$$

Gaussian (normal) distribution The distributions (1.88) and (1.92) are both special cases of the Gaussian (aka normal) distribution. In the general N-dimensional case this distribution has the following normalized form:

$$p(u_1, u_2, \ldots, u_N) = C \exp\left\{-\frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} g_{jk}(u_j - a_j)(u_k - a_k)\right\}, \tag{1.75}$$

Here a_j are real numbers; g_{jk} are the elements of the positive definite matrix $||g_{jk}||$; $C = g^{1/2}/(2\pi)^{N/2}$ is a constant that is given by the normalization condition for probability density (see Eq. (1.6); $g = |g_{jk}|$ is the determinant of the matrix $||g_{jk}||$. The constants a_j and g_{jk} are related to the first and second moments of the distribution (1.75) (see Eq. (1.46) and Eq. (1.47):

$$\langle u_j \rangle = a_j; \quad b_{jk} = \langle (u_j - \langle u_j \rangle)(u_k - \langle u_k \rangle) \rangle = \frac{g_{jk}}{g} \tag{1.76}$$

Here $G_{jk} = \partial g/\partial g_{jk}$ is the algebraic complement of the element g_{jk} in the determinant g. It means that the matrices $||g_{jk}||$ and $||b_{jk}||$ are mutually inverse.

The Gaussian distribution can also represented in the matrix form:

$$p(\boldsymbol{u}) = \frac{1}{(2\pi)^{N/2} b^{1/2}}\left\{-\frac{1}{2}(\boldsymbol{u} - \langle \boldsymbol{u} \rangle)^T \boldsymbol{b}^{-1}(\boldsymbol{u} - \langle \boldsymbol{u} \rangle)\right\}, \tag{1.77}$$

where $\boldsymbol{b} = ||b_{jk}||$ and $b = |b_{jk}|$.

The ordinary second-order moments B_{jk} can be expressed in terms of the normal distribution parameters according to Eq. (1.76):

$$B_{jk} = \langle u_j u_k \rangle = \frac{G_{jk}}{G} + a_j a_k. \tag{1.78}$$

We see from Eq. (1.76) and Eq. (1.77) that the first two moments completely determine the PDF, and thereby the entire statistics of random variables, for a normal distribution. Hence the knowledge of mean values and correlation functions provides a complete statistical description of a random Gaussian field $\boldsymbol{u}(M) = [u_1(M), u_2(M), \ldots, u_N(M)]$. Central moments can be obtained from the property of normal distributions, which states that all central moments of an odd order are zero, whereas central moments of an even order are expressed through central moments of the second order:

$$\begin{aligned} b_{k_1 k_2 \ldots k_N} &= \left\langle (u_1 - \langle u_1 \rangle)^{k_1}(u_2 - \langle u_2 \rangle)^{k_2} \ldots (u_N - \langle u_N \rangle)^{k_N} \right\rangle \\ &= \sum b_{i_1 i_2} b_{i_3 i_4} \ldots b_{i_{2K-1}} b_{i_{2K}}, \end{aligned} \tag{1.79}$$

where $k_1 + k_2 + \ldots + k_N = 2K$ and subscript pairs are formed from numbers 1, 2, … $2K$ so that the first index is less than the second, for example,

$$b_{1111} = \langle (u_1 - \langle u_1 \rangle)(u_2 - \langle u_2 \rangle)(u_3 - \langle u_3 \rangle)(u_4 - \langle u_4 \rangle) \rangle$$
$$= b_{12}b_{34} + b_{13}b_{24} + b_{14}b_{23}.$$

When studying random variables described by a normal distribution it is convenient to use characteristic functions because of their simple form (see Eq. (1.37)):

$$\varphi(\rho_1, \rho_2, \ldots, \rho_N) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{ i\sum_{k=1}^{N} \rho_k u_k \right\} p(u_1, u_2, \ldots, u_N) du_1 du_2 \ldots du_N$$
$$= \exp\left\{ i\sum_{k=1}^{N} a_k \rho_k - \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} b_{jk}\rho_j\rho_k \right\}. \qquad (1.80)$$

Cumulants can be obtained from Eq. (1.53), using Eq. (1.80). For a Gaussian distribution, the cumulants of the first and second orders are respectively equal to a_j and b_{jk} whereas cumulants of higher orders are identically equal to zero. By using characteristic functions, one can prove that any linear combination of Gaussian random variables will also result in a Gaussian distribution.

Gaussian distributions are of great importance in applications due to a number of reasons. First, the behavior of many random variables is well approximated by a Gaussian distribution. Secondly, according to the central limit theorem, a random variable that is a sum of a large number of independent components with arbitrary distributions (which is the most common situation in statistical mechanics) is Gaussian.

Let us consider a one-dimensional Gaussian distribution

$$p(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{u^2}{2\sigma^2} \right\}, \quad (\sigma^2 = \langle u^2 \rangle). \qquad (1.81)$$

The characteristic function follows from the relations (1.37) and (1.52):

$$\varphi(\rho) = \exp\left\{ -\frac{\rho^2\sigma^2}{2} \right\}, \quad \psi(\rho) = -\frac{\rho^2\sigma^2}{2} \qquad (1.82)$$

Then Eqs. (1.57)–(1.58) yield

$$B_1 = S_1 = 0; \quad B_2 = S_2 = \sigma^2; \quad S_{n>2} = 0. \qquad (1.83)$$

The recurrent relation (1.59) takes the form

$$B_n = (n-1)\sigma^2 B_{n-2}, \qquad (1.84)$$

from which there follows

$$B_{2n+1} = 0; \quad B_{2n} = (2n-1)!!\sigma^{2n}. \qquad (1.85)$$

Consider yet another average value $\langle Xf(X)\rangle$, which is helpful in many applications. Here X is a Gaussian random variable given by Eq. (1.81) and $f(X)$ is an arbitrary deterministic function. We make a further assumption that $f(X)\exp(-X^2/2\sigma^2)\to 0$ for $X\to\pm\infty$ (i.e. the exponent dominates at large values of X). Then

$$\langle X f(X)\rangle = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} X f(X)\exp\left\{-\frac{X^2}{2\sigma^2}\right\}dX.$$

Integrating by parts, we arrive at the following expression:

$$\langle X f(X)\rangle = \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{df(X)}{dX}\exp\left\{-\frac{X^2}{2\sigma^2}\right\}dX = \sigma^2\left\langle\frac{df(X)}{dX}\right\rangle. \tag{1.86}$$

A similar expression can be obtained for a Gaussian random vector $\boldsymbol{X}=(X_1, X_2, \ldots, X_N)$ with a multidimensional distribution given by Eq. 1.75:

$$\langle X_i f(\boldsymbol{X})\rangle = B_{ij}\left\langle\frac{df(\boldsymbol{X})}{d\boldsymbol{X}}\right\rangle, \tag{1.87}$$

where $B_{ij}=\langle X_iX_j\rangle$ are components of the correlation matrix.

1.9
Stationary Random Functions, Homogeneous Random Fields

When discussing the problem of random variable averaging in Section 4, we mentioned the ergodic hypothesis, which states that as we increase the temporal or spatial averaging interval to infinity, the corresponding mean values tend to the ensemble average. For the ergodic hypothesis to be valid, some necessary conditions should be satisfied. We thus arrive to a special class of random fields $u(\boldsymbol{X}, t)$ satisfying the ergodicity conditions. These fields are frequently encountered in Statistical Mechanics, in particular, in problems that involve Brownian motion and turbulence.

Consider first the time averaging of a function $u(t)$, written for simplicity as a function of one variable because its dependence on space coordinates $\boldsymbol{X}$ is of no relevance to the problem. The time average will be denoted by $\langle u\rangle_T$. Then, in accordance with Eq. (1.22),

$$\langle u(t)\rangle_T = \frac{1}{T}\int_{-T/2}^{T/2} u(t+\tau)d\tau. \tag{1.88}$$

According to the ergodic hypothesis, $\langle u(t)\rangle_T$ should tend to the ensemble average $\langle u(t)\rangle$ at $T \to \infty$. For this to happen, the following simple relation must take place:

$$\langle u(t)\rangle_T = U = const. \tag{1.89}$$

This condition can be derived by considering the difference between the average values of the random variable calculated at different moments, t and t_1, where $t_1 > t$:

$$\begin{aligned}\langle u(t)\rangle_T - \langle u(t_1)\rangle_T &= \frac{1}{T}\left\{\int_{-T/2}^{T/2} u(t+\tau)d\tau - \int_{-T/2}^{T/2} u(t_1+\tau)d\tau\right\}\\ &= \frac{1}{T}\left\{\int_{-T/2+t}^{-T/2+t_1} u(s)ds - \int_{-T/2+t}^{T/2+t_1} u(s)ds\right\}.\end{aligned} \tag{1.90}$$

At $T \to \infty$ the right-hand side of Eq. (1.90) goes to zero, thus giving rise to the condition (1.89).

Similarly, by time-averaging the product $u(t)u(t_1) = u(t)u(t+s)$, where $s = t_1 - t$, and letting T go to ∞, we conclude that the time average of the correlation function $(B_{uu}(t, t_1))_T$ can be equal to its ensemle average $(B_{uu}(t, t_1)) = \langle u(t)u(t_1)\rangle$ only if for any two instants of time t_1 and t_2, where $t_2 > t_1$, the following condition is satisfied:

$$B_{uu}(t_2, t_1) = B_{uu}(t_2 - t_1). \tag{1.91}$$

For a moment of N-th order, this condition takes the form

$$B_{uu\ldots u}(t_1, t_2, \ldots, t_N) = B_{uu}(t_2 - t_1, \ldots, t_N - t_1). \tag{1.92}$$

In order for the ensemble averages of random values $u(t_1)$, $u(t_2)$, ..., $u(t_N)$ to be obtainable by time averaging, it is necessary to consider only those random functions $u(t)$ for which the N-dimensional PDF (at any N and t_1, t_2, ..., t_N) will depend on $N-1$ parameters $t_2 - t_1$, $t_3 - t_1$, ..., $t_N - t_1$, rather than on N parameters t_1, t_2, ..., t_N. In other words, the PDF must satisfy the condition

$$p_{i_1,\ldots,i_N}(u_1, u_2, \ldots, u_N) = p_{t_2-t_1,\ldots,t_N-t_1}(u_1, u_2, \ldots, u_N). \tag{1.93}$$

It should be noted that the condition (1.93) leads to the conditions (1.98), (1.91) and (1.92), and if the random function is Gaussian, then from Eq. (1.89) and Eq. (1.91) one can derive the properties (1.92) and (1.93).

The condition (1.93) describes a class of random functions whose PDF does not vary as we shift the time t_i by any time interval. Such functions are called stationary random functions or stationary random processes. One example is a steady turbulent flow, whose average characteristics (velocity, pressure, temperature, etc.) do not change with time. Any hydrodynamic parameter $\boldsymbol{u}(u_1(t), u_2(t), \ldots, u_N(t))$, for example, flow velocity at different space points, whose PDF for any set of $u_{i_1}(t_1), u_{i_2}(t_2), \ldots, u_{i_N}(t_N)$ does not vary as we shift all the instants of time $t_1, t_2, \ldots, t_N$ by one and the same value, represents a multidimensional stationary random process. Then all the mixed moments of functions $u(t)$ will also depend only on the differences between the corresponding instants of time. For example, all mutual correlation functions $B_{jk}(t_1, t_2) = \langle u_j(t_1)u_k(t_2)\rangle$ depend only on the time difference $\tau = t_2 - t_1$.

Consider now the space averaging of a random function $u(\boldsymbol{X})$, where $\boldsymbol{X}(X_1, X_2, X_3)$ is a space point. The space average (recall the definition (1.23)) is equal to

$$\langle u(\boldsymbol{X})\rangle_{ABC} = \frac{1}{ABC}\int_{-A/2}^{A/2}\int_{-B/2}^{B/2}\int_{-C/2}^{C/2} u(X_1+\xi_1, X_2+\xi_2, X_3+\xi_3)d\xi_1 d\xi_2 d\xi_3. \tag{1.94}$$

By analogy with time averaging, we can find the conditions that must hold in order for $\langle u(\boldsymbol{X})\rangle_{ABC}$ to coinside with the ensemble average $\langle u(\boldsymbol{X})\rangle$ at $A\to\infty$, $B\to\infty$, $C\to\infty$ (or when at least one of the intervals A,B,C goes to the limit). It is evident that the necessary conditions would be relations similar to (1.89), (1.91)–(1.93) with t replaced by $\boldsymbol{X}$:

$$\langle u(\boldsymbol{X})\rangle = U = const, \tag{1.95}$$

$$B_{uu}(\boldsymbol{X}_1, \boldsymbol{X}_2) = B_{uu}(\boldsymbol{X}_2 - \boldsymbol{X}_1), \tag{1.96}$$

$$p_{X_1,X_2,\ldots,X_N}(u_1, u_2, \ldots, u_N) = p_{X_2-X_1,\ldots,X_N-X_1}(u_1, u_2, \ldots, u_N). \tag{1.97}$$

where $B_{uu}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \langle u(\boldsymbol{X}_1)u(\boldsymbol{X}_2)\rangle$.

A random field $u(\boldsymbol{X})$ satisfying the conditions (1.95)–(1.97) is called a statistically homogeneous field.

Thus, in order for the space averaging of a function of random variables to produce the same results as ensemble averaging, it is necessary for the field $u(\boldsymbol{X})$ to be homogeneous. Parameters of a homogeneous turbulent flow (velocity, pressure, temperature, etc.), which do not depend on spatial coordinates, are good examples. It is clear that homogeneity of the flow cannot be realized in the entire flow region, because any flow is always restricted by boundaries, and the flow near the boundary is essentially inhomogeneous. In reality, the property of homogeneity can be realized only far enough from the boundary.

It should be mentioned that generally speaking, the conditions of stationarity and homogeneity are not sufficient for the convergence of time and space averages to ensemble averages. The necessary and sufficient conditions are formulated by ergodic theorem. Namely, it is necessary and sufficient to ensure the fulfillment of the following condition for the correlation function of fluctuations $b_{uu}(\tau)$:

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T b_{uu}(\tau)d\tau = 0. \tag{1.98}$$

The necessary averaging interval T can be estimated from the corresponding correlation time T_1, which is given by

$$T_1 = \frac{1}{b_{uu}(0)} \int_0^T b_{uu}(\tau) d\tau. \tag{1.99}$$

For sufficiently large T, the following asymptotic formula for root-mean-square deviation of the time average from the ensemble average is valid:

$$\left\langle \left| \langle u \rangle_T - \langle u \rangle \right|^2 \right\rangle \approx 2 \frac{T_1}{T} b_{uu}(0). \tag{1.100}$$

Eq. (1.99) allows us to determine the minimum averaging time for a given deviation of $\langle u \rangle_T$ from $\langle u \rangle$. In the case of spatial averaging, a similar estimation for the root-mean-square deviation of $\langle u \rangle_V$ from $\langle u \rangle$ gives

$$\left\langle \left| \langle u \rangle_V - \langle u \rangle \right|^2 \right\rangle \approx 2 \frac{V_1}{V} b_{uu}(0). \tag{1.101}$$

Here $\langle u \rangle_V$ is the spatial (volume) average, and V_1 is the correlation volume equal to

$$V_1 = \frac{1}{b_{uu}(0)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b_{uu}(\boldsymbol{r}) dr_1 dr_2 dr_3. \tag{1.102}$$

1.10
Isotropic Random Fields. Spectral Representation

A scalar random field $u(\boldsymbol{X})$ is called isotropic when all finite-dimensional PDFs $p_{X_1, X_1, \ldots, X_N}(u_1, u_2, \ldots, u_N)$ corresponding to this field are invariant under rotations of points $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ around the axis passing through the origin of the coordinate system and under mirror reflections of this set of points relative to planes passing through the origin. In applications, random fields that are both homogeneous and isotropic present the greatest interest. Henceforth these fields will be called simply isotropic. Thus the term "isotropic field" will imply a field whose PDF $p_{X_1, X_2, \ldots, X_N}(u_1, u_2, \ldots, u_N)$ is invariant under parallel translations, rotations, and specular reflections of the set of points $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$.

The homogeneity condition (1.95) for the field $u(\boldsymbol{X})$ means that its average value $\langle u(\boldsymbol{X}) \rangle$ should be constant. This constant is often made equal to zero by replacing the initial field $u(\boldsymbol{X})$ with the field $u'(\boldsymbol{X}) = u(\boldsymbol{X}) - \langle u(\boldsymbol{X}) \rangle$.

The correlation function $B(\boldsymbol{X}, \boldsymbol{X}') = \langle u(\boldsymbol{X}) u(\boldsymbol{X}') \rangle$ of an isotropic field has the same values at any pair of points $(\boldsymbol{X}, \boldsymbol{X}')$ and $(\boldsymbol{X}_1, \boldsymbol{X}_1')$ that would coincide after a combination of parallel translation and rotation. If the distance between the points $\boldsymbol{X}$ and

$\boldsymbol{X}'$ is the same as the distance between $\boldsymbol{X}_1$ and $\boldsymbol{X}_1'$, then $B(\boldsymbol{X},\boldsymbol{X}') = B(\boldsymbol{X}_1,\boldsymbol{X}_1')$. Hence the correlation function $B(\boldsymbol{X}, \boldsymbol{X}')$ depends only on the distance r between the points $\boldsymbol{X}$ and $\boldsymbol{X}' = \boldsymbol{X} + \boldsymbol{r}$. Here $r = |\boldsymbol{X}' - \boldsymbol{X}| = |\boldsymbol{r}|$, and correlation function can be written as

$$\langle u(\boldsymbol{X})u(\boldsymbol{X}')\rangle = B(r). \tag{1.103}$$

Application of harmonic (Fourier) analysis to random processes and random fields, that is, expansion of random functions as Fourier series (for functions defined on a finite domain) or Fourier integrals (for functions defined on an infinite domain) has proved to be a very successful approach. For any stationary random functions or homogeneous random fields, which by their definition cannot decay on the infinity, it is possible to carry out Fourier expansion (another common term is "spectral representation" or "spectral expansion"). It has a clear physical meaning: superposition of harmonic oscillations (for stationary random processes) or plane waves (for homogeneous random fields). The integral representation of the correlation function of a homogeneous random field is

$$B(\boldsymbol{r}) = \int e^{i\boldsymbol{kr}} F(\boldsymbol{k})d\boldsymbol{k}, \tag{1.104}$$

$$F(\boldsymbol{k}) = \frac{1}{8\pi^3}\int e^{-i\boldsymbol{kr}} B(\boldsymbol{r})d\boldsymbol{r}, \tag{1.105}$$

where $F(\boldsymbol{k})$ is called the spectrum of the homogeneous field, and $\boldsymbol{k}$ is the wave vector.

For an isotropic field, the condition (1.103) holds, so the spectrum depends on $k = |\boldsymbol{k}|$ rather than on $\boldsymbol{k}$. If we represent x, y, z in terms of spherical coordinates as $x = r \sin\theta \cos\Phi$, $y = r \sin\theta \sin\Phi$, $z = r\cos\theta$, the relation (1.105) will take the following form:

$$\begin{aligned} F(\boldsymbol{k}) &= \frac{1}{8\pi^3}\int e^{-i\boldsymbol{kr}} B(\boldsymbol{r})d\boldsymbol{r} = \frac{1}{8\pi^3}\int_0^{\infty}\int_{-\pi}^{\pi}\int_0^{\pi} e^{-ikr\cos\theta} B(r)r^2 \sin\theta d\theta d\Phi dr \\ &= \frac{1}{2\pi^2}\int_{-\infty}^{\infty} \frac{\sin(kr)}{kr} B(r)r^2 dr = F(k). \end{aligned} \tag{1.106}$$

Similarly,

$$B(r) = 4\pi\int_{-\infty}^{\infty} \frac{\sin(kr)}{kr} F(k)k^2 dk. \tag{1.107}$$

Instead of looking at $F(\boldsymbol{k})$, we can consider the following statistical characteristic:

$$E(\boldsymbol{k}) = \int\int_{|\boldsymbol{k}|=k} F(\boldsymbol{k})dS(\boldsymbol{k}), \tag{1.108}$$

where $S(\boldsymbol{k})$ is a surface element of the sphere $|\boldsymbol{k}| = k$.

Putting $\boldsymbol{r}=0$ into Eq. (1.104) and recalling Eq. (1.103), we get

$$B(0) = \left\langle [u(\boldsymbol{X})]^2 \right\rangle = \int_0^\infty E(k)dk. \tag{1.109}$$

If u is the velocity (for instance, the velocity of a turbulent flow), then $B(0)$ stands for the total energy of the field $u(\boldsymbol{X})$. Therefore $E(k)dk$ has the meaning of the energy of plane waves with wave numbers in the interval $(k, k+dk)$.

The derivations above can be generalized for the case of an isotropic multidimensional random field $\boldsymbol{u}(\boldsymbol{X}) = (u_1(\boldsymbol{X}), u_2(\boldsymbol{X}), \ldots, u_N(\boldsymbol{X}))$ characterized by the correlation matrix

$$||B_{ij}|| = \langle u_i(\mathrm{X})u_j(\mathrm{X}+\mathbf{r})\rangle \tag{1.110}$$

The components of such a matrix are functions of $r=|\boldsymbol{r}|$. Hence the spectral representation of this field will be written as

$$B_{ij}(r) = 4\pi\int_0^\infty \frac{\sin(kr)}{kr} F_{ij}(k)k^2dk, \tag{1.111}$$

$$F_{ij}(k) = \frac{1}{2\pi^2}\int_0^\infty \frac{\sin(kr)}{kr} B_{ij}(r)r^2dr.$$

The above-formulated definition of an isotropic random field is valid for scalar random functions, for example, pressure $p(\boldsymbol{X})$, temperature $\vartheta(\boldsymbol{X})$, one-dimensional velocity $u(\boldsymbol{X})$ and so on. In the case of vector random fields such as three-dimensional velocity, as well as for the fields given by a set of vector and scalar hydrodynamic parameters (for example, a field of three-dimensional velocity, pressure, and temperature), isotropy is defined in the following way. A random vector field $\boldsymbol{u}(\boldsymbol{X})$ is called isotropic if the PDF of the components of the vector $\boldsymbol{u}(\boldsymbol{X})$ taken at an arbitrary set of points $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ is invariant under parallel translations, rotations, and mirror reflections of this set of points accompanied by rotation or mirror reflection of the coordinate system. Using the theory of invariants of the rotation-reflection group, we may conclude from this definition that the correlation tensor $B_{ij}(\boldsymbol{r})$ should be a linear combination of the constant invariant tensor δ_{ij} ("Kroneckers delta function") and the tensor r_ir_j. The coefficients in this linear combination will depend on the only invariant that can be built from components of the vector $\boldsymbol{r}$, that is, on the length $r=|\boldsymbol{r}|$:

$$B_{ij}(\boldsymbol{r}) = A_1(r)r_ir_j + A_2(r)\delta_{ij}. \tag{1.112}$$

1.11
Stochastic Processes. Markovian Processes. The Chapman–Kolmogorov Integral Equation

The term "stochastic process" implies that the time evolution of a system is described probabilistically. This means that there is a certain time-dependent random variable. Examples of stochastic processes include Brownian motion of a particle driven by a random force and the motion of particles suspended in a turbulent flow. The random variable is the spatial position $\boldsymbol{X}$ of the particle at different instants of time. One can measure the values $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, ... at the instants of time t_1, t_2, t_3, ... and assume that there should exist a joint PDF $p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2; \boldsymbol{X}_3, t_3; \ldots)$ such that $p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2; \boldsymbol{X}_3, t_3; \ldots)d\boldsymbol{X}_1, d\boldsymbol{X}_2, \ldots$ would give the probability for the particle to be located in the interval $(\boldsymbol{X}_1 + d\boldsymbol{X}_1)$ at the instant of time t_1, in the interval $(\boldsymbol{X}_2 + d\boldsymbol{X}_2)$ at the instant t_2, and so on. When the particle moves under the action of a rapidly fluctuating random force (in the case of Brownian motion this random force is the sum of interaction forces between the particle and the molecules of the surrounding fluid, or, to use a more casual term, the sum of collision forces), it can change its direction millions times per second. In this context, when considering two successive particle positions $\boldsymbol{X}_i$ and $\boldsymbol{X}_{i+1}$ at the instants t_i and t_{i+1} such that the time increment $\Delta t_i = t_{i+1} - t_i$ is much smaller than the characteristic time of the process but large as compared to the time between successive collisions of the particle with the surrounding molecules, it is natural to suggest a model where the particles position $\boldsymbol{X}_{i+1}$ at the instant t_{i+1} is determined by its position $\boldsymbol{X}_i$ at the previous instant t_i and does not depend on the earlier instants of time $\ldots, t_{i-2}$, t_{i-1}. In other words, in the proces of a chaotic small-scale random walk, the particle forgets its past very quickly. Such processes are known as Markovian processes.

Hence a Markovian process is a stochastic process characterized by the independence of the future from the past, where the past is defined as the set of all events observed up to the present instant of time t. In other words, one has to deal with random functions whose variations are statistically independent from one another.

A Markovian process can be described by using the concept of conditional probability. Let us consider an ordered sequence of times $t_1 \geq t_2 \geq t_3 \geq \ldots \geq \tau_1 \geq \tau_2 \geq \ldots$, where τ_1, τ_2, ... belong to the past and ... t_3, t_2, t_1 – to the future. Let $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ... denote the values of a random variable at the past instants of time τ_1, τ_2, ... and ... $\boldsymbol{X}_3$, $\boldsymbol{X}_2$, $\boldsymbol{X}_1$ denote its values at the future instants of time ... t_3, t_2, t_1. The PDF of the events $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ... under the condition that the events $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, ... have already occurred (the conditional PDF) is then written as

$$p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2; \boldsymbol{X}_3, t_3; \ldots | Y_1, \tau_1; Y_2, \tau_2; \ldots).$$

In accordance with the Markovian principle, we demand that the conditional probability must be completely determined by the state of the system at the most recent instant of time, that is, by the knowledge of the random variable at τ_1. Then the following equality must be valid:

$$p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2; \boldsymbol{X}_3, t_3; \ldots | Y_1, \tau_1; Y_2, \tau_2; \ldots)$$
$$= p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2; \boldsymbol{X}_3, t_3; \ldots | Y_1, \tau_1). \quad (1.113)$$

This relation means that any conditional probability can be expressed through an ordinary conditional probability of the type $p(\boldsymbol{X}_1, t_1|Y_1, \tau_1)$. Indeed, from the definition (1.2) of conditional probability we have:

$$p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2 | Y_1, \tau_1) = p(\boldsymbol{X}_1, t_1 | \boldsymbol{X}_2, t_2; Y_1, \tau_1)\, p(\boldsymbol{X}_2, t_2 | Y_1, \tau_1).$$

Applying the postulate (1.113) to the first factor on the right-hand side, we express the joint PDF through ordinary conditional PDFs:

$$p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2; | Y_1, \tau_1) = p(\boldsymbol{X}_1, t_1 | \boldsymbol{X}_2, t_2)\, p(\boldsymbol{X}_2, t_2 | Y_1, \tau_1). \quad (1.114)$$

Continuing this procedure, we obtain for N successive events:

$$p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2; \ldots; \boldsymbol{X}_N, t_N) = p(\boldsymbol{X}_1, t_1 | \boldsymbol{X}_2, t_2)\, p(\boldsymbol{X}_2, t_2 | \boldsymbol{X}_3, t_3) \times$$
$$\ldots \times p(\boldsymbol{X}_{N-1}, t_{N-1} | \boldsymbol{X}_N, t_N). \quad (1.115)$$

As one could expect, the Markovian principle results in the independence of conditional pairs of successive events.

From the consistency property (see property 4 in Section 4) of the PDF for two successive events $\boldsymbol{X}_2$ and $\boldsymbol{X}_1$ and from the relation (1.2) for the conditional probability there follows:

$$p(\boldsymbol{X}_1, t_1) = \int p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2) d\boldsymbol{X}_2 = \int p(\boldsymbol{X}_1, t_1 | \boldsymbol{X}_2, t_2) p(\mathbf{X}_2, t_2) d\boldsymbol{X}_2. \quad (1.116)$$

A similar relation can be written for the conditional probability:

$$p(\boldsymbol{X}_1, t_1 | \boldsymbol{X}_3, t_3) = \int p(\boldsymbol{X}_1, t_1; \boldsymbol{X}_2, t_2 | \boldsymbol{X}_3, t_3) d\boldsymbol{X}_2$$
$$= \int p(\boldsymbol{X}_1, t_1 | \boldsymbol{X}_2, t_2; \boldsymbol{X}_3, t_3)\, p(\boldsymbol{X}_2, t_2 | \boldsymbol{X}_3, t_3) d\boldsymbol{X}_2.$$

As far as $t_1 \geq t_2 \geq t_3$, the Markovian principle allows to drop the dependence on $\boldsymbol{X}_3$ in the first factor of the integrand:

$$p(\boldsymbol{X}_1, t_1 | \boldsymbol{X}_3, t_3) = \int p(\boldsymbol{X}_1, t_1 | \boldsymbol{X}_2, t_2)\, p(\boldsymbol{X}_2, t_2 | \boldsymbol{X}_3, t_3) d\boldsymbol{X}_2. \quad (1.117)$$

The integral equation (1.117) is called the Chapman–Kolmogorov equation. This equation forms the basis of the theory of stochastic processes.

When considering a Markovian process, it is important to know whether the range of the random value is continuous or discrete and whether the trajectory $\boldsymbol{X}(t)$ is a continuous function of t. As an example, consider rarefied gas molecules characterized by the velocity $\boldsymbol{V}(t)$ and by the position $\boldsymbol{X}(t)$. In this example, the velocity range is obviously continuous, but the function $\boldsymbol{V}(t)$ can be discontinuious, which happens when the interactions between molecules are modeled by the elastic collisions of rigid spheres. However, even in such a model, the position of a gas molecule $\boldsymbol{X}(t)$ remains a continuous function. In reality, molecules do not interact as rigid spheres. There exists a molecular interaction between them that is characterized by some interaction potential (for example, the Lennard–Jones potential). If we account for this potential, we will find that the molecules trajectory deflects continuously in the process of collision. The characteristic time of molecular collisions is extremely short. It is much shorter then the time intervals that make up a Markovian chain. It can be said that the Markovian method circumnavigates the issue of continuity of a random variable by approximating the real process on a large-scale time grid. Hence, irrespective of how the collision process is modeled, on large-scale time grid, the collision will always be marked by a velocity jump. By the same token, the trajectories are not necessarily continuous on this time grid. Another example is a chemical reaction that involves production consumption of molecules of a certain substance. The characteristic time of a chemical reaction is also very short as a rule. Therefore the random value, for example, molecular concentration, changes discontinuously on the large-scale time grid during the reaction.

In this context, the following continuity condition looks quite self-intuitive: if for any $\varepsilon > 0$ uniformly in $\boldsymbol{Z}$, t and Δt there holds

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|\boldsymbol{X}-\boldsymbol{Z}|>\varepsilon} p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)d\boldsymbol{X} = 0, \tag{1.118}$$

then the realization of $\boldsymbol{X}(t)$ is continuous function of t, with the probability 1. It means that the probability that the position $\boldsymbol{X}$ differs from $\boldsymbol{Z}$ by a finite amount at $\Delta t \to 0$ goes to zero faster than Δt. This is known as the Lindenberg continuity condition for a random function $\boldsymbol{X}(t)$.

One can show that Einsteins solution of the Brownian motion problem, which is a Gaussian PDF written as

$$p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t) = \frac{1}{(4\pi D\Delta t)^{1/2}} \exp\left\{ -\frac{(X-Z)^2}{4D\Delta t} \right\}, \tag{1.119}$$

satisfies the condition (1.118). On the other hand, the PDF

$$p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t) = \frac{\Delta t}{\pi[(X-Z)^2 + \Delta t^2]}, \tag{1.120}$$

which describes a Cauchy process, does not satisfy this condition. Both distributions tend to the delta function $\delta(\boldsymbol{X} - \boldsymbol{Z})$ at $\Delta t \to 0$ (see Eq. (1.9) and Eq. (1.16) and satisfy

Eq. (1.117). So the Chapman–Kolmogorov equation allows for both continuous and discontinuous solutions (PDFs).

1.12 The Chapman–Kolmogorov, Chapman–Feller, Fokker–Planck, and Liouville Differential Equations

1.12.1 Derivation of the Differential Chapman–Kolmogorov Equation

When solving concrete problems, one uses the differential form of the Chapman–Kolmogorov equation, which can be derived from the integral equation (1.117) under some additional assumptions. An additional assumption of continuity of the random process leads us to the Fokker–Planck equation. But discontinuous processes can also take place, as was mentioned in the previous section. Thus the Chapman–Kolmogorov differential equation should be able to describe both continuous and discontinuous processes. To satisfy this general requirement, we shall demand realization of the following conditions:

1. $$\lim_{\Delta t \to 0} \frac{1}{\Delta t} p(\boldsymbol{X}, t + \Delta t | \boldsymbol{Z}, t) = \boldsymbol{W}(\boldsymbol{X} | \boldsymbol{Z}, t) \tag{1.121}$$

   should take place in the region $|\boldsymbol{X} - \boldsymbol{Z}| \geq \varepsilon$ uniformly for all $\boldsymbol{X}$, $\boldsymbol{Z}$, t, and the limit should not depend on ε;

2. $$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|\boldsymbol{X}-\boldsymbol{Z}|<\varepsilon} (X_i - Z_i)\, p(\boldsymbol{X}, t + \Delta t | \boldsymbol{Z}, t) d\boldsymbol{X} = A_i(\boldsymbol{Z}, t) + 0(\varepsilon); \tag{1.122}$$

3. $$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int\limits_{|\boldsymbol{X}-\boldsymbol{Z}|<\varepsilon} (X_i - Z_i)(X_j - Z_j)\, p(\boldsymbol{X}, t + \Delta t | \boldsymbol{Z}, t) d\boldsymbol{X} = D_{ij}(\boldsymbol{Z}, t) + 0(\varepsilon). \tag{1.123}$$

The conditions 2 and 3 assume a uniform convergence with respect to $\boldsymbol{Z}$, ε, and t. Condition 1 is responsible for the continuity of the process. If $\boldsymbol{W}(\boldsymbol{X}|\boldsymbol{Z}, t) = 0$, the process can be described by continuous trajectories; otherwise the trajectories are discontinuous.

To derive the differential equation, let us consider how the average value of some (twice differentiable) function $f(\boldsymbol{X})$ varies with time. According to Eq. (1.7), the average is written as

$$\langle f(\boldsymbol{X}) \rangle = \int f(\boldsymbol{X})\, p(\boldsymbol{X}, t | \boldsymbol{Y}, t') d\boldsymbol{X}.$$

Then

$$\frac{\partial\langle f\rangle}{\partial t}=\left\{\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int f(\boldsymbol{X})[p(\boldsymbol{X},t+\Delta t|\boldsymbol{Y},t')-p(\boldsymbol{X},t|\boldsymbol{Y},t')]d\boldsymbol{X}\right\}.$$

Let us now put $\boldsymbol{X}_1=\boldsymbol{X}$, $t_1=t+\Delta t$, $\boldsymbol{X}_2=\boldsymbol{Z}$, $t_2=t$, $\boldsymbol{X}_3=\boldsymbol{Y}$, $t_3=t'$ into the Chapman–Kolmogorov integral equation (1.110). It means that in addition to the point $\boldsymbol{Y}$ at the time t' and the point $\boldsymbol{X}$ at the time $t+\Delta t$, we take yet another point $\boldsymbol{Z}$ (between these two) on the trajectory at the intermediate time t ($t'<t<t+\Delta t$). Then the last relation transforms into

$$\begin{aligned}\frac{\partial}{\partial t}\left\{\int f(\boldsymbol{X})\,p(\boldsymbol{X},t|\boldsymbol{Y},t')d\boldsymbol{X}\right\}&=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left\{\int d\boldsymbol{X}\int f(\boldsymbol{X})\right.\\ p(\boldsymbol{X},t+\Delta t|\boldsymbol{Z},t)p(\boldsymbol{Z},t|\boldsymbol{Y},t')d\boldsymbol{Z}&-\left.\int f(\boldsymbol{Z})\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')d\boldsymbol{Z}\right\}.\end{aligned}\tag{1.124}$$

We replaced the integration variable $\boldsymbol{X}$ by $\boldsymbol{Z}$ in the last term of the right-hand side. Let us now subdivide the region of integration over $\boldsymbol{X}$ into two regions: $|\boldsymbol{X}-\boldsymbol{Z}|\geq\varepsilon$ and $|\boldsymbol{X}-\boldsymbol{Z}|<\varepsilon$. In the latter region, we shall perform a Taylor series expansion of the function $f(\boldsymbol{X})$ that appears in the integrand:

$$\begin{aligned}f(\boldsymbol{X})&=f(\boldsymbol{Z})+\sum_i\frac{\partial f(\boldsymbol{Z})}{\partial Z_i}(X_i-Z_i)\\&+\sum_i\sum_j\frac{1}{2}\frac{\partial^2 f(\boldsymbol{Z})}{\partial Z_i\partial Z_j}(X_i-Z_i)(X_j-Z_j)+|\boldsymbol{X}-\boldsymbol{Z}|^2R(\boldsymbol{X},\boldsymbol{Z}),\end{aligned}\tag{1.125}$$

where the last term on the right-hand side of (1.125) is the residual term that obeys the condition $R(\boldsymbol{X},\boldsymbol{Z})\to 0$ at $|\boldsymbol{X}-\boldsymbol{Z}|\to 0$. Now, substituting Eq. (1.125) into the right-hand side of Eq. (1.124) and grouping the terms, we get:

$$\begin{aligned}&\frac{\partial}{\partial t}\left\{\int f(\boldsymbol{X})\,p(\boldsymbol{X},t|\boldsymbol{Y},t')d\boldsymbol{X}\right\}\\&=\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left\{\iint\limits_{|\boldsymbol{X}-\boldsymbol{Z}<\varepsilon}\left[\sum_i\frac{\partial f(\boldsymbol{Z})}{\partial Z_i}(X_i-Z_i)+\sum_i\sum_j\frac{1}{2}\frac{\partial^2 f(\boldsymbol{Z})}{\partial Z_i\partial Z_j}(X_i-Z_i)(X_j-Z_j)\right]\right.\\&\times p(\boldsymbol{X},t+\Delta t|\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|Y,t')d\boldsymbol{X}d\boldsymbol{Z}\\&+\iint\limits_{|\boldsymbol{X}-\boldsymbol{Z}|<\varepsilon}|\boldsymbol{X}-\boldsymbol{Z}|^2R(\boldsymbol{X},\boldsymbol{Z})\,p(\boldsymbol{X},t+\Delta t|\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')d\boldsymbol{X}d\boldsymbol{Z}\\&+\iint\limits_{|\boldsymbol{X}-\boldsymbol{Z}|\geqslant\varepsilon}f(\boldsymbol{X})\,p(\boldsymbol{X},t+\Delta t|\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')d\boldsymbol{X}d\boldsymbol{Z}\\&+\left.\iint\limits_{|\boldsymbol{X}-\boldsymbol{Z}|<\varepsilon}f(\boldsymbol{Z})\,p(\boldsymbol{X},t+\Delta t|\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')d\boldsymbol{X}d\boldsymbol{Z}-\int f(\boldsymbol{Z})\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')d\boldsymbol{Z}\right\}.\end{aligned}\tag{1.126}$$

Consider the terms in the left-hand side of Eq. (1.126) in the consecutive order. Since $p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)$ is the PDF, we note that

$$\int p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)d\boldsymbol{X} = 1.$$

With this in mind, and using the condition of uniform convergence that allows us to take the limit of the integrand, the last term in Eq. (1.126) can be rewritten as

$$\begin{aligned}&\int f(\boldsymbol{Z})\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')d\boldsymbol{Z}\\ &= \int f(\boldsymbol{Z})\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')d\boldsymbol{Z} \int p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)d\boldsymbol{X}\\ &= \iint f(\boldsymbol{Z})\, p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')d\boldsymbol{X}d\boldsymbol{Z}.\end{aligned}$$

To transform the first term, it is necessary to exploit the conditions 2 and 3:

$$\begin{aligned}&\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left\{\iint_{|\boldsymbol{X}-\boldsymbol{Z}|<\varepsilon}\left[\sum_i \frac{\partial f(\boldsymbol{Z})}{\partial Z_i}(X_i - Z_i) + \sum_i\sum_j \frac{1}{2}\frac{\partial^2 f(\boldsymbol{Z})}{\partial Z_i \partial Z_j}(X_i - Z_i)(X_j - Z_j)\right]\right.\\ &\left.\times p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')d\boldsymbol{X}d\boldsymbol{Z}\right\}\\ &= \int\left[\sum_i A_i(\boldsymbol{Z}, t)\frac{\partial f}{\partial Z_i} + \sum_i\sum_j \frac{1}{2}D_{ij}(\boldsymbol{Z}, t)\frac{\partial^2 f}{\partial Z_i \partial Z_j}\right] p(\boldsymbol{Z}, t|Y, t')d\boldsymbol{Z} + 0(\varepsilon).\end{aligned}$$

The second term tends to zero because of the condition $\boldsymbol{R}(\boldsymbol{X}, \boldsymbol{Z}) \to 0$ at $\varepsilon \to 0$, since $|\boldsymbol{X} - \boldsymbol{Z}| \to 0$.

Carrying out integration in the last term over two subdomains, $|\boldsymbol{X} - \boldsymbol{Z}| \geq \varepsilon$ and $|\boldsymbol{X} - \boldsymbol{Z}| < \varepsilon$ and taking into account Property 1, one can reduce the last three terms to

$$\begin{aligned}&\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left\{\iint_{|\boldsymbol{X}-\boldsymbol{Z}|\geqslant\varepsilon} f(\boldsymbol{X})\, p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')d\boldsymbol{X}d\boldsymbol{Z}\right.\\ &+ \iint_{|\boldsymbol{X}-\boldsymbol{Z}|<\varepsilon} f(\boldsymbol{Z})\, p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')d\boldsymbol{X}d\boldsymbol{Z}\\ &- \iint_{|\boldsymbol{X}-\boldsymbol{Z}|\geqslant\varepsilon} f(\boldsymbol{Z})\, p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')d\boldsymbol{X}d\boldsymbol{Z}\\ &\left.- \iint_{|\boldsymbol{X}-\boldsymbol{Z}|<\varepsilon} f(\boldsymbol{Z})\, p(\boldsymbol{X}, t+\Delta t|\boldsymbol{Z}, t)\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')d\boldsymbol{X}d\boldsymbol{Z}\right\}\\ &= \iint_{|\boldsymbol{X}-\boldsymbol{Z}|\geqslant\varepsilon}[f(\boldsymbol{X})\, W(\boldsymbol{X}|\boldsymbol{Z}, t)\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t') - f(\boldsymbol{Z})\, W(\boldsymbol{X}|\boldsymbol{Z}, t)\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')]d\boldsymbol{X}d\boldsymbol{Z}\end{aligned}$$

Nothing will be changed if we swap the variables $\boldsymbol{X}$ and $\boldsymbol{Z}$ in the first term:

$$\iint\limits_{|\boldsymbol{X}-\boldsymbol{Z}|\geqslant\varepsilon} f(\boldsymbol{Z})[W(\boldsymbol{Z}|\boldsymbol{X},t)\,p(\boldsymbol{X},t|\boldsymbol{Y},t')-W(\boldsymbol{X}|\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')]d\boldsymbol{X}d\boldsymbol{Z}.$$

Going to the limit $\varepsilon\to 0$ in Eq. (1.126), one obtains the following relation:

$$\begin{aligned}
&\frac{\partial}{\partial t}\left\{\int f(\boldsymbol{X})\,p(\boldsymbol{X},t|\boldsymbol{Y},t')d\boldsymbol{X}\right\}\\
&=\int\left[\sum_i A_i(\boldsymbol{Z},t)\frac{\partial f}{\partial Z_i}+\sum_i\sum_j\frac{1}{2}D_{ij}(\boldsymbol{Z},t)\frac{\partial^2 f}{\partial Z_i\partial Z_j}\right]p(\boldsymbol{Z},t|\boldsymbol{Y},t')d\boldsymbol{Z}\\
&\quad+\int f(\boldsymbol{Z})d\boldsymbol{Z}\int[W(\boldsymbol{Z}|\boldsymbol{X},t)\,p(\boldsymbol{X},t|\boldsymbol{Y},t')-W(\boldsymbol{X}|\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')]d\boldsymbol{X},
\end{aligned} \tag{1.127}$$

which is valid when the integral

$$\int W(\mathrm{Z}|\boldsymbol{X},t)\,p(\boldsymbol{X},t|\boldsymbol{Y},t^{c})d\boldsymbol{X} \tag{1.128}$$

exists.

The condition (1.121) determines the function $W(\boldsymbol{Z}|\boldsymbol{X}, t)$ only if $\boldsymbol{X}\neq\boldsymbol{Z}$. But there are situations when $\boldsymbol{X}=\boldsymbol{Z}$, for example, a Cauchy process given by Eq. (1.120). In this case the value of the integral (1.128) should be interpreted as the principal integral value. Because such singular cases are rare, we will not be using the principal integral value symbol in the discussion below.

Integration by parts of the first term on the right-hand side of Eq. (1.127) yields

$$\begin{aligned}
&\int f(\mathrm{Z})\frac{\partial p(\boldsymbol{Z},t|\boldsymbol{Y},t')}{\partial t}d\boldsymbol{Z}\\
&=\int f(\mathrm{Z})\left\{-\sum_i\frac{\partial}{\partial Z_i}[A_i(\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')]d\boldsymbol{Z}\right.\\
&\quad+\sum_i\sum_j\frac{1}{2}\frac{\partial^2 f}{\partial Z_i\partial Z_j}[D_{ij}(\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')]\\
&\quad\left.+\int[W(\boldsymbol{Z}|\boldsymbol{X},t)\,p(\boldsymbol{X},t|\boldsymbol{Y},t')-W(\boldsymbol{X}|\boldsymbol{Z},t)\,p(Z,t|\boldsymbol{Y},t')]d\boldsymbol{X}\right\}+\ldots,
\end{aligned}$$

where the dots denote the surface integrals over the boundary enclosing the considered region. As far as the function $f(\boldsymbol{Z})$ was chosen arbitrarily with the only requirement that it should be at least twice differentiable, we can impose an additional requirement that this function should vanish on the regions boundary. Then all surface integrals also vanish, and finally, we obtain the Chapman–Kolmogorov differential equation:

$$\begin{aligned}\frac{\partial p(\boldsymbol{Z},t|\boldsymbol{Y},t')}{\partial t}\\ &= -\sum_i \frac{\partial}{\partial Z_i}[A_i(\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')]\\ &+\sum_i\sum_j \frac{1}{2}\frac{\partial^2 f}{\partial Z_i \partial Z_j}[D_{ij}(\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')]\\ &+\int[W(\boldsymbol{Z}|\boldsymbol{X},t)\,p(\boldsymbol{X},t|\boldsymbol{Y},t')-W(\boldsymbol{X}|\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')]d\boldsymbol{X}.\end{aligned} \tag{1.129}$$

Consider now some particular cases of the Chapman–Kolmogorov equation.

1.12.2 Discontinuous ("Jump") Processes. The Kolmogorov–Feller Equation

This equation follows from the Chapman–Kolmogorov equation at $A_i = D_{ij} = 0$:

$$\frac{\partial p(\boldsymbol{Z},t|\boldsymbol{Y},t')}{\partial t} = \int[W(\boldsymbol{Z}|\boldsymbol{X},t)\,p(\boldsymbol{X},t|\boldsymbol{Y},t')-W(\boldsymbol{X}|\boldsymbol{Z},t)\,p(\boldsymbol{Z},t|\boldsymbol{Y},t')]d\boldsymbol{X}. \tag{1.130}$$

If we take $p(\boldsymbol{Z}, t|\boldsymbol{Y}, t') = \delta(\boldsymbol{Y} - \boldsymbol{Z})$ at $t = t'$ as the initial condition, then for small values of Δt the solution will be approximately equal to

$$p(\boldsymbol{Z}, t+\Delta t|Y,t') \approx \delta(\boldsymbol{Y}-\boldsymbol{Z})[1-\int W(\boldsymbol{X}|\boldsymbol{Z},t)\Delta t d\boldsymbol{X}] + W(\boldsymbol{Z}|\boldsymbol{X},t)\Delta t.$$

It is implied by this solution that for any Δt, there is a finite probability

$$1-\int W(\boldsymbol{X}|\boldsymbol{Z},t)\Delta\, t d\boldsymbol{X}$$

to find the particle at the initial position $\boldsymbol{Y}$, and the distribution of particles leaving $\boldsymbol{Y}$ is given by the function $W(\boldsymbol{Z}|\boldsymbol{Y}, t)$. Hence, the trajectory $\boldsymbol{X}(t)$ consists of linear segments $\boldsymbol{X} = const$ alternating with jumps whose distribution is given by the function $W(\boldsymbol{Z}|\boldsymbol{Y}, t)$. That is why the process has discontinuous character and the trajectories have discontinuities in a discrete set of points.

1.12.3 Diffusion Processes. The Fokker–Planck Equation

If the process is continuous, then $W(\boldsymbol{Z}|\boldsymbol{Y}, t) = 0$, and the Chapman–Kolmogorov equation is reduced to the Fokker–Planck equation:

$$\frac{\partial p(\mathbf{Z},t|\mathbf{Y},t')}{\partial t} = -\sum_i \frac{\partial}{\partial Z_i}[A_i(\mathbf{Z},t)\,p(\mathbf{Z},t|\mathbf{Y},t')] + \sum_i\sum_j \frac{1}{2}\frac{\partial^2}{\partial Z_i \partial Z_j}[D_{ij}(\mathbf{Z},t)\,p(\mathbf{Z},t|\mathbf{Y},t')] \tag{1.131}$$

Such a process is called the diffusion process. The vector $\boldsymbol{A}(\boldsymbol{Z}, t)$ is called the drift vector. It is similar to the velocity vector in the convective term of the transport equation. The matrix $\mathbf{D}(\mathbf{Z}, t) = ||D_{ij}(\mathbf{Z}, t)||$ is called the dispersion matrix. According to its definition (see Eq. (1.123)), it is non-negative, definite and symmetric. It can be shown that $\boldsymbol{D}$ is a tensor. It is known as the dispersion tensor.

To understand the physical meaning of $\boldsymbol{A}$ and $\boldsymbol{D}$, consider the initial phase of the process in the same manner as we just did for the Kolmogorov–Feller equation, with the same initial condition $p(\mathbf{Z}, t|\mathbf{Y}, t') = \delta(\mathbf{Y} - \mathbf{Z})$ at $t = t'$.

Assuming that during a small time $\Delta t \ll 1$ the values of A_j and D_{ij}. will not change much as compared to p, the equation transforms to

$$\frac{\partial p(\mathbf{Z},t|\mathbf{Y},t')}{\partial t} = -\sum_i A_i(\mathbf{Z},t)\frac{\partial}{\partial Z_i}[p(\mathbf{Z},t|\mathbf{Y},t')] + \sum_i\sum_j \frac{1}{2}D_{ij}(\mathbf{Z},t)\frac{\partial^2}{\partial Z_i \partial Z_j}[p(\mathbf{Z},t|\mathbf{Y},t')], \tag{1.132}$$

where $t - t' = \Delta t \ll 1$. On this small time interval, $A_i(\boldsymbol{Y},\boldsymbol{t})$ and $D_{ij}(\boldsymbol{Y},\boldsymbol{t})$ are regarded as dependent on the initial position $\boldsymbol{Y}$ but independent of time t. The solution of Eq. (1.132) has the form

$$p(\mathbf{Z},t+\Delta t|\mathbf{Y},t) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{D}(\mathbf{Y},t)|^{1/2}\Delta t^{1/2}} \times \exp\left\{-\frac{1}{2}\frac{[\mathbf{Z}-\mathbf{Y}-\boldsymbol{A}(\mathbf{Y},t)\Delta t]^T[\mathbf{D}(\mathbf{Y},t)]^{-1}[\mathbf{Z}-\mathbf{Y}-\boldsymbol{A}(\mathbf{Y},t)\Delta t]}{\Delta t}\right\} \tag{1.133}$$

where $D = |\boldsymbol{D}|$ is the determinant of the matrix $\boldsymbol{D}$.

Eq. (1.333) indicates that at the initial stage, the diffusion process is described by the Gaussian law (see Eq. 1.77) and that fluctuations with the correlation matrix $\boldsymbol{D}(\boldsymbol{Y}, t)\Delta t$ are superimposed on the regular drift with the velocity $\boldsymbol{A}(\boldsymbol{Y}, t)$. It means that at the initial stage, the systems trajectory can be represented as

$$\boldsymbol{Z}(t+\Delta t) = \boldsymbol{Y}(t) + \boldsymbol{A}(\boldsymbol{Y}(t),t)\Delta t + \boldsymbol{\eta}(t)(\Delta t)^{1/2}, \tag{1.134}$$

where $\boldsymbol{\eta}(t)$ is a random vector with the mean value and the correlation matrix given by

$$\langle\boldsymbol{\eta}\rangle = 0, \quad \langle\boldsymbol{\eta}(t)\boldsymbol{\eta}^T(t)\rangle = \boldsymbol{D}(\boldsymbol{Y},t). \tag{1.135}$$

In a diffusion process, trajectories are continuous everywhere because $Z(t+\Delta t) \rightarrow Z(t)$ at $\Delta t \rightarrow 0$. They are also non-differentiable at any point because of the term proportional to $(\Delta t)^{1/2}$. Since $Z(t+\Delta t) - Y(t) = \Delta Z$ is a random increment of the particles position, we can divide both parts of Eq. (1.134) by Δt, obtaining

$$\frac{\Delta Z}{\Delta t} = \boldsymbol{A}(\boldsymbol{Y}(t), t) + \boldsymbol{\eta}(t)(\Delta t)^{-1/2}. \tag{1.136}$$

Eq. (1.136) is a stochastic differential equation that has a fundamental role in describing the motion of particles driven by an external random force.

In a three-dimensional case, the Gaussian distribution (1.133) can be written in a simpler form. Let us direct the Cartesian axes Z_1, Z_2, Z_3, so that they would coincide with the principal directions of the dispersion tensor $\boldsymbol{D}$. Let D_{ij} be the principal values of the dispersion tensor matrix $||D_{ij}||$. In this coordinate system, the distribution (1.133) transforms into

$$p(\mathbf{Z}, t+\Delta t|\mathbf{Y}, t) = \frac{1}{(2\pi)^{3/2}[D_{11}(t)D_{22}(t)D_{33}(t)]^{1/2}} \times \exp\left\{-\frac{(Z_1-Y_1)^2}{2D_{11}(t)} - \frac{(Z_2-Y_2)^2}{2D_{22}(t)} - \frac{(Z_3-Y_3)^2}{2D_{33}(t)}\right\}. \tag{1.137}$$

Let us now introduce the probability flux with components

$$J_i(\mathbf{Z}, t) = A_i(\mathbf{Z}, t)\, p(\mathbf{Z}, t+\Delta t|\mathbf{Y}, t') - \frac{1}{2}\sum_j \frac{\partial}{\partial Z_j}[D_{ij}(\mathbf{Z}, t)\, p(\mathbf{Z}, t|\mathbf{Y}, t')].$$

Then the Fokker–Planck equation can be written in a compact, universally accepted form of a conservation equation:

$$\frac{\partial p(\mathbf{Z}, t|\mathbf{Y}, t')}{\partial t} + \sum_i \frac{\partial J_i(\mathbf{Z}, t)}{\partial Z_i} = 0. \tag{1.138}$$

Introduction of the probability flux allows us to formulate the boundary conditions for the Fokker–Planck equation. Consider the process in a region R bounded by the surface S. One can see the following possible types of boundary conditions.

a) Absorbing boundary

It is assumed that as soon as a particle reaches the boundary, it vanishes, that is, it leaves the system, for example, adheres to the surface or reacts with the boundary surface. Hence the probability to find the particle at the boundary is equal to zero, and the boundary condition for the PDF is

$$p(\mathbf{Z}, t|\mathbf{Y}, t') = 0 \quad \text{at} \quad \mathbf{Z} \in S. \tag{1.139}$$

b) Reflecting boundary

If the particle cannot leave the region R, then the probability flux in the $\boldsymbol{n}$ direction at the boundary surface should be equal to zero, that is,

$$\boldsymbol{n}\cdot\boldsymbol{J}(\boldsymbol{Z},t)=0 \quad \text{at} \quad \boldsymbol{Z}\in S, \tag{1.140}$$

where $\boldsymbol{n}$ is the normal to the boundary.

c) Surface of discontinuity

Suppose that the coefficients A_i and D_{ij} experience a jump at the surface S, but particles can cross the surface freely. Such a behavior is possible, when the surface is an interface between two media with different properties. At such a surface, the probabilities and the normal components of probability fluxes should be equal at both sides of the boundary surface:

$$p(\boldsymbol{Z},t|\boldsymbol{Y},t')|_{S+}=p(\boldsymbol{Z},t|\boldsymbol{Y},t')|_{S-}, \quad \boldsymbol{n}\cdot\boldsymbol{J}(Z,t)|_{S+}=\boldsymbol{n}\cdot\boldsymbol{J}(\boldsymbol{Z},t)|_{S-}. \tag{1.141}$$

d) Conditions at infinity

If the process is considered in an infinite region, then, depending on the problem under consideration, one of the following two boundary conditions must be valid:

$$p(\boldsymbol{Z},t)\to 0 \quad \text{or} \quad 1 \quad \text{and} \quad \frac{\partial p(\boldsymbol{Z},t)}{\partial Z}\to 0 \quad \text{at} \quad |\boldsymbol{Z}|\to\infty. \tag{1.142}$$

Of special interest is the one-dimensional case of the Fokker–Planck equation, in which the drift and dispersion coefficients are become scalar quantities A and D:

$$\frac{\partial p(Z,t|Y,t')}{\partial t}=-\frac{\partial}{\partial Z}[A(Z,t)\,p(Z,t|Y,t')]+\frac{1}{2}\frac{\partial^2}{\partial Z^2}[D(Z,t)\,p(Z,t|Y,t')].$$

This equation can be rewritten as

$$\frac{\partial p(Z,t|Y,t')}{\partial t}=\frac{\partial}{\partial Z}\left(\frac{1}{2}D(Z,t)\frac{\partial}{\partial Z}p(Z,t|Y,t')\right)$$
$$-\frac{\partial}{\partial Z}\left(\left(A(Z,t)-\frac{1}{2}\frac{\partial D}{\partial Z}\right)p(Z,t|Y,t')\right). \tag{1.143}$$

Now it is possible to compare it with the molecular diffusion equation, which describes the change of concentration C of a substance in the solution due to the thermal motion of solvent molecules:

$$\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{t}} = \frac{\partial(\boldsymbol{vC})}{\partial \boldsymbol{Z}} + \frac{\partial}{\partial \boldsymbol{Z}}\left(\boldsymbol{D}\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{Z}}\right), \tag{1.144}$$

where v is velocity of the substance under the action of external force. A comparison of equations (1.143) and (1.144) shows that $D/2$ has the meaning of diffusion coefficient D and the drift A

$$A = v + \frac{1}{2}\frac{\partial D}{\partial Z}$$

has the meaning of average velocity of particle displacement. The latter consists of two terms. The first term is the drift caused by external forces, and the second term is the drift caused by the inhomogeneneity of the medium. Another difference between equations (1.143) and (1.144) is the difference between the unknown variables: probability density p in (1.143) and concentration C in (1.144). But concentration can be obtained from the probability density by multiplying p by the number of particles N in a unit volume. Therefore if C/N is taken instead of C, then we can also take C/N instead of p under the condition that $N = const$.

A process described by the one-dimensional Fokker–Planck equation with $A = 0$ and $D = 1$,

$$\frac{\partial p(Z,t|Y,t')}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial Z^2}[p(Z,t|Y,t')]. \tag{1.145}$$

is called the Wiener process. Under the initial condition

$$p(\boldsymbol{Z},t|\boldsymbol{Y},t) = \delta(\boldsymbol{Y}-\boldsymbol{Z}) \quad \text{at} \quad t = t'$$

its solution is

$$p(Z,t|Y,t') = \frac{1}{(2\pi)^{1/2}}\exp\left\{-\frac{(Z-Y)^2}{2(t-t')}\right\}. \tag{1.146}$$

A multidimensional Wiener process is described by the multidimensional Fokker–Planck equation

$$\frac{\partial p(\boldsymbol{Z},t|\boldsymbol{Y},t')}{\partial t} = \frac{1}{2}\sum_i \frac{\partial^2}{\partial Z_i^2}[p(\boldsymbol{Z},t|\boldsymbol{Y},t')], \tag{1.147}$$

whose solution is

$$p(\boldsymbol{Z},t|\boldsymbol{Y},t') = \frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{(Z-Y)^2}{2(t-t')}\right\}. \tag{1.148}$$

Sometimes the Wiener process is causally referred to as ''Brownian motion''.

1.12.4
Deterministic Processes. The Liouville Equation

The Liouville equation follows from the Chapman–Kolmogorov equation at $W(\boldsymbol{Z}|\boldsymbol{Y}, t) = 0$ and $D_{ij} = 0$:

$$\frac{\partial p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')}{\partial t} = -\sum_i \frac{\partial}{\partial Z_i}[A_i(\boldsymbol{Z}, t)\, p(\boldsymbol{Z}, t|\boldsymbol{Y}, t')]. \tag{1.149}$$

Eq. (1.149) describes deterministic motion, which is given by a single-valued function and is determined by the initial conditions. Indeed, consider a trajectory $\boldsymbol{X}(t)$ that is a solution of the characteristic equation

$$\frac{d\boldsymbol{X}(t)}{dt} = \boldsymbol{A}(\boldsymbol{X}(t), t)$$

with the initial condition $\boldsymbol{X}(\boldsymbol{Y}, t') = \boldsymbol{Y}$. Here $\boldsymbol{A}$ is a vector with components A_i. Let us show that

$$p(\boldsymbol{Z}, t|\boldsymbol{Y}, t') = \delta(\boldsymbol{Z} - \boldsymbol{X}(\boldsymbol{Y}, t))$$

is the solution of Eq. (1.149) with the initial condition

$$p(\boldsymbol{Z}, t'|\boldsymbol{X}, t') = (\boldsymbol{Z} - \boldsymbol{X}).$$

Substituting it into Eq. (1.149), one obtains:

$$\begin{aligned}
&\frac{\partial \delta(\boldsymbol{Z} - \boldsymbol{X}(\boldsymbol{Y}, t))}{\partial t} \\
&= -\sum_i \left\{ \frac{\partial \delta(\boldsymbol{Z} - \boldsymbol{X}(\boldsymbol{Y}, t))}{\partial Z_i} \frac{dX_i(\boldsymbol{Y}, t)}{dt} \right\} = -\sum_i \left\{ \frac{\partial \delta(\boldsymbol{Z} - \boldsymbol{X}(\boldsymbol{Y}, t))}{\partial Z_i} A_i(\boldsymbol{X}(\boldsymbol{Y}, t), t) \right\} \\
&= -\sum_i \frac{\partial}{\partial Z_i}\{A_i(\boldsymbol{X}(\boldsymbol{Y}, t), t)\delta(\boldsymbol{Z} - \boldsymbol{X}(\boldsymbol{Y}, t))\} = -\sum_i \frac{\partial}{\partial Z_i}\{A_i(\boldsymbol{Z}, t)\delta(\boldsymbol{Z} - \boldsymbol{X}(\boldsymbol{Y}, t))\}.
\end{aligned}$$

The example above considers the motion of a single particle. In the case of many particles, the Liouville equation has a somewhat different form. Statistical Mechanics uses the Liouville equation to describe the motion of a particle ensemble as a set of mass points moving in accordance with Newtons second law,

$$\ddot{\boldsymbol{X}}_i = \frac{d^2\boldsymbol{X}}{dt^2} = \boldsymbol{F}_i \quad \text{or} \quad \dot{\boldsymbol{V}}_i = \frac{d\boldsymbol{V}_i}{dt} = \boldsymbol{F}_i; \quad \dot{\boldsymbol{X}}_i = \frac{d\boldsymbol{X}_i}{dt} = \boldsymbol{V}_i, \tag{1.150}$$

where $\boldsymbol{X}_i$, $\boldsymbol{V}_i$, $\boldsymbol{F}_i$ are, respectively, the radius vector, the velocity, and the force exerted on a unit mass of i-th particle by other particles and the environment. Let

$$\boldsymbol{X}_i(0) = \boldsymbol{X}_i^0, \quad \boldsymbol{V}_i(0) = \boldsymbol{V}_i^0. \tag{1.151}$$

be the initial positions and velocities of the points. Then in order to describe the time evolution of the state of an N-particle system, one has to integrate the system of equations (1.150) with the initial conditions (1.151). For $N \gg 1$, this is practically impossible, so one has to use statistical methods. Statistical Mechanics usually studies the dynamics of a mass point system by introducing generalized coordinates, which are either the ordinary coordinates $\boldsymbol{X}_i$ and velocities $\boldsymbol{V}_i$ or the ordinary coordinates $\boldsymbol{X}_i$ and momenta $m_i\boldsymbol{V}_i$ of particles.

The state of an N-particle system is characterized by the joint probability density taken at one given instant of time, which determines the chance to find particle 1 in the generalized coordinate interval $(\boldsymbol{X}_1 + d\boldsymbol{X}_1,\ \boldsymbol{V}_1 + d\boldsymbol{V}_1)$, AND to find particle 2 in the interval $(\boldsymbol{X}_2 + d\boldsymbol{X}_2,\ \boldsymbol{V}_2 + d\boldsymbol{V}_2)$, ..., and to find particle N in the interval $(\boldsymbol{X}_N + d\boldsymbol{X}_N,\ \boldsymbol{V}_N + d\boldsymbol{V}_N)$. If particle trajectories are known, that is, the functions

$$\boldsymbol{X}_i = \boldsymbol{X}_i(t), \quad \boldsymbol{V}_i = \boldsymbol{V}_i(t)$$

are given, then the probability density is equal to zero if $\boldsymbol{X}_i \neq \boldsymbol{X}_i(t)$ or $\boldsymbol{V}_i \neq \boldsymbol{V}_i(t)$ for even one value of i. It means that probability reduces to certainty and the probability density is given by a product of delta functions:

$$p(\boldsymbol{X}, \boldsymbol{V}, t) = \prod_{i=1}^{N} \delta(\boldsymbol{X}_i - \boldsymbol{X}_i(t))\delta(\boldsymbol{V}_i - \boldsymbol{V}_i(t)). \tag{1.152}$$

It is easy to show that this PDF satisfies the Liouville equation. Taking the derivative of the product of functions and using properties (1.10), (1.18), (1.21) of the delta function, we write:

$$\begin{aligned}
\frac{\partial p(\boldsymbol{X}, \boldsymbol{V}, t)}{\partial t}
&= -\sum_{j=1}^{N} \left(\prod_{\substack{k=1 \\ k \neq j}}^{N} \delta(\boldsymbol{X}_k - \boldsymbol{X}_k(t))\delta(\boldsymbol{V}_k - \dot{\boldsymbol{X}}_k(t)) \right) \dot{\boldsymbol{X}}_j \\
&\quad \times \frac{\partial \delta(\boldsymbol{X}_j - \boldsymbol{X}_j(t))}{\partial \boldsymbol{X}_j} \delta(\boldsymbol{V}_j - \dot{\boldsymbol{X}}_j(t)) - \sum_{j=1}^{N} \left(\prod_{\substack{k=1 \\ k \neq j}}^{N} \delta(\boldsymbol{X}_k - \boldsymbol{X}_k(t))\delta(\boldsymbol{V}_k - \dot{\boldsymbol{X}}_k(t)) \right) \\
&\quad \times \delta(\boldsymbol{X}_j - \boldsymbol{X}_j(t)) \ddot{\boldsymbol{X}}_j \cdot \frac{\partial \delta(\boldsymbol{V}_j - \dot{\boldsymbol{X}}_j(t))}{\partial \boldsymbol{V}_j}.
\end{aligned}$$

Now, using the relation (1.14) and the first equation (1.150), we get

$$\frac{\partial p(\boldsymbol{X},\boldsymbol{V},t)}{\partial t} = -\sum_{j=1}^{N}\boldsymbol{V}_j\cdot\frac{\partial\delta(\boldsymbol{X}_j-\boldsymbol{X}_j(t))}{\partial\boldsymbol{X}_j}\delta(\boldsymbol{V}_k-\dot{\boldsymbol{X}}_k(t))$$

$$\times\left(\prod_{\substack{k=1\\k\neq j}}^{N}\delta(\boldsymbol{X}_k-\boldsymbol{X}_k(t))\delta(\boldsymbol{V}_k-\dot{\boldsymbol{X}}_k(t))\right)-\sum_{j=1}^{N}\boldsymbol{F}_j\cdot\frac{\partial\delta(\boldsymbol{V}_j-\dot{\boldsymbol{X}}_j(t))}{\partial\boldsymbol{V}_j}$$

$$\times\delta(\boldsymbol{X}_k-\boldsymbol{X}_k(t))\left(\prod_{\substack{k=1\\k\neq j}}^{N}\delta(\boldsymbol{X}_k-\boldsymbol{X}_k(t))\delta(\boldsymbol{V}_k-\dot{\boldsymbol{X}}_k(t))\right).$$

It is easy to verify that

$$\frac{\partial p(\boldsymbol{X},\boldsymbol{V},t)}{\partial\boldsymbol{X}_j} = \frac{\partial\delta(\boldsymbol{X}_j-\boldsymbol{X}_j(t))}{\partial\boldsymbol{X}_j}\delta(\boldsymbol{V}_k-\dot{\boldsymbol{X}}_k(t))\times\left(\prod_{\substack{k=1\\k\neq j}}^{N}\delta(\boldsymbol{X}_k-\boldsymbol{X}_k(t))\delta(\boldsymbol{V}_k-\dot{\boldsymbol{X}}_k(t))\right),$$

$$\frac{\partial p(\boldsymbol{X},\boldsymbol{V},t)}{\partial\boldsymbol{V}_j} = \frac{\partial\delta(\boldsymbol{V}_j-\dot{\boldsymbol{X}}_j(t))}{\partial\boldsymbol{V}_j}\times\delta(\boldsymbol{X}_k-\boldsymbol{X}_k(t))\times\left(\prod_{\substack{k=1\\k\neq j}}^{N}\delta(\boldsymbol{X}_k-\boldsymbol{X}_k(t))\delta(\boldsymbol{V}_k-\dot{\boldsymbol{X}}_k(t))\right)$$

The Liouville equation finally reduces to

$$\frac{\partial p(\boldsymbol{X},\boldsymbol{V},t)}{\partial t} = -\sum_{i=1}^{N}\boldsymbol{V}_i\cdot\frac{\partial p(\boldsymbol{X},\boldsymbol{V},t)}{\partial\boldsymbol{X}_i}-\sum_{i=1}^{N}\boldsymbol{F}_i\cdot\frac{\partial p(\boldsymbol{X},\boldsymbol{V},t)}{\partial\boldsymbol{V}_i}. \tag{1.153}$$

If the generalized coordinates $\boldsymbol{Z}_i$ are defined as coordinates $\boldsymbol{X}_i$ and momenta per unit particle mass $\boldsymbol{V}_i$, then Eq. (1.153) will assume the following compact form:

$$\frac{\partial p(\boldsymbol{Z},t)}{\partial t} = -\sum_{i=1}^{N}\boldsymbol{A}_i\cdot\frac{\partial p(\boldsymbol{Z},t)}{\partial\boldsymbol{Z}_i}, \tag{1.154}$$

where $\boldsymbol{A}$ is the generalized vector that includes $\boldsymbol{V}_i$ and $\boldsymbol{F}_i$.

1.13
Stochastic Differential Equations. The Langevin Equation

When a randomly and rapidly fluctuating function of time and (or) spatial coordinates appears in a differential equation, this equation is called stochastic. The presence of a random component in the equation means that the solution will also be a random function. One example of a stochastic differential equation is the Langevin equation describing random trajectories of a particle driven by a random force. Another example is the equation of diffusion, which takes into account chemical reactions that are responsible for fluctuations. Let us consider these equations in more detail.

1.13.1
The Langevin Equation

In the theory of Brownian motion one frequently encounters the following Langevin equation:

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{a}(\boldsymbol{X}, t) + b(\boldsymbol{X}, t)\boldsymbol{\xi}(t), \qquad (1.155)$$

where $\boldsymbol{a}(\boldsymbol{X}, t)$ and $b(\boldsymbol{X}, t)$ are known functions and $\boldsymbol{\xi}(t)$ is a random fluctuating function.

The problem of Brownian motion of a particle driven by a fluctuating external force (that is, the force resulting from collisions with molecules of the surrounding fluid) reduces to this equation. Every second, the particle experiences millions of collisions, each collision resulting in a microscopic motion in the direction of impact. Therefore the particles position at the time t is determined only by its position at previous instant of time and does not depend on the motion prehistory. The particles motion can be considered as a Markowian process that is taking place under the action of a random force $\boldsymbol{\xi}(t)$ with the average value of zero:

$$\langle \boldsymbol{\xi}(t) \rangle = 0 \qquad (1.156)$$

The force $\boldsymbol{\xi}(t)$ is a "pseudoforce" that models the real impact forces. There is no correlation between the forces at different instants of time, in other words, $\boldsymbol{\xi}(t)$ and $\boldsymbol{\xi}(t')$ are mutually independent if $t \neq t'$. Or, to use another term, they are delta-correlated:

$$\langle \boldsymbol{\xi}(t)\boldsymbol{\xi}(t') \rangle = \boldsymbol{\xi}\delta(t - t'). \qquad (1.157)$$

Equations (1.155) and (1.157) serve as idealizations of Brownian motion. The actual equation of motion of a particle is given by Newtons second law:

$$m\frac{d^2\boldsymbol{X}}{dt^2} = -h\frac{d\boldsymbol{X}}{dt} + \boldsymbol{F} + \boldsymbol{f}, \qquad (1.158)$$

where $-hd\boldsymbol{X}/dt$ is the drag force exerted on the particle by the surrounding fluid (see Example 4 for more details), $\boldsymbol{F}$ is a systematic force such as gravitational, centrifugal, or electrostatic force (in other words, an external force), and $\boldsymbol{f}$ is the stochastic force. The small size of the particle allows us to neglect its inertia in the first approximation, in other words, the left-hand side of Eq. (1.158) is set to zero. This immediately results in Eq. (1.155). The condition (1.157) is also an idealization, because at $t = t'$ it gives an infinitely large dispersion, which is an obvious impossibility. The assumption (1.157) is similar to the idealization inherent in the concept of white noise in electrical engineering. The delta function representation of the correlation function is a natural idealization that helps us make the transition from a small time scale to a large scale typical for Brownian motion. Note, however, that differential equations that include perturbating terms given by random delta-correlated functions must be treated carefully, because the usual calculation rules will not always apply.

1.13.2
The Diffusion Equation

A deterministic description of processes in continuous media requires the use of conservation equations (conservation of mass, momentum, energy, and so on). But in order to get concrete solutions, the system of conservation equations must be complemented by constitutive relations or equations. As examples of such relations, we can mention Ficks law for diffusion (mass transport) processes, Fouriers law for heat transport processes, and the Navier–Stokes law describing the hydromechanics of viscous fluids. You may ask how these equations account for fluctuations of the relevant quantities. We shall answer this question by taking the diffusion equation as an example.

According to Ficks law, the diffusive flux of the substance $\boldsymbol{j}(\boldsymbol{X}, t)$ is proportional to the concentration gradient:

$$\boldsymbol{j}(\boldsymbol{X}, t) = -D\nabla C(\boldsymbol{X}, t). \tag{1.159}$$

On the other hand, we have the equation of conservation of mass (the continuity equation):

$$\frac{\partial C}{\partial t} + \nabla \cdot \boldsymbol{j}(\boldsymbol{X}, t) = 0. \tag{1.160}$$

Substitution of $\boldsymbol{j}(\boldsymbol{X}, t)$ from (1.159) results in the standard diffusion equation:

$$\frac{\partial C}{\partial t} = \nabla \cdot [D\nabla C(\boldsymbol{X}, t)]. \tag{1.161}$$

Fluctuations could be taken into account by adding a fluctuating term to the right-hand side Eq. (1.159):

$$\boldsymbol{j}(\boldsymbol{X}, t) = -D\nabla C(\boldsymbol{X}, t) + \boldsymbol{j}^{fl}(\boldsymbol{X}, t). \tag{1.162}$$

Just as we did for the fluctuating term in the Langevin equation, we assume that this term has the following statistical properties:

$$\left\langle \boldsymbol{j}^{fl}(\boldsymbol{X},t)\right\rangle = 0, \left\langle \boldsymbol{j}_i^{fl}(\boldsymbol{X},t)j_k^{fl}(\boldsymbol{X}',\boldsymbol{t}')\right\rangle \\ = K(\boldsymbol{X},t)\delta_{jk}\delta(\boldsymbol{X}-\boldsymbol{X}')\delta(t-t'). \tag{1.163}$$

The second property says that different components of the fluctuating flux vector $\boldsymbol{j}^{fl}$ taken at one and the same spatial point, as well as the values of one and the same component taken at different instants of time and/or at different points are assumed to be statistically independent. Or, to say it in fewer words, this property states that fluctuations have local behavior.

Eqs. (1.160) and (1.162) give us a diffusion equation where the additional term is expressed as the divergence of a vector:

$$\frac{\partial C}{\partial t} = \nabla\cdot[D\nabla C(\boldsymbol{X},t)] - \nabla\cdot\boldsymbol{j}^{fl}(\boldsymbol{X},t) \tag{1.164}$$

It can be shown that this term possesses the following statistical properties:

$$\left\langle \nabla\cdot\boldsymbol{j}^{fl}(\boldsymbol{X},t)\right\rangle = 0, \\ \left\langle \nabla\cdot\boldsymbol{j}^{fl}(\boldsymbol{X},t)\nabla'\cdot\boldsymbol{j}^{fl}(\boldsymbol{X}',t')\right\rangle = \nabla\cdot\nabla'[K_1(\boldsymbol{X},t)\delta(\boldsymbol{X}-\boldsymbol{X}')]\delta(t-t'). \tag{1.165}$$

1.13.2.1 *The Diffusion Equation with Chemical Reactions Taken into Account*

The deterministic diffusion equation with a source/sink term arising due to chemical reactions has the form

$$\frac{\partial C}{\partial t} + \nabla\cdot\boldsymbol{j}(\boldsymbol{X},t) = F[C(\boldsymbol{X},t)], \tag{1.166}$$

where the rate of substance production/consumption in the course of the chemical reaction appears in the right-hand side. This term usually depends on the substance concentrationl; depending on the reaction kinetics, it could be a linear or a nonlinear function of concentration. Production/consumption of matter gives rise to fluctuations, which can be taken into consideration by adding a fluctuating term $g^{fl}(\boldsymbol{X}, t)$ to the systematic term $F[C(\boldsymbol{X}, t)]$. The new term must satisfy the following statistical conditions:

$$\left\langle g^{fl}(\boldsymbol{X},t)\right\rangle = 0, \left\langle g^{fl}(\boldsymbol{X},t)g^{fl}(\boldsymbol{X}',t')\right\rangle \\ = K_2(\boldsymbol{X},t)\delta(\boldsymbol{X}-\boldsymbol{X}')\delta(t-t'). \tag{1.167}$$

The second condition expresses locality (lack of correlation between fluctuations at different points) as well as the Markovian character of chemical reactions. Eqs. (1.162), (1.166), and (1.167) give us the stochastic diffusion equation that accounts for chemical reactions:

$$\frac{\partial C}{\partial t} + \nabla \cdot \boldsymbol{j}(\boldsymbol{X}, t) = F[C(\boldsymbol{X}, t)] + G^{fl}(\boldsymbol{X}, t), \tag{1.168}$$

where

$$G^{fl}(\boldsymbol{X}, t) = -\nabla \cdot \boldsymbol{j}^{fl}(\boldsymbol{X}, t) + g^{fl}(\boldsymbol{X}, t).$$

The fluctuating term has the following properties:

$$\begin{aligned}&\left\langle G^{fl}(\boldsymbol{X}, t) G^{fl}(\boldsymbol{X}, t)\right\rangle = 0; \quad \left\langle G^{fl}(\boldsymbol{X}, t) G^{fl}(\boldsymbol{X}', t')\right\rangle \\ &= \{K_2(\boldsymbol{X}-\boldsymbol{X}')\delta(\boldsymbol{X}-\boldsymbol{X}') + \nabla \cdot \nabla'[K_1(\boldsymbol{X}, t)\delta(\boldsymbol{X}-\boldsymbol{X}')]\}\delta(t-t').\end{aligned}$$

Eqs. (1.164) and (1.168) are both Langevin equations. The main shortcoming of these equations is that the functions $K_1(\boldsymbol{X}, t)$ and $K_2(\boldsymbol{X}, t)$ are not known in advance. Information about these functions can be obtained by studying the process on the microscopic level, using the same approach that helped us derive the Chapman–Kolmogorov equation. This approach allows to interpret diffusion coefficients as components of a dispersion tensor or a correlation matrix. The difference between equations (1.168) and (1.164) is that $F[C(\boldsymbol{X}, t)]$ that appears in Eq. (1.168) is not a function but a functional.

1.13.2.2 *Brownian Motion of a Particle in a Hydrodynamic Medium*

Slow motion of a particle in a fluctuating hydrodynamic medium (the medium is assumed to be at rest at the infinity) under the action of the average viscous force exerted by the surrounding fluid $\langle \boldsymbol{F}(t) \rangle$ and the fluctuating force $\boldsymbol{F}^{fl}(t)$ (a force induced by thermal hydrodynamic fluctuations in the fluid or by some other source of random forces) is described by the Langevin equation (a stochastic analogue of Newtons second law):

$$m\frac{d\boldsymbol{U}}{dt} = -\langle \boldsymbol{F}(t) \rangle + \boldsymbol{F}^{fl}, \quad \boldsymbol{U} = \frac{d\boldsymbol{X}}{dt}. \tag{1.169}$$

To find the statistical properties of the random force $\boldsymbol{F}^{fl}$, it is necessary to use the hydrodynamic equations describing the fields of velocity $\boldsymbol{u}$ and pressure p in the fluid. The particle motion occurs at low Reynolds numbers, so one can use Navier–Stokes equations in the inertialess approximation, (i.e., Stokes equations):

$$\nabla \cdot \boldsymbol{u} = 0; \quad \nabla \cdot \boldsymbol{T} = 0, \tag{1.170}$$

where $\boldsymbol{T}(\boldsymbol{r}, t) = \boldsymbol{T}^s + \boldsymbol{T}^{fl}$ is the stress tensor in the fluid, which consists of a systematic component $\boldsymbol{T}^s = -p\boldsymbol{I} + 2\mu_e \boldsymbol{E}$ and a fluctuating component $\boldsymbol{T}^{fl}$; $-p\boldsymbol{I}$ is the spherical tensor (the isotropic part of the stress tensor); $2\mu_e\boldsymbol{E}$ is the deviator (the deviatoric part of the stress tensor); $\boldsymbol{E} = 0.5(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T)$ is the symmetric rate-of-strain tensor, μ_e is the coefficient of dynamic viscosity of the carrier (external) fluid.

In a statistically homogeneous medium, the fluctuating stress tensor $\boldsymbol{T}^{fl}$ is symmetric and its trace is equal to zero, that is, $T_{ii}^{fl} = 0$. The components of $\boldsymbol{T}^{fl}$ are usually assumed to have a Gaussian distribution with the average value $\left\langle T_{ik}^{fl} \right\rangle = 0$ and the correlation matrix

$$\left\langle \boldsymbol{T}_{ik}^{fl}(\boldsymbol{r}_1, \boldsymbol{t}_1)\boldsymbol{T}_{lm}^{fl}(\boldsymbol{r}_2, t_2) \right\rangle = 2k\theta\mu_e(\delta_{il}\delta_{km} + \delta_{im}\delta_{kl} - \frac{2}{3}\delta_{ik}\delta_{lm}) \times \delta(\boldsymbol{r}_1 - \boldsymbol{r}_2)\delta(t_1 - t_2). \tag{1.171}$$

Equations (1.170) can then be rewritten in the coordinate form:

$$\frac{\partial u_i}{\partial X_i} = 0; \quad \frac{\partial T_{ij}^s}{\partial X_j} = -\frac{\partial T_{ij}^{fl}}{\partial X_j}. \tag{1.172}$$

Following the method of small perturbations, we shall be looking for the solutions in the form

$$u_i = \langle u_i \rangle + \tilde{u}_i; \quad T_{ij}^s = \left\langle T_{ij}^s \right\rangle + \tilde{T}_{ij}^s, \tag{1.173}$$

where $\langle \boldsymbol{u} \rangle$, $\langle \boldsymbol{T}^s \rangle$ are the average values and $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{T}}^s$ are small fluctuating "additions" (perturbations).

Substituting (1.173) into (1.172) and neglecting small terms of higher orders, we get the equations for the average values and for the fluctuating terms:

$$\frac{\partial \langle u_i \rangle}{\partial X_i} = 0; \quad \frac{\partial \left\langle T_{ij}^s \right\rangle}{\partial X_j} = 0, \quad \frac{\partial \tilde{u}_i}{\partial X_i} = 0; \quad \frac{\partial \tilde{T}_{ij}^s}{\partial X_j} = -\frac{\partial \tilde{T}_{ij}^{fl}}{\partial X_j}. \tag{1.174}$$

Now, we must introduce two boundary conditions. First, the relative velocity at the interface between the fluid and the solid particle should be zero: $\langle u_i \rangle = U_i$, $\tilde{u}_i = 0$. Secondly, the fluid must be at rest at the infinity: $\langle u_i \rangle = \tilde{u}_i = 0$.

As a rule, the main purpose of hydrodynamic calculations is to find the force the surrounding fluid exerts on a moving particle. This force has systematic and fluctuating parts $\boldsymbol{F}$ and $\boldsymbol{F}^{fl}$, whose components are

$$\langle F_i \rangle = \int_s \left\langle T_{ij}^s \right\rangle n_j ds; \quad F_i^{fl} = \int_s \tilde{T}_{ij}^s n_j ds, \tag{1.175}$$

where n_i are components of the outer normal vector and S is the surface of the particle.

Solving the first two equations (1.174), we get the relation between the particles velocity and the drag force:

$$\langle F_i \rangle = R_{ij} U_j, \tag{1.176}$$

where R_{ij} are components of the resistance tensor $\boldsymbol{R}$.

In order to determine statistical characteristics of the random force $\tilde{F}_i$, let us use the relation that follows in a self-obvious way from Gausss theorem, equations (1.174), and the above-mentioned boundary conditions:

$$\int_V \left(\langle u_i \rangle \frac{\partial \tilde{T}^s_{ij}}{\partial X_j} - \tilde{u}_i \frac{\partial \left\langle T^s_{ij} \right\rangle}{\partial X_j} \right) dV = \int_V \frac{\partial}{\partial X_j} (\langle u_i \rangle \tilde{T}^s_{ij} - \tilde{u}_i \left\langle T^s_{ij} \right\rangle) dV$$

$$= \int_s (\langle u_i \rangle \widetilde{T}^s_{ij} - \tilde{u}_i \left\langle \widetilde{T}^s_{ij} \right\rangle) n_j ds = - \int_s \langle u_i \rangle \frac{\partial \widetilde{T}^s_{ij}}{\partial X_j} ds = \int_s U_i \tilde{T}^s_{ij} n_j ds = U_i \tilde{F}_i,$$

where integration is carried out over the volume V occupied by the fluid. Using this relation together with the property (1.171), we finally get the two-time correlation between components of the fluctuating force:

$$\left\langle \tilde{F}_i(t) \tilde{F}_j(t') \right\rangle = \frac{1}{U_i U_j} 2kT\delta(t-t') \int_s \langle u_i \rangle \langle T_{ij} \rangle n_j ds = 2kTR_{ij}\delta(t-t'). \qquad (1.177)$$

1.14
Variational (Functional) Derivatives

When considering random processes and random fields, one can see that the PDF depends on random variables, which, in turn, are functions of time and (or) spatial coordinates. Therefore, the variables in the PDF are random functions rather than random variables.

It was shown in Sections 1.5 and 1.6 that knowing the characteristic function φ and the characteristic functional, one can determine statistical parameters of random quantities such as moments and cumulants from the formulas (1.50) and (1.54). These formulas contain derivatives of φ and $\ln\varphi$. When considering random processes and random fields, we treat these derivatives as derivatives with respect to functions and not as derivatives with respect to random variables. Hence, instead of ordinary differentiation we have to perform functional, or, to use another term, variational, differentiation. The corresponding derivatives are called functional (variational) derivatives.

Let us recal the general definition of a functional. We say that a functional is given when there exists a rule that assigns a definite number to each function belonging to some set of functions. Some examples are given below:

- *linear functional*

$$F[\varphi(X)] = \int_{X_1}^{X_2} a(X)\varphi(X)dX,$$

where $a(X)$ is a given function and the limits X_1 and X_2 can be either finite or infinite;

- *quadratic functional*

$$F[\varphi(X)] = \int_{X_1}^{X_2}\int_{X_1}^{X_2} B(\tau_1, \tau_2)\varphi(\tau_1)\varphi(\tau_2)d\tau_1 d\tau_2,$$

where $B(\tau_1, \tau_2)$ is a given function;

- *function of functional*

$$F[\varphi(X)] = f(\Phi[\varphi])$$

where $f(X)$ is a given function and $\Phi[\varphi(X)]$ is a functional.

Consider the difference in values of one and the same functional taken for two functions $\varphi(\tau)$ and $\varphi(\tau)+\delta\varphi(\tau)$, where t lies in the interval $\tau \in (X-0.5\Delta X;\ X+0.5\Delta X)$. The difference (more strictly, the linear (with respect to $\delta\varphi(\tau)$) part of that difference) is called variation of the functional:

$$\delta F[\varphi] = \{F[\varphi+\delta\varphi] - F[\varphi]\}.$$

The variational (functional) derivative is defined as

$$\frac{\delta F[\varphi]}{\delta\varphi(X)} = \lim_{\Delta X \to 0} \frac{\delta F[\varphi]}{\int_{\Delta X} \varphi(\tau)d\tau}. \tag{1.178}$$

The variational derivative of a functional $F[\varphi]$ is itself a functional of $\varphi(\tau)$ that also depends on the point X as a parameter. Thus the variational derivative itself has two different derivatives. One can differentiate it in the ordinary way with respect to the parameter X; or one can take the variational derivative with respect to $\varphi(\tau)$ at the point $\tau = \tilde{X}$. The latter would be the second variational derivative of the initial functional $F[\varphi]$:

$$\frac{\delta}{\delta\varphi(\tilde{X})}\left[\frac{\delta F[\varphi]}{\delta\varphi(X)}\right] = \frac{\delta^2 F[\varphi]}{\delta\varphi(\tilde{X})\delta\varphi(X)}. \tag{1.179}$$

The second variational derivative is also a functional of $\varphi(\tau)$, but, in contrast to the first variational derivative, it now depends on two points: X and $\tilde{X}$. Variational derivatives of higher orders can be defined in a similar way. As examples, consider variational derivatives of the above-mentioned functionals.

$$\delta F = F[\varphi+\delta\varphi] - F[\varphi] = \int_{X_1}^{X_2} a(\tau)\delta\varphi(\tau)d\tau = \int_{X-0.5\Delta X}^{X+0.5\Delta X} a(\tau)\delta\varphi(\tau)d\tau.$$

If the function $a(\tau)$ is continuous on the interval ΔX, then, according to the mean value theorem,

$$\delta F[\varphi] = a(X') \int_{\Delta X} \delta\varphi(\tau)d\tau,$$

where $X' \in (X - 0.5\Delta X;\ X + 0.5\Delta X)$. The definition (1.178) yields

$$\frac{\delta F[\varphi]}{\delta\varphi(X)} = \lim_{\Delta X\to 0} a(X') \left(\int_{\Delta X} \delta\varphi(\tau)d\tau \Big/ \int_{\Delta X} \delta\varphi(\tau)d\tau \right) = a(X). \tag{1.180}$$

Applying the same approach to the quadratic functional, we get

$$\frac{\delta F[\varphi]}{\delta\varphi(X)} = \int_{X_1}^{X_2} (B(\tau, X) + B(X, \tau))\varphi(\tau)d\tau; \quad (X_1 < \tau < X_2). \tag{1.181}$$

Note that in many cases the function $B(\tau, X)$ is a symmetric function of its arguments, that is, $B(\tau, X) = B(X, \tau)$.

Finally, for a function of a functional,

$$\begin{aligned} F[\varphi + \delta\varphi] &= f(\Phi[\varphi + \delta\varphi]) = f(\Phi[\varphi] + \delta\Phi) \\ &= f(\Phi[\varphi]) + \frac{\partial f(\Phi[\varphi])}{\partial\Phi}\delta\Phi + \ldots = F[\varphi] + \frac{\partial f(\Phi[\varphi])}{\partial\Phi}\delta\Phi + \ldots \end{aligned}$$

and

$$\frac{\delta}{\delta\varphi(X)} f(\Phi[\varphi]) = \frac{\partial f(\Phi[\varphi])}{\partial\Phi} \frac{\delta\Phi[\varphi]}{\delta\varphi(X)}. \tag{1.182}$$

Consider now some properties of variational derivatives. Let a functional be a product of two functionals: $\Phi[\varphi] = F_1[\varphi]F_2[\varphi]$. The variation and variational derivative of this functional are:

$$\begin{aligned} \delta\Phi = \Phi[\varphi + \delta\varphi] - \Phi[\varphi] &= F_1[\varphi + \delta\varphi]F_2[\varphi + \delta\varphi] - F_1[\varphi]F_2[\varphi] \\ &= F_1[\varphi]\delta F_2[\varphi] + F_2[\varphi]\delta F_1[\varphi], \end{aligned}$$

$$\frac{\delta}{\delta\varphi(X)}[F_1[\varphi]F_2[\varphi]] = F_1[\varphi]\frac{\delta F_2[\varphi]}{\delta\varphi(X)} + F_2[\varphi]\frac{\delta F_1[\varphi]}{\delta\varphi(X)}. \tag{1.183}$$

Of special interest is the functional of a Gaussian distribution,

$$F[\varphi] = \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(\tau - \tau_0)^2}{2\sigma^2} \right\} \varphi(\tau)d\tau. \tag{1.184}$$

Eq. (1.180) gives us its variatlional derivative:

$$\frac{\delta F[\varphi]}{\delta\varphi(\tau)} = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\tau-\tau_0)^2}{2\sigma^2}\right\}. \tag{1.185}$$

Going to the limit $\sigma \to 0$ in Eqs. (1.183), (1.184) and using the definition of the delta function Eq. (1.9) and its property (1.11), we obtain:

$$\lim_{\sigma\to 0} F[\varphi] = \varphi(\tau_0), \quad \lim_{\sigma\to 0} = \frac{\delta F[\varphi]}{\delta\varphi(\tau)} = \delta(\tau-\tau_0).$$

Therefore,

$$\frac{\delta\varphi[\tau_0]}{\delta\varphi(\tau)} = \delta(\tau-\tau_0). \tag{1.186}$$

The relation (1.186) facilitates the derivation of formulas for some variational derivatives. As an example, let us derive the formula (1.181) for the quadratic functional:

$$\begin{aligned}
&\frac{\delta}{\delta\varphi(X)}\left\{\int_{X_1}^{X_2}\int_{X_1}^{X_2} B(\tau_1,\tau_2)\varphi(\tau_1)\varphi(\tau_2)d\tau_1 d\tau_2\right\} \\
&= \int_{X_1}^{X_2}\int_{X_1}^{X_2} B(\tau_1,\tau_2)\frac{\delta}{\delta\varphi(X)}[\varphi(\tau_1)\varphi(\tau_2)]d\tau_1 d\tau_2 \\
&= \int_{X_1}^{X_2}\int_{X_1}^{X_2} B(\tau_1,\tau_2)\left[\frac{\delta\varphi(\tau_1)}{\delta\varphi(X)}\varphi(\tau_2) + \varphi(\tau_1)\frac{\delta\varphi(\tau_2)}{\delta\varphi(X)}\right]d\tau_1 d\tau_2 \\
&= \int_{X_1}^{X_2}\int_{X_1}^{X_2} B(\tau_1,\tau_2)[\delta(\tau_1-X)\varphi(\tau_2) + \varphi(\tau_1)\delta(\tau_2-X)]d\tau_1 d\tau_2 \\
&= \int_{X_1}^{X_2}[B(\tau,X) + B(X,\tau)]\varphi(\tau)d\tau, \quad (X_1 < \tau < X_2).
\end{aligned}$$

Another example is the variational derivative of the functional

$$F[\varphi(\tau)] = \int_{X_1}^{X_2} L[(\tau,\varphi(\tau),\dot{\varphi}(\tau)]d\tau, \quad \dot{\varphi}(\tau) = \frac{d\varphi(\tau)}{d\tau}.$$

We have:

$$\frac{\delta F[\varphi]}{\delta\varphi(X)} = \int_{X_1}^{X_2} \left[\frac{\partial L}{\partial\varphi} + \frac{\partial L}{\partial\dot{\varphi}}\frac{d}{d\tau}\right] \frac{\delta\varphi(\tau_0)}{\delta\varphi(X)} d\tau = \int_{X_1}^{X_2} \left[\frac{\partial L}{\partial\varphi} + \frac{\partial L}{\partial\dot{\varphi}}\frac{d}{d\tau}\right] \delta(\tau - X) d\tau$$
$$= \left[-\frac{d}{dX}\frac{\partial}{\partial\dot{\varphi}} + \frac{\partial}{\partial\varphi}\right] L[(\tau, \varphi(\tau), \dot{\varphi}(\tau)], \quad X \in (X_1, X_2).$$

The functional $F[\varphi(\tau) + \eta(\tau)]$ can be expanded as a functional Taylor series in the vicinity of the point $\eta \sim 0$:

$$F[\varphi(\tau) + \eta(\tau)] = F[\varphi(\tau)] + \int \frac{\delta F[\varphi]}{\delta\varphi(X)} \eta(X) dX$$
$$+ \frac{1}{2!}\int\int \frac{\delta F[\varphi]}{\delta\varphi(X_1)\delta\varphi(X_2)} \eta(X_1)\eta(X_2) dX_1 dX_2 + \ldots. \tag{1.187}$$

Here and later, when the range of integration is not pointed out, it is assumed to be infinite.

The Taylor series (1.187) could be written in the compact form

$$F[\varphi(\tau) + \eta(\tau)] = \exp\left\{\int dX \eta(X) \frac{\delta}{\delta\varphi(X)}\right\} F[\varphi(\tau)]$$

using the following operator notation:

$$\exp\left\{\int dX \eta(X) \frac{\delta}{\delta\varphi(X)}\right\} = 1 + \int dX \eta(X) \frac{\delta}{\delta\varphi(X)}$$
$$+ \frac{1}{2!}\int\int dX_1 dX_2 \eta(X_1)\eta(X_2) \frac{\delta^2}{\delta\varphi(X_1)\delta\varphi(X_2)} + \ldots$$
$$= 1 + \int dX \eta(X) \frac{\delta}{\delta\varphi(X)} + \frac{1}{2!}\left\{\int dX \eta(X) \frac{\delta}{\delta\varphi(X)}\right\}^2 + \ldots.$$

Consider the transformation of variational derivatives as we change the functional variables. Let us replace the function $\varphi(t)$ by a new function $\psi(t)$ given by the equality $\varphi(t) = \Psi[\psi(t); t]$, where ψ is the functional of a function $\psi(t)$, which also depends on t. Then the functional $F[\varphi(\tau)]$ is a composite functional of $\psi(\tau)$:

$$F[\varphi(\tau)] = F[\Psi[\psi(\tau); \tau]] \equiv F_1[\psi(\tau)].$$

For such functional, there exists the following expression for the variational derivative:

$$\frac{\delta F_1[\psi(\tau)]}{\delta\varphi(t)} = \int dt' \frac{\delta F[\varphi(\tau)]}{\delta\varphi(t')} \frac{\delta\Psi[\psi(\tau); X']}{\delta\psi(t)}. \tag{1.188}$$

A functional change of variables plays an important role in Fourier transforms:

$$\varphi(X) = \int \psi(\omega) e^{i\omega X} d\omega = \varphi[\psi(\omega); X].$$

According to the formula (1.173), we have

$$\frac{\delta\varphi[\psi(\omega); X]}{\delta\psi(\omega')} = e^{i\omega' X}$$

and from (1.180), there follows

$$\frac{\delta F_1[\varphi[\psi(\omega); X]]}{\delta\varphi(\omega')} = \int \frac{\delta F[\varphi(X)]}{\delta\varphi(X')} e^{i\omega' X'} dX'. \tag{1.189}$$

1.15 The Characteristic Functional

It was shown in Section 1.5 that a random value u is completely determined by its characteristic function $\varphi(\rho) = \langle \exp(i\rho u) \rangle$, which allows us to use the inverse Fourier transform to get the PDF,

$$p(u) = \frac{1}{2\pi} \int e^{-i\rho u} d\rho$$

the moments,

$$B_n = \langle u^n \rangle = \left[\frac{1}{i} \frac{d}{d\rho} \right]^n \varphi(\rho)|_{\rho=0},$$

the cumulants,

$$S_n = \left[\frac{1}{i} \frac{d}{d\rho} \right]^n [\ln\varphi(\rho)]_{\rho=0}$$

and other statistical parameters of the PDF.

For a multidimensional random quantity $\boldsymbol{u} = (u_1, u_2, \ldots, u_N)$, the complete description is contained in the characteristic function

$$\varphi(\boldsymbol{\rho}) = \varphi(\rho_1, \rho_2, \ldots, \rho_N) = \langle \exp(i\boldsymbol{\rho}\cdot\mathbf{u}) \rangle = \left\langle \exp(i \sum_k \rho_k u_k) \right\rangle. \tag{1.190}$$

The corresponding joint PDF of the random values $u_1, u_2, \ldots, u_N$ is the Fourier transform of the characteristic function $\varphi(\rho_1, \rho_2, \ldots, \rho_N)$:

$$\begin{aligned} p(X_1, X_2, \ldots, X_N) &= \frac{1}{(2\pi)^N}\int e^{-i\boldsymbol{\rho}\cdot \boldsymbol{X}}\varphi(\boldsymbol{\rho})d\boldsymbol{\rho} \\ &= \frac{1}{(2\pi)^N}\int \exp(-i\sum_k \rho_k X_k)\varphi(\rho_1, \rho_2, \ldots, \rho_N)d\rho_1 d\rho_2\ldots d\rho_N \\ &= \delta(u_1 - X_1)\delta(u_2 - X_2)\ldots\delta(u_N - X_N). \end{aligned} \tag{1.191}$$

Consider now a random function $u(X)$. For its complete statistical description, it is sufficient to know the characteristic functional

$$\Phi[\rho] = \left\langle \exp\left\{ i\int \rho(\tau)u(\tau)d\tau \right\} \right\rangle, \tag{1.192}$$

where the function $\rho(\tau)$ is an arbitrary function replacing the set of numbers ρ_1, ρ_2, $\ldots$, ρ_N in (1.190).

Given $\Phi[\rho]$, we can find statistical characteristics of the random function $u(X)$, for example, the average value $\langle u(X)\rangle$, N-point moments $B_{uu\ldots u} = \langle u(X_1)\ \ldots\ u(X_N)\rangle$, etc.

To find the variational derivative, let us use the results obtained in Section 1.14, keeping in mind that the averaging operator commutes with the operator $\delta/\delta\rho(X)$:

$$\begin{aligned} \frac{\delta\Phi[\rho]}{\delta\rho(X)} = \frac{\delta}{\delta\varphi(X)}\exp\left\langle \left\{ i\int \rho(\tau)u(\tau)d\tau \right\} \right\rangle &= \left\langle \frac{\delta}{\delta\varphi(X)}\exp\left\{ i\int \rho(\tau)u(\tau)d\tau \right\} \right\rangle \\ &= i\left\langle u(X)\exp\left\{ i\int \rho(\tau)u(\tau)d\tau \right\} \right\rangle. \end{aligned}$$

Similarly, we write

$$\left(\frac{1}{i}\frac{\delta}{\delta\rho(X_1)}\right)\ldots\left(\frac{1}{i}\frac{\delta}{\delta\rho(X_n)}\right)\Phi = \left\langle u(X_1)\ldots u(X_n)\exp\left\{ i\int \rho(\tau)u(\tau)d\tau \right\} \right\rangle.$$

Setting $\rho = 0$ in the last relation, we obtain:

$$\frac{1}{i^n}\frac{\delta^n}{\delta\rho(X_1)\ldots\delta\rho(X_n)}\Phi\bigg|_{\rho=0} = \langle u(X_1)\ldots u(X_n)\rangle = B_{uu\ldots u}. \tag{1.193}$$

So, it is possible to find the multi-point moments for a given characteristic functional. Expanding the functional $\Phi[\rho]$ into a functional Taylor series according to Eq. (1.187) and taking into account Eq. (1.193), one can express the characteristic functional in terms of moments:

$$\begin{aligned} \Phi[\rho] = \sum_{n=0}^{\infty}\frac{i^n}{n!}\int\ldots\int & B_{uu\ldots u}(X_1, X_2, \ldots, X_n) \\ & \times \rho(X_1)\rho(X_2)\ldots\rho(X_n)dX_1 dX_2\ldots dX_n. \end{aligned} \tag{1.194}$$

If the functional has the form

$$\Phi[\rho] = \exp(\psi[\rho]),$$

then ψ is called the cumulant-generating function (see Section 1.6), and the cumulants themselves given by

$$S_{11\ldots1}(X_1, X_2, \ldots, X_n) = (-i)^k \frac{\delta^n}{\delta\rho(X_1)\ldots\delta\rho(X_n)}\bigg|_{\rho=0}. \tag{1.195}$$

As an example, let us find the characteristic functional for a Gaussian random process $u(X)$ with $\langle u(X)\rangle = 0$, under the additional assumption that the joint distribution for any two given values of X is also Gaussian. Let the random variable

$$\langle A\rangle = \int_{-\infty}^{\infty} \rho(\tau)u(\tau)d\tau, (\rho(\pm\infty) = 0,$$

have a Gaussian distribution

$$p(A) = \frac{1}{\sqrt{2\pi}\sigma_A} \exp\left\{-\frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right\}$$

with the parameters

$$A = \int_{-\infty}^{\infty} \rho(\tau)\langle u\rangle(\tau)d\tau = 0,$$

$$\sigma_A^2 = \langle A^2\rangle - (\langle A\rangle)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(\tau_1, \tau_2)\rho(\tau_1)\rho(\tau_2)d\tau_1 d\tau_2,$$

where $B(\tau_1, \tau_2) = \langle u(\tau_1)u(\tau_2)\rangle$ is a two-point correlation function. The characteristic functional of the random quantity A is

$$\langle \exp(e^{iA})\rangle = \exp\left(-\frac{1}{2}\sigma_A^2\right).$$

Therefore the characteristic functional of a Gaussian random process is

$$\Phi[\rho] = \exp\left\{-\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(\tau_1, \tau_2)\rho(\tau_1)\rho(\tau_2)d\tau_1 d\tau_2\right\}. \tag{1.196}$$

Characteristic functionals and variational derivatives are widely used in problems pertaining to particle motion under the action of random fluctuating forces, for example, in the theories of Brownian motion, diffusion, turbulence, etc. Such problems are the subject of the statistical theory of dynamic systems with fluctuating parameters. These parameters (e.g., coordinates and velocities of particles suspended in a fluid), are described by ordinary or partial differential equations (stochastic equations). The main challenge is to obtain and then solve a closed system of equations. It turns out that many processes can be treated as Markowian processes. Also, the distribution of fluctuating parameters (random variables) has proved to be Gaussian.

As an example, consider a simplified form of the Langevin equation (see Section 1.13) that describes the time rate of change of velocity $V(t)$ of a particle driven by a fluctuating force $\xi(t)$:

$$\frac{dV}{dt} = -hV + \xi(t); \quad X(0) = 0, \tag{1.197}$$

where h is a constant and $\xi(t)$ is a random function.

For a given $\xi(t)$, the solution of Eq. (1.197) has the form

$$V(t) = \int_0^i \xi(t)\exp(-h(t-\tau))d\tau. \tag{1.198}$$

From Eq. (1.198), one can derive all statistical characteristics of the random process $V(t)$. On the other hand, all statistical characteristics are contained in the characteristic function

$$\varphi(\rho) = (\exp(i\rho V(t)) = \left\langle \exp\left\{ i\rho\int_0^t \xi(t)\exp(-h(t-\tau))d\tau \right\} \right\rangle. \tag{1.199}$$

Taking into account the relation (1.192), the last equation can be rewritten as

$$\varphi(\rho) = \Phi\{\rho\exp(-h(t-\tau))\}. \tag{1.200}$$

Finally, we should mention the Furutsu–Donsker–Novikov correlation formula for the product of a Gaussian random function $X(t)$ and a functional $R[X(\tau)]$ that may depend (either explicitly or implicitly) on $X(\tau)$:

$$\langle X(t)R[X(\tau)]\rangle = \int_{-\infty}^{\infty} B(t,\tau_1)\left\langle \frac{\delta R[X(\tau)]}{\delta X(\tau_1)} \right\rangle d\tau_1. \tag{1.201}$$

This formula is the functional analogue of Eq. (1.86).

References

1 Gardiner, C.W. (1985) *Handbook of Stochastic Methods*, 2nd ed., Springer–Verlag.

2 Klyatskin, V.I. (1975) *Statistical Description of Dynamical Systems with Fluctuating Parameters*, Nauka, Moscow, (in Russian).

3 Klyatskin, V.I. (1980) *Stochastic Equations and Waves in Random Inhomogeneous Media*, Nauka, Moscow, (in Russian).

4 Landau, L.D. and Lifshitz E.M. (1964) *Statistical Physics*, Nauka, Moscow, (in Russian).

5 Leontovitch, M.A. (1983) *Introduction to Thermodynamics. Statistical Physics*, Nauka, Moscow, (in Russian).

6 Monin, A.C. and Yaglom, A.M. (1971) *Statistical Fluid Mechanics: Mechanics of Turbulence*, Vol. **1**, MIT Press, Cambridge, MA.

7 Monin, A.C. and Yaglom, A.M. (1975) *Statistical Fluid Mechanics: Mechanics of Turbulence*, Vol. **2**, MIT Press, Cambridge, MA.

7 Feller, W. (1974) *An Introduction to Probability Theory and Its Applications*, Vol. **I, II**, Wiley.

8 Chandrasekhar, S. (1943) Stochastic Problems in Physics and Astronomy, *Rev. Mod. Phys.*, **15**, 1–89.

9 Cercignani, C. (1969) *Mathematical Methods in Kinetic Theory*, Macmillan.

10 Einstein, A. and Smoluchowski, M. (1936) *Brownian Motion*, ONTI, Moscow, (in Russian).

11 Pope, S.B. (2000) *Turbulent Flows*, Cambridge Univ. Press.

12 Saffman, P.G. (1971) On the Boundary Condition at the Surface of a Porous Medium, *Studies in Appl. Math.*, **50** (2), 93–101.