Grundlehren der mathematischen Wissenschaften 260

A Series of Comprehensive Studies in Mathematics

For further volumes:
www.springer.com/series/138

Mark I. Freidlin • Alexander D. Wentzell

Random Perturbations of Dynamical Systems

Third Edition

Translated by Joseph Szücs

With 46 Illustrations

Springer

Mark I. Freidlin
Department of Mathematics
University of Maryland
College Park, MD
USA

Alexander D. Wentzell
Department of Mathematics
Tulane University
New Orleans, LA
USA

Translator
Joseph Szücs
Texas A&M University at Galveston
Galveston, TX
USA

ISSN 0072-7830
ISBN 978-3-642-25846-6 ISBN 978-3-642-25847-3 (eBook)
DOI 10.1007/978-3-642-25847-3

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012937993

Mathematics Subject Classification: 60F10, 34E10, 60H10, 60J60

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface to the Third Edition

Main innovations in this edition concern the averaging principle. A new section on deterministic perturbations of one-degree-of-freedom systems was added in Chap. 8. We show there that pure deterministic perturbations of an oscillator may lead to a stochastic, in a certain sense, long-time behavior of the system, if the corresponding Hamiltonian has saddle points. To give a rigorous meaning to this statement, one should, first, regularize the system by the addition of small random perturbations. It turns out that the stochasticity of long-time behavior is independent of the regularization. The stochasticity is an intrinsic property of the original system related to the instability of saddle points. This shows usefulness of a joint consideration of classical theory of deterministic perturbations together with stochastic perturbations.

We added a new Chap. 9 where deterministic and stochastic perturbations of systems with many degrees of freedom are considered. Because of the resonances, stochastic regularization in this case is even more important.

Small changes in the chapters where long-time behavior of the perturbed system is determined by large deviations were made. Most of these changes, actually, concern the terminology. In particular, we explained that the notion of sub-limiting distribution for a given initial point and a time scale is identical to the notion of metastability. We also explained that the stochastic resonance is a manifestation of metastability and the theory of this effect is a part of the large deviation theory. We also made some comments on the notion of quasi-potential which we introduced more than forty years ago. One should say that many of notions and results presented in this book became quite popular in applications, and many of them were later rediscovered in applied papers.

We also added references to recent papers where the proofs of some conjectures included in previous editions were obtained.

College Park, Maryland — M.I. Freidlin
New Orleans, Louisiana — A.D. Wentzell

Preface to the Second Edition

The first edition of this book was published in 1979 in Russian. Most of the material presented was related to large-deviation theory for stochastic processes. This theory was developed more or less at the same time by different authors in different countries. This book was the first monograph in which large-deviation theory for stochastic processes was presented. Since then a number of books specially dedicated to large-deviation theory have been published, including S. R. S. Varadhan [4], A. D. Wentzell [10], J.-D. Deuschel and D. W. Stroock [1], A. Dembo and O. Zeitouni [1]. Just a few changes were made for this edition in the part where large deviations are treated. The most essential is the addition of two new sections in the last chapter. Large deviations for infinite-dimensional systems are briefly considered in one new section, and the applications of large-deviation theory to wave front propagation for reaction-diffusion equations are considered in another one.

Large-deviation theory is not the only class of limit theorems arising in the context of random perturbations of dynamical systems. We therefore included in the second edition a number of new results related to the averaging principle. Random perturbations of classical dynamical systems under certain conditions lead to diffusion processes on graphs. Such problems are considered in the new Chap. 8. Some new results concerning fast oscillating perturbations of dynamical systems with conservation laws are included in Chap. 7. A few small additions and corrections were made in the other chapters as well. We would like to thank Ruth Pfeiffer and Fred Torcaso for their help in the preparation of the second edition of this book.

College Park, Maryland — M.I. Freidlin
New Orleans, Louisiana — A.D. Wentzell

Preface

Asymptotical problems have always played an important role in probability theory. In classical probability theory dealing mainly with sequences of independent variables, theorems of the type of laws of large numbers, theorems of the type of the central limit theorem, and theorems on large deviations constitute a major part of all investigations. In recent years, when random processes have become the main subject of study, asymptotic investigations have continued to play a major role. We can say that in the theory of random processes such investigations play an even greater role than in classical probability theory, because it is apparently impossible to obtain simple exact formulas in problems connected with large classes of random processes.

Asymptotical investigations in the theory of random processes include results of the types of both the laws of large numbers and the central limit theorem and, in the past decade, theorems on large deviations. Of course, all these problems have acquired new aspects and new interpretations in the theory of random processes.

One of the important schemes leading to the study of various limit theorems for random processes is dynamical systems subject to the effect of random perturbations. Several theoretical and applied problems lead to this scheme. It is often natural to assume that, in one sense or another, the random perturbations are small compared to the deterministic constituents of the motion. The problem of studying small random perturbations of dynamical systems has been posed in the paper by Pontrjagin, Andronov, and Vitt [1]. The results obtained in this article relate to one-dimensional and partly two-dimensional dynamical systems and perturbations leading to diffusion processes. Other types of random perturbations may also be considered; in particular, those arising in connection with the averaging principle. Here the smallness of the effect of perturbations is ensured by the fact that they oscillate quickly.

The contents of the book consists of various asymptotic problems arising as the parameter characterizing the smallness of random perturbations converges to zero. Of course, the authors could not consider all conceivable schemes of small random perturbations of dynamical systems. In particular, the book does not consider at all dynamical systems generated by random vector fields. Much attention is given to the study of the effect of perturbations on large time intervals. On such intervals small perturbations essentially influence the behavior of the system in general. In order to take account of this influence, we have to be able to estimate the probabilities of rare events, i.e., we need theorems on the asymptotics of probabilities of large deviations for random processes. The book studies these asymptotics and their applications to problems of the behavior of a random process on large time intervals, such as the problem of the limit behavior of the invariant measure, the problem of exit of a random process from a domain, and the problem of stability under random perturbations. Some of these problems have been formulated for a long time and others are comparatively new.

The problems being studied can be considered as problems of the asymptotic study of integrals in a function space, and the fundamental method used can be considered as an infinite-dimensional generalization of the well-known method of Laplace. These constructions are linked to contemporary research in asymptotic methods. In the cases where, as a result of the effect of perturbations, diffusion processes are obtained, we arrive at problems closely connected with elliptic and parabolic differential equations with a small parameter. Our investigations imply some new results concerning such equations. We are interested in these connections and as a rule include the corresponding formulations in terms of differential equations.

We would like to note that this book is being written when the theory of large deviations for random processes is just being created. There have been a series of achievements but there is still much to be done. Therefore, the book treats some topics that have not yet taken their final form (part of the material is presented in a survey form). At the same time, some new research is not reflected at all in the book. The authors attempted to minimize the deficiencies connected with this.

The book is written for mathematicians but can also be used by specialists of adjacent fields. The fact is that although the proofs use quite intricate mathematical constructions, the results admit a simple formulation as a rule.

Contents

Introduction

Let $b(x)$ be a continuous vector field in R^r. First we discuss nonrandom perturbations of a dynamical system

$$\dot{x}_t = b(x_t). \tag{1}$$

We may consider the perturbed system

$$\dot{X}_t = b(X_t, \psi_t), \tag{2}$$

where $b(x, y)$ is a function jointly continuous in its two arguments and turning into $b(x)$ for $y = 0$. We shall speak of *small* perturbations if the function ψ giving the perturbing effect is small in one sense or another.

We may speak of problems of the following kind: the convergence of the solution X_t of the perturbed system to the solution x_t of the unperturbed system as the effect of the perturbation decreases, approximate expressions of various accuracies for the deviations $X_t - x_t$ caused by the perturbations, and the same problems for various functionals of a solution (for example, the first exit time from a given domain D).

To solve the kind of problems related to a finite time interval we require less of the function $b(x, y)$ than in problems connected with an infinite interval (or a finite interval growing unboundedly as the perturbing effect decreases). The simplest result related to a finite interval is the following: if the solution of the system (1) with initial condition x_0 at $t = 0$ is unique, then the solution X_t of system (2) with initial condition X_0 converges to x_t uniformly in $t \in [0, T]$ as $X_0 \to x_0$ and $\|\psi\|_{0T} = \sup_{0 \le t \le T} |\psi_t| \to 0$. If the function $b(x, y)$ is differentiable with respect to the pair of its arguments, then we can linearize it near the point $x = x_t$, $y = 0$ and obtain a linear approximation δ_t of $X_t - x_t$ as the solution of the linear system

$$\dot{\delta}_t^i = \sum_j \frac{\partial b^i}{\partial x^j}(x_t, 0)\delta_t^j + \sum_k \frac{\partial b^i}{\partial y^k}(x_t, 0) \cdot \psi_t^k; \tag{3}$$

under sufficiently weak conditions, the norm $\sup_{0 \le t \le T} |X_t - x_t - \delta_t|$ of the remainder will be $o(|X_0 - x_0| + \|\psi\|_{0T})$. If $b(x, y)$ is still smoother, then we have the decomposition

$$X_t = x_t + \delta_t + \gamma_t + o(|X_0 - x_0|^2 + \|\psi\|_{0T}^2), \tag{4}$$

in which γ_t depends quadratically on perturbations of the initial conditions and the right side (the function γ_t can be determined from a system of linear differential equations with a quadratic function of ψ_t, δ_t on the right side), etc.

We may consider a scheme

$$\dot{X}_t^\varepsilon = b(X_t^\varepsilon, \varepsilon\psi_t) \tag{5}$$

depending on a small parameter ε, where ψ_t is a given function. In this case for the solution X_t^ε with initial condition $X_0^\varepsilon = x_0$ we can obtain a decomposition

$$x_t + \varepsilon Y_t^{(1)} + \varepsilon^2 Y_t^{(2)} + \cdots + \varepsilon^n Y_t^{(n)} \tag{6}$$

in powers of ε with the remainder infinitely small compared with ε^n, uniformly on any finite interval $[0, T]$.

Under more stringent restrictions on the function $b(x, y)$, results of this kind can be obtained for perturbations ψ, which are not small in the norm of uniform convergence but rather, for example, in some $\mathbf{L}^p$-norm or another.

As far as results connected with an infinite time interval are concerned, stability properties of the unperturbed system (1) as $t \to \infty$ are essential.

Let x_* be an equilibrium position of system (1), i.e., let $b(x_*) = 0$. Let this equilibrium position be asymptotically stable, i.e., for any neighborhood $U \ni x_*$ let there exist a small neighborhood V of x_* such that for any $x_0 \in V$ the trajectory x_t starting at x_0 does not leave U for $t \geq 0$ and converges to x_* as $t \to \infty$. Denote by G_* the set of initial points x_0 from which there start solutions converging to x_* as $t \to \infty$. For any neighborhood U of x_* and any point $x_0 \in G_*$ there exist $\delta > 0$ and $T > 0$ such that for

$$|X_0 - x_0| < \delta, \quad \sup_{0 \leq t < \infty} |\psi_t| < \delta$$

the solution X_t of system (2) with initial condition x_0 does not go out of U for $t \geq T$. This holds uniformly in x_0 within any compact subset of G_* (i.e., δ and T can be chosen the same for all points x_0 of this compactum). This also implies the uniform convergence of X_t to x_t on the infinite interval $[0, \infty)$ provided that $X_0 \to x_0$, $\sup_{0 \leq t < \infty} |\psi_t| \to 0$.

On the other hand, if the equilibrium position x_* does not have the indicated stability properties, then by means of arbitrarily small perturbations, the solution X_t of the perturbed system can be "carried away" from x_* for sufficiently large t even if the initial point X_0 equals x_*. In particular, there are cases where the solution x_t of the unperturbed system cannot leave some domain D for $t \geq 0$, but the solution X_t of the system obtained from the initial one by an arbitrarily small perturbation leaves the domain in finite time.

Some of these results also hold for trajectories attracted not to a point x_* but rather a compact set of limit points, for example, for trajectories winding on a limit cycle.

There are situations where besides the fact that the perturbations are small, we have sufficient information on their statistical character. In this case it is appropriate to develop various mathematical models of small random perturbations.

The consideration of random perturbations extends the notion of perturbations considered in classical settings at least in two directions. Firstly, the requirements of smallness become less stringent: instead of absolute smallness for all t (or in integral norm) it may be assumed that the perturbations are small only in mean over the ensemble of all possible perturbations. Small random perturbations may assume large values but the probability of these large values is small. Secondly, the consideration of random processes as perturbations extends the notion of the stationarity character of perturbations. Instead of assuming that the perturbations themselves do not change with time, we may assume that the factors which form the statistical structure of the perturbations are constant, i.e., the perturbations are stationary as random processes.

Such an extension of the notion of a perturbation leads to effects not characteristic of small deterministic perturbations. Especially important new properties occur in considering a long lasting effect of small random perturbations.

We shall see what models of small random perturbations may be like and what problems are natural to consider concerning them. We begin with perturbations of the form

$$\dot{X}_t^\varepsilon = b(X_t^\varepsilon, \varepsilon\psi_t), \tag{7}$$

where ψ_t is a given random process, for example, a stationary Gaussian process with known correlation function. (Nonparametric problems connected with arbitrarily random processes which belong to certain classes and are small in some sense are by far more complicated.) For the sake of simplicity, let the initial point X_0 not depend on ε: $X_0^\varepsilon = x_0$. If the solution of system (7) is unique, then the random perturbation $\psi(t)$ leads to a random process X_t^ε.

The first problem which arises is the following: Will X_t^ε converge to the solution x_t of the unperturbed system as $\varepsilon \to 0$? We may consider various kinds of probabilistic convergence: convergence with probability 1, in probability, and in mean. If $\sup_{0\le t\le T} |\psi_t| < \infty$ with probability 1, then, ignoring the fact that the realization of ψ_t is random, we may apply the results presented above to perturbations of the form $\varepsilon\psi_t$ and obtain, under various conditions on $b(x, y)$, that $X_t^\varepsilon \to x_t$ with probability 1, uniformly in $t \in [0, T]$ and that

$$X_t^\varepsilon = x_t + \varepsilon Y_t^{(1)} + o(\varepsilon) \tag{8}$$

or

$$X_t^\varepsilon = x_t + \varepsilon Y_t^{(1)} + \cdots + \varepsilon^n Y_t^{(n)} + o(\varepsilon^n) \tag{9}$$

($o(\varepsilon)$ and $o(\varepsilon^n)$ are understood as being satisfied with probability 1 uniformly in $t \in [0, T]$ as $\varepsilon \to 0$).

Nevertheless, it is not convergence with probability 1 which represents the main interest from the point of view of possible applications. In considering small random perturbations, perhaps we shall not have to do with X_t^ε for various ε simultaneously but only for one small ε. We shall be interested in questions such as: Can we guarantee with practical certainty that for a small ε the value of X_t^ε is close to x_t? What will the order of the deviation $X_t^\varepsilon - x_t$ be? What can be said about the distribution of the values of the random process X_t^ε and functionals thereof? etc. Fortunately, convergence with probability 1 implies convergence in probability, so that X_t^ε will converge to x_t in probability, uniformly in $t \in [0,T]$ as $\varepsilon \to 0$:

$$\mathsf{P}\Big\{\sup_{0\le t\le T}|X_t^\varepsilon - x_t| \ge \delta\Big\} \to 0 \tag{10}$$

for any $\delta > 0$.

For convergence in mean we have to impose still further restrictions on $b(x,y)$ and ψ_t; we shall not discuss this.

From the sharper result (8) it follows that the random process

$$Y_t^\varepsilon = \frac{X_t^\varepsilon - x_t}{\varepsilon}$$

converges to a random process $Y_t^{(1)}$ in the sense of distributions as $\varepsilon \to 0$ (this latter process is connected with the random perturbing effect ψ_t through linear differential equations). In particular, this implies that if ψ_t is a Gaussian process, then in first approximation, the random process X_t^ε will be Gaussian with mean x_t and correlation function proportional to ε^2. This implies the following result: if f is a smooth scalar-valued function in R^r and grad $f(x_{t_0}) \ne 0$, then

$$\mathsf{P}\left\{\frac{f(X_{t_0}^\varepsilon) - f(x_{t_0})}{\varepsilon} \le x\right\} = \Phi\left(\frac{x}{\sigma}\right) + o(1) \tag{11}$$

as $\varepsilon \to 0$, where $\Phi(y) = \int_{-\infty}^{y}(1/\sqrt{2}\pi)e^{-z^2/2}\,dz$ is the Laplace function and σ is determined from grad $f(x_{t_0})$ and the value of the correlation function of $Y_t^{(1)}$ at the point (t_0,t_0). We may obtain sharper results from (9): an expansion of the remainder $o(1)$ in powers of ε. We may also obtain results relative to asymptotic distributions of functionals of Y_t^ε, $0 \le t \le T$, and sharpenings of them, connected with asymptotic expansions.

Hence for random perturbations of the form (7) we may pose and solve a series of problems characteristic of the limit theorems of probability theory. Results on the convergence in probability of a random solution of the perturbed system to a nonrandom function correspond to laws of large numbers for sums of independent random variables. We can speak of the limit distribution under a suitable normalization; this corresponds to results of the type of the central limit theorem. Also as in sharpenings of the central limit theorem, we may obtain asymptotic expansions in powers of the parameter.

In the limit theorems for sums of independent random variables there is still another direction: the study of probabilities of *large deviations* (after normalization) of a sum from the mean. Of course, all these probabilities converge to zero.

Nevertheless we may study the problem of finding simple expressions equivalent to them or the problem of sharper (or rougher) asymptotics of them. The first general results concerning large deviations for sums of independent random variables have been obtained by Cramér [1]. These results have to do with asymptotics, up to equivalence, of probabilities of the form

$$\mathsf{P}\left\{\frac{\xi_1 + \cdots + \xi_n - nm}{\sigma\sqrt{n}} > x\right\} \tag{12}$$

as $n \to \infty$, $x \to \infty$ and also asymptotic expansions for such probabilities (under more stringent restrictions).

We may be interested in analogous problems for a family of random processes X_t^ε arising as a result of small random perturbations of a dynamical system. For example, let A be a set in a function space on the interval $[0, T]$, which does not contain the unperturbed trajectory x_t (and is at a positive distance from it). Then the probability

$$\mathsf{P}\{X^\varepsilon \in A\} \tag{13}$$

of the event that the perturbed trajectory X_t^ε belongs to A, of course, converges to 0 as $\varepsilon \to 0$, but what is the asymptotics of this infinitely small probability?

It may seem that such digging into extremely rare events contradicts the general spirit of probability theory, which ignores events of small probability. Nevertheless, it is exactly this determination of which almost unlikely events related to the random process X_t^ε on a finite interval are “more improbable” and which are “less improbable,” that, in several cases, serves as a key to the question of what the behavior, with probability close to 1, of the process X_t^ε will be on an infinite time interval (or on an interval growing with decreasing ε).

Indeed, for the sake of definiteness, we consider the particular case of perturbations of the form (7):

$$\dot{X}_t^\varepsilon = b(X_t^\varepsilon) + \varepsilon\psi_t. \tag{14}$$

Furthermore, let ψ_t be a stationary Gaussian process. Assume that the trajectories of the unperturbed system (1), beginning at points of a bounded domain D, do not leave this domain for $t \geq 0$ and are attracted to a stable equilibrium position x_* as $t \to \infty$. Will the trajectories of the perturbed system (14) also have this property with probability near 1? The results above related to small nonrandom perturbations cannot help us answer this question, since the supremum of $|\psi_t|$ for $t \in [0, \infty)$ is infinite with probability 1 (if we do not consider the case of “very degenerate” processes ψ_t). We have to approach this question differently. We divide the time axis $[0, \infty)$ into a countable number of intervals of length T. On each of these intervals, for small ε, the most likely behavior of X_t^ε is such that the supremum of $|X_t^\varepsilon - x_t|$ over the interval is small. (For intervals with large indices, X_t^ε will be simply close to x_* with overwhelming probability.) All other ways of behavior, in particular, the exit of X_t^ε from D on a given time interval, will have small probabilities for small ε. Nonetheless, these probabilities are positive for any $\varepsilon > 0$. (Again, we exclude from our considerations the class of "very degenerate" random processes ψ_t.) For a given $\varepsilon > 0$ the probability

$$\mathsf{P}\{X_t^\varepsilon \notin D \text{ for some } t \in [kT, (k+1)T]\} \tag{15}$$

will be almost the same for all intervals with large indices. If the events involving the behavior of our random process on different time intervals were independent, we would obtain from this that sooner or later, with probability 1, the process X_t^ε leaves D and the first exit time τ^ε has an approximately exponential distribution with parameter $T^{-1}\mathsf{P}\{X_t^\varepsilon \text{ exits from } D \text{ for some } t \in [kT, (k+1)T]\}$. The same will happen if these events are not exactly independent but the dependence between them decreases for distant intervals in a certain manner. This can be ensured by some weak dependence properties of the perturbing random process ψ_t.

Hence for problems connected with the exit of X_t^ε from a domain for small ε, it is essential to know the asymptotics of the probabilities of improbable events ("large deviations") involving the behavior of X_t^ε on finite time intervals. In the case of small Gaussian perturbations it turns out that these probabilities have asymptotics of the form $\exp\{-C\varepsilon^2\}$ as $\varepsilon \to 0$ (rough asymptotics, i.e., not up to equivalence but logarithmic equivalence). It turns out that we can introduce a functional $S(\varphi)$ defined on smooth functions (which are smoother than the trajectories of X_t^ε), such that

$$\mathsf{P}\{\rho(X^\varepsilon, \varphi) < \delta\} \approx \exp\{-\varepsilon^{-2} S(\varphi)\} \tag{16}$$

for small positive δ and ε, where ρ is the distance in a function space (say, in the space of continuous functions on the interval from T_1 to T_2; for the precise meaning of formula (16), cf. Chap. 3). The value of the functional at a given function characterizes the difficulty of the passage of X_t^ε near the function. The probability of an unlikely event consists of the contributions $\exp\{-\varepsilon^{-2}S(\varphi)\}$ corresponding to neighborhoods of separate functions φ; as $\varepsilon \to 0$, only the summand with smallest $S(\varphi)$ becomes essential. Therefore, it is natural that the constant C providing the asymptotics is determined as the infimum of $S(\varphi)$ over the corresponding set of functions φ. Thus for the probability in formula (15) the infimum has to be taken over smooth functions φ_t leaving D for $t \in [kT, (k+1)T]$. (Exact formulations and the form of the functional $S(\varphi)$ may be found in Sect. 5, Chap. 4; there we discuss its application to finding the asymptotics of the exit time τ^ε as $\varepsilon \to 0$.)

Another problem related to the behavior of X_t^ε on an infinite time interval is the problem of the limit behavior of the stationary distribution μ^ε of X_t^ε as $\varepsilon \to 0$. This limit behavior is connected with the limit sets of the dynamical system (1). Indeed, the stationary distribution shows how much time the process spends in one set or another. It is plausible to expect that for small ε the process X_t^ε will spend an overwhelming amount of time near limit sets of the dynamical system and, most likely, near stable limit sets. If system (1) has only one stable limit set K, then the measure μ^ε converges weakly to a measure concentrated on K as $\varepsilon \to 0$ (we do not formulate our assertions in so precise a way that we take account of the possibility of the existence of distinct limits μ^{ε_i} for different sequences $\varepsilon_i \to 0$). However, if there are several stable sets, even if there are at least two, K_1 and K_2, then the situation becomes unclear; it depends on the exact form of small perturbations.

The problem of what happens to the stationary distribution of a random process arising as an effect of random perturbations of a dynamical system when these perturbations decrease has been posed in the paper of Pontrjagin, Andronov, and Vitt [1]. The approach applied in this article does not relate to perturbations of the form (14) but rather perturbations under whose influence there arise diffusion processes (given by formulas (19) and (20) below). This approach is based on solving the Fokker–Planck differential equation; in the one-dimensional case the problem of finding the asymptotics of the stationary distribution has been solved completely (cf. also Bernstein's article [1] which appeared in the same period). Some results involving the stationary distribution in the two-dimensional case have also been obtained.

Our approach is not based on equations for the probability density of the stationary distribution but rather the study of probabilities of improbable events. We outline the scheme of application of this approach to the problem of asymptotics of the stationary distribution.

The process X_t^ε spends most of the time in neighborhoods of the stable limit sets K_1 and K_2, it occasionally moves to a significant distance from K_1 or K_2 and returns to the same set, and it very seldom passes from K_1 to K_2 or conversely. If we establish that the probability of the passage of X_t^ε from K_1 to K_2 over a long time T (not depending on ε) converges to 0 with rate

$$\exp\{-V_{12}\varepsilon^{-2}\}$$

as $\varepsilon \to 0$, and the probability of passage from K_2 to K_1 has the order

$$\exp\{-V_{21}\varepsilon^{-2}\}$$

and $V_{12} < V_{21}$, then it becomes plausible that for small ε the process spends most of the time in the neighborhood of K_2. This is so since a successful "attempt" at passage from K_1 to K_2 will fall on a smaller number of time intervals $[kT, (k+1)T]$ spent by the process near K_1, than a successful attempt at passage from K_2 to K_1 with respect to the number of time intervals of length T spent near K_2. Then μ^ε will converge to a measure concentrated on K_2. The constants V_{12} and V_{21} can be determined as the infima of the functional $S(\varphi)$ over the smooth functions φ passing from K_1 to K_2 and conversely on an interval of length T (more precisely, they can be determined as the limits of these infima as $T \to \infty$).

The program of the study limit behavior which we have outlined here is carried out not for random perturbations of the form (14) but rather perturbations leading to Markov processes; the exact formulations and results are given in Sect. 4, Chap. 6.

As we have already noted, random perturbations of the form (14) do not represent the only scheme of random perturbations which we shall consider (and not even the scheme to which we shall pay the greatest attention). An immediate generalization of it may be considered, in which the random process ψ_t is replaced by a generalized random process, a "white noise," which can be defined as the derivative (in the sense of distributions) of the Wiener process w_t:

$$\dot{X}_t^\varepsilon = b(X_t^\varepsilon) + \varepsilon \dot{w}_t. \tag{17}$$

Upon integrating (17), it takes the following form which does not contain distributions:

$$X_t^\varepsilon = X_0 + \int_0^t b(X_s^\varepsilon)\, ds + \varepsilon(w_t - w_0). \tag{18}$$

For perturbations of this form we can solve a larger number of interesting problems than for perturbations of the form (14), since they lead to a Markov process X_t^ε.

A further generalization is perturbations which depend on the point of the space and are of the form

$$X_t^\varepsilon = b(X_t^\varepsilon) + \varepsilon\sigma(X_t^\varepsilon)\dot{w}_t, \tag{19}$$

where $\sigma(x)$ is a matrix-valued function. The precise meaning of (19) can be formulated in the language of stochastic integrals in the following way:

$$X_t^\varepsilon = X_0 + \int_0^t b(X_s^\varepsilon)\, ds + \varepsilon \int_0^t \sigma(X_s^\varepsilon)\, dw_s. \tag{20}$$

Every solution of (20) is also a Markov process (a diffusion process with drift vector $b(x)$ and diffusion matrix $\varepsilon^2\sigma(x)\sigma^*(x)$). For perturbations of the white noise type, given by formulas (19), (20), we can also obtain results on convergence to the trajectories of the unperturbed system, of the type (10), and results on expansions of the type (9) in powers of ε, from which we can obtain results on asymptotic Gaussian character (for example, of the type (11)). Of course, since the white noise is a generalized process whose realizations are not bounded functions in any sense, these results cannot be obtained from the results concerning nonrandom perturbations mentioned at the beginning of the introduction; they have to be obtained independently (cf. Sect. 2, Chap. 2).

For perturbations of the white noise type we establish results concerning probabilities of large deviations of the trajectory X_t^ε from the trajectory x_t of the dynamical system (cf. Sect. 1, Chap. 4 and Sect. 3, Chap. 5). Moreover, because of the Markovian character of the processes, they become even simpler; in particular, the functional $S(\varphi)$ indicating the difficulty of passage of a trajectory near a function takes the following simple form:

$$S(\varphi) = \frac{1}{2} \int \sum_{i,j} a_{ij}(\varphi_t)(\dot{\varphi}_t^i - b^i(\varphi_t))(\dot{\varphi}_t^j - b^j(\varphi_t))\, dt,$$

where $(a_{ij}(x)) = (\sigma(x)\sigma^*(x))^{-1}$.

What other schemes of small random perturbations of dynamical systems shall we consider? What families of random processes will arise in our study? The generalizations may go in several directions and it is not clear which of these directions are preferred to others. Nevertheless, the problem may be posed in a different way: In what case may a given family of random processes be considered as a result of a random perturbation of the dynamical system (1)?

First, in the same way as we may consider the trajectory of a dynamical system, issued from any point, we have to be able to begin the random process at any point x of the space at any time t_0. Further the random process under consideration should depend on a parameter h characterizing the smallness of perturbations. For the sake of simplicity, we shall assume h is a positive numerical parameter converging to zero (in Sect. 3, Chap. 5 families depending on a two-dimensional parameter are considered). Hence for every real t_0, $x \in R^r$ and $h > 0$, $X_t^{t_0,x;h}$ is a random process with values in R^r, such that $X_{t_0}^{t_0,x;h} = x$. We shall say that $X_t^{t_0,x;h}$ is a result of small random perturbations of system (1) if $X_t^{t_0,x;h}$ converges in probability to the solution $x_t^{t_0,x}$ of the unperturbed system (1) with the initial condition $x_{t_0}^{t_0,x} = x$ as $h \downarrow 0$.

This scheme incorporates many families of random processes, arising in various problems naturally but not necessarily as a result of the "distortion" of some initial dynamical system.

EXAMPLE 0.1. Let $\{\xi_n\}$ be a sequence of independent identically distributed r-dimensional random vectors. For $t_0 \in R^1$, $x \in R^r$, $h > 0$ we put

$$X_t^{t_0,x;h} = x + h \sum_{k=[h^{-1}t_0]}^{[h^{-1}t]-1} \xi_k. \tag{21}$$

It is easy to see that $X^{t_0,x;h}$ converges in probability to $x^{t_0,x} = x + (t - t_0)m$, uniformly on every finite time interval as $h \downarrow 0$ (provided that the mathematical expectation $m = \mathsf{M}\xi_k$ exists), i.e., it converges to the trajectory of the dynamical system (1) with $b(x) \equiv m$.

EXAMPLE 0.2. For every $h > 0$ we construct a Markov process on the real line in the following way. Let two nonnegative continuous functions $l(x)$ and $r(x)$ on the real line be given. Our process, beginning at a point x, jumps to the point $x - h$ with probability $h^{-1}l(x)\,dt$ over time dt, to the point $x + h$ with probability $h^{-1}r(x)\,dt$, and it remains at x with the complementary probability. An approximate calculation of the mathematical expectation and variance of the increment of the process over a small time interval Δt shows that as $h \downarrow 0$, the random process converges to the deterministic, nonrandom process described by (1) with $b(x) = r(x) - l(x)$ (the exact results are in Sect. 2, Chap. 5).

Still another class of examples: ξ_t is a stationary random process and $X_t^h = X_t^{t_0,x;h}$ is the solution of the system

$$\dot{X}_t^h = b(X_t^h, \xi_{h^{-1}t}) \tag{22}$$

with initial condition x at time t_0. It can be proved under sufficiently weak assumptions that X_t^h converges to a solution of (1) with $b(x) = \mathsf{M}b(r, \xi_s)$ as $h \downarrow 0$ ($\mathsf{M}b(x, \xi_s)$ does not depend on s; the exact results may be found in Sect. 2, Chap. 7).

In the first example, the convergence in probability of $X_t^{t_0,x;h}$ as $h \downarrow 0$ is a law of large numbers for the sequence $\{\xi_n\}$. Therefore, in general we shall speak of results establishing the convergence in probability of random processes of a given family to the trajectories of a dynamical system as of results of the type of the law of large numbers. Similarly, results involving the convergence, in the sense of distributions, of a family of random processes $X_t^{t_0,x;h} - x_t^{t_0,x}$ after an appropriate normalization to a Gaussian process are results of the type of the central limit theorem. Results involving large deviations are results involving the asymptotics of probabilities of events that the realization of a random process falls in some sets of functions, not containing the trajectory $x_t^{t_0,x}$ of the unperturbed dynamical system. We say a few words on results of the last kind.

For the random step function (21) constructed from the independent random variables ξ_k, the results of the type of large deviations are connected, of course, with the asymptotics, as $n \to \infty$, of probabilities of the form

$$\mathsf{P}\left\{\frac{\xi_1 + \cdots + \xi_n}{n} > x\right\}. \tag{23}$$

The results concerning the asymptotics of probabilities (23) can be divided into two groups: for rapidly decreasing "tails" of the distribution of the terms ξ_i, the principal term of the probability is due to uniformly not too large summands and the asymptotics has the form $\exp\{-Cn\}$ (up to logarithmic equivalence); if, on the other hand, the "tails" of the ξ_i decrease slowly, then the principal part of probability (23) is due to one or more summands of order nx and the probability has the same order as $n\mathsf{P}\{\xi_i > nx\}$. The first general results concerning large deviations were obtained by Cramér under the assumption that the exponential moments $\mathsf{M}e^{z\xi_i}$ are finite, at least for all sufficiently small z; they belong to the first group of results. The results, considered in this book, on large deviations for families of random processes are also generalizations of results belonging to the first group. The assumptions under which they are obtained include analogues of the Cramér condition $\mathsf{M}e^{z\xi_i} < \infty$. Moreover, approximately half of the devices used in obtaining these results is a generalization of Cramér's method (cf. Sects. 2 and 3, Chap. 3 and Sects. 1 and 2, Chap. 5).

Furthermore, in this book we only consider rough results on large deviations, which hold up to logarithmic equivalence. In connection with this we introduce a notation for rough (logarithmic) equivalence:

$$A_h \asymp B_h \quad (h \downarrow 0), \tag{24}$$

if $\ln A_h \sim \ln B_h$ as $h \downarrow 0$.

Cramér's results and a great many subsequent results are not rough but sharp (up to equivalence and even sharper). Nevertheless, we have to take into consideration that random processes are more complicated objects than sums of independent variables. One may try to obtain sharp results on the asymptotics of large deviations for families of random processes; some results have indeed been obtained in this direction. However, in this respect there is an essentially different direction of research: from theorems on large deviations one tries to obtain various other interesting results on the asymptotic behavior of families of random processes which are deterministic in the limit (which may be considered as a result of small random perturbations of a dynamical system). In the authors' opinion, one can deduce more interesting rough consequences from rough theorems on large deviations than sharp consequences from sharp theorems.

Hence we shall consider results of three kinds: results of the type of the law of large numbers, of the type of the central limit theorem, and rough results of the type of large deviations (and, of course, all sorts of consequences of these results). The results of the first type are the weakest; they follow from results of the second or third type. Sometimes we shall speak of them in the first place because it is easier to obtain them and because they are a sort of test of the correctness of a family of random processes to appear in general as a result of small perturbations of a dynamical system.

The results of the second and third types are independent of each other and neither is stronger than the other. Therefore, in some cases we do not consider results of the type of the central limit theorem but rather discuss large deviations immediately (and in the process of obtaining results in this area, we obtain results of the type of the law of large numbers automatically).

The random perturbations are said to be homogeneous in time if the distributions of the values of the arising random process at any finite number of moments of time does not change if we simultaneously shift these moments and the initial moment t_0 along the time axis. In this case all that can be said about perturbations can be formulated naturally in terms of the family $X_t^{x,h}$ of random processes beginning at the point x at time 0: $X_0^{x,h} = x$. Among the schemes of random perturbations we consider, only (21) is not homogeneous in time.

We discuss the content of the book briefly. First we note that we consider problems in probability theory in close connection with problems of the theory of partial differential equations. To the random processes arising as a result of small random perturbations there correspond problems connected with equations containing a small parameter. We study the random perturbations by direct probabilistic methods and then deduce consequences concerning the corresponding problems for partial differential equations. The problems involving the connection between the theory of Markov processes and that of partial differential equations are discussed in Chap. 1. There we recall the necessary information from the theory of random processes.

In Chap. 2 we consider mainly schemes of random perturbations of the form $\dot{X}_t^\varepsilon = b(X_t^\varepsilon, \varepsilon\xi_t)$ or $\dot{X}_t^\varepsilon = b(X_t^\varepsilon) + \varepsilon\sigma(X_t^\varepsilon)\dot{w}_t$, where $\dot{w}_t$ is a white noise process. We discuss results of the type of the law of large numbers in Sect. 1, we discuss sharper results, connected with asymptotic expansions in Sect. 2, and the application of these results to partial differential equations in Sect. 3.

In Chap. 3, for the first time in this book, we consider results involving large deviations for a very simple family of random processes, namely, for the Wiener process w_t multiplied by a small parameter ε. The rough asymptotics of probabilities of large deviations can be described by means of the action functional. The action functional appears in all subsequent chapters. The general questions involving the description of large deviations by means of such functionals constitute the content of Sect. 3 of this chapter. We calculate the action functional for families of Gaussian processes in Sect. 4.

Chapter 4 is devoted mainly to the study of perturbations of dynamical systems by a white noise process. We determine the action functional for the corresponding family of random processes. We study the problem of exit from a neighborhood of a stable equilibrium position of a dynamical system, due to random perturbations, and we determine the asymptotics of the average exit time of the neighborhood and the position at the first exit time. In the same chapter we study the asymptotics of the invariant measure for a dynamical system with one equilibrium position. The problems to be considered are closely connected with the behavior, as $\varepsilon \to 0$, of the solution of problems for elliptic equations with a small parameter at the derivatives of the highest order. The limit behavior of the solution of Dirichlet's problem for an elliptic equation of the second order with a small parameter at the derivatives of the highest order in the case where the characteristics of the corresponding degenerate equation go out to the boundary was studied by Levinson [1]. In Chap. 4 this limit behavior is studied in the case where the characteristics are attracted to a stable equilibrium position inside the domain. (The case of a more complicated behavior of the characteristics is considered in Chap. 6.) We consider Gaussian perturbations of the general form in the last section of Chap. 4.

In Chap. 5 we generalize results of Chap. 4 to a sufficiently large class of families of Markov processes (including processes with discontinuous trajectories). Here the connection with theorems on large deviations for sums of independent random variables becomes clearer; in particular, there appears the apparatus of Legendre transforms of convex functions, which is a natural tool in this area (a separate section is devoted to Legendre transforms).

In Chap. 6 the generalization goes in a different direction: from problems for systems with one equilibrium position to systems with a more complicated structure of equilibrium positions, limit sets, etc. Here an essential role is played by sets of points equivalent to each other in the sense of a certain equivalence relation connected with the system and the perturbations. In the case of a finite number of critical sets, the perturbed system can be approximated in some sense by a finite Markov chain with transition probabilities depending on the small parameter. For the description of the limit behavior of such chains a peculiar apparatus of discrete character, connected with graphs, is developed. A large portion of the results of this chapter admits a formulation in the language of differential equations.

In Chap. 7 we consider problems connected with the averaging principle. Principally, we consider random processes defined by equations of the form $\dot{X}_t^\varepsilon = b(X_t^\varepsilon, \xi_{t/\varepsilon})$, where ξ_t is a stationary process with sufficiently good mixing properties. For the family of random processes X_t^ε we establish theorems of the type of the law of large numbers, the central limit theorem, and finally, of large deviations. Special attention is paid to the last group of questions. In Sect. 6, Chap. 7 we study the behavior of X_t^ε on large time intervals. Here we also consider examples and the corresponding problems of the theory of partial differential equations. In Chap. 7 we also consider systems of differential equations in which the velocity of the fast motion depends on the "slow" variables.

White noise perturbations of Hamiltonian systems are considered in Chap. 8. Let $b(x)$ in (17) be a Hamiltonian vector field in $R^2 : b(x) = (\partial H(x)/\partial x^2, -\partial H(x)/\partial x^1)$, where $H(x)$, $x \in R^2$, is a smooth function. Let $\lim_{|x|\to\infty} H(x) = \infty$. Then all the points of the phase space R^2 are equivalent for the process X_t^ε defined by (17) from the large-deviation point of view. This means, roughly speaking, that for any $x \in R^2$ and any open set $\mathscr{E} = R^2$ one can find a nonrandom $t = t(\varepsilon)$ such that the probability that X_t^ε goes from x to $\mathscr{E}$ in the time $t(\varepsilon)$ is, at least, not exponentially small as $\varepsilon \downarrow 0$. Here the averaging principle rather than large-deviation theory allows us to calculate asymptotic behavior of many interesting characteristics of the process X_t^ε as $\varepsilon \downarrow 0$. One can single out a fast and a slow component of X_t^ε as $\varepsilon \downarrow 0$: The fast component is close to the deterministic motion with a fixed slow component. The slow component $H(X_{t/\varepsilon}^\varepsilon)$, at least locally, is close to the one-dimensional diffusion governed by an operator with the coefficients defined by averaging with respect to the fast component. But if the Hamiltonian $H(x)$ has more than one critical point, the slow component $H(X_{t/\varepsilon}^\varepsilon)$ does not converge in general to a Markov process as $\varepsilon \downarrow 0$. To have a Markov process the limit should consider a projection of X_t^ε on the graph homeomorphic to the set of all connected components of the level sets of $H(x)$, provided with the natural topology. We describe in Chap. 8 the diffusion processes on graphs and calculate the process which is the limit of the slow component of X_t^ε as $\varepsilon \downarrow 0$. As usual, such a result concerning the diffusion process implies a new result concerning the PDEs with a small diffusion coefficient and the Hamiltonian field as a drift. The limit of the solutions of a Dirichlet problem for such an equation is found as the solution of an appropriate boundary problem in a domain on the graph.

In the last section of Chap. 8, pure deterministic perturbations of Hamiltonian systems with one degree of freedom are considered. If the Hamiltonian has saddle points, the classical averaging does not work, and a stochastic process on the graph corresponding to the Hamiltonian should be considered as the limiting slow motion for this pure deterministic system. To give a rigorous meaning to this statement, various stochastic regularizations of the system are considered; but the resulting limiting slow motion is independent of the type of regularization.

Perturbations of Hamiltonian systems with many degrees of freedom are studied in Chap. 9. As is known, in the many-degrees-of-freedom case, because of the resonances, the limit of the slow component for the system with a fixed initial point may not exist. Assuming that the set of resonance frequencies is "small enough", the limit of the slow component exists in the sense of convergence in the Lebesgue measure for the initial points. This is equivalent to regularization of the system by stochastic perturbations of the initial point. But even in the case of one degree of freedom, examples show that if the Hamiltonian has more than two saddle points, this type of regularization does not provide the convergence of the slow motion; this convergence can be provided by stochastic perturbations of the equation. We show in Chap. 9 that in the many-degrees-of-freedom case, the averaging principle in a domain without critical points can be regularized by stochastic perturbations of the equation, if the resonance set has the Lebesgue measure zero. We apply these results to a system of weakly coupled oscillators. The phase space of the slow motion is in this case an open book Π—a multidimensional counterpart of a graph. We calculate the limiting slow motion, which is a stochastic process on Π, deterministic inside the "pages" and having a stochastic behavior at the "binding" of the book.

Chapter 10 contains the applications of the results obtained in the preceding chapters to the study of stability with respect to small random perturbations. We introduce a certain numerical characteristic of stability, which is connected with the action functional. A series of optimal stabilization problems is considered.

The last, eleventh, chapter has the character of a survey. We discuss sharpenings of theorems on large deviations, large deviations for random measures, results concerning the action functional for diffusion processes with reflection at the boundary, random perturbations of infinite-dimensional systems, and applications of large-deviation theory to asymptotic problems for reaction-diffusion equations.

Since the first edition of this book was published, many papers have appeared that generalize and refine some of the results included there. Some further results and references can be found in Day [1], [2] (exit problem for processes with small diffusion). The averaging principle is studied in Kifer [4], [5], Liptser [1], [2], and Gulinsky and Veretennikov [1]. Problems concerning infinite-dimensional systems are considered in Da Prato and Zabczyk [1]. Further references can be found in these publications and in the monographs on large-deviation theory mentioned in the Preface to the Second Edition.