The Annals of Probability
1993, Vol. 21, No. 4, 2215–2245

DIFFUSION PROCESSES ON GRAPHS AND THE AVERAGING PRINCIPLE

By Mark I. Freidlin and Alexander D. Wentzell

University of Maryland

A number of asymptotic problems for "classical" stochastic processes leads to diffusion processes on graphs. In this paper we study several such examples and develop a general technique for these problems. Diffusion in narrow tubes, processes with fast transmutations and small random perturbations of Hamiltonian systems are studied.

1. Introduction. Let $X^\varepsilon(t)$, $\varepsilon > 0$, be a family of Markov processes on a space M. It is possible that as $\varepsilon \to 0$ the process $X^\varepsilon(t)$ moves faster and faster in some directions, whereas the motion in other directions does not accelerate. This is the situation where one can expect that the so-called averaging principle works: We can identify the points of the space M in the "fast" directions, obtaining a new space $Y(M)$ (Y is the mapping effecting the identification). The "fast" motion "across" $Y(M)$ is not a Markov process in general, but in its "fast" time it is nearly one because the characteristics of the "fast" motion depend on the "slow" variables and vary slowly compared to the "fast" motion itself. The slow process $Y(X^\varepsilon(t))$ also is not a Markov one, but the averaging principle means that it converges in some sense to a Markov process $Y(t)$ on $Y(M)$ as $\varepsilon \to 0$, and the characteristics of this limiting process are obtained by averaging the characteristics of the process $Y(X^\varepsilon(t))$ over the "fast" directions with respect to the stationary distribution of the "fast" Markov process.

As a first example suppose $M = \{1, \dots, n\} \times [z_1, z_2]$; $X^\varepsilon(t) = (\nu^\varepsilon(t), Z^\varepsilon(t))$ is a right-continuous Markov process on M such that its z-coordinate is a diffusion process on the segment $\{i\} \times [z_1, z_2]$, corresponding to a second-order differential operator

$$L_i = \frac{1}{2} a_i(z) \frac{d^2}{dz^2} + b_i(z) \frac{d}{dz}, \tag{1.1}$$

$a_i(z) > 0$, with reflection at the ends while the process is on the ith segment; and the process $X^\varepsilon(t)$ jumps from time to time from one segment $\{i\} \times [z_1, z_2]$ to the point with the same z-coordinate on another segment $\{j\} \times [z_1, z_2]$ according to jump densities $(1/\varepsilon) c_{ij}(z)$. This means that if the process is at the point $(i, z) \in M$ at some time, then after a small time Δt it will be on the jth segment, $j \neq i$, with probability $(1/\varepsilon) c_{ij}(z)\, \Delta t + o(\Delta t)$.

Received October 1991; revised July 1992.
AMS 1991 *subject classifications*. 60H10, 58G32.
Key words and phrases. Averaging principle, diffusion on graphs, perturbations of Hamiltonian systems.

The process $X^\varepsilon(t)$ can be described as corresponding to the generator A^ε defined by

$$A^\varepsilon f(i,z) = L_i f(i,z) + \frac{1}{\varepsilon}\sum_{j=1}^{n} c_{ij}(z) f(j,z), \tag{1.2}$$

defined on smooth functions f satisfying the boundary conditions $f'(i,z_1) = f'(i,z_2) = 0$. The function $c_{ii}(z)$ is defined as $-\sum_{j\neq i} c_{ij}(z)$, and it is non-positive, whereas $c_{ij}(z) \geq 0$ for $j \neq i$. Here the "fast" direction is that of the i-coordinate and it is nearly a continuous-time Markov chain with transition densities $c_{ij}(z)$ speeded up by a factor $1/\varepsilon$ times, if Z^ε is near the point z. Of course, the averaging principle works only if this Markov chain is ergodic. In this case $Y(M) = [z_1, z_2]$, and the corresponding identifying mapping is $Y(i,z) = z$. The process $Y(X^\varepsilon(t))$ converges weakly as $\varepsilon \to 0$ to the diffusion process $Y(t)$ on the interval $[z_1, z_2]$ corresponding to the averaged operator

$$\bar{L} = \frac{1}{2}\bar{a}(z)\frac{d^2}{dz^2} + \bar{b}(z)\frac{d}{dz}, \tag{1.3}$$

with reflection at z_1, z_2, where

$$\bar{a}(z) = \sum_{i=1}^{n} q_i(z) a_i(z), \qquad \bar{b}(z) = \sum_{i=1}^{n} q_i(z) b_i(z), \tag{1.4}$$

$[q_i(z),\ i = 1,\dots,n]$ being the stationary distribution of the continuous-time Markov chain with transition densities $c_{ij}(z)$. This means that $(q_i(z))$ is the unique solution of the system

$$\begin{aligned} &\sum_{i=1}^{n} q_i(z) c_{ij}(z) = 0, \qquad j = 1,\dots,n,\\ &\sum_{i=1}^{n} q_i(z) = 1. \end{aligned} \tag{1.5}$$

Of course, the process $X^\varepsilon(t) = (\nu^\varepsilon(t), Z^\varepsilon(t))$ is related to some systems of partial differential equations, and its behavior as $\varepsilon \to 0$ is related to asymptotic problems for these systems.

Another example was considered by Khasminskii [4]. Let M be the ring described, in polar coordinates, as $M = [r_1, r_2] \times [0, 2\pi)$. The process $X^\varepsilon(t)$ is the diffusion corresponding to the operator

$$\begin{aligned} L^\varepsilon = {} & \frac{1}{2}a_{11}(r,\varphi)\frac{\partial^2}{\partial r^2} + \frac{1}{2}a_{22}(r,\varphi)\frac{\partial^2}{\partial \varphi^2} + b_1(r,\varphi)\frac{\partial}{\partial r}\\ & + \frac{1}{\varepsilon}b_2(r,\varphi)\frac{\partial}{\partial\varphi}, \end{aligned} \tag{1.6}$$

the coefficients being periodic in φ and smooth, and b_2, a_{11}, a_{22} strictly positive. Here the "fast" motion is that in the φ-coordinate and $Y(M) =$

$[r_1, r_2]$. According to [4], the limiting process on $Y(M)$ is the diffusion process corresponding to the averaged operator

$$\bar{L} = \frac{1}{2}\bar{a}(r)\frac{d^2}{dr^2} + \bar{b}(r)\frac{d}{dr}, \tag{1.7}$$

$$\bar{a}(r) = \int_0^{2\pi} a_{11}(r,\varphi)m(r,\varphi)\,d\varphi, \qquad \bar{b}(r) = \int_0^{2\pi} b_1(r,\varphi)m(r,\varphi)\,d\varphi. \tag{1.8}$$

Here $m(r,\varphi)$ is the density of the normalized invariant measure on the unit circle for the dynamical system defined by the fast motion: $\dot{\varphi}^{(r)}(t) = b_2(r, \varphi^{(r)}(t))$, where r is fixed; that is,

$$m(r,\varphi) = \frac{b_2(r,\varphi)^{-1}}{\int_0^{2\pi} b_2(r,\varphi)^{-1}\,d\varphi}. \tag{1.9}$$

Of course, one must also specify the behavior of the process after reaching the ends of the interval $[r_1, r_2]$; this can be done by imposing boundary conditions restricting the domain of definition of the operators L^ε and $\bar{L}$. For example, if the process $X^\varepsilon(t)$ in the ring undergoes reflection at the boundary, the limiting process also will be the process with reflection at z_1, z_2 (the boundary conditions for $\bar{L}$ are those of zero first derivative at the points z_1, z_2).

A similar situation arises when we consider nondegenerate diffusion processes in narrow tubes with reflection at the boundary. Let $M^\varepsilon = R^1 \times (\varepsilon\Gamma)$, where Γ is a closed region in R^{d-1} with a piecewise smooth boundary $\partial\Gamma$. Let us consider the process $X^\varepsilon(t) = (Y^\varepsilon(t), Z^\varepsilon(t))$ in M^ε corresponding to the operator

$$L^\varepsilon f(y,z) = \tfrac{1}{2}D(z/\varepsilon)\,\Delta f(y,z), \qquad y \in R^1,\ z \in \varepsilon\Gamma,$$

with normal reflection at $R^1 \times (\varepsilon\partial\Gamma)$. Here the state space of the process $X^\varepsilon(t)$ depends on ε; the identifying mapping is defined by $Y(y,z) = y$. The "fast" motion $Z^\varepsilon(t)$ is not fast absolutely, but only as compared to the size of the cross section of the tube. The process $Z^\varepsilon(t)$ is the same as the diffusion process $Z^1(t)$ in Γ corresponding to the operator $(1/2)D(z)\Delta$, with normal reflection at $\partial\Gamma$ and with changed time and space scales. It is not difficult to prove that the "slow" process $Y^\varepsilon(t)$ converges weakly to the diffusion process on R^1 corresponding to the operator $(1/2)\bar{D}(d^2/dy^2)$, where

$$\bar{D} = \frac{|\Gamma|}{\int_\Gamma D(z)^{-1}\,dz},$$

$|\Gamma|$ being the $(d-1)$-dimensional volume of Γ. The coefficient $\bar{D}$ is the result of averaging $D(z)$ with respect to the invariant measure of the process $Z^1(t)$, which has a density proportional to $D(z)^{-1}$.

But there are situations when the averaging principle leads to a space $Y(M)$ on which it is unusual to consider differential operators.

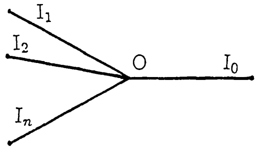

FIG. 1.

Let us consider the first example, but assume that the transition densities are nonzero only for $z \in [z_*, z_2]$, $z_1 < z_* < z_2$ and $c_{ij}(z) \equiv 0$ for $z \in [z_1, z_*)$. The fast transitions between different copies of the interval $[z_1, z_2]$ exist only on the right part of it. So it is natural to take as $Y(M)$ the space consisting of points (i, z), $i = 1, \dots, n$, $z \in [z_1, z_*]$, forming the segments $I_1, \dots, I_n$, and of points $(0, z)$, $z \in [z_*, z_2]$, forming the segment I_0; and all points $(0, z_*), (1, z_*), \dots, (n, z_*)$ are identified. The limiting process should be a diffusion process on the graph $Y(M)$ (see Figure 1). On each of the segments $I_1, \dots, I_n$ it is governed by the operator L_i, and it undergoes reflection at the point z_1. On the segment I_0 the limiting process is governed by the operator $L_0 = \overline{L}$ defined by (1.3) and (1.4) with reflection at z_2. To determine the process uniquely, we have only to add some information about its behavior at the point O where all our segments $I_0, I_1, \dots, I_n$ meet.

We face a similar situation if we take a slight generalization of the second example. Suppose that the fast motion dynamical system $\dot{X} = (1/\varepsilon)b(X)$ in the plane has trajectories shown in Figure 2. The ∞-shaped curve divides the phase plane into three parts G_1, G_2, G_3. Consider the diffusion process that is the result of perturbation of this system by a white noise which is small compared to the deterministic motion:

$$\dot{X}^\varepsilon(t) = \dot{W}(t) + \frac{1}{\varepsilon} b(X^\varepsilon(t)). \tag{1.10}$$

It is easy to see that the mapping $Y(x)$ should identify all points of each closed trajectory, taking the regions G_1, G_2, G_3 to three segments I_1, I_2, I_3 and the ∞-shaped curve to their common point O. Again the limiting process should be a diffusion process on a graph. The characteristics of this process in an interior point of a segment are obtained by averaging along the corresponding periodic

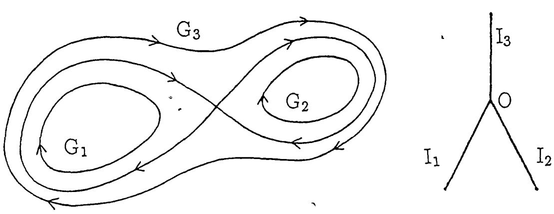

FIG. 2.

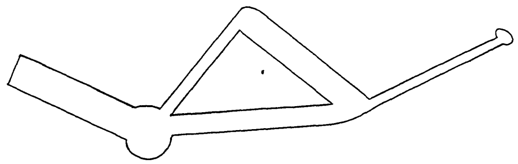

FIG. 3.

trajectory as described above. However, the question of describing the behavior of the limiting process at the vertex of the graph arises. Some results concerning this example were obtained by Wolansky [9, 10]. The study of this problem is made easier if we assume that the vector field $b(x)$ is divergence-free. The most interesting example of such fields is that of Hamiltonian systems. A graph can be connected with any good enough Hamiltonian system, and small random perturbations lead to a stochastic process on the graph. Under some natural additional assumptions, the characteristics of the limiting process everywhere on the graph except at its vertices can be evaluated by a procedure of averaging over the constant-energy manifolds. To determine the process uniquely we should add characteristics of its behavior at the vertices.

Here is one more example, which is a generalization of the above mentioned diffusion in a narrow tube. Consider a graph in R^d consisting of n rectilinear segments $I_1, \dots, I_n$ and m vertices $O_1, \dots, O_m$. Let M^ε be a closed region in R^d consisting of m small neighborhoods of the points O_k, interconnected by n narrow tubes going along the segments I_i (see Figure 3). The sizes of the neighborhoods and of the cross sections of the tubes are supposed to be proportional to ε (e.g., one can take the union of εa_i-neighborhoods of the segments I_i). Consider the d-dimensional Wiener process in M^ε with normal reflection at the boundary. One can expect that as $\varepsilon \to 0$, the process converges in some sense to a one-dimensional process on the graph. Of course, in the interior parts of the segments I_i the limiting process is a Wiener process, but its behavior at the vertices $O_1, \dots, O_m$ is yet to be found.

In these examples diffusion processes on graphs arise as limits as $\varepsilon \to 0$. Of course, they can be of interest in their own right, providing mathematical models of some applied problems having nothing to do with the averaging principle.

The sense in which we understand the convergence of $Y(X^\varepsilon(t))$ to $Y(t)$ is the weak convergence of distributions in the space of continuous functions. The tool we use is that of martingale problems (see [6]). We set forth general results on convergence in Section 2.

Diffusion processes on graphs are studied in Section 3. Section 4 gives a general approach to proving the convergence towards a diffusion on a graph. In Sections 5 and 6 these results are applied to the above mentioned examples of the processes on $\{1, \dots, n\} \times [z_1, z_2]$ and in a region with narrow tubes. In Section 7 we reformulate the results in the language of partial differential equations.

In the last section we discuss in brief the problem of random perturbations of Hamiltonian systems.

Throughout the paper we will use the following notation: If the letter P with some subscripts or superscripts denotes the probability measure, the corresponding expectation will be denoted by E with the same subscripts and superscripts. The indicator function of a set A will be denoted by 1_A.

2. Martingale-problem method of proving weak convergence. The tool we will use to establish weak convergence of distributions in a functional space is martingale problems; see [6]. But in this book martingale problems are formulated in terms of the space $\mathbb{C}^\infty$ of infinitely differentiable functions, whereas we will use some other sets of functions.

Let X be a complete separable metric space; $\mathbb{C}[0,\infty)$, the space of all continuous functions on $[0,\infty)$ with values in X; $\mathscr{F}_{[0,t]}$ $[\mathscr{F}_{[0,\infty)}]$, the σ-algebra of subsets of $\mathbb{C}[0,\infty)$ generated by the sets $\{x(\cdot) \in \mathbb{C}[0,\infty)\colon x(s) \in B\}$, where $s \in [0,t]$ $[s \in [0,\infty)]$ and B is a Borel set. By $\mathbb{C}(X)$ we will denote the space of all bounded continuous real-valued functions on X.

Let A be a linear operator in $\mathbb{C}(X)$ defined on a set $D \subseteq \mathbb{C}(X)$. We will say that a probability measure P on $(\mathbb{C}[0,\infty), \mathscr{F}_{[0,\infty)})$ is a solution of the martingale problem corresponding to the operator A, starting from the point $x_0 \in X$, if for any $f \in D$ the random function defined on the probability space $(\mathbb{C}[0,\infty), \mathscr{F}_{[0,\infty)}, P)$ by

$$f(x(t)) - \int_0^t Af(x(s))\,ds, \qquad t \in [0,\infty), \tag{2.1}$$

is a martingale with respect to the nondecreasing family of σ-algebras $(\mathscr{F}_{[0,t]})$; and if

$$P\{x(\cdot)\colon x(0) = x_0\} = 1. \tag{2.2}$$

The general plan of using martingale problems to prove the weak convergence of probability measures P^ε, $\varepsilon \to 0$, on $\mathbb{C}[0,\infty)$ is as follows.

First, we establish the weak pre-compactness of the family $\{P^\varepsilon,\ \varepsilon > 0\}$; that is, in any sequence $\varepsilon_n \to 0$ there exists a subsequence ε_{n_k} such that $P^{\varepsilon_{n_k}}$ converges weakly to some probability measure P. Then we prove that for any $f \in D$, any $\lambda > 0$, and any $t_0 \ge 0$,

$$\begin{aligned}\Delta(\varepsilon) = \operatorname{ess\,sup} \Bigg| E^\varepsilon \Bigg[\int_{t_0}^\infty e^{-\lambda t}[\lambda f(x(t)) - Af(x(t))]\,dt \\ - e^{-\lambda t_0} f(x(t_0)) \Bigg| \mathscr{F}_{[0,t_0]} \Bigg] \Bigg| \to 0\end{aligned} \tag{2.3}$$

as $\varepsilon \to 0$. This means that for any n, any $0 \le t_1 < \cdots < t_n \le t_0$, and any bounded measurable $G(x_1, \dots, x_n)$, $x_i \in X$,

$$\left| E^\varepsilon G(x(t_1)), \dots, x(t_n)) \cdot \left[\int_{t_0}^{\infty} e^{-\lambda t} [\lambda f(x(t)) - Af(x(t))]\, dt - e^{-\lambda t_0} f(x(t_0)) \right] \right| \le \sup|G| \cdot \Delta(\varepsilon). \tag{2.4}$$

If we take a continuous G, the functional under the expectation sign is continuous, and the limit transition from (2.4) with $\varepsilon = \varepsilon_{n_k}$ yields

$$\int_{t_0}^{\infty} \lambda e^{-\lambda t} EG(x(t_1), \dots, x(t_n)) \left[f(x(t)) - f(x(t_0)) - \int_{t_0}^{t} Af(x(s))\, ds \right] dt = 0. \tag{2.5}$$

Since a continuous function is determined uniquely by its Laplace transform, this means that $EG \cdot [f(x(t)) - f(x(t_0)) - \int_{t_0}^{t} Af(x(s))\, ds] = 0$ for all n and $0 \le t_1 < \cdots < t_n \le t_0$; and the random function (2.1) is a martingale with respect to P. If the measure on X defined by

$$\mu^\varepsilon(A) = P^\varepsilon\{x(\cdot): x(0) \in A\} \tag{2.6}$$

converges weakly as $\varepsilon \to 0$ to the measure δ_{x_0} concentrated at a point $x_0 \in X$, we obtain that the limiting measure P is a solution of the martingale problem corresponding to A, starting from x_0.

The last step is to prove the uniqueness of the solution of the martingale problem. Let P_{x_0} be the unique solution of the problem, starting from x_0. Then the family of measures P^ε converges weakly to P_{x_0} as $\varepsilon \to 0$: otherwise there would exist another subsequence $P^{\varepsilon'_{n_k}}$ converging to a different limit, which is impossible.

In order to establish precompactness, we will be using the following result:

THEOREM 2.1. *Let $\{P^\varepsilon, \varepsilon > 0\}$ be a family of probability distributions on the space $\mathbb{C}[0,\infty)$. Let there exist for any $\rho > 0$, a constant A_ρ such that for any $a \in X$ there exists a function $f_\rho^a(x)$ on X such that $f_\rho^a(a) = 1$, $f_\rho^a(x) = 0$ for $\rho(x,a) \ge \rho$, $0 \le f_\rho^a \le 1$ everywhere, and $f_\rho^a(x(t)) + A_\rho t$, $x(\cdot) \in \mathbb{C}[0,\infty)$, is a submartingale with respect to each of the probabilities P^ε. Then the family $\{P^\varepsilon\}$ is precompact in the sense of weak convergence on $\mathbb{C}[0,\infty)$.*

The proof is that of Theorem 1.4.6 in [6] (but the formulation is freed from any reference to $\mathbb{C}^\infty$ or to other specific structures on R^d).

To ensure uniqueness, we use the following:

THEOREM 2.2. *Let A be a linear operator in $\mathbb{C}(X)$ defined on a set D. Let a set $\Psi \subseteq \mathbb{C}(X)$ be such that for measures μ_1, μ_2 on X the equality $\int_X f\, d\mu_1 = \int_X f\, d\mu_2$ for all $f \in \Psi$ implies $\mu_1 = \mu_2$. Let, for every $f \in \Psi$ and every $\lambda > 0$, there exist a solution $F \in D$ of the equation $\lambda F - AF = f$. Assume also that for* *every $x \in X$ there exists a solution P_x of the martingale problem corresponding to A, starting from x; and let $(x(t), P_x)$, $t \geq 0$, $x \in X$, be a homogeneous Markov process. Then the solution of the martingale problem is unique.*

This is a variant of Theorem 6.3.2 [with condition (ii)] of [6], only freed from $\mathbb{C}^\infty$ and adapted to time-homogeneous processes to which we will apply it. It follows also from the corollary to Theorem 5.1, Chapter IV of [11], because the existence of solution implies that the expectation of $\int_0^\infty e^{-\lambda t} f(x(t))\, dt$ is determined uniquely.

In the proof of convergence, the most essential part is to check (2.3). We will prove a general result on this, which can be applied to different averaging problems, in Section 4; and particular results, in Sections 5 and 6.

3. Diffusion processes on a graph. Let $Y(M)$ be a graph consisting of a finite number of segments I_i, $i = 1, \dots, n$, that are homeomorphic to an interval of the real line, with ends O_k, $k = 1, \dots, m$. Several segments can meet at a vertex O_k; we will write $I_i \sim O_k$ if the segment I_i has the vertex O_k as its end. Suppose a coordinate, y_i, is chosen on each of the segments I_i, changing between finite limits. The distance on $Y(M)$ will be the minimal length of a path connecting two points, the length being measured using the coordinate y_i on I_i. For a function $f(y)$, $y \in Y(M)$, and for a segment $I_i \sim O_k$, let us denote by $(df/dy_i)_{in}(O_k)$ the derivative of the function f with respect to the coordinate on I_i, taken at the point O_k in the direction inside I_i (that is, with plus sign if the coordinate has its minimum at the end O_k, and with minus if it has its maximum).

For every segment I_i, let $m_i = m_i(y_i)$ be an increasing function with finite limits at the ends of I_i. Let L_i be the operator of the generalized second derivative on I_i: $L_i f = (d/dm_i)(df/dy_i)$ [at the points of jumps of m_i, $L_i f(y)$ is defined as the jump of df/dy_i divided by the jump of m_i].

Let α_k, p_{ki}, $k = 1, \dots, m$, $i = 1, \dots, n$, $I_i \sim O_k$, be some nonnegative constants, $\alpha_k + \sum_{i:\, I_i \sim O_k} p_{ki} > 0$. Let us define a linear operator A in the space $\mathbb{C}(Y(M))$ of continuous functions by

$$Af(y) = L_i f(y) \quad \text{for } y \in I_i, \tag{3.1}$$

with the domain of definition $D(A)$ consisting of all functions that have a continuous generalized derivative on every I_i, satisfy the condition

$$\alpha_k Af(O_k) = \sum_{i:\, I_i \sim O_k} p_{ki} \left(\frac{df}{dy_i} \right)_{in} (O_k) \tag{3.2}$$

for each vertex O_k, $k = 1, \dots, m$, and such that $Af \in \mathbb{C}(Y(M))$.

THEOREM 3.1. *The operator A is an infinitesimal operator of a strongly continuous semigroup of linear operators on $\mathbb{C}(Y(M))$ corresponding to a conservative Markov process $(Y(t), P_y)$ on $Y(M)$ with continuous paths. Also* *the following statements are true*:

(a) *Before this process leaves some segment I_i it coincides with the diffusion process with generator L_i.*

(b) *If $\alpha_k = 0$, this process almost surely spends zero time at the vertex O_k.*

(c) *If $\alpha_1 = \cdots = \alpha_m = 0$, then for any $t > 0$ the distribution of $Y(t)$ has a density with respect to a measure m that is equal to 0 for all points O_k and has a distribution function $c_i m_i$ on each of the segments I_i, c_i being positive constants.*

Conversely, let $(Y(t), P_y)$ be a conservative Markov process on $Y(M)$ with continuous paths that coincides, before it leaves I_i, with the diffusion process with generator L_i, and let the corresponding semigroup of linear operators $P^t f(y) = E_y f(Y(t))$ take $\mathbb{C}(Y(M))$ to itself. Then its infinitesimal operator is defined by $Af(y) = L_i f(y)$, $y \in I_i$, with domain of definition described by (3.2).

If the process $Y(t)$ almost surely spends zero time at the points O_k, then $\alpha_k = 0$.

The same is true for operators L_i of the form

$$L_i f(y) = \frac{1}{2} a_i(y) \frac{d^2 f}{dy_i^2} + b_i(y) \frac{df}{dy_i}$$

(*where a_i, b_i are continuous functions on I_i including the ends, and a_i is strictly positive*).

The proof of every statement but that of the existence of a density consists in verifying the fulfillment of the conditions of the Hille–Yosida theorem, and so on. It can be carried out similarly to the proof of corresponding results for diffusions on an interval: see [5] or the original paper by Feller [1]. The existence of a density can be proved by using the method of [7]. The idea of the proof can be outlined as follows: First we prove that the resolvent $R_\lambda f(y) = \int_0^\infty e^{-\lambda t} P^t f(y)\, dt$ can be represented as

$$R_\lambda f(y) = \int_{Y(M)} r_\lambda(y, \eta) f(\eta) m(d\eta). \tag{3.3}$$

This part of the proof does not differ from that in [5]. Then, if $Y(M)$ has no loops, the constants c_i can be chosen so that the operators A, R_λ, P^t are self-adjoint with respect to the measure m. Then the existence of a density is proved as in [7], using (3.3), the inversion formula

$$P^t f(y) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{\lambda t} R_\lambda f(y)\, d\lambda$$

and estimates of the norm of the operator R_λ for complex λ deduced from self-adjointness. If the graph does have cycles, we undo them at some points, and prove the existence of a density before reaching these points; the existence of a density on a graph with cycles is proved using the strong Markov property.

Theorem 3.1 is important to understand the meaning of the results on convergence; but only the part concerning the existence of a process with given conditions (3.2) is used to obtain them. The existence of a density is used to obtain reformulations of the results in the language of partial differential equations.

Now we will formulate some results concerning partial differential equations on graphs. We start with parabolic equations.

THEOREM 3.2. *The initial-boundary value problem*

$$\frac{\partial u}{\partial t}(t,y) = L_i u(t,y) \quad \textit{for } y \in I_i,\ i = 1,\dots,n; \tag{3.4}$$

$$\alpha_k Au(t,O_k) = \sum_{i:\, I_i \sim O_k} p_{ki}\left(\frac{\partial u}{\partial y_i}\right)_{in}(t,O_k), \qquad k = 1,\dots,m; \tag{3.5}$$

$$u(0,y) = f(y) \quad \textit{for } y \in Y(M), \tag{3.6}$$

where $Au(t,O_k)$ is the common value of $L_i u(t,O_k)$ for all $I_i \sim O_k$, has a unique solution for any continuous initial condition f on $Y(M)$.

Some of the conditions (3.5) *can be replaced by*

$$u(t,O_k) = \varphi(t,O_k), \tag{3.5$'$}$$

where $\varphi(t,O_k)$ are continuous functions.

If $\alpha_1 = \cdots = \alpha_m = 0$, the initial function f can be taken bounded, and continuous only inside the segments I_i but not at vertices O_k, and (3.6) *is replaced by*

$$u \textit{ is bounded}, \qquad u(0,y) = f(y) \quad \textit{for } y \in Y(M) \setminus \{O_1,\dots,O_m\}. \tag{3.6$'$}$$

PROOF. Let $(Y(t), P_y)$ be the Markov process of Theorem 3.1. Then the function $u(t,y) = E_y f(Y(t))$ solves the problem (3.4)–(3.6) or (3.4)–(3.5) and (3.6′). If τ is the first time $Y(t)$ hits any of vertices where conditions (3.5) are replaced by (3.5′),

$$u(t,y) = E_y\left[1_{\{\tau \le t\}}\varphi(t-\tau, Y(\tau)) + 1_{\{\tau > t\}} f(Y(t))\right]$$

is the solution of the corresponding initial-boundary value problem. The uniqueness can be proved in the usual martingale way, and in the case of $\alpha_k = 0$ and condition (3.6′), with the use of the existence of a density. □

Now we proceed to stationary problems.

THEOREM 3.3. *The problem*

$$L_i u(y) - c(y)u(y) + g(y) = 0 \quad \textit{for } y \in I_i,\ i = 1,\dots,n; \tag{3.7}$$

$$\alpha_k Au(O_k) = \sum_{i:\, I_i \sim O_k} p_{ki}\left(\frac{du}{dy_i}\right)_{in}(O_k), \qquad k = 1,\dots,m, \tag{3.8}$$

where $Au(O_k)$ is the common limit of $L_i u(y)$ as $y \to O_k$ along each of the segments $I_i \sim O_k$ and $c(y)$ is a strictly positive continuous function, has a unique solution for any continuous g.

If we replace at least some of conditions (3.8) by Dirichlet conditions

$$u(O_k) = \varphi(O_k), \tag{3.8'}$$

if the graph is connected, and if all p_{ki} are positive for $I_i \sim O_k$, then the solution exists and is unique for a continuous $c(y)$ that is only nonnegative.

In the case of $c(y) \equiv 0$ the problem (3.7) and (3.8) is solvable only if the integral of the function f with respect to the invariant measure of the process $Y(t)$ is equal to 0; and the solution is unique only up to an additive constant.

If $\alpha_1 = \cdots = \alpha_m = 0$, we can take the function f bounded and continuous only inside I_i, and require the fulfillment of (3.7) only inside I_i but not at the ends. Then a bounded solution exists and is unique.

The solutions are given correspondingly by

$$u(y) = E_y \int_0^\infty \exp\left\{ - \int_0^t c(Y(s))\, ds \right\} g(Y(t))\, dt,$$

$$u(y) = E_y \int_0^\tau \exp\left\{ - \int_0^t c(Y(s))\, ds \right\} g(Y(t))\, dt$$

and, in the case of $c(y) \equiv 0$,

$$u(y) = \int_0^\infty E_y g(Y(t))\, dt.$$

4. Convergence to a diffusion process on a graph. Let, for any $\varepsilon > 0$, M^ε be a metric space; and let Y^ε be a continuous mapping of M^ε into some graph $Y(M)$.

Let g^ε be a closed set in M^ε, its image $Y^\varepsilon(g^\varepsilon)$ being, for small ε, closer to $\{O_1, \dots, O_m\}$ than some $l(\varepsilon)$, $l(\varepsilon) \to 0$ as $\varepsilon \to 0$. Denote by g_k^ε the part of the set g^ε such that $Y^\varepsilon(g_k^\varepsilon)$ is near the vertex O_k.

For a small positive $\delta > l(\varepsilon)$, let G^δ be the set of all points $x \in M^\varepsilon$ such that $Y^\varepsilon(x)$ lies closer than δ to the set $\{O_1, \dots, O_m\}$; Γ^δ, the set of points at exactly the distance δ from this set (of course, G^δ and Γ^δ depend also on ε). The set Γ^δ is the union of mutually disjoint sets Γ_{ki}^δ of points $x \in M^\varepsilon$ such that $Y^\varepsilon(x)$ lies at distance δ from O_k, on the segment I_i; similarly, $G^\delta = \bigcup_{k=1}^m G_k^\delta$, where G_k^δ is the part of G^δ near O_k.

Suppose that $(X^\varepsilon(t), P_x^\varepsilon)$ is a strong Markov process on M^ε. Let us denote by τ^ε the time when this process reaches g^ε; by σ^δ, the time it reaches Γ^δ. We do not suppose $X^\varepsilon(t)$ to be continuous, but we will suppose that $Y^\varepsilon(X^\varepsilon(t))$ is continuous.

THEOREM 4.1. *Let L_i, $i = 1, \dots, n$, be second-order differential operators on I_i, as in Section 3. Suppose that for any function f on I_i belonging to some* *set D_i and for any $\lambda > 0$,*

$$E_x\left[e^{-\lambda\tau^\varepsilon}f(Y^\varepsilon(X^\varepsilon(\tau^\varepsilon))) - f(Y^\varepsilon(x)) + \int_0^{\tau^\varepsilon} e^{-\lambda t}(\lambda f(Y^\varepsilon(X^\varepsilon(t))) - L_i f(Y^\varepsilon(X^\varepsilon(t))))\,dt\right] = O(k(\varepsilon)) \tag{4.1}$$

as $\varepsilon \to 0$, uniformly with respect to x such that $Y^\varepsilon(x) \in I_i$, where $\lim_{\varepsilon\to 0} k(\varepsilon) = 0$. [This requirement means that the non-Markov process $Y^\varepsilon(X^\varepsilon(t))$ before the time τ^ε of reaching g^ε converges in some sense as $\varepsilon \to 0$ to some diffusion process $Y_i(t)$ on each of the segments I_i.] For any segment I_i of the graph, let any solution of the equation $L_i u = \lambda u$ on I_i belong to the set D_i.

Let $\delta = \delta(\varepsilon) \to 0$, $\delta(\varepsilon)/l(\varepsilon) \to \infty$, $\delta(\varepsilon)/k(\varepsilon) \to \infty$ as $\varepsilon \to 0$. Suppose that for $\lambda > 0$,

$$E_x^\varepsilon\int_0^\infty e^{-\lambda t}1_{G^\delta}(X^\varepsilon(t))\,dt \to 0, \qquad \text{as } \varepsilon \to 0, \tag{4.2}$$

uniformly in the initial point; and that

$$P_x^\varepsilon\{X^\varepsilon(\sigma^\delta) \in \Gamma_{ki}^\delta\} \to p_{ki} \tag{4.3}$$

uniformly in x in the set g_k^ε, where $\sum_{i:\, I_i \sim O_k} p_{ki} = 1$.

Let A be a linear operator defined as in Section 3 for the differential operators L_i, $\alpha_k = 0$ and p_{ki} given by (4.3). Let f be a function on $Y(M)$ belonging to $D(A)$ such that its restriction to each of the segments I_i belongs to D_i.

Then for every fixed $t_0 \geq 0$ and $\lambda > 0$,

$$\Delta(\varepsilon) = \operatorname{ess\,sup}\left|E_x^\varepsilon\left[\int_{t_0}^\infty e^{-\lambda t}[\lambda f(Y^\varepsilon(X^\varepsilon(t))) - Af(Y^\varepsilon(X^\varepsilon(t)))]\,dt - f(Y^\varepsilon(X^\varepsilon(t_0)))\,\middle|\,\mathscr{F}_{[0,t_0]}\right]\right| \to 0, \qquad \text{as } \varepsilon \to 0, \tag{4.4}$$

uniformly with respect to x.

PROOF. By the Markov property of the process $(X^\varepsilon(t), P_x^\varepsilon)$, it is sufficient to prove that

$$E_x^\varepsilon\int_0^\infty e^{-\lambda t}(\lambda f(Y^\varepsilon(X^\varepsilon(t))) - Af(Y^\varepsilon(X^\varepsilon(t))))\,dt - f(Y^\varepsilon(X^\varepsilon(0))) \to 0, \qquad \text{as } \varepsilon \to 0, \tag{4.5}$$

uniformly with respect to $x \in M^\varepsilon$.

Let us introduce a sequence of Markov times $\sigma_0 = 0 \leq \tau_0 < \sigma_1 < \tau_1 < \cdots < \sigma_n < \tau_n < \cdots$ putting $\tau_k = \min\{t \geq \sigma_k\colon X^\varepsilon(t) \in g^\varepsilon\}$, $\sigma_{k+1} = \min\{t \geq \tau_k\colon X^\varepsilon(t) \in \Gamma^\delta\}$. We will need the following lemma.

LEMMA 4.1. *Let $u_i(y)$, $y \in I_i$, be the solution of the boundary value problem*

$$L_i u_i(y) = \lambda u_i(y), \qquad y \in I_i,$$
$$u_i = 1 \quad \text{at the ends of } I_i.$$

Then, uniformly in x such that $Y^\varepsilon(x) \in I_i$,

$$E_x^\varepsilon e^{-\lambda \tau^\varepsilon} = u_i(Y^\varepsilon(x)) + O(k(\varepsilon)) + O(l(\varepsilon)).$$

PROOF. Apply (4.1) to the function u_i:

$$E_x^\varepsilon e^{-\lambda \tau^\varepsilon} u_i(Y^\varepsilon(X^\varepsilon(\tau^\varepsilon))) = u_i(Y^\varepsilon(x)) + O(k(\varepsilon)),$$

and take into account that

$$X^\varepsilon(\tau^\varepsilon) \in g^\varepsilon, \rho(Y^\varepsilon(X^\varepsilon(\tau^\varepsilon)), O_k) \le l(\varepsilon)$$

for some end O_k of the segment I_i, and $|u_i(Y^\varepsilon(X^\varepsilon(\tau^\varepsilon))) - 1| = O(l(\varepsilon))$. □

We can rewrite the left-hand side of (4.5) as

$$\begin{aligned}
&\sum_{n=0}^{\infty} E_x^\varepsilon\Bigg[e^{-\lambda\tau_n} f(Y^\varepsilon(X^\varepsilon(\tau_n))) - e^{-\lambda\sigma_n} f(Y^\varepsilon(X^\varepsilon(\sigma_n))) \\
&\qquad + \int_{\sigma_n}^{\tau_n} e^{-\lambda t}(\lambda f(Y^\varepsilon(X^\varepsilon(t))) - Af(Y^\varepsilon(X^\varepsilon(t))))\,dt\Bigg] \\
&\qquad + \sum_{n=0}^{\infty} E_x^\varepsilon\Bigg[e^{-\lambda\sigma_{n+1}} f(Y^\varepsilon(X^\varepsilon(\sigma_{n+1}))) \\
&\qquad\qquad - e^{-\lambda\tau_n} f(Y^\varepsilon(X^\varepsilon(\tau_n))) \\
&\qquad\qquad + \int_{\tau_n}^{\sigma_{n+1}} e^{-\lambda t}(\lambda f(Y^\varepsilon(X^\varepsilon(t))) \\
&\qquad\qquad\qquad - Af(Y^\varepsilon(X^\varepsilon(t))))\,dt\Bigg].
\end{aligned} \tag{4.6}$$

From Lemma 4.1 we can easily deduce that

$$\sum_{n=0}^{\infty} E_x^\varepsilon e^{-\lambda\sigma_n} = O(1/\delta) \tag{4.7}$$

(this is, roughly speaking, the average number of nonnegligible summands in the sum $\sum_{n=0}^{\infty} E_x^\varepsilon e^{-\lambda\sigma_n}$). The sum $\sum_{n=0}^{\infty} E_x^\varepsilon e^{-\lambda\tau_n}$, of course, is smaller still. Using in each term of the first sum in (4.6) the strong Markov property with respect to σ_n, we obtain

$$\begin{aligned}
&\sum_{n=0}^{\infty} E_x^\varepsilon e^{-\lambda\sigma_n} E_{x'}^\varepsilon\Bigg[e^{-\lambda\tau^\varepsilon} f(Y^\varepsilon(X^\varepsilon(\tau^\varepsilon))) - f(Y^\varepsilon(x')) \\
&\qquad - \int_0^{\tau^\varepsilon} e^{-\lambda t}(\lambda f(Y^\varepsilon(X^\varepsilon(t))) \\
&\qquad\qquad - Af(Y^\varepsilon(X^\varepsilon(t))))\,dt\Bigg]\Bigg|_{x'=X^\varepsilon(\sigma_n)}.
\end{aligned} \tag{4.8}$$

The inner expectation is equal to the expression (4.1) for some $i \in \{1, \dots, n\}$, and it is $O(k(\varepsilon))$. Estimate (4.7) implies that the first sum in (4.6) converges to 0.

From the second sum we separate the sum of the integrals, which does not exceed

$$E_x^\varepsilon \int_0^\infty e^{-\lambda t} 1_{G^\delta}(X^\varepsilon(t))\, dt \cdot \max |\lambda f - Af|;$$

it is estimated by (4.2), and it converges to 0. In the nth term of what remains we subtract and add $e^{-\lambda \tau_n} f(Y^\varepsilon(X^\varepsilon(\sigma_{n+1})))$. The first sum that results is equal to

$$\sum_{n=0}^{\infty} E_x^\varepsilon (e^{-\lambda \sigma_{n+1}} - e^{-\lambda \tau_n}) f(Y^\varepsilon(X^\varepsilon(\tau_n))),$$

and it does not exceed

$$E_x^\varepsilon \int_0^\infty \lambda e^{-\lambda t} 1_{G^\delta}(X^\varepsilon(t))\, dt \cdot \max |f| \to 0$$

by (4.2). In the nth term of the second sum,

$$\sum_{n=0}^{\infty} E_x^\varepsilon e^{-\lambda \tau_n} \big(f(Y^\varepsilon(X^\varepsilon(\sigma_{n+1}))) - f(Y^\varepsilon(X^\varepsilon(\tau_n))) \big),$$

we use the strong Markov property with respect to τ_n and obtain

$$\sum_{n=0}^{\infty} E_x^\varepsilon e^{-\lambda \tau_n} E_{x'}^\varepsilon \big[f(Y^\varepsilon(x^\varepsilon(\sigma^\delta))) - f(Y^\varepsilon(x')) \big] \big|_{x' = X^\varepsilon(\tau_n)}. \tag{4.9}$$

For $x' \in g_k^\varepsilon$ the inner expectation is equal to

$$\begin{aligned}
&\sum_{i:\, I_i \sim O_k} \left[f(O_k) + \frac{df}{dy_i}(O_k) \cdot \delta + o(\delta) - f(Y^\varepsilon(x')) \right] P_{x'}^\varepsilon \{ X^\varepsilon(\sigma^\delta) \in \Gamma_{ki}^\delta \} \\
&\quad = \sum_{i:\, I_i \sim O_k} \left[\frac{df}{dy_i}(O_k) \cdot \delta + o(\delta) - O(l(\varepsilon)) \right] \cdot [p_{ki} + o(1)] \\
&\quad = \delta \cdot \sum_{i:\, I_i \sim O_k} p_{ki} \frac{df}{dy_i}(O_k) + o(\delta).
\end{aligned}$$

For functions satisfying the condition (3.2) this is $o(\delta)$. By the estimate (4.7), the sum (4.6) tends to zero. □

The theorem is proved. If we want to apply it in order to establish the weak convergence of the distribution of $Y^\varepsilon(X^\varepsilon(t))$ to that of the diffusion process $Y(t)$ on the graph, we have to verify the conditions of Theorems 2.1 and 2.2. Since we prove (4.4) only for functions $f \in D(A)$ belonging to D_i on each of the segments I_i, in Theorem 2.2 we take as D the set of functions f belonging to $D(A)$ such that f belongs to D_i on each of I_i. So we have to prove that the solution $F \in D(A)$ of the equation $\lambda F - AF = f$ belongs to D_i on I_i, and this for an everywhere dense set of functions $f \in \mathbb{C}(Y(M))$. Then the general plan of proving weak convergence set forth in Section 2 works. If $x(\varepsilon)$, $\varepsilon > 0$, is a family of points of M^ε such that $Y^\varepsilon(x(\varepsilon)) \to y$ as $\varepsilon \to 0$, we obtain that the probability distribution of $Y^\varepsilon(X^\varepsilon(\cdot))$ in $\mathbb{C}(Y(M))$ corresponding to the probability $P^\varepsilon_{x(\varepsilon)}$ converges weakly to the solution of the martingale problem corresponding to the operator A, starting from the point y; that is, to the probability P_y corresponding to the diffusion process on the graph.

5. Averaging for processes with discrete fast component. Now let z_* be an inner point of a finite interval $[z_1, z_2]$. Let $M^\varepsilon = M = \{1, \dots, n\} \times [z_1, z_2]$; $Y(M)$, the space consisting of points (i, z) with $i = 1, \dots, n$, $z \in [z_1, z_*]$ (forming the segments $I_1, \dots, I_n$), and of points $(0, z)$, $z \in [z_*, z_2]$ (forming the segment I_0); and all couples $(0, z_*), (1, z_*), \dots, (n, z_*)$ are identified. The mapping $Y^\varepsilon = Y$ is defined by $Y(i, z) = (i, z)$ for $z \le z_*$ and by $Y(i, z) = (0, z)$ for $z > z_*$.

Let L_i, $i = 1, \dots, n$, be second-order differential operators:

$$L_i = \frac{1}{2} a_i(z) \frac{d^2}{dz^2} + b_i(z) \frac{d}{dz}, \tag{5.1}$$

where $a_i(z)$, $b_i(z)$ are twice continuously differentiable functions on $[z_1, z_2]$. Let $c_{ij}(z)$ be twice continuously differentiable functions on $[z_*, z_2]$ such that $c_{ij}(z) > 0$ for $j \neq i$, $\sum_{j=1}^n c_{ij}(z) = 0$; and $c_{ij}(z) \equiv 0$ for $z < z_*$.

Let $(X^\varepsilon(t), P^\varepsilon_x)$ be a Markov process on M with generating operator $A^\varepsilon f(x) = A^\varepsilon f(i, z)$ defined by

$$A^\varepsilon f(i, z) = L_i f(i, z) + \frac{1}{\varepsilon} \sum_{j=1}^n c_{ij}(z) f(j, z) \tag{5.2}$$

on the set of functions f that are twice continuously differentiable with respect to $z \in [z_1, z_2]$ and satisfy the boundary conditions $(d/dz) f(i, z_1) = (d/dz) f(i, z_2) = 0$. Let us denote the i- and z-coordinates of this process by $\nu^\varepsilon(t)$, $Z^\varepsilon(t)$ correspondingly. The process $X^\varepsilon(t) = (\nu^\varepsilon(t), Z^\varepsilon(t))$ can be described as follows: while on the ith segment $\{i\} \times [z_1, z_2]$, it is the diffusion corresponding to the operator L_i, with reflection at the boundaries z_1, z_2; and it jumps from one segment to another according to the jump densities $(1/\varepsilon) c_{ij}(z)$. This means that if the process is near a point (i, z), $z \neq z_*$, it jumps to the jth segment during an infinitely small time interval Δt with probability $(1/\varepsilon) c_{ij}(z) \, \Delta t + o(\Delta t)$.

For fixed $z \ge z_*$, the matrix $(c_{ij}(z), i, j = 1, \dots, n)$ is the matrix of transition densities of an ergodic continuous-time Markov chain. Let us denote by $(q_i(z), i = 1, \dots, n)$ the stationary distribution of this Markov chain; that is, $(q_i(z))$ for any $z \ge z_*$ is the unique solution of

$$\begin{aligned} \sum_{i=1}^n q_i(z) c_{ij}(z) &= 0, \qquad j = 1, \dots, n, \\ \sum_{i=1}^n q_i(z) &= 1. \end{aligned} \tag{5.3}$$

Let us define the differential operator L_0 on $[z_*, z_2]$ by

$$L_0 = \frac{1}{2}a_0(z)\frac{d^2}{dz^2} + b_0(z)\frac{d}{dz}, \tag{5.4}$$

$$a_0(z) = \sum_{i=1}^{n} q_i(z)a_i(z), \qquad b_0(z) = \sum_{i=1}^{n} q_i(z)b_i(z). \tag{5.5}$$

THEOREM 5.1. *The distribution of the process $Y(X^\varepsilon(t))$ in the space $\mathbb{C}(Y(M))$ converges weakly as $\varepsilon \to 0$ to that of the diffusion process on $Y(M)$ as described in Theorem* 3.1, *with operators L_i on the segments $I_0, I_1, \dots, I_n$, with conditions*

$$\frac{d}{dz}f(i, z_1) = 0, \qquad i = 1, \dots, n, \tag{5.6}$$

$$\frac{d}{dz}f(0, z_2) = 0 \tag{5.7}$$

at the dead ends of the segments forming $Y(M)$, and with the following conditions at the point where all $(n+1)$ segments meet:

$$\left(\sum_{i=1}^{n} q_i(z_*)a_i(z_*)\right) \cdot \frac{d}{dz}f(0, z_*) = \sum_{i=1}^{n} q_i(z)a_i(z_*)\frac{d}{dz}f(i, z_*). \tag{5.8}$$

PROOF. We first have to verify the conditions of Theorem 2.1. Let $h(x)$ be a smooth function on $[0, \infty)$ such that $h(x) = 1$ for $0 \le x \le 1/4$, $0 \le h(x) \le 1$ everywhere, $h(x) = 0$ for $x \ge 1/2$. For $a = (i_a, z_a) \in Y(M)$ such that $|z_a - z_1|$, $|z_a - z_2|$, $|z_a - z_*| > \rho/2$ we define the function $f_\rho^a(y)$ of $y = (i, z) \in Y(M)$ by

$$f_\rho^a(i, z) = 0 \quad \text{for } i \ne i_a,$$
$$f_\rho^a(i_a, z) = h(|z - z_a|/\rho);$$

if $|z_a - z_j| \le \rho/2$, $j = 1$ or 2, we put

$$f_\rho^a(i, z) = 0 \quad \text{for } i \ne i_a,$$
$$f_\rho^a(i_a, z) = 1 \quad \text{for } |z - z_j| \le \rho/2,$$
$$f_\rho^a(i, z) = h\big(|z - z_j|/\rho - 1/2\big) \quad \text{for } |z - z_j| > \rho/2;$$

if $|z_a - z_*| \le \rho/2$,

$$f_\rho^a(i, z) = 1 \quad \text{for } |z - z_*| \le \rho/2,$$
$$f_\rho^a(i, z) = h(|z - z_*|/\rho - 1/2) \quad \text{for } |z - z_*| > \rho/2.$$

It is easy to see that the function $g_\rho^a(x) = f_\rho^a(Y(x))$, $x \in M$, is twice continuously differentiable on each segment $\{i\} \times [z_1, z_2]$, $i = 1, \dots, n$, and $(d/dz)g_\rho^a(i, z_1) = (d/dz)g_\rho^a(i, z_2) = 0$. We conclude that

$$g_\rho^a(\nu^\varepsilon(t), Z^\varepsilon(t)) - \int_0^t L_{\nu^\varepsilon(s)}g_\rho^a(\nu^\varepsilon(s), Z^\varepsilon(s))\,ds$$

is a martingale. But the first term is $f_\rho^a(Y(X^\varepsilon(t)))$, and the integrand does not exceed $\max_{i,z}|L_i g_\rho^a(i,z)| = A_\rho \le \text{const}\cdot\rho^{-2}$, so $f_\rho^a(Y(X^\varepsilon(t))) - A_\rho t$ is a submartingale.

Then we have to apply the construction of Theorem 4.1. The conditions (4.1), (4.2) and (4.3) will be dealt with in Lemmas 5.1, 5.2 and 5.3.

As $g^\varepsilon = g$ we take the set consisting of points $(i, z_1), (i, z_2), (i, z_*)$, $i = 1, \dots, n$. For a function f on $\{1,\dots,n\}\times[z_1, z_*]$ (or on $\{1,\dots,n\}\times[z_*, z_2]$) that is twice continuously differentiable with respect to z, for $\lambda > 0$, and for $(i,z)\in\{1,\dots,n\}\times[z_1,z_*]$ (or $\{1,\dots,n\}\times[z_*,z_2]$) we have

$$
\begin{aligned}
E^\varepsilon_{(i,z)}\Bigg[& e^{-\lambda\tau^\varepsilon} f(\nu^\varepsilon(\tau^\varepsilon), Z^\varepsilon(\tau^\varepsilon)) - f(i,z) \\
& + \int_0^{\tau^\varepsilon} e^{-\lambda t}\Bigg(\lambda f(\nu^\varepsilon(t), Z^\varepsilon(t)) - L_{\nu^\varepsilon(t)} f(\nu^\varepsilon(t), Z^\varepsilon(t)) \\
& \qquad - \frac{1}{\varepsilon}\sum_{j=1}^n c_{\nu^\varepsilon(t),j}(Z^\varepsilon(t))\, f(j, Z^\varepsilon(t))\Bigg)\,dt\Bigg] = 0.
\end{aligned}
\tag{5.9}
$$

The estimate (4.1) is easy in the case of $x = (i,z)$, $z \le z_*$, $Y(x)\in I_i$, with D_i consisting of all twice continuously differentiable functions and with 0 at the right-hand side. This follows from (5.9), since $c_{ij}(Z^\varepsilon(t)) = 0$ for all i, j and $t < \tau^\varepsilon$. For $i = 0$ [i.e., for $x = (i,z)$, $z \ge z_*$, $Y(x)\in I_0$] we take D_0 to be the set of all 4 times continuously differentiable functions on I_0, and $k(\varepsilon) = \varepsilon$.

LEMMA 5.1. *For any $\lambda > 0$ there exists a constant C such that for any $\varepsilon > 0$ and $x \in M$, and any subinterval $[z_3, z_4] \subset [z_1, z_2]$,*

$$
E_x \int_0^\infty e^{-\lambda t} 1_{[z_3,z_4]}(Z^\varepsilon(t))\,dt \le C\cdot(z_4 - z_3). \tag{5.10}
$$

PROOF. The proof can be based on the fact that for a smooth function $f(z)$ on $[z_1, z_2]$ with $f'(z_1) = f'(z_2) = 0$ the random function

$$
f(Z^\varepsilon(t)) - \int_0^t L_{\nu^\varepsilon(s)} f(Z^\varepsilon(s))\,ds
$$

is a martingale. This enables us to obtain estimates for the process $Z^\varepsilon(t)$ that do not depend on ε and are similar to those one obtains for diffusion processes on $[z_1, z_2]$. □

LEMMA 5.2. *Let $g(i,z)$ be a function on M such that $g(i,z) \equiv 0$ for $z < z_*$, $g(i,z)$ is twice continuously differentiable for $z \in [z_*, z_2]$, and*

$$
\sum_{i=1}^n q_i(z) g(i,z) = 0 \tag{5.11}
$$

for $z \in [z_*, z_2]$.

Then for any $\lambda > 0$,

$$\left| E_x \int_0^\infty e^{-\lambda t} g(\nu^\varepsilon(t), Z^\varepsilon(t))\, dt \right| \tag{5.12}$$
$$\leq \varepsilon \cdot \text{const} \cdot (\max |g(i,z)| + \max |g'(i,z)| + \max |g''(i,z)|).$$

PROOF. Denote by $\hat{C}(z)$ the $(n-1)\times(n-1)$ matrix $(c_{ij}(z),\ i, j = 1, \dots, n-1)$. By the ergodicity assumption, this matrix is uniformly nondegenerate for $z \in [z_*, z_2]$. Denote by $d_{ij}(z)$, $i, j = 1, \dots, n-1$, the elements of the inverse matrix $\hat{C}(z)^{-1}$. Let us put, for $i = 1, \dots, n-1$, $z \in [z_*, z_2]$,

$$G(i,z) = \sum_{j=1}^{n-1} d_{ij}(z) g(j,z),$$

and $G(n,z) = 0$. The functions $G(i,z)$ are twice continuously differentiable on $[z_*, z_2]$, and

$$\begin{aligned} |G(i,z)| &\leq \text{const} \cdot \|g\|, \\ |L_i G(i,z)| &\leq \text{const} \cdot (\|g\| + \|g'\| + \|g''\|), \end{aligned} \tag{5.13}$$

where the constants depend on the norms of the functions $d_{ij}(z)$, $d'_{ij}(z)$, $d''_{ij}(z)$. Let us extend the functions $G(i,z)$ on the interval $[z_1, z_*)$ in such a way that they are twice continuously differentiable on $[z_1, z_2]$, with $G'(i, z_1) = 0$ and (5.13) holding on the whole interval $[z_1, z_2]$.

Suppose at first that $G'(i, z_2) = 0$. Then the function $G(i,z)$, $(i,z) \in M$, belongs to the domain of definition of the generating operator A^ε of the process $X^\varepsilon(t)$. Therefore the random function

$$\eta^\varepsilon(t) = \varepsilon G(\nu^\varepsilon(t), Z^\varepsilon(t)) - \int_0^t \varepsilon A^\varepsilon G(\nu^\varepsilon(s), Z^\varepsilon(s))\, ds$$

is a martingale, so for $x_0 = (i_0, z_0)$,

$$\begin{aligned} &\frac{1}{\lambda}\varepsilon G(i_0, z_0) \\ &= E_{x_0}\int_0^\infty e^{-\lambda t}\eta^\varepsilon(0)\, dt = E_{x_0}\int_0^\infty e^{-\lambda t}\eta^\varepsilon(t)\, dt \\ &= E_{x_0}\int_0^\infty e^{-\lambda t}\left[\varepsilon G(\nu^\varepsilon(t), Z^\varepsilon(t)) - \frac{1}{\lambda}\varepsilon A^\varepsilon G(\nu^\varepsilon(t), Z^\varepsilon(t))\right] dt. \end{aligned} \tag{5.14}$$

We have

$$\varepsilon A^\varepsilon G(i,z) = \varepsilon L_i G(i,z) + \sum_{j=1}^{n-1} c_{ij}(z) G(j,z).$$

The last sum is equal to $g(i,z)$ for $i = 1, \dots, n-1$. As for $i = n$, we have

$$\sum_{i=1}^{n} q_i(z)\left[\sum_{j=1}^{n-1} c_{ij}(z) G(j,z)\right] = \sum_{j=1}^{n-1} G(j,z)\left[\sum_{i=1}^{n} q_i(z) c_{ij}(z)\right] = 0,$$

so

$$\sum_{j=1}^{n-1} c_{nj}(z)G(j,z) = -\frac{1}{q_n(z)}\sum_{i=1}^{n-1} q_i(z)g(i,z) = g(n,z)$$

by the condition (5.11). So we have

$$E^{\varepsilon}_{x_0}\int_0^\infty e^{-\lambda t}g(\nu^\varepsilon(t),Z^\varepsilon(t))\,dt$$

$$(5.15)\qquad = \varepsilon\cdot\Big[-G(i_0,z_0) + E^{\varepsilon}_{x_0}\int_0^\infty e^{-\lambda t}\big(\lambda G(\nu^\varepsilon(t),Z^\varepsilon(t)) - L_{\nu^\varepsilon(t)}G(\nu^\varepsilon(t),Z^\varepsilon(t))\big)\,dt\Big].$$

So we obtain the inequality (5.12).

If $G'(i,z_2)\neq 0$, we change the function $G(i,z)$ in an α-neighborhood of z_2, obtaining a function $G_\alpha(i,z)$ with $G'_\alpha(i,z)=0$. Define $g_\alpha(i,z)=\sum_{j=1}^{n-1}c_{ij}(z)G_\alpha(j,z)$ for $z\in[z_*,z_2]$, and $g_\alpha(i,z)=0$ for $z\in[z_1,z_*)$. We have

$$\Big|E^{\varepsilon}_{x_0}\int_0^\infty e^{-\lambda t}g(\nu^\varepsilon(t),Z^\varepsilon(t))\,dt\Big|$$

$$(5.16)\qquad \le\Big|E^{\varepsilon}_{x_0}\int_0^\infty e^{-\lambda t}g_\alpha(\nu^\varepsilon(t),Z^\varepsilon(t))\,dt\Big| + \Big|E_{x_0}\int_0^\infty e^{-\lambda t}(g-g_\alpha)(\nu^\varepsilon(t),Z^\varepsilon(t))\,dt\Big|.$$

By (5.15), the first expectation in the right-hand side is equal to

$$(5.17)\qquad \varepsilon\cdot\Big[-G_\alpha(i_0,z_0) + E^{\varepsilon}_{x_0}\int_0^\infty e^{-\lambda t}\lambda G_\alpha(\nu^\varepsilon(t),Z^\varepsilon(t))\,dt - \int_0^\infty e^{-\lambda t}L_{\nu^\varepsilon(t)}G_\alpha(\nu^\varepsilon(t),Z^\varepsilon(t))\,dt\Big].$$

The function G_α is bounded independently of α; but the function $L_iG_\alpha(i,z)$ is bounded everywhere except in the α-neighborhood of the point z_2, where it is of order $O((\|g\|+\|g'\|)/\alpha)$. So the absolute value of (5.17) does not exceed

$$\varepsilon\cdot\text{const}\cdot(\|g\|+\|g'\|+\|g''\|)\Big(1+\frac{1}{\alpha}E^{\varepsilon}_{x_0}\int_0^\infty e^{-\lambda t}1_{(z_2-\alpha,z_2]}(Z^\varepsilon(t))\,dt\Big).$$

By Lemma 5.1, the last factor is not greater than $1+C$.

The second term in the right-hand side is estimated using the fact that $\|g_\alpha-g\| = O(\alpha\|g\|+\|g'\|)$; taking α of order ε or smaller, we obtain (5.12).

Now let a function f on I_0 be 4 times continuously differentiable. Formula (5.9) yields, for $x = (i, z) \in M$, $z \geq z_*$,

$$E_x^\varepsilon \bigg[e^{-\lambda\tau^\varepsilon} f(Y(X^\varepsilon(\tau^\varepsilon))) - f(Y(x)) - \int_0^{\tau^\varepsilon} e^{-\lambda t} \big(\lambda f(Y(X^\varepsilon(t))) - L_{\nu^\varepsilon(t)} f(Z^\varepsilon(t)) \big)\, dt \bigg] = 0.$$

To obtain (4.1), we have to prove that

$$E_x^\varepsilon \int_0^{\tau^\varepsilon} e^{-\lambda t} \big[L_{\nu^\varepsilon(t)} f(Z^\varepsilon(t)) - L_0 f(Z^\varepsilon(t)) \big]\, dt = O(\varepsilon). \tag{5.18}$$

We define the function g on M by $g(i, z) = L_i f(z) - L_0 f(z)$ for $z \geq z_*$, $g(i, z) \equiv 0$ for $z < z_*$. This function is twice continuously differentiable for $z \in [z_*, z_2]$, and $\sum_{i=1}^n q_i(z) g(i, z) = 0$. Using the strong Markov property with respect to τ^ε, we obtain that the left-hand side of (5.18) is equal to

$$E_x^\varepsilon \int_0^\infty e^{-\lambda t} g(\nu^\varepsilon(t), Z^\varepsilon(t))\, dt - E_x^\varepsilon e^{-\lambda\tau^\varepsilon} E_{x'}^\varepsilon \int_0^\infty e^{-\lambda t} g(\nu^\varepsilon(t), Z^\varepsilon(t))\, dt \bigg|_{x' = X^\varepsilon(\tau^\varepsilon)},$$

which is $O(\varepsilon)$ by Lemma 5.1. □

Now we find the limits (4.3).

LEMMA 5.3. *For $\delta = \varepsilon^\beta$, $\beta \in (1/4, 1/2)$, we have as $\varepsilon \to 0$*

$$P_{(i, z_*)}^\varepsilon \{ Z^\varepsilon(\sigma^\delta) = z_* + \delta \} \to 1/2; \tag{5.19}$$

$$P_{(i, z_*)}^\varepsilon \{ Z^\varepsilon(\sigma^\delta) = z_* - \delta,\ \nu^\varepsilon(\sigma^\delta) = j \} \to \left(\frac{1}{2} \right) \frac{q_j(z_*) a_j(z_*)}{\sum_{j=1}^n q_j(z_*) a_j(z_*)}. \tag{5.20}$$

PROOF. Take an increasing additive functional of the process $(\nu^\varepsilon, Z^\varepsilon)$:

$$\varphi(t) = \int_0^t a_{\nu^\varepsilon(s)}(Z^\varepsilon(s))\, ds.$$

If we take the inverse function of φ, $\varphi(\tau(t)) = t$, and define $(\tilde{\nu}^\varepsilon(t), \tilde{Z}^\varepsilon(t)) = (\nu^\varepsilon(\tau(t)), Z^\varepsilon(\tau(t)))$, such a random change of time leads to a Markov process of the same kind, but with different coefficients: $\tilde{c}_{ij}(z) = c_{ij}(z)/a_i(z)$, $\tilde{b}_i(z) = b_i(z)/a_i(z)$ and $\tilde{a}_i(z) = 1$. The stationary distribution $(\tilde{q}_i(z))$ for the continuous-time Markov chain with transition densities $\tilde{c}_{ij}(z)$ ($z \in [z_*, z_2]$) is the solution of the system $\sum_{i=1}^n \tilde{q}_i(z) \tilde{c}_{ij}(z) = 0$, $j = 1, \ldots, n$, $\sum_{i=1}^n \tilde{q}_i(z) = 1$. It is easy to see that

$$\tilde{q}_i(z) = \frac{q_i(z) a_i(z)}{\sum_{i=1}^n q_i(z) a_i(z)}.$$

The exit probabilities (5.19) and (5.20) are not affected by a random time change; so without loss of generality we can assume that $a_i(z) = 1$, that is, that the generating operator A^ε of the process $(\nu^\varepsilon(t), Z^\varepsilon(t))$ is given by

$$A^\varepsilon f(i, z) = \frac{1}{2} \frac{d^2 f(i, z)}{dz^2} + b_i(z) \frac{bf(i, z)}{dz} + \frac{1}{\varepsilon} \sum_{j=1}^n c_{ij}(z) f(j, z)$$

for f twice continuously differentiable and satisfying the reflecting boundary conditions $f'(i, z_1) = f'(i, z_2) = 0$.

Denote by $(\bar{\nu}^\varepsilon(t), \bar{Z}(t))$ the Markov process on M with generating operator

$$\bar{A}^\varepsilon f(i,z) = \frac{1}{2}\frac{d^2 f(i,z)}{dz^2} + \frac{1}{\varepsilon}\sum_{j=1}^{n} \bar{c}_{ij}(z) f(j,z),$$

defined on the same set of functions, where $\bar{c}_{ij}(z) = c_{ij}(z_*)$ for $z \geq z_*$ and $\bar{c}_{ij}(z) = 0$ for $z < z_*$. The second coordinate of this process does not depend on ε: it is the standard one-dimensional Wiener process with reflection at the ends z_1, z_2 of the interval. The first coordinate can be represented as a continuous-time Markov chain $\bar{N}(t)$ with transition densities $c_{ij}(z_*)$, taken at time equal to $(1/\varepsilon)\int_0^t 1_{[z_*, z_2]}(\bar{Z}(s))\,ds$.

Let us denote by $\bar{\sigma}^\delta$ the time when $\bar{Z}(t)$ reaches one of the points $z_* \pm \delta$. It is clear that

$$\begin{aligned}
&P^\varepsilon_{(i,z_*)}\{\bar{Z}(\bar{\sigma}^\delta) = z_* - \delta, \bar{\nu}^\varepsilon(\bar{\sigma}^\delta) = j\}\\
&\quad = P^\varepsilon_{(i,z_*)}\left\{\bar{Z}(\bar{\sigma}^\delta) = z_* - \delta, \bar{N}\left(\frac{1}{\varepsilon}\int_0^{\bar{\sigma}^\delta} 1_{[z_*, z_2]}(\bar{Z}(s))\,ds\right) = j\right\}\\
&\quad\times E_{z_*}\left[1_{\{\bar{Z}(\bar{\sigma}^\delta) = z_* - \delta\}}\bar{p}_{ij}\left(\frac{1}{\varepsilon}\int_0^{\bar{\sigma}^\delta} 1_{[z_*, z_2]}(\bar{Z}(s))\,ds\right)\right],
\end{aligned}\tag{5.21}$$

where E_{z_*} is the expectation corresponding to the Wiener process $\bar{Z}(t)$ starting from the point z_*, and $\bar{p}_{ij}(t)$ are the transition probabilities of the continuous-time Markov chain $\bar{N}(t)$.

By the self-similarity of the Wiener process, the random variable $\int_0^{\bar{\sigma}^\delta} 1_{[z_*, z_2]}(\bar{Z}(s))\,ds$ has exactly the same distribution as δ^2 multiplied by some strictly positive random variable that does not depend on ε; so if $\delta^2/\varepsilon \to \infty$, the argument in $\bar{p}_{ij}$ tends to ∞ in probability, and $\bar{p}_{ij}$ of this random argument converges in probability to $q_j(z_*)$. So the expression (5.20) converges to $(1/2)q_j(z_*)$.

Now, the distribution in the space of trajectories on the time interval $[0, T]$ corresponding to the process $(\nu^\varepsilon(t), Z^\varepsilon(t))$ is absolutely continuous with respect to that corresponding to $(\bar{\nu}^\varepsilon(t), \bar{Z}(t))$ with density π^ε_T given by

$$\begin{aligned}
\pi_T(\bar{\nu}^\varepsilon, \bar{Z}) &= \prod_{0<t\leq T} \frac{c_{\bar{\nu}^\varepsilon(t-), \bar{\nu}^\varepsilon(t)}(\bar{Z}(t))}{\bar{c}_{\bar{\nu}^\varepsilon(t-), \bar{\nu}^\varepsilon(t)}(\bar{Z}(t))}\\
&\quad\times\exp\left\{\int_0^T b_{\bar{\nu}^\varepsilon(t)}(\bar{Z}(t))\,d\bar{Z}(t) - \frac{1}{2}\int_0^T b_{\bar{\nu}^\varepsilon(t)}(\bar{Z}(t))^2\,dt\right.\\
&\quad\left. - \frac{1}{\varepsilon}\int_0^T \sum_{j\neq\bar{\nu}^\varepsilon(t)} \left(c_{\bar{\nu}^\varepsilon(t), j}(\bar{Z}(t)) - \bar{c}_{\bar{\nu}^\varepsilon(t), j}(\bar{Z}(t))\right) dt\right\},
\end{aligned}\tag{5.22}$$

the product being taken only over the points t of jumps of $\bar{\nu}^\varepsilon$.

Using this density, we can represent the probability of every event having to do with the process $(\bar{\nu}^\varepsilon(t), \bar{Z}(t))$. In particular,

$$
\begin{aligned}
&P^\varepsilon_{(i,z_*)}\{\sigma^\delta \le T,\, Z^\varepsilon(\sigma^\delta) = z_* - \delta,\, \nu^\varepsilon(\sigma^\delta) = j\} \\
&\quad = E^\varepsilon_{(i,z_*)} 1_{\{\bar{\sigma}^\delta \le T,\, \bar{Z}(\bar{\sigma}^\delta) = z_* - \delta,\, \bar{\nu}^\varepsilon(\bar{\sigma}^\delta) = j\}} \pi_T(\bar{\nu}^\varepsilon, \bar{Z}) \qquad (5.23)\\
&\quad = E^\varepsilon_{(i,z_*)} 1_{\{\bar{\sigma}^\delta \le T,\, \bar{Z}(\bar{\sigma}^\delta) = z_* - \delta,\, \bar{\nu}^\varepsilon(\bar{\sigma}^\delta) = j\}} \pi_{\bar{\sigma}^\delta \wedge T}(\bar{\nu}^\varepsilon, \bar{Z})
\end{aligned}
$$

(the last equality because $\pi_t(\bar{\nu}^\varepsilon, \bar{Z})$, $t \ge 0$, is a martingale).

Now we take $T = \varepsilon^\gamma$, $\gamma < 2\beta$. Then the difference of the probability in the left-hand side of (5.20) and that in the left-hand side of (5.23) converges to 0 as $\varepsilon \to 0$. Also the difference of the indicator function in (5.23) and $1_{\{\bar{Z}(\bar{\sigma}^\delta) = z_* - \delta,\, \bar{\nu}^\varepsilon(\bar{\sigma}^\delta) = j\}}$ converges in probability to zero. In order to prove (5.20) it is sufficient to prove that if $\gamma > 1 - 2\beta$, the density $\pi_{\bar{\sigma}^\delta \wedge T}(\bar{\nu}^\varepsilon, \bar{Z})$ converges in the mean square sense to 1. We have $E^\varepsilon_{(i,z_*)} \pi_{\bar{\sigma}^\delta \wedge T}(\bar{\nu}^\varepsilon, \bar{Z}) = 1$; so it is sufficient to prove that the expectation of the square converges to 1. We have

$$
\begin{aligned}
&E^\varepsilon_{(i,z_*)} \pi_{\bar{\sigma}^\delta \wedge T}(\bar{\nu}^\varepsilon, \bar{Z})^2 \\
&\quad = E^\varepsilon_{(i,z_*)} \prod_{0 < t \le \bar{\sigma}^\delta \wedge T} \left(\frac{c_{\bar{\nu}^\varepsilon(t-), \bar{\nu}^\varepsilon(t)}(\bar{Z}(t))}{\bar{c}_{\bar{\nu}^\varepsilon(t-), \bar{\nu}^\varepsilon(t)}(\bar{Z}(t))} \right)^2 \\
&\qquad \times \exp\Bigg\{ 2\int_0^{\bar{\sigma}^\delta \wedge T} b_{\bar{\nu}^\varepsilon(t)}(\bar{Z}(t))\, d\bar{Z}(t) - 2\int_0^{\bar{\sigma}^\delta \wedge T} b_{\bar{\nu}^\varepsilon(t)}(\bar{Z}(t))^2\, dt \\
&\qquad\qquad - \frac{1}{\varepsilon} \int_0^{\bar{\sigma}^\delta \wedge T} \sum_{j \ne \bar{\nu}^\varepsilon(t)} \left(\frac{c_{\bar{\nu}^\varepsilon(t), j}(\bar{Z}(t))^2}{\bar{c}_{\bar{\nu}^\varepsilon(t), j}(\bar{Z}(t))} - \bar{c}_{\bar{\nu}^\varepsilon(t), j}(\bar{Z}(t)) \right) dt \Bigg\} \\
&\qquad \times \exp\Bigg\{ \int_0^{\bar{\sigma}^\delta \wedge T} b_{\bar{\nu}^\varepsilon(t)}(\bar{Z}(t))^2\, dt \\
&\qquad\qquad + \frac{1}{\varepsilon} \int_0^{\bar{\sigma}^\delta \wedge T} \sum_{j \ne \bar{\nu}^\varepsilon(t)} \frac{\left(c_{\bar{\nu}^\varepsilon(t), j}(\bar{Z}(t)) - \bar{c}_{\bar{\nu}^\varepsilon(t), j}(\bar{Z}(t))\right)^2}{\bar{c}_{\bar{\nu}^\varepsilon(t), j}(\bar{Z}(t))}\, dt \Bigg\}.
\end{aligned}
$$

[We have added and subtracted some integrals in the exponent; the ratio in the last integral is taken to be equal to 0 for $\bar{Z}(t) < z_*$.] The product over the jumps and the first exponential term forms the density of the distribution of some process $(\tilde{\nu}^\varepsilon, \tilde{Z}^\varepsilon)$ with respect to that of the process $(\bar{\nu}^\varepsilon, \bar{Z})$, namely, of the process of the same kind with drift $\tilde{b}_i(z) = 2b_i(z)$ and jump densities

$(1/\varepsilon)c_{ij}(z)^2/\bar{c}_{ij}(z)$, $j \neq i$, $z \geq z_*$; the expectation of this density is equal to 1. So

$$E^{\varepsilon}_{(i,z_*)}\pi_{\bar{\sigma}^\delta\wedge T}(\bar{\nu}^\varepsilon, \bar{Z})^2$$
$$\leq \exp\left\{T \cdot \max |b_i(z)|^2 + \frac{T}{\varepsilon} \max_{i,\, z:\, z_* \leq z \leq z_* + \delta} \sum_{j \neq i} \frac{(c_{ij}(z) - c_{ij}(z_*))^2}{c_{ij}(z_*)}\right\}.$$

The first term between the braces is $O(\varepsilon^\gamma)$; the second is $O(\varepsilon^{\gamma-1} \cdot \delta^2)$ [we have used the smoothness of $c_{ij}(z)$], and it tends to 0 as $\varepsilon \to 0$ if $\gamma > 1 - 2\beta$.

If $1/4 < \beta < 1/2$, we can take $\gamma = 1/2$, and obtain

$$\Big| P^{\varepsilon}_{(i,z_*)}\{\sigma^\delta \leq T,\, Z^\varepsilon(\sigma^\delta) = z_* - \delta,\, \nu^\varepsilon(\sigma^\delta) = j\}$$
$$- P^{\varepsilon}_{(i,z_*)}\{\bar{\sigma}^\delta < T,\, \bar{Z}(\bar{\sigma}^\delta) = z_* - \delta,\, \bar{\nu}^\varepsilon(\bar{\sigma}^\delta) = j\}\Big|$$
$$= \Big| E_{(i,z_*)} 1_{\{\bar{\sigma}^\delta \leq T,\, \bar{Z}(\bar{\sigma}^\delta) = z_* - \delta,\, \bar{\nu}^\varepsilon(\bar{\sigma}^\delta) = j\}} \big(\pi_{\bar{\sigma}^\delta \wedge T}(\bar{\nu}^\varepsilon, \bar{Z}) - 1\big)\Big|$$
$$\leq \sqrt{E^{\varepsilon}_{(i,z_*)}\pi_{\bar{\sigma}^\delta\wedge T}(\bar{\nu}^\varepsilon, \bar{Z})^2 - 1} \to 0.$$

Formula (5.20) is proved. □

6. The limiting process for diffusion in narrow branching tubes. Now let $Y(M)$ be a system of rectilinear segments $I_1, \dots, I_n$ in R^d, with ends $O_1, \dots, O_m$. As the coordinate on each segment we take the length. For $I_i \sim O_k$, let e_{ki} be the unit vector directed from the vertex O_k into I_i.

Let $g_1, \dots, g_m$ be closed regions in R^d with piecewise smooth boundaries; Γ_i, for every $i = 1, \dots, n$, a closed region in the orthogonal complement of e_{ki} with a piecewise smooth boundary. Let c_{ki} be some numbers corresponding to each pair of a vertex O_k and a segment $I_i \sim O_k$. We suppose that for all I_i and O_k with $I_i \sim O_k$, the set $c_{ki}e_{ki} + \Gamma_i$ forms a part of the boundary of g_k.

Let us denote, for $i = 1, \dots, n$, by I_i^ε the segment I_i without its portions of length εc_{ki} near its ends. For every positive ε, let us define

$$g_k^\varepsilon = O_k + \varepsilon g_k, \qquad k = 1, \dots, m, \qquad g^\varepsilon = \bigcup_{k=1}^m g_k^\varepsilon,$$

$$M^\varepsilon = g^\varepsilon \cup \bigcup_{i=1}^n (I_i^\varepsilon + \varepsilon\Gamma_i)$$

(see Figure 3).

Suppose that for any $\varepsilon > 0$ a mapping Y^ε of M^ε into $Y(M)$ is given such that $Y^\varepsilon(x) = x_0$ for $x = x_0 + \varepsilon x'$, $x_0 \in I_i^\varepsilon$, $x' \in \Gamma_i$ (i.e., on the cylindrical parts of M^ε); and that $\max\{|x - Y^\varepsilon(x)|: x \in M^\varepsilon\} = O(\varepsilon)$.

Let $(X^\varepsilon(t), P_x^\varepsilon)$ be the d-dimensional Wiener process in M^ε with normal reflection at the boundary.

THEOREM 6.1. *If $Y^\varepsilon(x) \to x_0$ as $\varepsilon \to 0$, then the distribution of the process $Y^\varepsilon(X^\varepsilon(t))$ in $\mathbb{C}(Y(M))$ converges weakly as $\varepsilon \to 0$ to that of the diffusion process described in Theorem* 3.1 *with operator $L_i = (1/2)d^2/dy_i^2$ and with gluing conditions*

$$\sum_{i:\, I_i \times O_k} |\Gamma_i| \left(\frac{df}{dy_i} \right)_{in} (O_k) = 0, \qquad k = 1, \dots, m, \tag{6.1}$$

where $|\Gamma_i|$ is the $(d-1)$-dimensional volume of the region Γ_i.

PROOF. We have to verify the conditions of Sections 2 and 4. Let us verify the fulfillment of the conditions of Theorem 2.1.

For ε such that $\varepsilon \cdot \max\{|x|: x \in g_k\}$, $\varepsilon \cdot \max\{|x|: x \in \Gamma_i\}$, $\max\{|Y^\varepsilon(x) - x|: x \in Y(M)\} < \rho/4$ we define functions $f_\rho^a(y)$, $a \in Y(M)$, on $Y(M)$ in the same way as in the proof of Theorem 5.1; the function $g_{\rho,\varepsilon}^a(x) = f_\rho^a(Y^\varepsilon(x))$, $x \in M^\varepsilon$, has zero normal derivative at the boundary of M^ε, so

$$f_\rho^a(Y^\varepsilon(X^\varepsilon(t))) - \int_0^t \tfrac{1}{2}\Delta g_{\rho,\varepsilon}^a(X^\varepsilon(s))\, ds$$

is a martingale. But $|(1/2)\Delta g_{\rho,\varepsilon}^a(x)| \le A_\rho = (1/2)\rho^{-2}\max|h''(x)|$; this implies the precompactness.

The estimate (4.1) is fulfilled with $k(\varepsilon) = 0$ because the motion of $X^\varepsilon(t)$ along each segment I_i is a one-dimensional Wiener process.

To prove estimate (4.2), we first prove that, for δ greater than some constant multiplied by ε, for $x \in g^\varepsilon$,

$$E_x^\varepsilon \sigma^\delta \le \mathrm{const} \cdot \delta^2, \tag{6.2}$$

the constant being independent of x or of ε. We use the construction of Theorem 4.1 with the Markov times $\sigma_0 < \tau_0 < \sigma_1 < \tau_1 < \cdots < \sigma_n < \tau_n < \cdots$, with $C\varepsilon$ instead of δ, where the constant $C > \max c_{ki}$. The expectation (6.2) does not exceed

$$E_x^\varepsilon \sum_{n:\, \sigma_n < \sigma^\delta} E_{x'}^\varepsilon(\tau^\varepsilon \wedge \sigma^\delta)\Big|_{x' = X^\varepsilon(\sigma_n)} + E_x^\varepsilon \sum_{n:\, \tau_n < \sigma^\delta} E_{x'}^\varepsilon \sigma^{C\varepsilon}\Big|_{x' = X^\varepsilon(\tau_n)}. \tag{6.3}$$

The first inner expectation has to do only with the standard one-dimensional Wiener process, and it is equal to $o(\varepsilon \cdot \delta)$. The second inner expectation is equal to ε^2 multiplied by the corresponding expectation for $\varepsilon = 1$: $E_{x''}^1 \sigma^C \le$ const $< \infty$ (because before time $\sigma^{C\varepsilon}/\varepsilon^2$ the process $(1/\varepsilon)[X^\varepsilon(\varepsilon^2 t) - O_k]$ is a d-dimensional Wiener process with reflection at the boundary in some region that does not depend on ε). This means that the expression (6.3) does not exceed $O(\varepsilon \cdot \delta)$ multiplied by the expected number of $\sigma_n < \sigma^\delta$. This expected number is estimated in the same way as in Theorem 4.1, and is $O(\delta/\varepsilon)$; so (6.2) is proved.

Now we use the times $\sigma_0 < \tau_0 < \sigma_1 < \tau_1 < \cdots < \sigma_n < \tau_n < \cdots$ not with $C\varepsilon$ but with δ, and we obtain

$$
\begin{aligned}
&E_x^\varepsilon \int_0^\infty e^{-\lambda t} 1_{G^\delta}(X^\varepsilon(t))\,dt \\
(6.4)\qquad &\le E_x^\varepsilon \sum_{n=0}^\infty e^{-\lambda\sigma_n} E_{x'}^\varepsilon \int_0^{\tau^\varepsilon} 1_{G^\delta}(X^\varepsilon(t))\,dt \bigg|_{x'=X^\varepsilon(\sigma_n)} \\
&\quad + E_x^\varepsilon \sum_{n=0}^\infty e^{-\lambda\tau_n} E_x^\varepsilon \sigma^\delta \bigg|_{x'=X^\varepsilon(\tau_n)}.
\end{aligned}
$$

Both inner expectations are $O(\delta^2)$: For the first one this is a simple computation because before τ^ε the process along the segment is the one-dimensional Wiener; the second follows by (6.2). The expectation of $\sum_{n=0}^\infty e^{-\lambda\sigma_n}$ is $O(1/\delta)$, so

$$
E_x^\varepsilon \int_0^\infty e^{-\lambda t} 1_{G^\delta}(X^\varepsilon(t))\,dt = O(\delta). \tag{6.5}
$$

To find the limits (4.3) we proceed as follows. Considering transitions between the boundaries of small neighborhoods of O_k, it is easy to prove that if $\delta/\varepsilon \to \infty$,

$$
\max_{x\in g_k^\varepsilon} P_x^\varepsilon\{X^\varepsilon(\sigma^\delta) \in \Gamma_{ki}^\delta\} - \min_{x\in g_k^\varepsilon} P_x^\varepsilon\{X^\varepsilon(\sigma^\delta) \in \Gamma_{ki}^\delta\} \to 0.
$$

Let us denote by $p_{ki}(\varepsilon,\delta)$ arbitrary numbers between these minima and maxima.

We will use the relation

$$
\mu^\varepsilon(A) = \int_{g^\varepsilon} \nu^{\varepsilon\delta}(dx) E_x^\varepsilon \int_0^{\tau_0} 1_A(X^\varepsilon(t))\,dt \tag{6.6}
$$

between the invariant measures μ^ε of the process $X^\varepsilon(t)$ on M^ε, and $\nu^{\varepsilon\delta}$ of the Markov chain $X^\varepsilon(\tau_n)$ on g^ε (see [3]; also [8]). It is easy to see that the first invariant measure is, up to a constant factor, the Lebesgue measure on M^ε.

Let I_i be a segment with ends O_{k_1}, O_{k_2}. If we take as the set A the part $A_{k_1 i}$ of the cylindrical tube between the cross sections $\Gamma_{k_1 i}^\delta$ and $\Gamma_{k_1 i}^{2\delta}$, only the parts of the integral in (6.6) over $g_{k_1}^\varepsilon$ and $g_{k_2}^\varepsilon$ are positive. The expectation in this formula can be taken only over the event $\{X^\varepsilon(\sigma^\delta) \in \Gamma_{k_j i}^\delta\}$, and the integral from 0 to τ_0 can be replaced by that from $\sigma_0 = \sigma^\delta$ to τ_0:

$$
\begin{aligned}
\mu^\varepsilon(A_{k_1 i}) &= \int_{g_{k_1}^\varepsilon} \nu^{\varepsilon\delta}(dx) E_x^\varepsilon 1_{\{X^\varepsilon(\sigma^\delta)\in\Gamma_{k_1 i}^\delta\}} \int_{\sigma^\delta}^{\tau_0} 1_{A_{k_1 i}}(X^\varepsilon(t))\,dt \\
(6.7)\qquad &\quad + \int_{g_{k_2}^\varepsilon} \nu^{\varepsilon\delta}(dx) E_x^\varepsilon 1_{\{X^\varepsilon(\sigma^\delta)\in\Gamma_{k_2 i}^\delta\}} \int_{\sigma^\delta}^{\tau_0} 1_{A_{k_1 i}}(X^\varepsilon(t))\,dt.
\end{aligned}
$$

Using the strong Markov property with respect to σ^δ, we can rewrite the expectations here in the form

$$E_x^\varepsilon 1_{\{X^\varepsilon(\sigma^\delta)\in\Gamma_{k_j i}^\delta\}} E_{x'}^\varepsilon \int_0^{\tau^\varepsilon} 1_{A_{k_1 i}}(X^\varepsilon(t))\,dt\Big|_{x'=X^\varepsilon(\sigma^\delta)}. \tag{6.8}$$

Since the motion along the cylindrical tubes is a one-dimensional standard Wiener process, the inner expectation is easily evaluated; it is equal to $\delta^2 + o(\delta^2)$ for $x' \in \Gamma_{k_1 i}^\delta$ and to $O(\delta^3)$ for $x' \in \Gamma_{k_2 i}^\delta$. So the expectation (6.8) is equal to $p_{k_1 i}(\varepsilon,\delta)\delta^2 + o(\delta^2)$ for $j = 1$ and to $o(\delta^2)$ for $j = 2$.

The left-hand side of (6.7) is equal to the Lebesgue measure of $A_{k_1 i}$, that is, $\delta\cdot\varepsilon^{d-1}|\Gamma_i|$. The formula (6.7) yields

$$\delta\cdot\varepsilon^{d-1}\cdot|\Gamma_i| = \nu^{\varepsilon\delta}(g_{k_1}^\varepsilon)\cdot\left[p_{k_1 i}(\varepsilon,\delta)\delta^2 + o(\delta^2)\right] + \nu^{\varepsilon\delta}(g_{k_2}^\varepsilon)\cdot o(\delta^2). \tag{6.9}$$

From (6.9) it can easily be seen that $\nu^{\varepsilon\delta}(g_k^\varepsilon)$ are of order $\delta^{-1}\varepsilon^{d-1}$; and that

$$\nu^{\varepsilon\delta}(g_k^\varepsilon)\cdot p_{ki}(\varepsilon,\delta) \sim \delta^{-1}\varepsilon^{d-1}|\Gamma_i|.$$

Dividing this by $\nu^{\varepsilon\delta}(g_k^\varepsilon) \sim \delta^{-1}\varepsilon^{d-1}\sum_{i:\,I_i\times O_k}|\Gamma_i|$, we obtain

$$p_{ki}(\varepsilon,\delta) \to \frac{|\Gamma_i|}{\sum_{i:\,I_i\times O_k}|I_i|}.$$

This proves the theorem. □

7. Formulations in the language of partial differential equations. In the previous sections we established a number of results concerning the convergence of some families of Markov processes after proper identification with diffusion processes on graphs. As is known, solutions of different boundary-value or initial-boundary value problems for second-order partial differential equations can be represented as expectations of appropriate functionals of the corresponding processes. Using such representations one can more or less easily deduce convergence of the solutions from that of the processes. We present some results of this kind in this section.

THEOREM 7.1. *Let functions $a_i(z), b_i(z), c_{ij}(z)$ satisfy the conditions of Theorem 5.1. Then the solution of the initial-boundary value problem for the system of n equations*

$$\begin{aligned}
&\frac{\partial u^\varepsilon}{\partial t}(t,i,z) = L_i u^\varepsilon(t,i,z) + \frac{1}{\varepsilon}\sum_{j=1}^n c_{ij}(z)u^\varepsilon(t,j,z),\\
&\frac{\partial u^\varepsilon}{\partial z}(t,i,z_1) = \frac{\partial u^\varepsilon}{\partial z}(t,i,z_2) = 0,\\
&u^\varepsilon(0,i,z) = f(i,z)
\end{aligned} \tag{7.1}$$

with continuous initial conditions has a limit as $\varepsilon \to 0$:

$$(7.2)\qquad \begin{aligned} u^\varepsilon(t,i,z) &\to u^0(t,i,z) \quad \text{for } z_1 \le z < z_*, \\ u^\varepsilon(t,i,z) &\to u^0(t,0,z) \quad \text{for } z_* \le z \le z_2, \end{aligned}$$

and this limit is the unique bounded solution of the problem

$$(7.3)\qquad \begin{aligned} &\frac{\partial u^0}{\partial t}(t,i,z) = L_i u^0(t,i,z), \qquad i = 1,\dots,n,\ z_1 \le z \le z_*, \\ &\frac{\partial u^0}{\partial t}(t,0,z) = L_0 u^0(t,0,z), \qquad z_* \le z \le z_2; \\ &\frac{\partial u^0}{\partial z}(t,i,z_1) = 0, \qquad i = 1,\dots,n, \\ &\frac{\partial u^0}{\partial z}(t,0,z_2) = 0, \\ &\frac{\partial u^0}{\partial z}(t,0,z_*) = \frac{\sum_{i=1}^n a_i(z_*)q_i(z_*)(\partial u^0/\partial z)(t,i,z_*)}{\sum_{i=1}^n a_i(z_*)q_i(z_*)}; \\ &\quad u^0(0,i,z) = f(i,z), \qquad i = i,\dots,n,\ z_1 \le z < z_*, \\ &\quad u^0(0,0,z) = \bar{f}(z), \qquad z_1 \le z < z_*, \end{aligned}$$

where the operator L_0 *is defined by formulas* (5.4), (5.5), *and* $\bar{f}$ *by*

$$(7.4)\qquad \bar{f}(z) = \sum_{i=1}^n q_i(z) f(i,z).$$

PROOF. We use a result similar to, but stronger in some respects than, Lemma 5.2:

LEMMA 7.1. *Let* $g(i,z)$ *be a function on* M *such that* $g(i,z) \equiv 0$ *for* $z < z_*$, $g(i,z)$ *is continuous for* $z \in [z_*, z_2]$ *and* $\sum_{i=1}^n q_i(z) g(i,z) = 0$. *Then for any* $t > 0$,

$$(7.5)\qquad E_x^\varepsilon g(\nu^\varepsilon(t), Z^\varepsilon(t)) \to 0 \quad \text{as } \varepsilon \to 0.$$

The solution of problem (7.1) is represented as

$$(7.6)\qquad u^\varepsilon(t,i,z) = E_{(i,z)}^\varepsilon f(\nu^\varepsilon(t), Z^\varepsilon(t)).$$

We define a function $\tilde{f}$ on $Y(M)$ by $\tilde{f}(i,z) = f(i,z)$ on the segments $I_1,\dots,I_n$ and by $\tilde{f}(0,z) = \bar{f}(z)$ on the 0th segment. We have

$$(7.7)\qquad u^\varepsilon(t,i,z) = E_{(i,z)}^\varepsilon \tilde{f}(Y(X^\varepsilon(t))) + E_{(i,z)}^\varepsilon g(X^\varepsilon(t)),$$

where $g(x) = f(x) - \tilde{f}(Y(x))$, $x \in M$. Lemma 7.1 yields that the second expectation converges to 0 as $\varepsilon \to 0$. The function $\tilde{f}$ is bounded and continuous except at the point 0 where the segments of the graph meet. But for the limiting process on the graph $(Y(t), P_y)$ we have by Theorem 3.1 that $P_y\{Y(t) = 0\} = 0$; so the weak convergence yields

$$u^\varepsilon(t,i,z) \to E_{Y(i,z)} \tilde{f}(Y(t)) = u^0(t, Y(i,z)),$$

where the function $u^0(t,y) = E_y \tilde{f}(Y(t))$ satisfies (7.3). □

THEOREM 7.2. *Under the conditions of Theorem* 5.1, *the solution of the boundary-value problem*

$$L_i u^\varepsilon(i,z) + \frac{1}{\varepsilon} \sum_{j=1}^{n} c_{ij}(z) u^\varepsilon(j,z) + g(i,z) = 0, \tag{7.8}$$
$$i = 1, \dots, n,\ z_1 \le z \le z_2;$$
$$u^\varepsilon(i,z_1) = \varphi_{i1}, \qquad u^\varepsilon(i,z_2) = \varphi_{i2}$$

with continuous $g(i,z)$ *has a limit as* $\varepsilon \to 0$:

$$\begin{aligned} &u^\varepsilon(i,z) \to u^0(i,z) \quad \text{for } z_1 \le z < z_*, \\ &u^\varepsilon(i,z) \to u^0(0,z) \quad \text{for } z_* \le z \le z_2; \end{aligned} \tag{7.9}$$

and the limit is the unique bounded solution of

$$\begin{aligned} &L_i u^0(i,z) + g(i,z) = 0, \qquad i = 1, \dots, n,\ z_1 \le z < z_*, \\ &L_0 u^0(0,z) + \bar{g}(z) = 0, \qquad z_* < z \le z_i; \\ &\frac{du^0}{dz}(0, z_*) = \frac{\sum_{i=1}^n a_i(z_*) q_i(z_*) (du^0/dz)(i, z_*)}{\sum_{i=1}^n a_i(z_*) q_i(z_*)}; \\ &u^0(i, z_1) = \varphi_{i1}, \qquad i = 1, \dots, n, \\ &u^0(0, z_2) = \bar{\varphi}_2, \end{aligned} \tag{7.10}$$

where

$$\bar{g}(z) = \sum_{i=1}^{n} q_i(z) g(i,z), \tag{7.11}$$

$$\bar{\varphi}_2 = \frac{\sum_{i=1}^n a_i(z_2) q_i(z_2) \varphi_{i2}}{\sum_{i=1}^n a_i(z_2) q_i(z_2)}. \tag{7.12}$$

The proof combines the ideas of Theorem 7.1 and Lemma 5.3.

THEOREM 7.3. *Let* M^ε *be a family of regions formed by a system of narrow branching tubes as in Theorem* 6.1. *Let* $g(x)$ *and* $c(x)$ *be continuous functions in* $\bigcup_{\varepsilon \le \varepsilon_0} M^\varepsilon$, $c(x) > 0$. *Then the solution of the boundary-value problem*

$$\begin{aligned} &\tfrac{1}{2} \Delta u^\varepsilon(x) - c(x) u^\varepsilon(x) + g(x) = 0, \qquad x \in M^\varepsilon; \\ &\frac{\partial u^\varepsilon}{\partial n}(x) = 0, \qquad x \in \partial M^\varepsilon, \end{aligned} \tag{7.13}$$

converges as $\varepsilon \to 0$ *to that of the problem*

$$(7.14)\qquad \begin{aligned} &\frac{1}{2}\frac{d^2u^0}{dy^2} - c(y)u^0(y) + g(y) = 0, \qquad y \in I_i, \qquad i = 1,\dots,n,\\ &\sum_{i:\, I_i \sim O_k} |\Gamma_i| \left(\frac{du^0}{dy_i}\right)_{in} (O_k) = 0, \qquad k = 1,\dots,m. \end{aligned}$$

Results on some other boundary-value problems and on initial-boundary value problems for equations associated with families of processes considered in Sections 5 and 6 can also be obtained.

8. Averaging principle for perturbed Hamiltonian systems and diffusion processes on graphs. Let us consider a Hamiltonian system with one degree of freedom

$$(8.1)\qquad \dot{x} = \overline{\nabla} H(x), \qquad x = (p,q) \in R^2,$$

where the Hamiltonian H is bounded from below, is smooth enough, and has compact level sets $S_h = \{x \in R^2\colon\ H(x) = h\}$, $\overline{\nabla} H(p,q) = (\partial H/\partial q,\ -\partial H/\partial p)$. Suppose for simplicity that H has a finite number of stationary points, that all of these points are nondegenerate and that each set S_h contains at most one stationary point.

Any nonempty level set S_h corresponding to a noncritical value h consists of one or several components formed by periodic trajectories of the system. If h is a critical value of $H(x)$, one of the components of S_h consists of the equilibrium point O_k and, perhaps, some trajectories that tend to O_k as $t \to \pm\infty$. Let us identify all points belonging to the same connected component of S_h. The result of this identification $Y(M)$ is homeomorphic to a graph with a finite number of vertices $O_1,\dots,O_m$ and a finite number of segments $I_1,\dots,I_n$. Each vertex corresponds to a critical point (together, perhaps, with some trajectories identified with it). The interior points of the segments correspond to periodic trajectories (see Fig. 2). As the coordinate on each segment, we can take the value h of the Hamiltonian on its points.

Consider perturbations of (8.1) by a small white noise. After proper rescaling of time the perturbed system has the form

$$(8.2)\qquad \dot{X}^{\varepsilon}(t) = \frac{1}{\varepsilon}\overline{\nabla} H(X^{\varepsilon}(t)) + \dot{W}(t),$$

where W_t is a two-dimensional Wiener process, $0 < \varepsilon \ll 1$. The motion $X^{\varepsilon}(t)$ consists of fast motion along the periodic orbits of the nonperturbed system and of slow motion across the periodic orbits. To describe the slow motion let us apply Itô's formula to $H(X^{\varepsilon}(t))$:

$$(8.3)\qquad \begin{aligned} H(X^{\varepsilon}(t)) - H(X^{\varepsilon}(t_0)) &= \int_{t_0}^{t} (\nabla H(X^{\varepsilon}(s)), d\dot{W}(s))\\ &\quad + \int_{t_0}^{t} \tfrac{1}{2}\,\Delta H(X^{\varepsilon}(s))\, ds. \end{aligned}$$

Assume that $X^\varepsilon(t_0)$ belongs to the ith component S_h^i of S_h, h being a noncritical value for $H(x)$. During a small time interval $(t_0, t_0 + \Delta t)$ the change of $H(X^\varepsilon(t))$ will be small, but the trajectory $X^\varepsilon(t)$ will make many rotations along the periodic trajectory S_h^i. Simple calculations show that the stochastic integral in (8.3) for $t = t_0 + \Delta t$, $0 < \varepsilon$, $\Delta t \ll 1$, will have the variance $a_i(h) \cdot \Delta t$, where

$$a_i(h) = c^{-1} \int_{S_h^i} |\nabla H(x)|\, dl, \qquad c = \int_{S_h^i} |\nabla H(x)|^{-1}\, dl,$$

where dl means integration with respect to the length of the curve. The expectation of the other integral in (8.3) for $0 < \varepsilon$, $\Delta t \ll 1$ is close to $b_i(h) \cdot \Delta t$, where

$$b_i(h) = c^{-1} \int_{S_h^i} \frac{\Delta H(x)\, dl}{2|\nabla H(x)|}.$$

This implies that the process $H(X^\varepsilon(t))$ starting on a segment $I_i \subset Y(M)$, considered before leaving this segment, converges weakly as $\varepsilon \downarrow 0$ to the diffusion corresponding to the operator

$$L_i = \frac{a_i(h)}{2} \frac{d^2}{dh^2} + b_i \frac{d}{dx}, \qquad h_{i1} < h < h_{i2}$$

[the ends h_{ij} correspond to critical points O_{k_j} of $H(x)$]. The process $Y(X_t^\varepsilon)$ on the graph converges weakly as $\varepsilon \to 0$ to a Markov process on the graph. This limiting process is governed by the operators L_i inside the corresponding segments. The diffusion coefficients $a_i(h)$ degenerate when h approaches the vertices. One can check that the vertices corresponding to minimum and maximum points of the Hamiltonian are inaccessible for the limiting process, and no additional conditions should be added at these points. The vertices corresponding to the saddle points of $H(x)$ turn out to be regular boundary points in the sense of [1] and [5], and to determine the limiting process uniquely we have to find gluing conditions at such O_k.

Since the operators L_i degenerate at the ends of the segments, it is better to rewrite them in the form of a generalized second derivative $D_{m_i} D_{y_i}$ and use the derivatives $[df/dy_i]_{in}(O_k)$ in the gluing conditions. But since y_i turns out to have a positive continuous derivative with respect to the h-coordinate, we can write the gluing conditions in the classical form

$$\sum_{i:\, I_i \sim O_k} p_{ki} \cdot \left[\frac{df}{dh_i}\right]_{in} (O_k) = 0.$$

The fact that the Lebesgue measure is invariant for X_t^ε helps us to find the constants p_{ki}, as in Section 6. Suppose that the connected component of the level set S_h, $h = H(O_k)$, containing the critical point corresponding to O_k, consists of several trajectories s_j tending to this point as $t \to \pm\infty$. Each segment I_i ending at O_k parametrizes a family of periodic trajectories; and as a point of I_i approaches O_k, the corresponding trajectories converge to the union of some s_j, say $\bigcup_{j \in M_{ki}} s_j$. (Each s_j appears in two such unions, corresponding to two sides of s_j.) It turns out that p_{ki} are proportional to

$$\sum_{j \in M_{ki}} \int_{s_j} |\nabla H(x)| \, dl.$$

We will return to the problem discussed in this section elsewhere.

REFERENCES

[1] FELLER, W. (1952). The parabolic differential equations and the associated semigroups of transformations. *Ann. Math.* **55** 468–519.

[2] FREIDLIN, M. I. and WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems.* Springer, New York.

[3] KHASMINSKII, R. Z. (1960). Ergodic properties of recurrent diffusion processes and stabilization of the solutions of the Cauchy problem for parabolic equations. *Teor. Veroyatnost. i Primenen.* **5** 179–196.

[4] KHASMINSKII, R. Z. (1963). Principle of averaging for parabolic and elliptic equations for Markov processes with small parameter. *Teor. Veroyatnost. i Primenen.* **8** 1–21.

[5] MANDL, P. (1968). *Analytic Treatment of One-dimensional Markov Processes.* Academia, Prague.

[6] STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes.* Springer, Berlin.

[7] WENTZELL, A. D. (1961). On absolute continuity of the transition probabilities of a one-dimensional diffusion process. *Teor. Veroyatnost. i Primenen.* **6** 439–446.

[8] WENTZELL, A. D. (1973). Formulae for eigenfunctions and eigenvalues associated with a Markov process. *Theory Probab. Appl.* **18** 1–26.

[9] WOLANSKY, G. (1988). Stochastic perturbations to conservative dynamical systems on the plane. I. Convergence of invariant distributions. *Trans. Amer. Math. Soc.* **309** 621–639.

[10] WOLANSKY, G. (1990). Limit theorems for a dynamical system in the presence of resonances and homoclinic orbits. *J. Differential Equations* **83** 300–335.

[11] IKEDA, N. and WATANABE S. (1981). *Stochastic Differential Equations and Diffusion Processes.* North-Holland, Amsterdam.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742